The protein folding 'speed limit'

Jan Kubelka, James Hofrichter and William A Eaton*

How fast can a protein possibly fold? This question has stimulated experimentalists to seek fast folding proteins and to engineer them to fold even faster. Proteins folding at or near the speed limit are prime candidates for all-atom molecular dynamics simulations. They may also have no free energy barrier, allowing the direct observation of intermediate structures on the pathways from the unfolded to the folded state. Both experimental and theoretical approaches predict a speed limit of approximately $N/100$ μs for a generic N-residue single-domain protein, with α proteins folding faster than β or αβ. The predicted limits suggest that most known ultrafast folding proteins can be engineered to fold more than ten times faster.

Addresses
Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Building 5, Room 104, Bethesda, MD 20892-0520, USA
*e-mail: eaton@helix.nih.gov

Abbreviation
FRET   Förster resonance energy transfer

Introduction
The introduction of pulsed laser techniques to trigger folding processes in nanoseconds [1] has had a major impact on experimental, theoretical and computational studies of protein folding. These new techniques have allowed the investigation of the mechanism of formation of the basic structural elements of proteins — α helices, β hairpins and loops — as well as mechanisms of formation for the fastest folding proteins [2–6]. Fast folding studies have raised new kinds of questions. One of these, first posed by Hagen *et al*. [7] and the subject of this review, is: how fast can a protein possibly fold or what is the folding 'speed limit'?

The obvious significance of establishing speed limits is related to computer simulations of protein folding. In principle, much of what one would like to know about the mechanism of folding for a particular protein is contained in folding trajectories calculated using all-atom molecular dynamics. However, simulation of folding is computa-

tionally intensive because many long trajectories must be calculated in order to obtain sufficient statistical sampling to describe kinetics. Proteins that fold in the shortest possible time are therefore prime candidates for such studies. The notion of a speed limit and the possibility of direct simulation by molecular dynamics have motivated several groups both to search for and design ultrafast folding proteins ($\tau_{folding} < 100$ μs), and to reengineer them to make them fold even faster [8–20]. The discovery of such proteins has led to direct comparisons of simulated and experimental folding kinetics [13,15,16,21–23]. Comparisons have also been made between unfolding simulations at experimentally inaccessible high temperatures and experiments at much lower temperatures [24,25].

The not-so-obvious significance is that at the speed limit the free energy barrier to folding may disappear — the 'downhill' folding scenario of Bryngelson *et al*. [14,26–29]. Downhill folding presents the possibility of obtaining much more information on the folding mechanism. For small proteins (<100 residues), it is commonly found that only two populations of molecules are observable at equilibrium — folded and unfolded [30]. In kinetic experiments on the folding of these so-called two-state proteins, no intermediates are observed and there is simply an exchange, with time, of the folded and unfolded populations (Figure 1). No direct information can be obtained on the microscopic pathways that connect the two states (in principle, this could be obtained if it were possible to resolve the state-to-state transitions in single-molecule measurements [31–33]). However, indirect information on the structure of the transition state ensemble can be obtained by measuring the relative effects of single mutations on the folding equilibrium and kinetics, and interpreting the results in terms of the extent to which the interactions of the residue in the transition state resemble those in the native state (φ-value analysis [34]). By contrast, in downhill folding there is the prospect that intermediate structures could be directly observed all along the reaction coordinate by spectroscopic methods [28] (Figure 1).

In this review we describe both experimental and theoretical approaches to predicting a protein folding speed limit, and then briefly compare results for the fastest folding proteins with each other and with molecular dynamics simulations of their kinetics.

Estimates from measured rates for structural elements
An easy method of estimating the speed limit is to assert that a protein cannot fold faster than the rate at which its

Figure 1
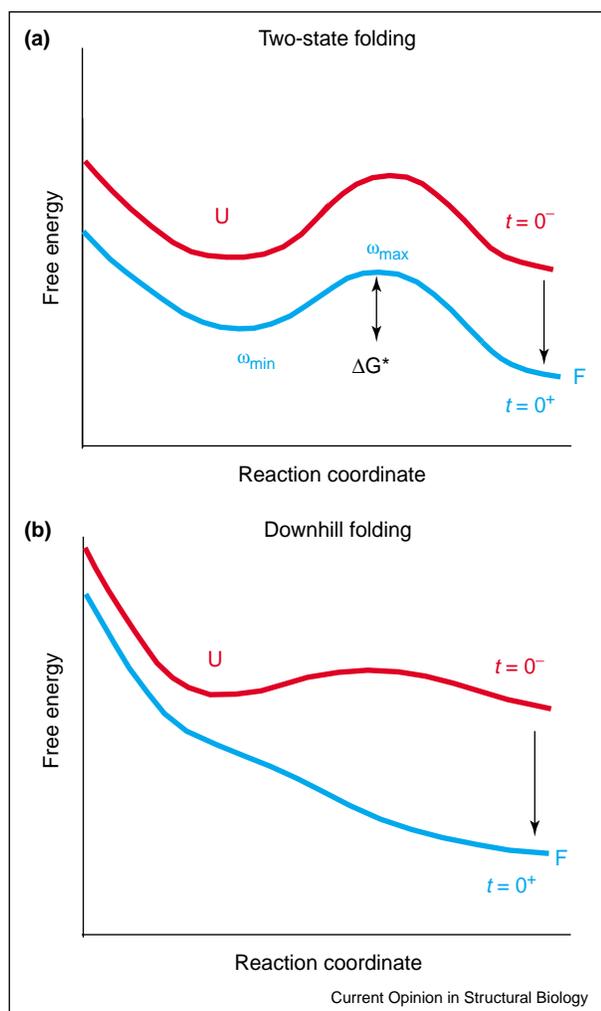
Schematic free energy profiles before ($t = 0^-$) and after ($t = 0^+$) a temperature jump for **(a)** two-state folding and **(b)** downhill folding. It is assumed that the protein cold denatures, so that the initial temperature is lower than the temperature of maximum stability.

slowest forming structural element folds. With this postulate we can estimate the speed limit by considering the data for peptide fragments on the rates of forming α helices, β hairpins and loops (Table 1).

There is now a significant body of data on diffusion-limited loop formation for peptides of varying composition. The experimental methods are based on measuring the rate of contact formation from triplet-triplet energy transfer [35,36] and fluorescence quenching [37] using extrinsic probes, or by quenching of the triplet state of tryptophan by cysteine [38–42]. There is general agreement among the methods, although some differences in the results remain to be explained, such as the apparently longer times observed using the fluorescence quenching technique [37], suggesting that these measured rates are

not purely diffusion limited. Nevertheless, the vast majority of measured loop formation times are less than 0.1 μs, much faster than the formation of α helices or β hairpins of comparable length (Table 1). From this we conclude that loop formation is not rate limiting and therefore cannot by itself be a useful determinant of the speed limit.

When sequences corresponding to helices and hairpins are excised from proteins, they are almost invariably unstable [43–45]. In the case of the α helix, most folding studies have been carried out on alanine-rich peptides, because the helix propensity of alanine is the highest among the amino acids. According to helix-coil theory, the high propensity results from the smaller entropy cost of ordering alanine into a helical conformation. The consequence for the kinetics of α-helix formation is that the entropy decrease associated with nucleating a helical turn is smaller for alanine than for other amino acids [5,46]. As alanine-rich peptides are probably the fastest helix formers, the data in Table 1 suggest a speed limit of approximately 0.5 μs for α-helical proteins.

Not only is it difficult to identify isolated β hairpins from naturally occurring proteins [47,48], but also most designed hairpins do not have a tryptophan for fluorescence detection, the most common probe in kinetic studies. The first hairpin-forming peptide to be characterized in equilibrium and kinetic studies was the 16-residue fragment from the C terminus of protein GB1 [49]. This peptide exhibits two-state behavior and folds in 6 μs at 297 K. Because the GB1 hairpin forms very fast and addresses most issues concerning how a small protein folds, it has become a major benchmark for simulation studies (see the review by Gnanakaran *et al.* [50] for references to simulation and theoretical studies of structural elements).

Only quite recently have folding rates been reported for other hairpins. Using laser temperature jump and IR detection, Xu *et al.* [51] have measured the equilibrium constants and rates for two designed hairpin-forming peptides. A sequence similar to the GB1 hairpin, but containing four tryptophans instead of one (trpzip4), folds in 13 μs, whereas an unrelated sequence (peptide I) folds in 0.8 μs. Muñoz *et al.* (V Muñoz *et al.*, unpublished) have studied an engineered version of the N-terminal hairpin-forming peptide from ubiquitin [52]. This peptide aggregates, making the analysis more complex, but studies of the concentration dependence show that the formation rate is approximately 1/(20 μs) at 300 K.

Although the data for β hairpins are still sparse, our postulate suggests that the speed limit for all-β or αβ proteins is most probably slower (possibly tenfold) than the 0.5 μs estimated for α proteins. We should point out that the rates measured for α helices and β hairpins are for relatively stable secondary structures, unlike the vast majority of protein fragments. A frequent misconception

Table 1

Rate of formation of structural elements of proteins.

Sequence	$\tau_{contact}$(μs)		T (K)	References
Disordered loop				
Lip-(AARAA)$_4$-W-NH$_2$	**0.09**		293	[41]
C-(AGQ)$_n$-W-NH$_2$; n = 1–6	**0.02–0.1**		273	[40]
Xan-(GS)$_n$-NAla-GS; n = 1–28	**0.007–0.16**		296	[36]
Xan-S$_n$-NAla-GS; n = 2–11	**0.012–0.03**		296	[36]
W-X$_6$-DBO-NH$_2$; X = G,S,D,N,A,T,L,F,E,Q,H,R,K,V,I	**0.03–0.4**		Ambient (X=G) – (X=I)	[37]
Unfolded cytochrome c HTVEKGGKHKTGPNLH	**0.25**		295	[96]

Sequence	$\tau_{folding}$ (μs)[a]	$\Delta G_{folding}$ (kcal/mol)	T (K)	References
α Helix				
Ac-YG (AKA$_3$)$_2$ AG-NH$_2$	**2.0**	1.2	284	[97]
Ac-YGSPEA$_3$KA$_4$-^{D}R-NH$_2$	**0.7**	0.5	284	[97]
Ac-YGG (KA$_4$)$_3$ K-NH$_2$	**0.7**	0.2	290	[98]
Ac-YGSPEA$_3$ (KA$_4$)$_2$ ^{D}R-NH$_2$	**0.6**	0	288	[99]
Ac-WA$_3$H$^+$ (A$_3$RA)$_3$ A-NH$_2$	**0.4**	0	300	[46]
A$_5$ (A$_3$RA)$_3$ A	**0.8**	0.7	310	[100]
β Hairpin				
(GB1) GEWTYDDATKTFTVTE	**6**	0	297	[49]
(trpzip4) GEWTWDDATKTWTWTE	**13**		300	[51]
(peptide I) SESYINPDGTWTVTE	**0.8**	0	300	[51]
(ubiquitin) MQIWVKNPDGTITLEVK	**~20**	~−1	303	b

[a]Calculated assuming a two-state system from $\tau_{folding} = \tau_{obs} (1 + \exp(\Delta G_{folding}/k_B T))$, where τ_{obs} is the measured relaxation time. [b]V Muñoz et al., unpublished. DBO, 2,3-diazabicyclo[2,2,2]oct-2-ene; NAla, naphtylalanine; Xan, xanthone.

in discussions of protein folding is that helices and hairpins have been shown to fold very fast, so that slow folding must arise from sources other than secondary structure formation. However, unstable secondary structure elements form slowly [6,53]. The scaling of the rate with stability is not yet known, but the data in Table 1 suggest a rough scaling of $\tau_{folding} \propto K_{eq}^{-1/2}$ for α helices.

Estimates from polymer collapse theory

In order to obtain a deeper understanding of the speed limit, one would like to have a theory, as for simple bimolecular reactions in solution. Smoluchowski showed, almost 100 years ago, that the upper limit for the rate coefficient k is determined by the rate at which reactants diffuse together and derived the simple formula $k = 4\pi Da$, where D is the sum of the diffusion coefficients of the two reactants and a is the sum of their radii at contact [54]. Debye later showed how this rate is either decreased or increased for charged reactants, depending on whether they have the same or opposite charge, respectively [54]. The analogous question for protein folding would be to ask: how fast can a polypeptide chain collapse from an unfolded, random coil configuration to an object as compact as the native protein? This remains an important unsolved problem in polymer physics. De Gennes [55] investigated the collapse dynamics of homopolymers, but used a model that applies to polymers much longer than those considered here. Subsequently, Pitard and Orland [56] derived an expression from an approximate solution to the (Langevin) equation of motion for uncharged homopolymers of lengths relevant to proteins. The model describes collapse induced by changing the solvent and includes hydrodynamic interaction [57]. Because collapse in their model is driven by nonspecific hydrophobic forces, the authors argue that homopolymer and heteropolymer (i.e. polypeptide) collapse will be similar. The collapse time (at long times) is given by $\tau_{collapse} \sim \eta a_o^3 n/T$, where η is the solvent viscosity, a_o is the length of the chain segment, n is the number of segments and T is the absolute temperature [57]. The proportionality factor in this equation is complex and was not given, so no estimate of the collapse time has yet been made with this model. However, this relation does show that, in theoretically estimating speed limits for polypeptides of varying length, we should use a linear length scaling.

Estimates from reaction rate theory

Another theoretical approach to estimating the speed limit is to use Kramers' theory of unimolecular reaction rates in solution [58–60]. Kramers' theory assumes that the dynamics can be described by one-dimensional diffusion along a reaction coordinate in which both the reactant well (in this case, the minimum in the free energy versus reaction coordinate profile corresponding to the unfolded state) and barrier top are parabolic (Figure 1). The folding time (the reciprocal of the folding rate coefficient) is given by:

$$\tau_{\text{folding}} = \frac{2\pi k_B T}{\omega_{\min}\omega_{\max}D_{\max}}\exp\left(\frac{\Delta G^*}{k_B T}\right) \approx 2\pi\tau_{\text{corr}}\exp\left(\frac{\Delta G^*}{k_B T}\right)$$

(Equation 1)

where $\omega_{\min}$ and $\omega_{\max}$ are frequencies that characterize the curvature of the free energy profile at the unfolded well and (inverted) barrier top, respectively, $D_{\max}$ is the diffusion constant at the barrier top, ΔG^* is the height of the free energy barrier, k_B is Boltzmann's constant and T is the absolute temperature (a unit mass has been assumed for the fictitious particle diffusing on this surface) (Figure 1). If $\omega_{\min} \approx \omega_{\max}$ and $D_{\max} \approx D_{\min}$, the pre-exponential factor depends only on τ_{corr}, the decay time of the autocorrelation function for motion in the unfolded well ($\tau_{\text{corr}} = k_B T/D_{\min}\omega_{\min}^2$). The accuracy of this simpler form of Kramers' equation was tested by Socci et al. [60] using Monte Carlo simulations of folding a lattice representation of a protein, with the number of native contacts, Q, as the reaction coordinate. They obtained agreement within a factor of two between the mean folding time determined in the simulation and the mean (first passage) time calculated from $\tau_{\text{corr}}(Q)$ for the unfolded well and the barrier height for the free energy versus Q profile, for which $\omega_{\min}(Q) \approx \omega_{\max}(Q)$ is a good approximation.

The accuracy of Kramers' equation, even for barriers as low as 2 $k_B T$, suggests that the pre-exponential factor closely approximates the folding time in the absence of a free energy barrier. The folding time may decrease further as the free energy profile develops a steeper downhill gradient by increasing the stability (Figure 1). However, stability may come at the expense of increased transient trapping in local energy minima due to stronger non-native inter-residue interactions (landscape 'roughness' [26]; see also the review by Onuchic and Wolynes, this issue), with the result that the folding time may first decrease with increasing gradient and then increase [28]. This uncertainty suggests that we may tentatively identify the speed limit as the folding time when the free energy barrier first disappears and therefore simply use the predicted pre-exponential factor for this estimate.

Kramers' equation can be used in two ways to obtain the pre-exponential factor and therefore the speed limit. One is to combine the measured folding time with theoretical calculation of the barrier height. Muñoz and Eaton [61] used a very simple Ising-like model to calculate barrier heights from free energy profiles, with the number of ordered residues as the reaction coordinate. Their model considered only native inter-residue interactions, obtained from the contact map of the three-dimensional structure, and the entropy loss from ordering residues. A pre-exponential factor of approximately 10 µs gave the best agreement with experimental rates for 18 two-state proteins with an average length of 80 residues. Using a more extensive and rigorous analytical theory, Portman et al. [62,63] obtained a pre-exponential factor of 0.4 µs for the 80-residue monomeric λ repressor from their calculation of the free energy versus Q profile. The potentially most accurate theoretical estimate of the pre-exponential factor from barrier heights and measured folding times could be derived from folding free energy surfaces obtained by exhaustive sampling of conformational space using all-atom molecular dynamics simulations [64,65]. However, there has, as yet, been no such systematic analysis.

A second method is to obtain τ_{corr} directly from experiments. Q is not yet a measurable quantity. However, a related quantity can be measured — the decay time of the end-to-end distance (r) correlation function, $\tau_{\text{corr}}(r)$, for the denatured state. Socci has recently found that $\tau_{\text{corr}}(r) \approx \tau_{\text{corr}}(Q)$ for the same lattice simulation [60] that accurately reproduced the folding times (N Socci, personal communication). The advantage of $\tau_{\text{corr}}(r)$ is that it can be obtained from the measured relaxation time of the Förster resonance energy transfer (FRET) efficiency for the denatured protein. Sadqi et al. [66] have measured $\tau_{\text{corr}}(\text{FRET})$ for the acid-denatured state of the 40-residue protein BBL. In these experiments, a rapid change in temperature with a near-IR laser pulse causes a change in the equilibrium distribution of distances between donor and acceptor fluorophores attached to the N and C termini. Again assuming a Gaussian chain, $\tau_{\text{corr}}(r) \approx 2\tau_{\text{corr}}(\text{FRET})$ [67] = 0.12 µs at 305 K. This would lead to a speed limit for this 40-residue protein of $2\pi\tau_{\text{corr}}(r) \approx 0.8$ µs. Using the linear length scaling suggested by the homopolymer collapse theory [57] (also note: as $\tau_{\text{corr}}(r) = \langle r^2 \rangle/3D_{\min}(r) = C_N N l^2/3D_{\min}$ for a Gaussian chain, $\tau_{\text{corr}}(r) \propto N$), the predicted speed limit from these experiments for a 100-residue protein is ∼2 µs. Interestingly, the FRET efficiency increases with increasing temperature, indicating that the polypeptide is becoming more compact, presumably as a result of stronger hydrophobic interactions at the elevated temperature. These experiments also show that $\tau_{\text{corr}}(r)$ first decreases with increasing temperature and then increases. This effect is similar to that previously mentioned, whereby increasing the stability of the native protein in a downhill folding scenario may increase the folding time by decreasing the diffusion coefficient for motion on the free energy surface [28].

An estimate of the speed limit from the dynamics in the denatured well can be made by determining $D_{min}(r)$ in experiments on the rate of contact formation. These experiments exploit the theory of Szabo *et al.* [68] on the diffusion-limited rate k_{D+} of end-to-end contact formation measured from quenching of the tryptophan triplet state by cysteine in disordered peptides ($k_{D+} = 4\pi D_{min}(r)a/(2\pi\langle r^2\rangle/3)^{3/2}$). The dependence of the rate on viscosity yields a diffusion coefficient $D_{min}(r)$ of $\sim 2 \times 10^{-6}$ cm^{2}s^{-1} at 293 K for end-to-end contact formation of the disordered peptide C(AGQ)$_3$W [40]. We can use the empirical equation of Millett *et al.* [69] for R_g to estimate $\langle r^2\rangle$, assuming $\langle r^2\rangle = 6R_g^2 = 0.2N^{1.2}$ nm^2, and find $\tau_{corr}(r) = \langle r^2\rangle/3D_{min}(r) \approx 85$ ns for a 100-residue protein and therefore a speed limit of ~ 0.5 µs. There is no information yet on how $D_{min}(r)$ scales with polypeptide length, but one would expect that, if there is a dependence, it would be to decrease $D_{min}(r)$ (to produce a longer folding time) because the number of interacting residues will increase. Also, $D_{min}(r)$ would be expected to be lower for folding sequences compared to the peptide C(AGQ)$_3$W.

Finally, a potential method for measuring $\tau_{corr}(r)$ is from fluctuations in the FRET efficiency at equilibrium on small numbers of molecules (correlation spectroscopy) or on single molecules. Schuler *et al.* [31,32,67,70] have measured the variance in the FRET efficiency distribution in 1 ms snapshots of single freely diffusing unfolded protein molecules (Csp*Tm* — a 68-residue cold shock protein). By comparing the width of the FRET efficiency distribution to that of polyproline labeled with the same dyes as a control, they found that $\tau_{corr}(r) < 200$ µs. More recent measurements (B Schuler *et al.*, unpublished) using a shorter observation period of 100 µs show that $\tau_{corr}(r) < 30$ µs, consistent with our above estimates (the limit of 30 µs, the measured folding time of 12 ms and the activation energy of 5 k_BT place limits on the free energy and entropy of activation for this protein of >4 k_BT and <1 k_B, respectively).

Ultrafast folders and simulations

The preceding analysis suggests that the speed limit for folding a generic single-domain protein with N residues is about $N/100$ µs, with α-helical proteins folding slightly faster and β or αβ proteins more slowly. Table 2 summarizes the data for proteins that have been measured to fold in <100 µs, as well as proteins for which folding times have been extrapolated to <100 µs (see Ivankov *et al.* [71] for a summary of known folding rates). The fastest folding of these proteins are α-helical, as expected from the data on structural elements. In α-helical structures, most of the stabilizing interactions are local, corresponding to hydrogen-bonding and sidechain interactions between residues separated by three residues along the sequence. In the simplest model of protein folding, a high density of local inter-residue interactions results in a larger compensation of the entropy loss as the chain becomes

ordered, and consequently a lower free energy barrier and faster folding [61,72].

Calculations of folding rates from molecular dynamics simulations have been carried out using distributed computing for three ultrafast folding proteins — BBA5 [21], the subdomain of villin [16,22] and the tryptophan cage [13,15] (references to simulation papers that do not calculate rates, as well as theoretical papers related to the specific protein are also given in Table 2, but discussion of these studies is beyond the scope of this review). The idea of distributed computing is to use the enormous amount of idle time available on PCs around the world to calculate independent trajectories. The calculation takes advantage of the fact that, for an exponential process, the fraction of molecules that fold in time t is simply t/τ (for $t \ll \tau$), where τ is the folding time. The first protein to be studied by both distributed computing and experiment is the 23-residue designed miniprotein BBA5, consisting of a helix and a hairpin connected by a turn containing a D-proline. Using tens of thousands of volunteer processors and an implicit solvent model to reduce the computation time, Snow *et al.* [21] observed 16 folding events, as judged by the similarity to the known native structure. The agreement between simulated and measured rates and equilibrium constants is surprisingly good, with a measured folding time using laser temperature jump of 7.5 ± 3.5 µs and a simulated time calculated from the fraction of successful 20 ns trajectories of 6 ($-3,+7$) µs (the uncertainties reflect different choices for the definition of a folded structure), and a measured equilibrium constant of 0.25 compared to the simulated value of ~ 1 obtained from the ratio of folding and unfolding rates. The folding mechanism is not complex, consisting essentially of the docking of the helix and hairpin, with the intrinsically more stable helix usually forming first. Although there is some question as to whether this mini-protein truly folded in these trajectories (e.g. by showing that the folded conformation persists and that its mean energy is the same as that found in simulations starting with the known folded structure), the simulation produced the extremely important result that molecular dynamics simulations may soon, if they have already not done so, fold ultrafast proteins knowing only the amino acid sequence.

Distributed computing has also been used to calculate the folding rate of another designed mini-protein — the 20-residue tryptophan cage. Snow *et al.* determined a folding time of 2–7 µs [13], again in excellent agreement with the measured time of 4 µs at room temperature obtained by Qui *et al.* using laser temperature jump [15].

The 35-residue villin subdomain is the smallest naturally occurring polypeptide that folds autonomously without disulfide bonds or cofactors, and has equilibrium properties comparable to those of much larger singledomain proteins [73]. A double mutant of the subdomain of the

villin headpiece folds in 1.7 ± 0.3 μs ([16]; J Kubelka *et al.*, unpublished). Because of its small size and rapid folding, it has been the focus of numerous computer simulations (see references in Table 2). Using distributed computing, Zagrovic *et al.* [22] observed many folding events for the villin subdomain, as judged by the similarity to the known native structure. They observed 35 folding events in 6000 $\sim$30 ns trajectories and therefore obtained a folding time of $5 (+11, -3)$ μs at 300 K, compared to the measured time using laser temperature jump of 4.3 ± 0.6 μs for the wild-type subdomain at 300 K [16].

The difficult part of analyzing simulations is to extract information concerning the folding mechanism, which is much more complex and therefore more interesting than that of BBA5. From examination of the successful trajectories, Zagrovic *et al.* [22] proposed that breaking the C-terminal phenylalanine–core interaction is the rate-limiting step in folding the protein and therefore that removal of the phenyl group would eliminate this misfolded trap and increase the folding rate. However, Kubelka *et al.* [16] found no change in either the equilibrium constant or folding rate on replacing this phenylalanine with alanine. There are at least two possible explanations for the apparent discrepancy, in addition to the obvious criticism that the force-fields are not yet sufficiently accurate for mechanistic studies. One is that the observed folding trajectories are not representative of the shortest trajectories of an exponential distribution [74,75], because the starting structures were not taken from an equilibrium distribution in the denatured state. A second possibility is that the trajectories are representative, but the analysis of the trajectories is incomplete in that the entire unfolded ensemble was not considered [16].

There have also been comparisons of experiments and all-atom molecular dynamics simulations for the FBP28 WW domain and the engrailed homeodomain. At very high temperatures (e.g. 498 K) proteins unfold in nanoseconds in simulations, making it possible to calculate unfolding trajectories using all-atom molecular dynamics including explicit solvent [24,25,76]. The most complete study has been carried out on the engrailed homeodomain [17]. The simulations suggest that the transition states are similar at several temperatures (348 K, 373 K and 498 K) above the folding temperature (325 K). However, the unfolded structures are rather different [17,25], implying that the microscopic pathways from the unfolded state to the transition state may also be significantly different. Evidence supporting the relative insensitivity of transition state structures to temperature comes from studies on a much slower folding protein, CI2, for which ϕ-values have been calculated from the ensemble of structures identified as transition state structures and are in remarkably good agreement with experimental values [77–79]. This agreement is surprising, as the hydrophobic effect, a dominant force in folding, is highly temperature dependent, so it will be important to understand why transition states appear to be insensitive to temperature. The reader should consult the recent reviews by Fersht and Daggett for further discussion of the comparison of experiments and unfolding simulations [24,25].

Current record holders: potential downhill folders

As pointed out in the Introduction to this review, at the speed limit the free energy barrier to folding may disappear to produce downhill folding (Figure 1), opening up the possibility of directly observing intermediate structures all along the reaction coordinate [28]. This immediately raises two related questions concerning the ultrafast folders listed in Table 2. Which proteins are intrinsically the fastest folders and how close is each protein to its theoretical speed limit? To answer these questions, we must normalize the folding times for size and stability, as small proteins fold faster than large proteins and increasing stability almost always decreases the folding time. Normalization for stability is straightforward, assuming a linear free energy relation between rate and equilibrium constant. Using the average exponent (i.e. the ϕ-value) of 1/3 derived from mutation studies [80], the data can be normalized with the relation $\tau_{\text{folding}} \propto K_{\text{eq}}^{-1/3}$.

Normalization for size is more problematic. Attempts have been made to extract the length dependence of folding rates from experimental data [71,81], but the exact functional form and parameters are uncertain because of the narrow range of sizes, and the dominance of the folding rate by topology (e.g. contact order) and stability [82]. For the present, we must therefore rely on results from simulations of simplified representations of proteins, and on theoretical arguments for the length dependence of both the folding time and the speed limit. These studies predict that scaling of protein folding times with size will vary from a polynomial dependence to an exponential dependence [71,81,83–88]. Wolynes [85] has proposed that the size dependence will depend on the relation of the experimental temperature to both the folding temperature T_f and the glass transition temperature T_g (the temperature at which the kinetics become sluggish due to deep energy traps; see the review by Onuchic and Wolynes, this issue), and is polynomial for fast folders where $T \ll T_f$ and $T \gg T_g$. As the temperatures for almost all of the proteins in Table 2 fall below the folding temperature, we assume polynomial scaling with a folding time proportional to N^3, which is close to what has been found in several studies using lattice representations of proteins [81,84,87]. Semi-empirical relations proposed by Li *et al.* with $\tau \propto \exp(1.1N^{1/2})$ or $\tau \propto \exp(0.36N^{2/3})$ give slightly different results [71,81], but our qualitative conclusions are not affected.

To estimate how the folding speed limit depends on size, we used a linear length scaling. A simple argument
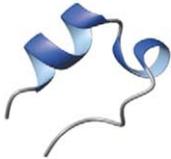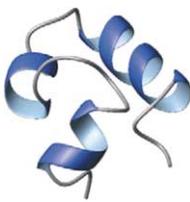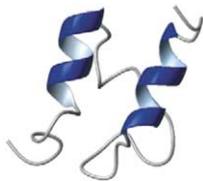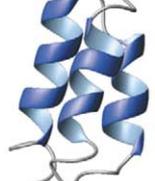
Table 2

Ultrafast folding proteins.

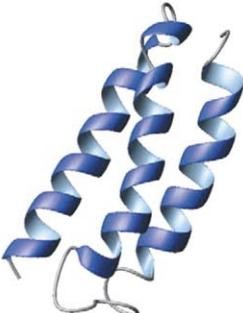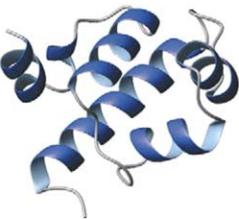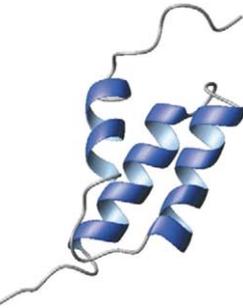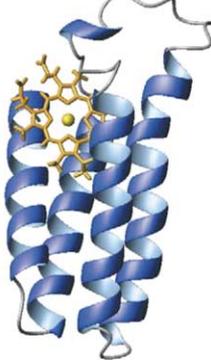| Protein name (PDB code) and mutation	Number residues (relative contact order)[a]	Structure	$\tau_{folding}$ (μs)	$\Delta G_{folding}$ (kcal/mol)	T (K)	Data reference	Simulation/theory references
Measured folding times							
Tryptophan cage (1L2Y)	20 (0.200)		**4.1**	−0.7	296	[15]	[13,23,101−106]
BBA5	23 (0.136)		**7.5**	+0.8	298	[21]	[21,107]
Villin headpiece subdomain (1VII)	35 (0.115)		**4.3**	−3.1	300	[16]	[22,91,108−119]
N68H/K65NIe	35		**1.7**	−3.6	300	b	
HP-36	36		**8**	−0.6	329	[18]	
WW domain Pin (1PIN)	34 (0.190)		**85**	−1.9	312	[11]	[120]
WW domain FBP28 (1E0L)	37 (0.170)		**∼30**[c]	−1.7	298	[12]	[120−122]
W30A			**24**	−1.2	298	[12]	
W30A			**47**	−0.8	313	[19]	
W30F			**19**	−2.1	313	[19]	
Peripheral subunit binding domain (2PDD)	41 (0.105)		**62**	−1.1	314	[8]	[91]
Albumin binding domain (1PRB)	47 (0.123)						
K5I			**2.5**	−2.6	350	[89]	[88,123]
K5I/K39V			**1**	−2.8	347		

Table 2 Continued

Protein name (PDB code) and mutation	Number residues (relative contact order)[a]	Structure	$\tau_{folding}$ (μs)	$\Delta G_{folding}$ (kcal/mol)	T (K)	Data reference	Simulation/theory references
Engrailed homeodomain (1ENH)	61 (0.121)		27	−2.1	298	[10]	[10,17,91]
α_3D (2A3D)	73 (0.096)		3	−1.9	323	[20]	[20]
λ-repressor (1LMB)	80 (0.092)		250	−3.0	310	[124]	[62,63,88, 125,126]
A37G			44	−0.4	330	[14]	
G46A/G48A			22	−1.0	335		
G46A/G48A/D14A			18	−1.5	335		
Extrapolated folding times[d]							
Protein A, B domain (1BDC)	58 (0.092)		8 (200)[e]	−4.3	310	[127]	[91,119,128–146]
Cytochrome b_{652} (1QQ3)	106 (0.075)		5 (670)[f]	−10	298	[147]	[88]
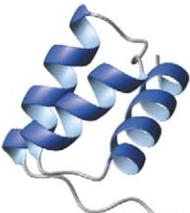
[a]The relative contact order (RCO), which is the mean separation in sequence by contacting atoms, is defined by $RCO \equiv \frac{1}{N\Lambda} \sum_{contacts} |i-j|$,

where N is the number of residues in the protein, Λ is the total number of atomic contacts and $|i-j|$ is the separation in sequence between residues i and j with distance between atoms less than 0.6 nm. [b]J Kubelka et al., unpublished. [c]Extrapolated from 2 M urea. [d]Measured folding time in parentheses. [e]In 2 M urea. [f]In 2.2 M GndHCl. Nle = norleucine (-HC$^\alpha$ -(CH$_2$)$_3$-CH$_3$).

justifying a linear N dependence, in addition to those given above for the collapse time and pre-exponential factor, is based on the assumption that, at the speed limit, folding proceeds down a steep free energy gradient (Figure 1). Using the number of ordered residues as the reaction coordinate, the time to move down a steep free energy hill is simply proportional to the length of the reaction coordinate and therefore the size of the protein, N, assuming also that the diffusion coefficient does not decrease with increasing size. A simple linear proportionality is also obtained with Q as the reaction coordinate, because for small proteins the number of native contacts is also approximately proportional to N.

Table 3 shows folding times normalized for size and stability, and the proximity of the measured folding time to the speed limit for each protein. The temperatures of the measured folding times are different (Table 2), but folding times are much less sensitive to temperature than unfolding times, and change little or often decrease with increasing temperature. The designed three-helix bundle α_3D [20] is the intrinsically fastest folding protein, with the naturally occurring three-helix bundle albumin binding domain [89] a close second (the same ranking is obtained with a weaker length dependence such as N^2). Although topology must play a role in determining the rate of the fastest folders, there is only a weak correlation between relative contact order and the logarithm of the folding times normalized for size and stability ($r = 0.67$, $p = 0.03$, not including data from extrapolated folding times). Another issue in comparing folding times is the amount of native structure that is already present in the unfolded state, which is not known for the three intrinsically fastest folding proteins. Engrailed homeodomain, for example, contains a considerable amount of helix in the

unfolded state [90], so folding involves mostly docking of prenucleated helices [17,25] rather than the more complex and slower process of organizing structure from a more random unfolded state without helices [6,53,91]. Similar studies of the structure of the denatured state for the other proteins in Table 2 will therefore be important for assessing the factors that contribute to ultrafast folding.

Table 3 shows the proximity to the speed limit as measured by the ratio of the measured folding time to the speed limit of $N/100$ μs. This analysis makes two interesting points. First, it shows that, for most of these ultrafast proteins, there is still considerable room to speed up folding, either by stabilizing the native state using protein engineering or by changing solvent conditions. In view of our expectation that proteins with β secondary structure will fold more slowly than α-helical proteins, the two WW domains are most probably closer to their speed limit than is indicated by our estimates. Second, the analysis predicts that three of the intrinsically fastest folding proteins are folding close to their speed limit — the 73-residue designed α_3D, the mutant of the 47-residue albumin binding domain and the mutant of the villin subdomain (if the extrapolated value of 5 μs for cytochrome b_{562} is confirmed by direct measurement, it should be added to this list). Each protein folds with exponential time courses, whereas one might expect non-exponential kinetics if the barrier is absent. However, downhill folding does not guarantee that non-exponential kinetics will be detected [92,93], raising the possibility that all three proteins may be folding without a barrier. It is interesting in this regard that most variations on the simple analytical model [61,94,95] predict barrierless folding for the villin subdomain and the albumin binding domain (α_3D has not been studied with these models) (ER Henry, WA Eaton, unpublished).

Table 3

Size- and stability-normalized folding times and proximity to speed limit.

Protein name (PDB code) and mutation	τ (normalized) (μs)[a]	$\dfrac{\tau_{(observed)}}{\tau_{(speed\,limit)}}$ [b]
From measured folding times		
α_3D (2A3D)	1	4
Albumin binding domain (1PRB) K5I/K39V	2	2
λ-repressor (1LMB) G46A/G48A/D14A	4	20
Villin headpiece subdomain (1VII) N68H/K65Nle	15	5
BBA5	20	30
Engrailed homeodomain (1ENH)	20	40
Tryptophan cage (1L2Y)	40	20
WW domain FBP28 (1E0L) W30F	60	50
Peripheral subunit binding domain (2PDD)	80	150
WW domain Pin (1PIN)	300	250
From extrapolated folding times		
Protein A, B domain (1BDC)	20	10
Cytochrome b_{562} (1QQ3)	60	5

[a]Calculated from $(50/N)^3 (K_{eq}/15)^{1/3} \tau$ (observed), where N is the number of residues, $K_{eq} = \exp(-\Delta G_{folding}/k_BT)$ and τ (observed) is $\tau_{folding}$, as given in Table 2. The average size for the ten proteins with measured folding times is $\sim$50 and the average stability, calculated as $\exp(-\langle\Delta G_{folding}/k_BT\rangle)$, is $\sim$15. [b]Calculated using τ (speed limit) = 0.01 N μs and the observed $\tau_{folding}$ given in Table 2.

Two recent experimental studies have directly addressed the issue of downhill folding. In laser temperature jump experiments, Yang *et al.* found that mutations of the 80-residue λ-repressor decrease the folding time to ∼20 μs and produce non-exponential kinetics and also introduce a new ∼2 μs kinetic phase [14]. The 2 μs phase is attributed to population of the region of the free energy barrier top, indicating that the barrier is very small and that the folding rate of these mutants is therefore approaching the speed limit. Interestingly, the analysis in Table 3 suggests that, of the ultrafast folding proteins, λ-repressor is folding at a rate relatively close to its estimated speed limit (with a remaining free energy barrier of only ∼3 k_BT), supporting the interpretation of Yang *et al.* [14].

Garcia-Mira *et al.* [29] have argued from equilibrium measurements that the 40-residue peripheral subunit binding domain from oxoglutarate dehydrogenase is a downhill folder. They coined the term 'one-state' folding to indicate that there is no barrier separating folded and unfolded states, but a continuum of structures from folded to unfolded. Their conclusion is based on the lack of coincidence in the temperature-dependent changes in circular dichroism, fluorescence and heat capacity. They also make the interesting speculation that downhill folders may be 'molecular rheostats', with biological function controlled by the distribution of partially unfolded conformations. No kinetic data were reported, but a close relative, the peripheral subunit binding domain from pyruvate dehydrogenase, is among the ultrafast folders (Table 2).

Conclusions

There are caveats to all of our estimates of a protein folding speed limit. There are still only limited data for basing a prediction on rates of secondary structure formation and the use of Kramers' reaction rate theory contains several assumptions that require additional testing. Nevertheless, both experimental and theoretical methods lead to very similar estimates of ∼N/100 μs for generic single-domain proteins, with α proteins folding faster than β or αβ proteins. Comparison of the measured folding times with the speed limits estimated for individual proteins indicates that there is still room to engineer most of the known ultrafast folding proteins to fold more than ten times faster. As we learn more about the relation between topology and folding mechanisms, we anticipate that the speed limit for each topology class will be somewhat different, as suggested by the difference in the rate of formation of α helices and β hairpins.

There are now several folding simulations of ultrafast folding proteins, with multiple trajectories from distributed computing that allow calculation of folding rates. There is still considerable information to be extracted from these trajectories, and we have just begun to see a glimmer of what can be learned about microscopic folding pathways from the unfolded to the folded state at atomic resolution. However, many more experimental tests must be made to assess the accuracy of these simulations. Nevertheless, it is clear that we are on the threshold of making major advances in our understanding of protein folding from a combination of experimental, theoretical and simulation studies of ultrafast folding proteins.

References

1. Jones CM, Henry ER, Hu Y, Chan CK, Luck SD, Bhuyan A, Roder H, Hofrichter J, Eaton WA: **Fast events in protein-folding initiated by nanosecond laser photolysis**. *Proc Natl Acad Sci USA* 1993, **90**:11860-11864.

2. Eaton WA, Muñoz V, Thompson PA, Chan CK, Hofrichter J: **Submillisecond kinetics of protein folding**. *Curr Opin Struct Biol* 1997, **7**:10-14.

3. Callender RH, Dyer RB, Gilmanshin R, Woodruff WH: **Fast events in protein folding: the time evolution of primary processes**. *Annu Rev Phys Chem* 1998, **49**:173-202.

4. Gruebele M: **The fast protein folding problem**. *Annu Rev Phys Chem* 1999, **50**:485-516.

5. Eaton WA, Muñoz V, Hagen SJ, Jas GS, Lapidus LJ, Henry ER, Hofrichter J: **Fast kinetics and mechanisms in protein folding**. *Annu Rev Biophys Biomol Struct* 2000, **29**:327-359.

6. Myers JK, Oas TG: **Mechanisms of fast protein folding**. *Annu Rev Biochem* 2002, **71**:783-815.

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

8. Spector S, Raleigh DP: **Submillisecond folding of the peripheral subunit-binding domain**. *J Mol Biol* 1999, **293**:763-768.

9. Crane JC, Koepf EK, Kelly JW, Gruebele M: **Mapping the transition state of the WW domain beta-sheet**. *J Mol Biol* 2000, **298**:283-292.

10. Mayor U, Johnson CM, Daggett V, Fersht AR: **Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation**. *Proc Natl Acad Sci USA* 2000, **97**:13518-13522.

11. Jager M, Nguyen H, Crane JC, Kelly JW, Gruebele M: **The folding mechanism of a beta-sheet: the WW domain**. *J Mol Biol* 2001, **311**:373-393.

12. Ferguson N, Johnson CM, Macias M, Oschkinat H, Fersht A: **Ultrafast folding of WW domains without structured aromatic clusters in the denatured state**. *Proc Natl Acad Sci USA* 2001, **98**:13002-13007.

13. Snow CD, Zagrovic B, Pande VS: **The Trp cage: folding kinetics and unfolded state topology via molecular dynamics simulations**. *J Am Chem Soc* 2002, **124**:14548-14549.

14. Yang WY, Gruebele M: **Folding at the speed limit**. *Nature* 2003, **423**:193-197.

15. Qiu LL, Pabit SA, Roitberg AE, Hagen SJ: **Smaller and faster: the 20-residue Trpcage protein folds in 4 μs**. *J Am Chem Soc* 2002, **124**:12952-12953.

16. Kubelka J, Eaton WA, Hofrichter J: **Experimental tests of villin subdomain folding simulations**. *J Mol Biol* 2003, **329**:625-630.

17. Mayor U, Guydosh NR, Johnson CM, Grossmann JG, Sato S, Jas GS, Freund SMV, Alonso DOV, Daggett V, Fersht AR: **The complete folding pathway of a protein from nanoseconds to microseconds**. *Nature* 2003, **421**:863-867.

18. Wang MH, Tang YF, Sato SS, Vugmeyster L, McKnight CJ, Raleigh DP: **Dynamic NMR line-shape analysis demonstrates that the villin headpiece subdomain folds on the microsecond time scale**. *J Am Chem Soc* 2003, **125**:6032-6033.

19. Nguyen H, Jager M, Moretto A, Gruebele M, Kelly JW: **Tuning the free-energy landscape of a WW domain by temperature, mutation, and truncation**. *Proc Natl Acad Sci USA* 2003, **100**:3948-3953.

20. Zhu Y, Alonso DOV, Maki K, Huang CY, Lahr SJ, Daggett V, Roder H, DeGrado WF, Gai F: **Ultrafast folding of α_3D: a *de novo* designed three-helix bundle protein**. *Proc Natl Acad Sci USA* 2003, **100**:15486-15491.

21. Snow CD, Nguyen N, Pande VS, Gruebele M: **Absolute comparison of simulated and experimental protein-folding dynamics**. *Nature* 2002, **420**:102-106.

22. Zagrovic B, Snow C, Shirts M, Pande V: **Simulation of folding of a small alphahelical protein in atomistic detail using worldwide distributed computing**. *J Mol Biol* 2002, **323**:927-937.

23. Zagrovic B, Pande V: **Solvent viscosity dependence of the folding rate of a small protein: distributed computing study**. *J Comput Chem* 2003, **24**:1432-1436.

24. Daggett V, Fersht AR: **Is there a unifying mechanism for protein folding?** *Trends Biochem Sci* 2003, **28**:18-25.

25. Fersht AR, Daggett V: **Protein folding and unfolding at atomic resolution**. *Cell* 2002, **108**:573-582.

26. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG: **Funnels, pathways, and the energy landscape of protein-folding - a synthesis**. *Proteins* 1995, **21**:167-195.

27. Sabelko J, Ervin J, Gruebele M: **Observation of strange kinetics in protein folding**. *Proc Natl Acad Sci USA* 1999, **96**:6031-6036.

28. Eaton WA: **Searching for 'downhill scenarios' in protein folding**. *Proc Natl Acad Sci USA* 1999, **96**:5897-5899.

29. Garcia-Mira MM, Sadqi M, Fischer N, Sanchez-Ruiz JM, Muñoz V: **Experimental identification of downhill protein folding**. *Science* 2002, **298**:2191-2195.

30. Jackson SE: **How do small single domain proteins fold?** *Fold Des* 1998, **3**:R81-R91.

31. Schuler B, Lipman EA, Eaton WA: **Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy**. *Nature* 2002, **419**:743-747.

32. Lipman EA, Schuler B, Bakajin O, Eaton WA: **Single-molecule measurement of protein folding kinetics**. *Science* 2003, **301**:1233-1235.

33. Rhoades E, Gussakovsky E, Haran G: **Watching proteins fold one molecule at a time**. *Proc Natl Acad Sci USA* 2003, **100**:3197-3202.

34. Fersht A: *Structure and Mechanism in Protein Science*. New York: WH Freeman and Company; 1998.

35. Bieri O, Wirz J, Hellrung B, Schutlowski M, Drewello M, Kiefhaber T: **The speed limit for protein folding measured by triplet-triplet energy transfer**. *Proc Natl Acad Sci USA* 1999, **96**:9597-9601.

36. Krieger F, Fierz B, Bieri O, Drewello M, Kiefhaber T: **Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding**. *J Mol Biol* 2003, **332**:265-274.

37. Huang F, Nau WM: **A conformational flexibility scale for amino acids in peptides**. *Angew Chem Int Ed Engl* 2003, **42**:2269-2272.

38. Lapidus LJ, Eaton WA, Hofrichter J: **Measuring the rate of intramolecular contact formation in polypeptides**. *Proc Natl Acad Sci USA* 2000, **97**:7220-7225.

39. Lapidus LJ, Eaton WA, Hofrichter J: **Dynamics of intramolecular contact formation in polypeptides: distance dependence of quenching rates in a roomtemperature glass**. *Phys Rev Lett* 2001, **87**:258101.

40. Lapidus LJ, Steinbach PJ, Eaton WA, Szabo A, Hofrichter J: **Effects of chain stiffness on the dynamics of loop formation in polypeptides. Appendix: testing a 1-dimensional diffusion model for peptide dynamics**. *J Phys Chem B* 2002, **106**:11628-11640.

41. Lapidus LJ, Eaton WA, Hofrichter J: **Measuring dynamic flexibility of the coil state of a helix-forming peptide**. *J Mol Biol* 2002, **319**:19-25.

42. Buscaglia M, Schuler B, Lapidus LJ, Eaton WA, Hofrichter J: **Kinetics of intramolecular contact formation in a denatured protein**. *J Mol Biol* 2003, **332**:9-12.

43. Rohl CA, Baldwin RL: **Deciphering rules of helix stability in peptides**. In *Energetics of Biological Macromolecules*, part B. Methods in Enzymology 1998, **295**:1-26.

44. Serrano L: **The relationship between sequence and structure in elementary folding units**. *Adv Protein Chem* 2000, **53**:49-85.

45. Scheraga HA, Vile JA, Ripoll DR: **Helix-coil transitions re-visited**. *Biophys Chem* 2002, **101**:255-265.

46. Thompson PA, Muñoz V, Jas GS, Henry ER, Eaton WA, Hofrichter J: **The helix-coil kinetics of a heteropeptide**. *J Phys Chem B* 2000, **104**:378-389.

47. Lacroix E, Kortemme T, de la Paz ML, Serrano L: **The design of linear peptides that fold as monomeric beta-sheet structures**. *Curr Opin Struct Biol* 1999, **9**:487-493.

48. Gellman SH: **Minimal model systems for beta sheet secondary structure in proteins**. *Curr Opin Chem Biol* 1998, **2**:717-725.

49. Muñoz V, Thompson PA, Hofrichter J, Eaton WA: **Folding dynamics and mechanism of beta-hairpin formation**. *Nature* 1997, **390**:196-199.

50. Gnanakaran S, Nymeyer H, Portman J, Sanbonmatsu KY, Garcia AE: **Peptide folding simulations**. *Curr Opin Struct Biol* 2003, **13**:168-174.

51. Xu Y, Oyola R, Gai F: **Infrared study of the stability and folding kinetics of a 15-residue β-hairpin**. *J Am Chem Soc* 2003, **125**:15388-15394.

52. Searle MS, Williams DH, Packman LC: **A short linear peptide derived from the N-terminal sequence of ubiquitin folds into a water-stable nonnative beta-hairpin**. *Nat Struct Biol* 1995, **2**:999-1006.

53. Karplus M, Weaver DL: **Protein-folding dynamics - the diffusion-collision model and experimental data**. *Protein Sci* 1994, **3**:650-668.

54. Steinfeld JI, Francisco JS, Hase WL: *Chemical Kinetics and Dynamics*, edn 2. Upper Saddle River, NJ: Prentice Hall; 1999.

55. de Gennes PG: **Kinetics of collapse for a flexible coil**. *Journal De Physique Lettres* 1985, **46**:L639-L642.

56. Pitard E, Orland H: **Dynamics of the swelling or collapse of a homopolymer**. *Europhys Lett* 1998, **41**:467-472.

57. Pitard E: **Influence of hydrodynamics on the dynamics of a homopolymer**. *Eur Phys J B* 1999, **7**:665-673.

58. Kramers HA: **Brownian motion in a field of force and the diffusion model of chemical reactions**. *Phys* 1940, **7**:284-304.

59. Berne BJ, Borkovec M, Straub JE: **Classical and modern methods in reaction-rate theory**. *J Phys Chem* 1988, **92**:3711-3725.

60. Socci ND, Onuchic JN, Wolynes PG: **Diffusive dynamics of the reaction coordinate for protein folding funnels**. *J Chem Phys* 1996, **104**:5860-5868.

61. Muñoz V, Eaton WA: **A simple model for calculating the kinetics of protein folding from three-dimensional structures**. *Proc Natl Acad Sci USA* 1999, **96**:11311-11316.

62. Portman JJ, Takada S, Wolynes PG: **Microscopic theory of protein folding rates. I. Fine structure of the free energy profile and folding routes from a variational approach**. *J Chem Phys* 2001, **114**:5069-5081.

63. Portman JJ, Takada S, Wolynes PG: **Microscopic theory of protein folding rates. II. Local reaction coordinates and chain dynamics**. *J Chem Phys* 2001, **114**:5082-5096.

64. Shea JE, Brooks CL: **From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding**. *Annu Rev Phys Chem* 2001, **52**:499-535.

65. Brooks CL: **Protein and peptide folding explored with molecular simulations**. *Acc Chem Res* 2002, **35**:447-454.

66. Sadqi M, Lapidus LJ, Muñoz V: **How fast is protein hydrophobic collapse?** *Proc Natl Acad Sci USA* 2003, **100**:12117-12122.

67. Gopich IV, Szabo A: **Single-macromolecule fluorescence resonance energy transfer and free-energy profiles**. *J Phys Chem B* 2003, **107**:5058-5063.

68. Szabo A, Schulten Z, Schulten K: **1st passage time approach to diffusion controlled reactions**. *J Chem Phys* 1980, **72**:4350-4357.

69. Millett IS, Doniach S, Plaxco KW: **Toward a taxonomy of the denatured state: Small angle scattering studies of unfolded proteins**. *Adv Protein Chem* 2002, **62**:241-262.

70. Schuler B, Lipman EA, Eaton WA: **Corrigendum: Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy**. *Nature* 2003, **421**:94.

71. Ivankov DN, Garbuzynskiy SO, Alm E, Plaxco KW, Baker D, Finkelstein AV: **Contact order revisited: influence of protein size on the folding rate**. *Protein Sci* 2003, **12**:2057-2062.

72. Plaxco KW, Simons KT, Baker D: **Contact order, transition state placement and the refolding rates of single domain proteins**. *J Mol Biol* 1998, **277**:985-994.

73. Frank BS, Vardar D, Buckley DA, McKnight CJ: **The role of aromatic residues in the hydrophobic core of the villin headpiece subdomain**. *Protein Sci* 2002, **11**:680-687.

74. Paci E, Cavalli A, Vendruscolo M, Caflisch A: **Analysis of the distributed computing approach applied to the folding of a small beta peptide**. *Proc Natl Acad Sci USA* 2003, **100**:8217-8222.

75. Fersht AR: **On the simulation of protein folding by short time scale molecular dynamics and distributed computing**. *Proc Natl Acad Sci USA* 2002, **99**:14122-14125.

76. Daggett V, Levitt M: **Protein folding–unfolding dynamics**. *Curr Opin Struct Biol* 1994, **4**:291-295.

77. Daggett V, Li AJ, Itzhaki LS, Otzen DE, Fersht AR: **Structure of the transition state for folding of a protein derived from experiment and simulation**. *J Mol Biol* 1996, **257**:430-440.

78. Ladurner AG, Itzhaki LS, Daggett V, Fersht AR: **Synergy between simulation and experiment in describing the energy landscape of protein folding**. *Proc Natl Acad Sci USA* 1998, **95**:8473-8478.

79. Kazmirski SL, Wong KB, Freund SMV, Tan YJ, Fersht AR, Daggett V: **Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution**. *Proc Natl Acad Sci USA* 2001, **98**:4349-4354.

80. Sanchez IE, Kiefhaber T: **Origin of unusual ϕ-values in protein folding: evidence against specific nucleation sites**. *J Mol Biol* 2003, **334**:1077-1085.

81. Li MS, Klimov DK, Thirumalai D: **Thermal denaturation and folding rates of single domain proteins: size matters**. *Polymers* 2004, in press.

82. Dinner AR, So SS, Karplus M: **Statistical analysis of protein folding kinetics**. In *Computational Methods for Protein Folding*. Advances in Chemical Physics 2002, **120**:1-34.

83. Thirumalai D: **From minimal models to real proteins - time scales for protein folding kinetics**. *J Phys I* 1995, **5**:1457-1467.

84. Gutin AM, Abkevich VI, Shakhnovich EI: **Chain length scaling of protein folding time**. *Phys Rev Lett* 1996, **77**:5433-5436.

85. Wolynes PG: **Folding funnels and energy landscapes of larger proteins within the capillarity approximation**. *Proc Natl Acad Sci USA* 1997, **94**:6170-6175.

86. Finkelstein AV, Badretdinov AY: **Rate of protein folding near the point of thermodynamic equilibrium between the coil and the most stable chain fold**. *Fold Des* 1997, **2**:115-121.

87. Cieplak M, Hoang TX, Li MS: **Scaling of folding properties in simple models of proteins**. *Phys Rev Lett* 1999, **83**:1684-1687.

88. Koga N, Takada S: **Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model**. *J Mol Biol* 2001, **313**:171-180.

89. Wang T, Zhu Y, Gai F: **Folding of a three-helix bundle at the folding speed limit**. *J Am Chem Soc* 2004, in press.

90. Mayor U, Grossmann JG, Foster NW, Freund SMV, Fersht AR: **The denatured state of engrailed homeodomain under denaturing and native conditions**. *J Mol Biol* 2003, **333**:977-991.

91. Islam SA, Karplus M, Weaver DL: **Application of the diffusion-collision model to the folding of three-helix bundle proteins**. *J Mol Biol* 2002, **318**:199-215.

92. Hagen SJ, Eaton WA: **Two-state expansion and collapse of a polypeptide**. *J Mol Biol* 2000, **301**:1019-1027.

93. Hagen SJ: **Exponential decay kinetics in 'downhill' protein folding**. *Proteins* 2003, **50**:1-4.

94. Alm E, Baker D: **Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures**. *Proc Natl Acad Sci USA* 1999, **96**:11305-11310.

95. Galzitskaya OV, Finkelstein AV: **A theoretical search for folding/unfolding nuclei in three-dimensional protein structures**. *Proc Natl Acad Sci USA* 1999, **96**:11299-11304.

96. Chang IJ, Lee JC, Winkler JR, Gray HB: **The protein-folding speed limit: intrachain diffusion times set by electron-transfer rates in denatured Ru(NH3)(5)(His-33)-Zn-cytochrome c**. *Proc Natl Acad Sci USA* 2003, **100**:3838-3840.

97. Wang T, Du DG, Gai F: **Helix-coil kinetics of two 14-residue peptides**. *Chem Phys Lett* 2003, **370**:842-848.

98. Werner JH, Dyer RB, Fesinmeyer RM, Andersen NH: **Dynamics of the primary processes of protein folding: helix nucleation**. *J Phys Chem B* 2002, **106**:487-494.

99. Huang CY, Getahun Z, Zhu YJ, Klemke JW, DeGrado WF, Gai F: **Helix formation via conformation diffusion search**. *Proc Natl Acad Sci USA* 2002, **99**:2788-2793.

100. Lednev IK, Karnoup AS, Sparrow MC, Asher SA: **Alpha-helix peptide folding and unfolding activation barriers: a nanosecond UV resonance Raman study**. *J Am Chem Soc* 1999, **121**:8074-8086.

101. Pitera JW, Swope W: **Understanding folding and design: replica-exchange simulations of 'Trp-cage' fly miniproteins**. *Proc Natl Acad Sci USA* 2003, **100**:7587-7592.

102. Chowdhury S, Lee MC, Xiong GM, Duan Y: ***Ab initio* folding simulation of the Trp-cage mini-protein approaches NMR resolution**. *J Mol Biol* 2003, **327**:711-717.

103. Simmerling C, Strockbine B, Roitberg AE: **All-atom structure prediction and folding simulations of a stable protein**. *J Am Chem Soc* 2002, **124**:11258-11259.

104. Zhou RH: **Trp-cage: folding free energy landscape in explicit water**. *Proc Natl Acad Sci USA* 2003, **100**:13280-13285.

105. Carnevali P, Toth G, Toubassi G, Meshkat SN: **Fast protein structure prediction using Monte Carlo simulations with modal moves**. *J Am Chem Soc* 2003, **125**:14244-14245.

106. Schug A, Herges T, Wenzel W: **Reproducible protein folding with the stochastic tunneling method**. *Phys Rev Lett* 2003, **91**:158102.

107. Wang L, Duan Y, Shortle R, Imperiali B, Kollman PA: **Study of the stability and unfolding mechanism of BBA1 by molecular dynamics simulations at different temperatures**. *Protein Sci* 1999, **8**:1292-1304.

108. Srinivas G, Bagchi B: **Study of pair contact formation among hydrophobic residues in a model HP-36 protein: relationship between contact order parameter and rate of folding and collapse**. *J Phys Chem B* 2003, **107**:11768-11773.

109. Mukherjee A, Bagchi B: **Correlation between rate of folding, energy landscape, and topology in the folding of a model protein HP-36**. *J Chem Phys* 2003, **118**:4733-4747.

110. Fernandez A, Shen MY, Colubri A, Sosnick TR, Berry RS, Freed KF: **Large-scale context in protein folding: villin headpiece**. *Biochemistry* 2003, **42**:664-671.

111. Pande VS, Baker I, Chapman J, Elmer SP, Khaliq S, Larson SM, Rhee YM, Shirts MR, Snow CD, Sorin EJ *et al.*: **Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing**. *Biopolymers* 2003, **68**:91-109.

112. Srinivas G, Bagchi B: **Foldability and the funnel of HP-36 protein sequence: use of hydropathy scale in protein folding**. *J Chem Phys* 2002, **116**:8579-8587.

113. Srinivas G, Bagchi B: **Folding and unfolding of chicken villin headpiece: energy landscape of a single-domain model protein**. *Curr Sci* 2002, **82**:179-185.

114. Shen MY, Freed KF: **All-atom fast protein folding simulations: the villin headpiece**. *Proteins* 2002, **49**:439-445.

115. Sullivan DC, Kuntz ID: **Protein folding as biased conformational diffusion**. *J Phys Chem B* 2002, **106**:3255-3262.

116. Duan Y, Kollman PA: **Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution**. *Science* 1998, **282**:740-744.

117. Duan Y, Wang L, Kollman PA: **The early stage of folding of villin headpiece subdomain observed in a 200-nanosecond fully solvated molecular dynamics simulation**. *Proc Natl Acad Sci USA* 1998, **95**:9897-9902.

118. Sullivan DC, Kuntz ID: **Conformation spaces of proteins**. *Proteins* 2001, **42**:495-511.

119. Jang SM, Kim E, Shin S, Pak Y: ***Ab initio* folding of helix bundle proteins using molecular dynamics simulations**. *J Am Chem Soc* 2003, **125**:14841-14846.

120. Karanicolas J, Brooks CL: **The structural basis for biphasic kinetics in the folding of the WW domain from a formin-binding protein: lessons for protein design?** *Proc Natl Acad Sci USA* 2003, **100**:3954-3959.

121. Ibragimova GT, Wade RC: **Stability of the beta-sheet of the WW domain: a molecular dynamics simulation study**. *Biophys J* 1999, **77**:2191-2198.

122. Ferguson N, Pires JR, Toepert F, Johnson CM, Pan YP, Volkmer-Engert R, Schneider-Mergener J, Daggett V, Oschkinat H, Fersht A: **Using flexible loop mimetics to extend Phi-value analysis to secondary structure interactions**. *Proc Natl Acad Sci USA* 2001, **98**:13008-13013.

123. Takada S: **Protein folding simulation with solvent-induced force field: folding pathway ensemble of three-helix-bundle proteins**. *Proteins* 2001, **42**:85-98.

124. Burton RE, Huang GS, Daugherty MA, Fullbright PW, Oas TG: **Microsecond protein folding through a compact transition state**. *J Mol Biol* 1996, **263**:311-322.

125. Portman JJ, Takada S, Wolynes PG: **Variational theory for site resolved protein folding free energy surfaces**. *Phys Rev Lett* 1998, **81**:5237-5240.

126. Burton RE, Myers JK, Oas TG: **Protein folding dynamics: quantitative comparison between theory and experiment**. *Biochemistry* 1998, **37**:5337-5343.

127. Myers JK, Oas TG: **Preorganized secondary structure as an important determinant of fast protein folding**. *Nat Struct Biol* 2001, **8**:552-558.

128. Linhananta A, Zhou YQ: **The role of sidechain packing and native contact interactions in folding: discontinuous molecular dynamics folding simulations of an all-atom G(o)over-bar model of fragment B of Staphylococcal protein A**. *J Chem Phys* 2002, **117**:8983-8995.

129. Zhou YQ, Linhananta A: **Thermodynamics of an all-atom off-lattice model of the fragment B of Staphylococcal protein A: implication for the origin of the cooperativity of protein folding**. *J Phys Chem B* 2002, **106**:1481-1485.

130. Favrin G, Irback A, Wallin S: **Folding of a small helical protein using hydrogen bonds and hydrophobicity forces**. *Proteins* 2002, **47**:99-105.

131. Kussell E, Shimada J, Shakhnovich EI: **A structure-based method for derivation of all-atom potentials for protein folding**. *Proc Natl Acad Sci USA* 2002, **99**:5343-5348.

132. Alonso DOV, Dagget OV, Daggett V: **Staphylococcal protein A: unfolding pathways, unfolded states, and differences between the B and E domains**. *Proc Natl Acad Sci USA* 2000, **97**:133-138.

133. Zhou YQ, Karplus M: **Interpreting the folding kinetics of helical proteins**. *Nature* 1999, **401**:400-403.

134. Zhou YQ, Karplus M: **Folding of a model three-helix bundle protein: a thermodynamic and kinetic analysis**. *J Mol Biol* 1999, **293**:917-951.

135. Kolinski A, Galazka W, Skolnick J: **Monte Carlo studies of the thermodynamics and kinetics of reduced protein models: application to small helical, beta, and alpha/beta proteins**. *J Chem Phys* 1998, **108**:2608-2617.

136. Zhou YQ, Karplus M: **Folding thermodynamics of a model three-helix-bundle protein**. *Proc Natl Acad Sci USA* 1997, **94**:14429-14432.

137. Guo ZY, Brooks CL, Boczko EM: **Exploring the folding free energy surface of a three-helix bundle protein**. *Proc Natl Acad Sci USA* 1997, **94**:10161-10166.

138. Olszewski KA, Kolinski A, Skolnick J: **Folding simulations and computer redesign of protein A three-helix bundle motifs**. *Proteins* 1996, **25**:286-299.

139. Boczko EM, Brooks CL: **First-principles calculation of the folding free-energy of a 3-helix bundle protein**. *Science* 1995, **269**:393-396.

140. Kolinski A, Skolnick J: **Monte-Carlo simulations of protein-folding. 2. Application to protein-a, Rop, and Crambin**. *Proteins* 1994, **18**:353-366.

141. Shea JE, Onuchic JN, Brooks CL: **Exploring the origins of topological frustration: Design of a minimally frustrated model of fragment B of protein A**. *Proc Natl Acad Sci USA* 1999, **96**:12512-12517.

142. Berriz GF, Shakhnovich EI: **Characterization of the folding kinetics of a threehelix bundle protein via a minimalist Langevin model**. *J Mol Biol* 2001, **310**:673-685.

143. Hu LH, Wang J, Wang W: **Folding simulations for a three-helix bundle protein**. *Int J Mod Phys B* 2003, **17**:49-54.

144. Garcia AE, Onuchic JN: **Folding a protein in a computer: an atomic description of the folding/unfolding of protein A**. *Proc Natl Acad Sci USA* 2003, **100**:13898-13903.

145. Vila JA, Ripoll DR, Scheraga HA: **Atomically detailed folding simulation of the B domain of staphylococcal protein A from random structures**. *Proc Natl Acad Sci USA* 2003, **100**:14812-14816.

146. Ghosh A, Elber R, Scheraga HA: **An atomically detailed study of the folding pathways of protein A with the stochastic difference equation**. *Proc Natl Acad Sci USA* 2002, **99**:10394-10398.

147. Wittung-Stafshede P, Lee JC, Winkler JR, Gray HB: **Cytochrome *b*(562) folding triggered by electron transfer: approaching the speed limit for formation of a four-helix-bundle protein**. *Proc Natl Acad Sci USA* 1999, **96**:6587-6590.