Article No. jmbi.1999.3278 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 294, 619–625

JMB

COMMUNICATION

First Principles Prediction of Protein Folding Rates

Derek A. Debe and William A. Goddard III*

Materials and Process Simulation Center (MSC) Beckman Institute (139-74) Division of Chemistry and Chemical Engineering California Institute of Technology, Pasadena, CA 91125, USA

**Corresponding author*

Experimental studies have demonstrated that many small, single-domain proteins fold *via* simple two-state kinetics. We present a first principles approach for predicting these experimentally determined folding rates. Our approach is based on a nucleation-condensation folding mechanism, where the rate-limiting step is a random, diffusive search for the native tertiary topology. To estimate the rates of folding for various proteins *via* this mechanism, we first determine the probability of randomly sampling a conformation with the native fold topology. Next, we convert these probabilities into folding rates by estimating the rate that a protein samples different topologies during diffusive folding. This topology-sampling rate is calculated using the Einstein diffusion equation in conjunction with an experimentally determined intra-protein diffusion constant. We have applied our prediction method to the 21 topologically distinct small proteins for which two-state rate data is available. For the 18 beta-sheet and mixed alpha-beta native proteins, we predict folding rates within an average factor of 4, even though the experimental rates vary by a factor of $\sim 4 \times 10^4$. Interestingly, the experimental folding rates for the three four-helix bundle proteins are significantly underestimated by this approach, suggesting that proteins with significant helical content may fold by a faster, alternative mechanism. This method can be applied to any protein for which the structure is known and hence can be used to predict the folding rates of many proteins prior to experiment.

Keywords: protein folding; kinetics; diffusion; fold topology; nucleation-condensation

One of the most important challenges in biology is to understand the relationship between the folded structure of a protein and its primary amino acid sequence. Consequently, there has been great interest in understanding how proteins fold. An important advance in 1991 was the experimental demonstration that stable intermediates were not present in the fast folding of chymotrypsin inhibitor 2 (Jackson & Fersht, 1991). Since then, two-state folding rates for 20 more small (<120 residues), topologically distinct proteins have been determined, providing sufficient rate data to begin testing quantitative aspects of proposed folding mechanisms (Jackson, 1998). Recently, Plaxco *et al.* 1998b) reported a statistically significant correlation between the natural log of the two-state folding rate, $\ln(k_f)$, and a measure of the native state topological complexity (contact order). This empirical observation suggests that the chemistry underlying the folding of simple, single-domain proteins may be universal, implying that a single mechanistic model might quantitatively account for the observed folding rates.

We recently proposed the Topomer-Sampling Model (TSM) of protein folding, wherein proteins fold by a two-state mechanism consisting of (Debe *et al.*, 1999a):

(i) Topomer diffusion: random, diffusive sampling to find the native topomer (topomers are tubes of topologically equivalent conformations), followed by

(ii) Intra-topomer ordering: non-random, local conformational changes within the native topology to find the unique native state.

Abbreviations used: TSM, topomer-sampling model; RGP, restrained generic protein; NTP, native topology probability.

E-mail address of the corresponding author: wag@wag.caltech.edu.

0022-2836/99/480619–7 $30.00/0

Assuming step (i) represents the rate-limiting step in folding, the TSM folding rate is given by:

$$k_f = P(N\text{top}) \times k_{\text{top}} \quad (1)$$

where $P(N\text{top})$ is the probability of randomly sampling a structure with the native tertiary topology (i.e. a structure in the native topomer), and k_{top} is the rate at which a protein samples different tertiary topologies as it folds (see (Debe *et al.*, 1999a, for a precise definition of native tertiary topology in this context). Previously, we developed a method to determine the total number of topomers, S_N, for a protein of length N, allowing us to estimate $P(N\text{top}) = (S_N)^{-1}$. Using this value of $P(N\text{top})$ in equation (1), we estimated that the rate for the topomer diffusion step (i) is $\sim 10\ s^{-1}$ for a 100-residue protein. This was an encouraging result, since this calculated rate is similar to experimentally observed folding rates. However, this calculation assumed that all topologies for a protein of length N have an equal probability of being sampled, and thus the predicted folding rate did not depend on the structure of the native fold. We now propose a quantitative first principles approach for predicting folding rates of specific proteins, where the probability of sampling the native topology is explicitly calculated from the native protein structure. This approach accurately predicts the folding rates of beta-sheet and mixed alpha-beta proteins.

We estimate $P(N\text{top})$ for a specific, native protein structure as follows. Consider choosing μ contacts (residue pairs whose alpha-carbons are within ~8 Å) uniformly distributed throughout the protein structure. Given these μ contacts, we may write,

$$P(N\text{top}) = P(\mu) \times P(N\text{top}|\mu), \quad (2)$$

where $P(\mu)$ is the probability of forming the μ contacts, and $P(N\text{top} \mid \mu)$ is the conditional probability of sampling the native topology when the μ contacts are satisfied. We will focus on solving the terms in equation (2) to determine $P(N\text{top})$ for a native protein structure.

The determination of $P(\mu)$ is not trivial, since the probability of forming various inter-residue contacts in a protein depends on the location of these contacts along the protein sequence. For contact pairs that overlap in the sequence, the probability of forming one contact pair is influenced by the presence of the other contact, and thus the correlation between contact pairs must be considered while determining $P(\mu)$. Flory determined $P(\mu)$ for an average polymer of length N with μ arbitrary cross-links (the mean field approximation) (Flory, 1956). This mean field result has also been obtained using replica calculations (Gutin & Shaklinovich, 1994). Less progress has been made determining $P(\mu)$ for a particular set of μ contacts (no mean field approximation). Chan & Dill (1990) have determined correlation functions for up to three overlapping contacts with a non-arbitrary sequence separation using a cubic lattice polymer representation. However, $P(\mu)$ has not been determined for specific protein or polymer contact configurations for $\mu > 3$.

We determine $P(\mu)$ as follows. Let $P(1)$ be the probability of sampling a conformation that satisfies one of the μ specified contacts. Then $P(\mu)$ can be written as:

$$P(\mu) = P(\mu|1) \times P(1), \quad (3)$$

where $P(\mu|1)$ is the conditional probability of sampling a conformation that satisfies all μ contacts when one of the contacts is already satisfied. Similarly, the first term in equation (3) can be written as:

$$P(\mu|1) = P(\mu|2) \times P(2|1). \quad (4)$$

By recursion, it follows that:

$$P(\mu) = P(\mu|\mu-1) \times P(\mu-1|\mu-2) \times \ldots \times P(2|1) \times P(1) \quad (5)$$

The individual probability terms in equation (5) can be solved using the Restrained Generic Protein (RGP) Direct Monte Carlo Method (Debe *et al.*, 1999b). The RGP method is a fast computational procedure for generating large ensembles of self-avoiding, off-lattice (ball-and-stick; Levitt, 1976) protein conformations that comply with a set of user-defined inter-residue distance restraints. The term $P(1)$ in equation (5) is given by the probability of satisfying one (or more) of the μ contacts in an unrestrained RGP ensemble of protein conformations. The next term, $P(2|1)$, is given by the probability of satisfying two (or more) of the μ contacts in an ensemble of conformations that already comply with the inter-residue contacts that were satisfied during the determination of $P(1)$. Hence, the i (or more) contacts satisfied in the conformations during the determination of $P(i|i-1)$ are saved and used as inter-residue restraints to be satisfied by the conformations generated during the determination of $P(i+1 \mid i)$. Each term in equation (5) can be determined by this approach, yielding $P(\mu)$.

Once $P(\mu)$ is determined, the remaining term required to solve for $P(N\text{top})$ by equation (2) is $P(N\text{top}|\mu)$, the probability of sampling the native topology when the μ contacts are satisfied. $P(N\text{top}|\mu)$ is determined by generating RGP conformations that satisfy all μ contacts and using the Native Topomer Test Procedure (Debe *et al.*, 1999a) to determine the fraction of the conformations possessing the native tertiary topology.

Thus, $P(N\text{top})$ can be calculated according to equation (2). The predicted rate of folding is computed using equation (1), where k_{top} is the rate a protein samples different topologies as it folds. We take this as the inverse of the time, τ_{top}, to diffuse from one topology to the next, approximated by

the Einstein diffusion equation (Einstein, 1905):

$$k_{\text{top}} = (\tau_{\text{top}})^{-1} = (\bar{x}^2/6D)^{-1} \quad (6)$$

where $\bar{x}$ is the average CRMS distance between two neighboring topologies for a protein of length N as from our previous calculations ($\bar{x}=a+b(N)^{1/3}$, where $a=-4.12$; $b=2.61$; Debe *et al.*, 1999a), and $D \approx 5 \times 10^{-7}$ cm^2/s is an experimentally determined intra-protein diffusion constant (Hagen *et al.*, 1996). While the Einstein diffusion equation is certainly an approximation for a finite, constrained system such as a protein, the proportionality $\tau \propto \bar{x}^2$ has been shown to hold for proteins using molecular dynamics (see for example, trajectories 11 and 16 in Figure 2(a) of Lazaridis & Karplus, 1997). Furthermore, since $(\bar{x}_{N=100}/\bar{x}_{N=50})^2 \approx 2$, the variation in predicted folding rates is dominated by the $P(N\text{top})$ term in equation (1), not by k_{top}. Hence, we expect that deviations from ideal diffusive behavior and chain length scaling errors do not significantly affect the rate predictions.

We refer to the overall procedure outlined above as the Native Topology Probability (NTP) method. Table 1 shows the predicted and experimental rate data for the 21 small proteins whose two-state folding rates have been determined. Figure 1 compares the experimentally determined $\ln(k_f)$ *versus* the predicted $\ln(k_f)$ for the set of 18 topologically distinct, beta-sheet and mixed alpha-beta proteins. The linear fit has a significant correlation ($R=0.78$), corresponding to an average prediction error of $e^{1.3} \approx 3.7$. This is similar to the error in rate, $e^{1.0} \approx 2.7$ arising from sequence changes in five structurally homologous protein families for which there is sufficient rate data (vertical error bars). Thus, the NTP method accurately predicts the general, sequence-independent rate for alpha-beta and beta protein folding.

The NTP procedure considerably underestimates the folding kinetics for the three α-helical bundle proteins (Table 1), where $\ln(k_{\text{exp}})$ exceeds $\ln(k_{\text{pred}})$ by $e^{6.6} \approx 735$, $e^{3.8} \approx 45$, and $e^{2.8} \approx 16$. We expect that sequence-local helical conformational biases probably cause these helical proteins to fold faster *via* a diffusion collision mechanism (Karplus & Weaver, 1976). Future work will focus on understanding the amount of early helix formation required to produce the observed helical bundle folding rates.

The NTP rate predictions are based on the mechanistic assumption that the rate-limiting step in protein folding is a random, diffusive search for the native topology (Debe *et al.*, 1999a). Thus we assume a nucleation-condensation mechanism (Fersht, 1995), where the transition-state required for condensation to the native state is the set of structures having the same tertiary topology as the native state. This is similar to the transition-state picture developed by the Fersht group from CI2 protein engineering studies (Itzhaki *et al.*, 1995). The accuracy of our first principles predicted rates provides evidence in favor of this nucleation-condensation mechanism. Furthermore, our calcu-

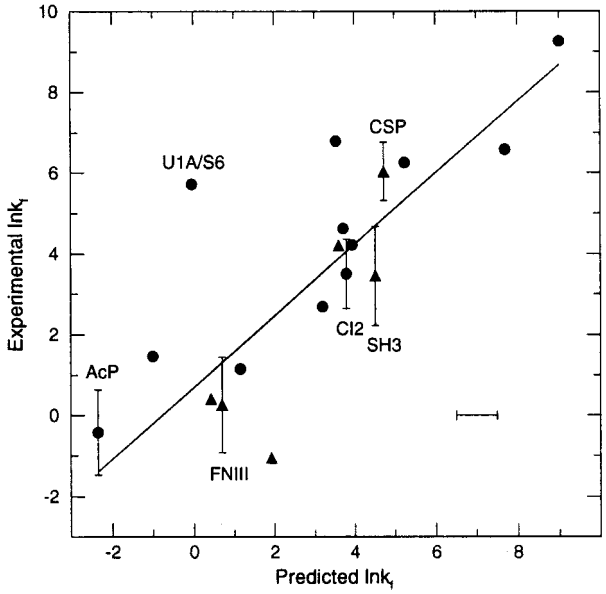

Figure 1. Experimental folding rate *versus* predicted folding rate for all 18 alpha-beta (●) and beta (▲) proteins for which there is rate data (Table 1). Over this data set the average error in the predicted rate is $e^{1.3} \approx 3.7$ ($R=0.78$; $R=0.87$ for the fit excluding the outlier U1A/S6). The vertical error bars show the average error due to sequence changes across five structurally homologous families (average error is $e^{1.0} \approx 2.7$). The horizontal error bar represents the average error in the NTP rate predictions, $e^{0.5} \approx 1.6$. The predicted folding rates were calculated from equation (1). $P(N\text{top})$ in equation (1) was determined using equation (2). The term $P(\mu)$ in equation (2) was determined from the individual terms of the form $P(i+1|i)$ in equation (5). These terms were determined from ensembles of protein conformations generated by the RGP method (see the text). The radius of gyration (Rg) of the RGP conformations used in the determination of each $P(i+1|i)$ term was limited to $Rg_{\min} < Rg < 2Rg_{\min}$, where $Rg_{\min} = -1.26 + 2.79(N)^{1/3}$ (Maiorov & Crippen, 1995). This ensured that overly compact and non-compact conformations would not be considered. Inter-residue contacts were considered satisfied if their alpha-carbons were within 9.5 Å. 200 conformations were generated for each $P(i+1|i)$ determination. $P(i+1|i)$ was typically ~0.2, yielding ~40 conformations out of the 200 that satisfied $i+1$ or more contacts. The ~40 sets of contacts satisfied during the determination of $P(i+1|i)$ were saved and used during the determination of $P(i+2|i+1)$. 200 new conformations were grown to determine $P(i+2 \mid i+1)$, so that on average, each of the ~40 different constraint sets was used to grow five of the new conformations (the algorithm cycles through the restraint sets in the order they were originally generated). Note that during the determination of $P(i+1 \mid i)$, more than $i+1$ contacts can be satisfied, for example $i+3$. In this case, all $i+3$ contacts are saved and used as restraints in the determination of the next probabilities, so that $i+2$ and $i+3$ contacts are necessarily satisfied when this contact set is used in the determination of $P(i+2 \mid i+1)$ and $P(i+3 \mid i+2)$, respectively. Note that the contact distance of 9.5 Å is the only adjustable parameter in our model and is the same for all of the proteins considered. A distance of 9.5 Å was chosen so that the calculated $P(N\text{top})$ values result in $\ln(k_f)$ predictions of the appropriate magnitude.

Table 1. Predicted and experimentally determined kinetic data for the 21 small, single domain, topologically distinct proteins (and protein families) that have been characterized

Protein	PDB Code	Length	Pred. $\ln k_f$	Pred. σ	No. Exp.	Exp. $\ln k_f$	Exp. σ
α *Proteins*							
Monomeric λ-repressor[a]	1LMB	80	3.67	0.52	3M	10.23	1.26
ACBP[b]	2ABD	86	2.84	1.59	4H	6.62	1.04
Im9[c]	1IMQ	86	4.47	0.74	1	7.31	-
β *Proteins*							
Tendamistat[d]	2AIT	74	3.60	0.13	1	4.20	-
Cold shock protein (A and B)[e]	1CSP 1MJC	67 69	4.73	0.83	4H 3M	6.04	0.72
SH3 domain (α-spectrin, src, fyn)[f]	1TUD 1NYF	64 67	4.51	0.24	3H	3.45	1.22
SH3 domain (Pl3 kinase)[g]	1PKS	84	1.92	0.06	1	−1.05	-
Twitchin[h]	1WIT	93	0.43	0.98	1	0.41	-
^{9}FN-III and Tenascin (short and long)[i]	1FNF 1TEN	90 92	0.70	0.67	2H	0.26	1.18
α/β *Proteins*							
Protein G B1 domain[j]	2GB1	56	5.23	0.26	2M	6.26	0.60
CI2[k]	1COA	64	3.80	0.58	90M	3.51	0.86
ADAh2[l]	1PBA	80	3.54	0.46	1	6.80	-
Arc repressor[m]	1ARR	53	9.01	0.51	1	9.27	-
Ubiquitin (Val26 → Ala)[n]	1UBQ	76	3.71	0.27	1	4.63	-
IgG binding domain of protein L[o]	2PTL	62	3.94	0.47	10M	4.22	0.64
Hpr[p]	1HDN	85	3.20	1.06	1	2.70	-
FKBP12[q]	1FKB	107	−1.00	0.45	1	1.46	-
AcP (Muscle and common type)[r]	1APS	98	−2.35	0.16	2H	−0.42	1.05
C-terminal domain of protein L9[s]	1DIV	93	1.15	1.00	1	1.15	-
N-terminal domain of protein L9[t]	1DIV	56	7.69	0.24	1	6.58	-
Spliceosomal U1A and ribosomal S6[u]	1URN 1RIS	102	−0.04	0.56	2T	5.73	0.09

The predicted folding rates have an average uncertainty of $\sim e^{0.5} \approx 1.6$, based on calculations using at least two different sets of contacts to determine $P(\mu)$ and $P(N\text{top}|\mu)$ for each protein (the contacts sets used for each protein are given in Supplemental Tables 1-1 through 1-21). Column 6 lists the number and type of experiments that have been done on structurally homologous proteins of different sequence (M denotes point mutation experiments; H denotes homologous sequence experiments; T denotes experiments on proteins that are structural, but not sequence homologs). Such experiments estimate the extent that sequence-specific effects influence the folding rate.

[a] Huang & Oas (1995); Burton *et al*. (1996); Ghaemmaghami *et al*. (1998).
[b] Kragelund *et al*. (1995), (1996).
[c] Ferguson *et al*. (1999).
[d] Schonbrunner *et al*. (1997).
[e] Schindler *et al*. (1995); Perl *et al*. (1998); Reid *et al*. (1998).
[f] Viguera *et al*. (l994), (l996); Grantcharova & Baker (1997); Plaxco *et al*. (1998a).
[g] Guijarro *et al*. (1998).
[h] S. J. Hamill, unpublished observations. Data from Jackson (1998).
[i] Plaxco *et al*. (1997); Clarke *et al*. (1997); Hamill *et al*. (1998).
[j] Smith *et al*. (1996).
[k] Jackson & Fersht (1991); Itzhaki *et al*. (1995); Ladurner *et al*. (1998).
[l] Villegas *et al*. (1995).
[m] Robinson & Sauer (1996).
[n] Khorasanizadeh *et al*. (1996).
[o] Scalley *et al*. (1997).
[p] Van Nuland *et al*. (1998a).
[q] S. E. Jackson, unpublished observations. Data from Jackson (1998).
[r] van Nuland *et al*. 1998a; Taddei *et al*. (1999).
[s] Sato *et al*. (1999).
[t] Kuhlman *et al*. (1998).
[u] Silow & Oliveberg, (1997); Otzen *et al*. (1999).

lations demonstrate how a nucleation-condensation mechanism accounts for the inverse relationship between folding rate and solvent viscosity recently observed for three small, two-state folding proteins (Jacob *et al*., 1997; Plaxco & Baker, 1998; Bhattacharyya & Sosnick, 1999). This relationship

is directly implied in our model by equation (6) given the inverse proportionality between the diffusion constant and solvent viscosity in Stokes' Law.

We did not include any information about the stability of the native fold to accurately predict folding rates. Thus, we do not expect that stability is a primary factor in determining the folding rate (Plaxco *et al.*, 1998b). However, a correlation could exist between native stability and folding rate for structurally homologous protein families (Plaxco *et al.*, 1998a). This correlation could arise from stabilizing sequence changes that increase the probability of trapping the protein once it has diffused into its native topomer.

The NTP predictions provide a useful framework for understanding factors that can change folding kinetics. The predicted folding rate is equivalent to the rate of randomly sampling a conformation in the transition-state ensemble. Thus, the folding rate is directly related to the difficulty of finding a conformation that can quickly condense to the native state. As the transition-state becomes more native-like (referred to as a tight transition state), the difficulty of finding a structure in the transition-state increases, and the folding rate decreases unless there is a mechanism to aid the search for a transition-state structure. Similarly, as the transition-state becomes less native-like (a loose transition state), the folding rate is expected to increase. Folding kinetics can also be affected by a change in the number of topologies that are available to the folding peptide. Faster folding rates might arise from sequence-specific conformation biases, such as helical formation, which could preclude a protein from sampling a significant fraction of its possible topologies.

Careful examination of Figure 1 reveals that the NTP procedure accurately predicts the folding rates of all reported two-state beta-sheet and mixed alpha-beta proteins, with only one exception. The exception is spliceosomal protein UIA (and its recently characterized structural homolog ribosomal protein S6), which folds some $e^{5.8} \approx 330$ times more rapidly than predicted by the NTP method. U1A and S6 are quite long for two-state folding proteins (102 and 101 residues, respectively) and exhibit exceptional folding behavior, in that the $\log(k_{\text{unfold}})$ *versus* denaturant concentration and $\log(k_{\text{fold}})$ *versus* denaturant plots exhibit equal and opposite curvature in many of the mutants studied (Silow & Oliveberg, 1997; Otzen *et al.*, 1999). This curvature is interpreted by Silow and Oliveberg to imply that there is a change in the position of the folding transition-state with denaturant concentration (many of the mutants exhibit a very loose folding transition state in water, which qualitatively implies they should indeed fold faster than the NTP procedure predicts). Removing U1A/S6 from our data set on these grounds greatly improves the overall correlation ($R = 0.87$), corresponding to an average prediction error of $e^{1.1} \approx 3.0$.

The NTP predictions apply to single domain proteins with single folding nuclei. Similar estimates can be made for multiple domain proteins if the nucleus formation events for different domains are independent (indeed, this assumption was used to predict the folding rates for the Arc repressor and the C-terminal domain of protein L9). Based on our previous estimates, we expect that single domain proteins longer than ~120 residues require more than a second to fold by a topomer sampling mechanism (Debe *et al.*, 1999a), which would expose them to proteolysis *in vivo*. Possibly, the mechanisms that speed up the folding kinetics in alpha-helical proteins and the U1A/S6 family also allow protein domains beyond ~120 residues to fold on the shorter time scale appropriate for *in vivo* folding. Strong local structure propensities (Baldwin, 1993), early helix formation (Viguera *et al.*, 1997), and beta sheet inducing mechanisms such as glycosylation (O'Connor & Imperiali, 1998) are likely to play an important role during the *in vivo* folding of large protein domains.

To predict the folding rates for specific native proteins, we have developed a procedure for determining the probability of satisfying a specific set of contacts in a native protein structure. In addition to accurate rate prediction, our method explains the observed dependence of folding rate on solvent viscosity and provides a satisfying structural definition of the folding transition-state that correlates well with a nucleation-condensation picture of folding. Our approach is quite different from correlated energy landscape (Plotkin *et al.*, 1996) and free energy functional (Shoemaker *et al.*, 1999) folding theories, which use mean-field approximations to estimate the conformational entropy. Our approach avoids a mean-field treatment of contact probability, allowing it to be applied to native proteins very easily. However, unlike these theories, our method lacks any quantitative estimate of the enthalpy of various conformations (often given by the interaction energies of various contacts in other folding models (Miyazawa & Jernigan, 1985)). Future work will seek to merge our approach into a theoretical framework that allows free energies to be estimated. We expect that our calculations can be used to tune the entropy estimates in mean-field approaches to specific native proteins (possibly using an interpolation between mean-field and specific contact probability formulations (Shoemaker *et al.*, 1999)), leading to unified theories that yield approximate free energy estimates simultaneously with accurate, experimentally verifiable rate predictions.

Acknowledgements

We thank Professors Sunney I. Chan, H. B. Gray, Kevin W. Plaxco and Joseph F. Danzer for helpful discussions. D. D. thanks the NIH for a Predoctoral Training in Biotechnology grant (GM 08346-10). We also

thank the NIH for funding (HD 36385). The computational facilities used in the research were partially supported by ARO DURIP, NSF (CHE 95-22179 and ASC 9217368), DOE-ASCI, ARO MURI (Dendrimers), ARO MURI (Olfaction), Beckman Institute, BP Chemical, Chevron, Exxon, Avery Dennison, Dow Chemical, Owens-Coming, Asahi Chemical, and Seiko-Epson.

References

Baldwin, R. L. (1993). Pulsed H/D-exchange studies of folding intermediates. *Curr. Opin. Struct. Biol.* **3**, 84-91.

Bhattacharyya, R. P. & Sosnick, T. R. (1999). Viscosity dependence of the folding kinetics of a dimeric and monomeric coiled coil. *Biochemistry,* **38**, 2601-2609.

Burton, R. E., Huang, G. S., Daugherty, M. A., Fullbright, P. W. & Oas, T. G. (1996). Microsecond protein folding through a compact transition state. *J. Mol. Biol.* **263**, 311-322.

Chan, H. S. & Dill, K. A. (1990). The effects of internal constraints on the configurations of chain molecules. *J. Chem. Phys.* **92**, 3118-3135.

Clarke, J., Hamill, S. J. & Johnson, C. M. (1997). Folding and stability of a fibronectin type III domain of human tenascin. *J. Mol. Biol.* **270**, 771-778.

Debe, D. A., Carlson, M. J. & Goddard, W. A. (1999a). The topomer-sampling model of protein folding. *Proc. Natl Acad. Sci. USA,* **96**, 2596-2601.

Debe, D. A., Carlson, M. J., Sadanobu, J., Chan, S. I. & Goddard, W. A. (1999b). Protein fold determination from sparse distance restraints: the restrained generic protein direct Monte Carlo method. *J. Phys. Chem. B,* **103**, 3001-3008.

Einstein, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys. (Leipzig),* **17**, 549-560.

Ferguson, N., Capaldi, A. P., James, R., Kleanthous, C. & Radford, S. E. (1999). Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J. Mol. Biol.* **286**, 1597-1608.

Fersht, A. R. (1995). Optimization of rates of protein-folding - the nucleation-condensation mechanism and its implications. *Proc. Natl Acad. Sci. USA,* **92**, 10869-10873.

Flory, P. J. (1956). Theory of elastic mechanisms in fibrous proteins. *J. Am. Chem. Soc.* **78**, 5222-5235.

Ghaemmaghami, S., Word, J. M., Burton, R. E., Richardson, J. S. & Oas, T. G. (1998). Folding kinetics of a fluorescent variant of monomeric lambda repressor. *Biochemistry,* **37**, 9179-9185.

Grantcharova, V. P. & Baker, D. (1997). Folding dynamics of the src SH3 domain. *Biochemistry,* **36**, 15685-15692.

Guijarro, J. I., Morton, C. J., Plaxco, K. W., Campbell, L. D. & Dobson, C. M. (1998). Folding kinetics of the SH3 domain of PI3 kinase by real-time NMR combined with optical spectroscopy. *J. Mol. Biol.* **276**, 657-667.

Gutin, A. M. & Shakhnovich, E. I. (1994). Statistical-mechanics of polymers with distance constraints. *J. Chem. Phys.* **100**, 5290-5293.

Hagen, S. J., Hofrichter, J., Szabo, A. & Eaton, W. A. (1996). Diffusion-limited contact formation in unfolded cytochrome c: estimating the maximum rate of protein folding. *Proc. Natl Acad. Sci. USA,* **93**, 11615-11617.

Hamill, S. J., Meekhof, A. E. & Clarke, J. (1998). The effect of boundary selection on the stability and folding of the third fibronectin type III domain from human tenascin. *Biochemistry,* **37**, 8071-8079.

Huang, G. S. & Oas, T. G. (1995). Structure and stability of monomeric lambda-reppressor - NMR evidence for 2-state folding. *Biochemistry,* **34**, 3884-3892.

Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1995). The structure of the transition-state for folding of chymotrypsin inhibitor-2 analyzed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein-folding. *J. Mol. Biol.* **254**, 260-288.

Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold Design,* **3**, R81-R91.

Jackson, S. E. & Fersht, A. R. (1991). Folding of chymotrypsin inhibitor-2.1. Evidence for a 2-state transition. *Biochemistry,* **30**, 10428-10435.

Jacob, M., Schindler, T., Balbach, J. & Schmid, F. X. (1997). Diffusion control in an elementary protein folding reaction. *Proc. Natl Acad. Sci. USA,* **94**, 5622-5627.

Karplus, M. & Weaver, D. L. (1976). Protein folding dynamics. *Nature,* **260**, 404-406.

Khorasanizadeh, S., Peters, I. D. & Roder, H. (1996). Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nature Struct. Biol.* **3**, 193-205.

Kragelund, B. B., Robinson, C. V., Knudsen, J., Dobson, C. M. & Poulsen, F. M. (1995). Folding of a 4-helix bundle - studies of acyl-coenzyme-A binding-protein. *Biochemistry,* **34**, 7217-7224.

Kragelund, B. B., Hojrup, P., Jensen, M. S., Schjerling, C. K., Juul, E., Knudsen, J. & Poulsen, F. M. (1996). Fast and one-step folding of closely and distantly related homologous proteins of a four-helix bundle family. *J. Mol. Biol.* **256**, 187-200.

Kuhlman, B., Luisi, D. L., Evans, P. A. & Raleigh, D. P. (1998). Global analysis of the effects of temperature and denaturant on the folding and unfolding kinetics of the N-terminal domain of the protein L9. *J. Mol. Biol.* **284**, 1661-1670.

Ladurner, A. G., Itzhaki, L. S., Daggett, V. & Fersht, A. R. (1998). Synergy between simulation and experiment in describing the energy landscape of protein folding. *Proc. Natl Acad. Sci. USA,* **95**, 8473-8478.

Lazaridis, T. & Karplus, M. (1997). ''New-View'' of protein folding reconciled with the old through multiple unfolding simulations. *Science,* **278**, 1928-1931.

Levitt, M. (1976). A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* **104**, 59-107.

Maiorov, V. N. & Crippen, G. M. (1995). Size-independent comparison of protein 3-dimensional structures. *Proteins: Struct. Funct. Genet.* **22**, 273-283.

Miyazawa, S. & Jernigan, R. L. (1985). Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules,* **218**, 534-552.

O'Connor, S. E. & Imperiali, B. (1998). A molecular basis for glycosylation-induced conformational switching. *Chem. Biol.* **5**, 427-437.

Otzen, D. E., Kristensen, O., Proctor, M. & Oliveberg, M. (1999). Structural changes in the transition state of protein folding: alternative interpretations of curved chevron plots. *Biochemistry,* **38**, 6499-6511.

Perl, D., Welker, C., Schindler, T., Schroder, K., Marahiel, M. A., Jaenicke, R. & Schmid, F. X. (1998). Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nature Struct. Biol.* **5**, 229-235.

Plaxco, K. W. & Baker, D. (1998). Limited internal friction in the rate-limiting step of a two-state protein folding reaction. *Proc. Natl Acad. Sci. USA,* **95**, 13591-13596.

Plaxco, K. W., Spitzfaden, C., Campbell, I. D. & Dobson, C. M. (1997). A comparison of the folding kinetics and thermodynamics of two homologous fibronectin type III modules. *J. Mol. Biol.* **270**, 763-770.

Plaxco, K. W., Guijarro, J. I., Morton, C. L., Pitkeathly, M., Campbell, I. D. & Dobson, C. M. (1998a). The folding kinetics and thermodynamics of the Fyn-SH3 domain. *Biochemistry,* **37**, 2529-2537.

Plaxco, K. W., Simons, K. T. & Baker, D. (1998b). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985-994.

Plotkin, S. S., Wang, J. & Wolynes, P. G. (1996). Correlated energy landscape model for finite, random heteropolymers. *Phys. Rev. E,* **53**, 6271-6296.

Reid, K. L., Rodriguez, H. M., Hillier, B. J. & Gregoret, L. M. (1998). Stability and folding properties of a model beta-sheet protein, *Escherichia coli* CspA. *Protein Sci.* **7**, 470-479.

Robinson, C. R. & Sauer, R. T. (1996). Equilibrium stability and sub-millisecond refolding of a designed single-chain arc repressor. *Biochemistry,* **35**, 13878-13884.

Sato, S., Kuhlman, B., Wu, W. J. & Raleigh, D. P. (1999). Folding of the multidomain ribosomal protein L9: the two domains fold independently with remarkably different rates. *Biochemistry,* **38**, 5643-5650.

Scalley, M. L., Yi, Q., Gu, H. D., McCormack, A., Yates, J. R. & Baker, D. (1997). Kinetics of folding of the IgG binding domain of peptostreptoccocal protein L. *Biochemistry,* **36**, 3373-3382.

Schindler, T., Herrler, M., Marahiel, M. A. & Schmid, F. X. (1995). Extremely rapid protein-folding in the absence of intermediates. *Nature Struct. Biol.* **2**, 663-673.

Schonbrunner, N., Koller, K. P. & Kiefhaber, T. (1997). Folding of the disulfide-bonded beta-sheet protein tendamistat: rapid two-state folding without hydrophobic collapse. *J. Mol. Biol.* **268**, 526-538.

Shoemaker, B. A., Wang, I. & Wolynes, P. G. (1999). Exploring structures in protein folding funnels with free energy functionals: the transition state ensemble. *J. Mol. Biol.* **287**, 675-674.

Silow, M. & Oliveberg, M. (1997). High-energy channeling in protein folding. *Biochemistry,* **36**, 7633-7637.

Smith, C. K., Bu, Z. M., Anderson, K. S., Sturtevant, J. M., Engelman, D. M. & Regan, L. (1996). Surface point mutations that significantly alter the structure and stability of a protein's denatured state. *Protein Sci.* **5**, 2009-2019.

Taddei, N., Chiti, F., Paoli, P., Fiaschi, T., Bucciantini, M., Stefani, M., Dobson, C. M. & Ramponi, G. (1999). Thermodynamics and kinetics of folding of common-type acylphosphatase: comparison to the highly homologous muscle isoenzyme. *Biochemistry,* **38**, 2135-2142.

van Nuland, N. A. L., Chiti, F., Taddei, N., Raugei, G., Ramponi, G. & Dobson, C. M. (1998a). Slow folding of muscle acylphosphatase in the absence of intermediates. *J. Mol. Biol.* **283**, 883-891.

van Nuland, N. A. J., Meijberg, W., Warner, L., Forge, V., Scheek, R. M., Robillard, G. T. & Dobson, C. M. (1998b). Slow cooperative folding of a small globular protein HPr. *Biochemistry,* **37**, 622-637.

Viguera, A. R., Martinez, J. C., Filimonov, V. V., Mateo, P. L. & Serrano, L. (1994). Thermodynamic and kinetic-analysis of the Sh3 domain of spectrin shows a 2-state folding transition. *Biochemistry,* **33**, 2142-2150.

Viguera, A. R., Serrano, L. & Wilmanns, M. (1996). Different folding transition states may result in the same native structure. *Nature Struct. Biol.* **3**, 874-880.

Viguera, A. R., Villegas, V., Aviles, F. X. & Serrano, L. (1997). Favourable native-like helical local interactions can accelerate protein folding. *Fold. Design,* **2**, 23-33.

Villegas, V., Azuaga, A., Catasus, L., Reverter, D., Mateo, P. L., Aviles, F. X. & Serrano, L. (1995). Evidence for a 2-state transition in the folding process of the activation domain of human procarboxypeptidase-A2. *Biochemistry,* **34**, 15105-15110.

Edited by F. E. Cohen

(Received 12 July 1999; received in revised form 5 October 1999; accepted 5 October 1999)

http://www.academicpress.com/jmb

Supplementary Material comprising one Table is available from JMB Online