How do small single-domain proteins fold?

Sophie E Jackson

Many small, monomeric proteins fold with simple two-state kinetics and show wide variation in folding rates, from microseconds to seconds. Thus, stable intermediates are not a prerequisite for the fast, efficient folding of proteins and may in fact be kinetic traps and slow the folding process. Using recent studies, can we begin to search for trends which may lead to a better understanding of the protein folding process?

Address: Chemical Laboratory, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK.
E-mail: sej13@cam.ac.uk

Folding & Design 01 August 1998, **3**:R81–R91
http://biomednet.com/elecref/13590278003R0081

For over 30 years many groups have tried to understand the principles that govern how a linear amino acid polypeptide chain folds into a unique three-dimensional structure. An understanding of this fundamental process would help in attempts to predict structure from sequence, in the rational design of proteins *de novo*, and in understanding how and why proteins misfold. Since Levinthal [1] first put forward the idea that proteins cannot fold by a random search of conformational space, researchers have been trying to characterise the pathways by which proteins fold. As a result, the field was dominated for many years by the study of proteins that folded slowly through intermediate states that were sufficiently stable to be characterised [2,3]. In 1991, it was demonstrated that stable intermediates were not prerequisites for fast folding: a small protein, chymotrypsin inhibitor 2 (CI2), could fold rapidly without populating any intermediate states [4]. Since then, many small proteins, with differing structures, stabilities and sequences, have been shown to fold with simple two-state kinetics (see Figure 1 and Table 1; [5–31]). In the past two years there have been many reports of such systems and we are now in a position to construct a database of such proteins and search for general features and correlations.

In this review, kinetic and thermodynamic data for the folding of proteins that fold with simple, two-state kinetics are summarised and compared with data for proteins that fold via stable intermediates and with data for dimeric proteins. Some general trends and correlations begin to emerge between the structural and thermodynamic properties of proteins and their folding kinetics.

Two-state behaviour

Two-state systems are the simplest models of protein folding. In these cases, only the unfolded state (U) and the folded, native state (F) are populated on the folding pathway. Both unfolding and refolding (taking into account heterogeneity in the unfolded state resulting from proline isomerisation) are monophasic processes.

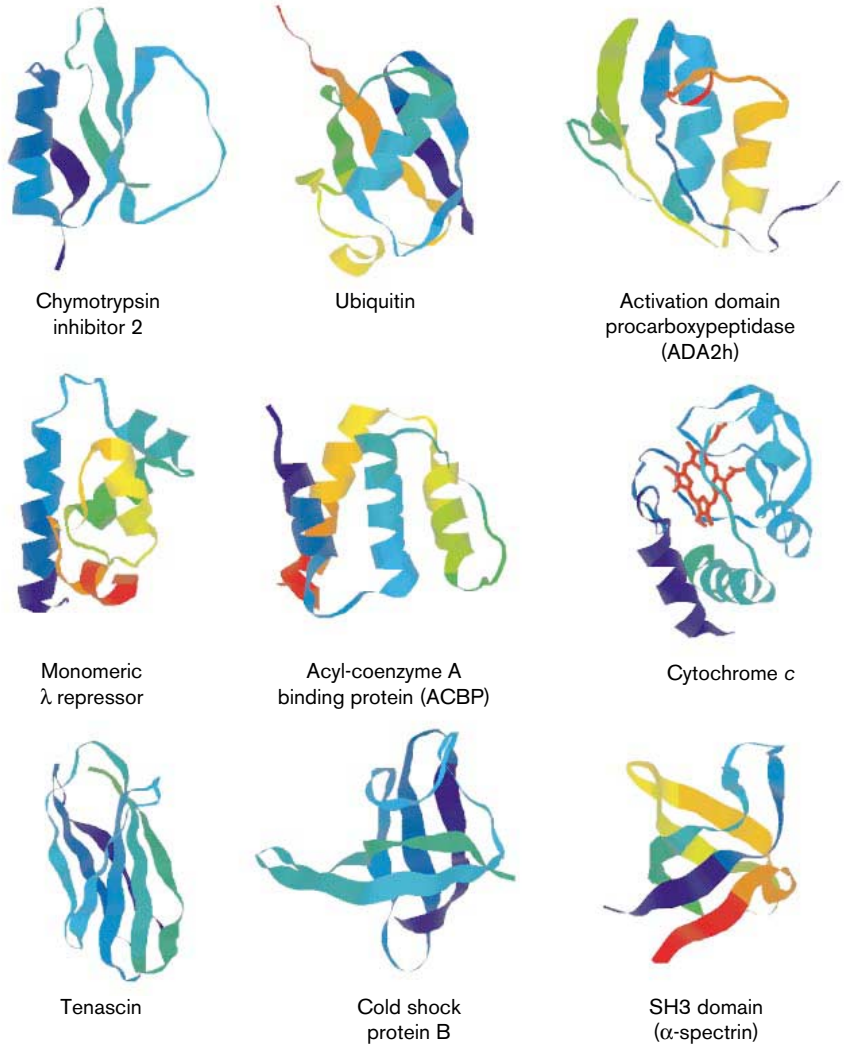
The criteria by which a protein can be shown to fold according to a two-state model have been well-documented [4]. For such proteins plots of the natural logarithm of the rate constants for unfolding, $\ln k_U$, and refolding, $\ln k_F$, versus denaturant concentration, [D], are frequently linear. More importantly, the thermodynamic parameters for the change in free energy of unfolding in the absence of denaturant, $\Delta G_{U-F}{}^{H_2O}$, and m, a constant related to the average fractional change in degree of exposure of residues on unfolding, calculated from the kinetic data, agree with those obtained directly from equilibrium data.

For a number of proteins, which have been shown to fold with two-state kinetics, plots of $\ln k_U$ versus denaturant concentration show slight deviations from linearity. This has been attributed to movements in the position of the transition state with denaturant concentration [32–34], but may also result from the fact that the denaturing activity of urea and guanidine hydrochloride (GdnHCl) does not depend linearly on denaturant concentration [35]. Non-linearity in plots of $\ln k_F$ versus denaturant concentration can result from the presence of a meta-stable intermediate on or off the folding pathway, but non-linearity may also be observed for proteins that fold without populating intermediate states, for the reasons given above. In these cases, the protein can still be shown to fold with two-state kinetics from a comparison of $\Delta G_{U-F}{}^{H_2O}$ and m obtained from kinetic and equilibrium data.

For other proteins, which have been shown to fold with two-state kinetics, burst kinetics, which are normally attributed to the formation of intermediate states, have been observed. Although this appears to be inconsistent with the kinetic data, it has been shown recently that these burst phases can arise from a change in the unfolded state between highly denaturing conditions and native conditions, and may not represent a structured intermediate state [9,26].

Table 1 shows the kinetic and thermodynamic data for the folding of proteins that have been shown to fold with two-state kinetics according to the criteria above. In several cases, conditions can be found such that proteins that normally fold through a populated intermediate state can fold with two-state kinetics. Mutations that destabilise the intermediate state relative to the unfolded state can switch the kinetics from three-state to two-state, as

Figure 1



Structures of nine proteins that fold with two-state kinetics (see Table 1).

with ubiquitin [25]. Changes in pH or temperature can have a similar effect [36,37]. Some of these examples are also included in Table 1. Care needs to be taken with some of these data. A protein can only be said to fold with two-state kinetics at the minimum concentration of denaturant for which the rate of folding has been measured. Although the rate of refolding has been measured at very low concentrations of denaturant, or in the absence of denaturant for many of the proteins in Table 1, some have only been measured in moderate concentrations of denaturant, which will tend to destabilise intermediates if present. For this reason, the lowest concentration of denaturant for which folding rates have been measured, $[D]_{min}$, is also given.

It is clear from Table 1 that, despite the fact that these are all small, monomeric proteins that fold with simple two-state kinetics, there is a large range in both the rates with which these proteins fold, and in the position of the transition state on the reaction coordinate (β_T). Thus, even for the simplest possible systems we observe an enormous diversity in folding behaviour.

Many theoretical studies have attempted to identify the most important factors in determining how a protein folds. Some have suggested that chain length is a crucial determinant in the rate of folding, whilst others have suggested that topology or stability are more important ([38]; see also [39], a recent review that provides an excellent introduction to the theoretical approaches used). A number of experimental studies have also attempted to address these issues. Plaxco *et al.* [38] have suggested that topology is the most important determinant, whilst studies of horse and yeast cytochrome *c* have suggested that stability is the

Table 1

Kinetic and thermodynamic parameters for the folding of small, monomeric proteins that fold with two-state kinetics.

Protein	References	PDB code	Contact order	Chain length	Structure	Denaturant	$[D]_{min}$ (M)	$\Delta G_{U-F}^{H_2O}$ (kcal mol^{-1})	m (kcal mol^{-1}M^{-1})	$[D]_{50\%}$ (M)	Temp. (°C)	$k_F^{H_2O}$ (s^{-1})	$m^{\ddagger-U}$ (kcal mol^{-1} M^{-1})	$k_U^{H_2O}$ (s^{-1})	$m^{\ddagger-F}$ (kcal mol^{-1} M^{-1})	[a]β_T
α-Helical proteins																
Monomeric λ repressor (λ_{6-85})																
wild type	[5,6]	1LMB	9.4	80	α	Urea	0	[b]3.0	1.1	2.7	37	[c]4900	0.45	30	0.70	0.39
Gly46→Ala,Gly48→Ala							2	4.8		4.4		88,000	0.89	36	0.18	0.83
[d]ACBP																
bovine	[7,8]	2ABD	14.0	86	α	GdnHCl	0.5	7.1	3.0	2.3	5	279	2.16	0.0001	1.40	0.61
rat							0.5	6.1	3.4	1.8		395	2.31	0.00676	1.06	0.69
yeast							2.2	6.8	2.0	3.8		4105	1.18	0.0015	0.88	0.57
[e]bovine							0.5	7.0			20	704	2.21	0.003	1.30	0.63
[f]Cytochrome *c*	[9]	1HRC	11.4	104	α	GdnHCl	0.25	6.9	2.4	2.9	20	2800	1.13	0.017	1.29	0.47
Cytochrome *c*																
Horse – oxidized (Fe^{III})	[10]	1HRC	11.4	104	α	GdnHCl	2.6	17.7	3.3	5.4	23	400	1.34	nd	nd	0.40
Yeast – oxidized (Fe^{II})		1YCC	11.6	103			0.8	14.6	3.1	4.7	23	15000	1.36			0.34
β Proteins																
Tendamistat	[11]	2AIT	21.6	74	β (2 S–S)	GdnHCl	0.5	8.1	1.3	6.2	25	67	0.83	4.5×10^{-5}	0.45	[g]0.65
Cold shock proteins																
CspB (*Bacillus subtilis*)	[12]	1CSP	16.4	67	β-barrel	Urea	0.5	3.0	0.76	3.9	25	1070	0.57	12	0.10	0.85
		1NMG	14.5													
CspB (*B. subtilis*)	[13]			67	β-barrel	GdnHCl	0.25	2.7	1.8	1.5	25	689	1.67	9.9	0.16	0.91
CspB (*B. caldolyticus*)	[13]			66	β-barrel	GdnHCl	0.5	4.8	1.8	2.7	25	1370	1.63	0.64	0.13	0.93
CspB (*Thermotoga maritima*)	[13]			68	β-barrel	GdnHCl	0.5	6.3	1.9	3.3	25	565	1.57	0.018	0.25	0.86
CspA	[14]	1MEF	17.7	69	β-barrel	Urea	0.75	3.0	0.7	4.2	25	200	0.7	4.2	0.05	0.94
		1MJC	16.0					2.9	0.57	4.9	10	188	0.69	3.3	0.07	0.91
SH3 domains																
SH3 domain (α-spectrin)	[15,16]	1AEY	19.9	62	β-barrel	Urea	0.2	2.9	0.8	3.9	25	8.1	0.57	0.045	0.20	0.74
		1SHG	19.1									4.1	0.53	0.0067	0.24	[h]0.69
SH3 domain (src)	[17]	1SRL	19.6	64	β-barrel	GdnHCl	0.25	4.1	1.6	2.6	22	57	0.99	0.1	[i]0.45	0.69
		1SRM	19.0													
SH3 domain (PI3 kinase)	[18]	1PKS	20.0	84	β-barrel	GdnHCl	0.15	3.4	2.3	1.5	20	0.35	1.42	0.00067	0.93	0.60
SH3 domain (fyn)	[19]	1AON	17.1	67	β-barrel	GdnHCl	0.75	6.0	1.4	4.3	20	94	1.09	0.00099	0.52	0.68
		1NYF	18.3													
		1SHF-A	18.3													
β-Sandwich domains																
[g]FN-III	[20]	1FNF	18.1	90	β-sandwich	GdnHCl	0.1	1.2	3	0.4	25	0.4	1.9	nd	nd	0.63
[j]Twitchin			19.7	93	β-sandwich	Urea	0	3.9	1.3	3.1	20	1.50	0.81	0.000203	0.35	0.70
Tenascin																
(short form)	[21]	1TEN	17.1	90	β-sandwich	Urea	0.2	4.8	1.3	3.8	20	2.9	0.82	0.0028	0.26	0.76
(long form)	[31]							6.7	1.3	5.3	25	6.0	0.92	0.000072	0.31	0.75
[k]10FN-III	[20]	1FNF	16.5	94	β-sandwich	GdnHCl	0.6	6.1	1.4	4.4	25	155	0.9	nd	nd	0.65
α/β Proteins																
CI2	[4]	1COA	15.5	64	α/β	GdnHCl	0	7.0	1.8	3.9	25	48	1.12	0.00018	0.73	0.61
		2CI2	15.3													
Activation domain procarboxypeptidaseA2 (ADAh2)	[22]	1PBA	16.6	80	α/β	Urea	1	4.1	1.0	4.2	25	897	0.75	0.65	0.26	0.74
Arc repressor (single chain)	[23]	1ARR		[l]106	α/β	Urea	0.8	6.3	1.3	4.6	25	10600	0.95	1.5	0.29	0.77
		1PAR														
[m]Ubiquitin											8	264	1.4	0.0004	0.8	0.64
wild type	[24]	1UBQ	15.1	76	α/β	GdnHCl	0.5	[n]7.1	1.9	3.7	25	1532	1.3	0.000437	0.9	0.59
Val26→Ala	[25]	1UBQ	15.1	76	α/β	GdnHCl		3.9	2.0	1.9	25	102	1.47	0.08	0.63	0.70
IgG binding domain of streptococcal protein L	[26]	2PTL	17.7	62	α/β	GdnHCl	0.2	4.6	1.9	2.4	22	60	1.5	0.02	0.5	0.75
Spliceosomal protein U1A	[27]	1URN	16.9	102	α/β	GdnHCl	0.5	9.3	2.3	4.1	25	316	[o]0.63	0.000063	[o]0.52	[o]0.55
Hpr (histidine-containing phosphocarrier protein)	[28]	1HDN	18.4	85	α/β	GdnHCl	0.5	4.6	2.2	2.1	20	14.9	1.65	0.0021	0.92	0.64
[p]FKBP12		1FKB	17.8	107	α/β	Urea	0	5.5	1.4	3.9	25	4.3	1.11	0.00017	0.54	0.67
Muscle AcP	[29]	1APS	21.7	98	α/β	Urea	0.2	5.4	1.25	4.3	28	0.23	0.96	0.00011	0.25	0.79
Villin 14T	[30]	1VIK	12.3	126	α/β	Urea	1.5	6.2	1.6	3.9	37	900	1.11	0.061	0.41	0.73

[a]See Equation 1. [b]The value for $\Delta G_{U-F}^{H_2O}$ is the average from CD and NMR experiments [5]. [c]This value is different to the value of 3600 s^{-1} used by Plaxco *et al.* [38], which is taken from [5] and involves a long extraploation from 1.4 M urea to H_2O. This value is taken from a more recent paper and involves a much shorter extrapolation to obtain $k_F^{H_2O}$ [6]. [d]Values are taken from [8] not [7]. The equilibrium values are averaged. [e]Birthe Kragelund, personal communication. [f]Bill Eaton, personal communication. [g]Values estimated from Figure 5 of [11]. [h]$\beta_T = 0.75$ if calculated using $m^{\ddagger-F}/m$. [i]Calculated using $\ln k_U = \ln k_U^{H_2O} - m^{\ddagger-F} - 0.014[\text{urea}]^2$ [16]. [j]S.J. Hamill, E. Cota, C. Chothia and J. Clarke, unpublished observations. [k]Detailed studies on the folding of [10]FN-III using GdnSCN as a denaturant suggest that the folding of this domain is not two-state (S.J. Hamill, E. Cota, C. Chothia and J. Clarke, unpublished observations). It has not been included in the analyses presented in this paper. [l]The single-chain Arc repressor has a 15 amino acid glycine-rich linker region which is not expected to be involved in structure or stability, the chain length given here is twice that of the monomer [23]. [m]Initial reports suggested that wild-type ubiquitin folded with two-state kinetics [24]. Subsequent studies over a wider range of denaturant concentrations showed that wild-type ubiquitin folds with three-state kinetics [25], the mutants Val26→Ala and Val26→Gly (in the presence of 0.4 M Na_2SO_4) fold with two-state kinetics. [n]Equilibrium data are the average of NMR, CD and fluorescence experiments. [o]There is some curvature in the plots of $\log k_F$ and $\log k_U$ versus [GdnHCl] such that the refolding rate is given by $\log k_F = 2.50 - 0.36[\text{GdnHCl}] - 0.070[\text{GdnHCl}]^2$, and the unfolding rate constant is defined as $\log k_U = -4.2 + 1.25[\text{GdnHCl}] - 0.061[\text{GdnHCl}]^2$, the values for $m^{\ddagger-F}$ and $m^{\ddagger-U}$ in the table are the slopes calculated at 4.0 M GdnHCl, the midpoint of unfolding from the equilibrium studies. It should be noted that the value for β_T decreases with decreasing [GdnHCl] such that at 0 M GdnHCl it is 0.26 and at high [GdnHCl] it is 0.84. [p]S.E.J., E.R.G. Main and K.F. Fulton, unpublished observations. nd, not determined.

most important factor (albeit for two proteins with very similar topology; [10]). A comparison of the results for four β-sandwich domain proteins (Table 1, ^{10}FN-III is not included because there is some evidence that this protein does not fold with two-state kinetics; S.J. Hamill, E. Cota, C. Chothia and J. Clarke, unpublished observations) also strongly suggests that for proteins with similar topology the primary influence on the rate of refolding is stability. But results from studies on the folding of homologues of cold shock protein B [13] and SH3 domains (Table 1) seem to disagree. The relative importance of local versus non-local interactions in determining the rate of folding has been studied [40]. In short, there are conflicting theoretical and experimental studies on the relative importance of size (chain length), topology and stability on folding. Thus, we are still far from understanding even some of the basic determinants of folding.

Characterisation of transition states for folding

Transition states for folding can only be characterised by studying the kinetics of unfolding and refolding. Many studies have now used a number of different approaches to both initiate unfolding or refolding, and to characterise the transition state. Commonly, refolding is initiated by a rapid shift in conditions from highly denaturing (low or high pH, or high concentrations of denaturant) to native. The subsequent folding process is followed using a spectroscopic technique such as fluorescence or CD or, for slowly folding proteins, NMR [41]. Mass spectrometry has also been used as a sensitive technique for studying folding [42–44]. Recently, very fast folding reactions have been initiated using electron-transfer [10]. Kinetic data have also been obtained from NMR line-shape analysis; this technique, however, is limited to the few proteins for which the unfolded and folded states are in fast-exchange [5,6].

Information on the thermodynamic nature of the transition state can be obtained by studying the temperature dependence of the unfolding and refolding rate constants. Using this method, values for the change in enthalpy (ΔH), entropy (ΔS), and heat capacity (ΔC_p), between the unfolded state (U), transition state (‡), and the folded state (F) can be calculated from an analysis of the Eyring plots. In general, the values for ΔH and ΔS are difficult to interpret because it is not possible to distinguish between the contribution from the protein and that from the solvent. The change in heat capacity on folding, $\Delta C_p^{‡–U}$, has been used as a measure of the extent of burial of hydrophobic residues between U and ‡. For some proteins this data is consistent with values obtained for the compactness of the transition state, as measured from the denaturant concentration dependence (see below); for example, CI2 and FKBP12 ([4]; S.E.J., E.R.G. Main and K.F. Fulton, unpublished observations). For other proteins, however, there are some discrepancies and the transition state appears to be less compact from measurements of $\Delta C_p^{‡–U}$ [38,45]. The exact nature of these differences remains unclear.

Information on the compactness of the transition state has also been obtained from studies on the dependence of the unfolding and refolding rates on the concentration of denaturant [D]. The rate of change of the natural logarithm of the unfolding rate, $m^{‡–F}$, or refolding rate, $m^{‡–U}$, on [D] is related to the average change in solvent-accessible surface area between initial states and transition states [4]. For two-state systems, these are related to the m value obtained from equilibrium experiments, $m = RT\,(m^{‡–F} - m^{‡–U})$. The relative values of $m^{‡–F}$ and $m^{‡–U}$ are thus a measure of the compactness of the transition state. In general $m^{‡–F}$ and $m^{‡–U}$ can be measured more accurately than the equilibrium m value, which is sensitive to baselines, so the position of the transition state on the reaction coordinate, β_T, for folding is defined as:

$$\beta_T = \frac{m^{‡–U}}{m^{‡–F} - m^{‡–U}} \quad (1)$$

such that a value of β_T close to 1 indicates a transition state that is very native-like with respect to solvent-accessible surface area, and lower values indicate transition states that are more like the unfolded state. In some cases, $m^{‡–F}$ is difficult to measure accurately and, in these cases, β_T is calculated from $m^{‡–U}/m$. As the value for β_T can be measured easily and with a high degree of accuracy it has been used as a measure of the position of the transition state on the folding pathway. Values are summarised in Table 1. β_T can also be calculated for proteins that fold through intermediate states from unfolding and equilibrium data:

$$\beta_T = 1 - \frac{m^{‡–F}}{m} \quad (2)$$

These values are shown in Table 2.

Recent studies have extended these experiments by studying the effect of sugars and alcohols on the rate of folding and unfolding [45]. These have yielded information on the extents of hydration and secondary structure, particularly α helices, in the transition state.

These methods can provide some structural information on the nature of the transition state. In general, however, they are only useful in determining the average properties of the transition state (e.g. the average solvent-accessible surface area) and cannot provide specific information on the structure of the transition state. In comparison, protein engineering techniques have been used to gain a much more detailed picture of the energetics and structure of the transition state [46]. Such experiments can provide high-resolution structural data, which is not obtainable by any other means. Results from protein engineering experiments are discussed later.

Table 2

Kinetic and thermodynamic parameters for the folding of proteins that fold with three-state kinetics.

Protein	Reference	PDB code	Contact order	Chain length	Structure	Denaturant	$\Delta G_{U-F}^{H_2O}$ (kcal mol^{-1})	m (kcal mol^{-1}M^{-1})	$[D]_{50\%}$ (M)	Temp. (°C)	$k_F^{H_2O}$ (s^{-1})	$\Delta G_{U-I}^{H_2O}$ (kcal mol^{-1})	$k_U^{H_2O}$ (s^{-1})	$m^{\ddagger-F}$ (kcal mol^{-1} M^{-1})	β_T
Ubiquitin															
wild type	[24]	1UBQ	15.1	67	α/β	GdnHCl	7.2	2.0	3.7	25	350	1.5	0.0012	0.82	0.59
wild type + 0.4 M Na_2SO_4							10.2	2.2	4.6		900	3.1	9×10^{-5}	1.05	0.52
Barstar	[54]	1BTA	12.2	89	α	Urea	4.8	1.3	3.9	25	31	1.9	0.068	0.17	0.87
[a]CD2, pH 7.0	[55]	1HNG	17.5	98	β-sandwich	GdnHCl				25	6.0	0.8	5.0×10^{-4}	1.60	[b]0.68
[a]CD2, pH 4.5						GdnHCl	8.7	5.0	1.7	25	14	1.8	1.0×10^{-4}	1.89	0.62
Barnase	[56]	1BNI	11.4	110	α/β	Urea	10.5	2.3	4.6	25	13	3.2	1.1×10^{-4}	0.27	0.88
[c]Suc 1				113	α/β	Urea	7.2	1.7	4.4	25	65	1.6	0.0001	0.78	0.54
Lysozyme (hen egg white)	[57]	1HEL	10.8	129	α/β	GdnHCl				20	4	2.7	6.2×10^{-7}	0.73	0.75
Lysozyme (hen egg white)	[58]	1HEL				GdnHCl	13.5	2.1	3.8	25	3.5	4.9	5.0 x 10-5	0.43	0.80
CheY	[59]	3CHY	9.0	129	α/β	Urea	5.2	1.6	3.3	25	2.7	[d]nd	0.012	0.47	0.71
[e]p16				148	α	Urea	3.1	1.7	1.9	25	33	1.5	1.4	0.08	0.95
GroEL apical domain (191-345)	[60]	1JON	15.7	154	α/β	Urea	[f]5.6	2.0	2.8	25	2.3	3.8	0.004	0.45	0.78
Ribonuclease H (*Escherichia coli*), pH 5.5	[61]	2RN2	12.4	155	α/β	Urea	[g]9.9	2.1	4.7	25	0.6	3.6	1.69×10^{-5}	0.42	0.80
Ribonuclease H (*E. coli*), pH 5.5	[62]	2RN2	12.4	155	α/β	GdnHCl	9.5	5.2	1.8	25	4.1	4.8	3.7×10^{-5}	1.91	0.63
N-terminal domain from PGK	[58]	1PHP	11.5	175	α/β	GdnHCl	8.4	7.6	1.1	25	9.5	5.2	0.03	1.24	0.84
[h]C-terminal domain from PGK	[63]	1PHP	8.0	219	α/β	GdnHCl	13.6	13	1.0	25	0.03	3.4	7.6×10^{-10}	7.1	0.45

[a]Some care must be taken in comparing these proteins as the values are calculated using denaturant activity not concentration of denaturant. [b]Calculated assuming the equilibrium m value does not vary significantly with pH. [c]F. Rousseau, J.W.H. Schymkowitz, M. Sánchez del Pino and L.S. Itzhaki, unpublished observations. [d]Not determined as it requires knowledge of the *cis–trans* isomerisation ratio in the unfolded state. [e]L. Itzhaki, personal communication. [f]Average of values obtained from CD and fluoresence experiments. [g]Calculated from the fit of the kinetic data to a three-state model. [h]The folding of the C-terminal domain of PGK is monitored in the presence of the N-terminal domain (fluorescence probe is only in C-terminal domain) – the domains act independently.

β_T has been used extensively as a measure of the position of the transition state on the reaction coordinate. Homologous proteins with similar β_T values are assumed to fold through very similar transition states. A recent protein engineering study on the transition state for folding of wild-type and circular permutants of an SH3 domain, however, have shown that structurally different transition states can result in similar values for β_T [16]. This illustrates the care that must be taken in interpreting these values, and also the need for higher resolution data on the transition state.

Nature of the transition state ensemble

Unlike a simple chemical reaction in which only a few, high-energy, covalent bonds are made and broken, the transition state for folding involves the simultaneous making and breaking of literally hundreds of weak non-covalent interactions. For this reason, and just as the native state is not a single structure, the transition state will be an ensemble of structures of similar energy. The exact nature of this ensemble has been the subject of much discussion. Some theoretical studies have predicted that there will be a wide range of structures present in the transition state, whilst others have suggested that there is only a fairly narrow range. Experimental approaches have been developed to study the nature of the transition state ensemble. Fersht *et al.* [46] have used a Brønsted analysis, which compares the refolding rates to changes in the free energies of unfolding of wild-type and mutant proteins. A linear relationship is observed and has been used to show that components of the transition state ensemble are close in structure. This has also been shown for FKBP12 (K.F. Fulton, E.R.G. Main and S.E.J., unpublished observations) and ADAh2 (L. Serrano, personal communication).

For many proteins, it has been found that mutations do not significantly affect the position of the transition state on the reaction coordinate. Changes could occur if a mutation destabilised one set of similar structures in the transition state ensemble, thereby favouring another set with different structures. Some movement of β_T with mutation, temperature and denaturant concentration has been observed and attributed to Hammond behaviour — the transition state becoming more native-like as the energy difference between the transition and the native state is decreased [32,33]. In general, these movements are not large. In contrast, the position of the transition state for the folding of the monomeric λ repressor is affected by mutation and large changes in β_T are observed (β_T varies between 0.39 and 0.83; [47]). This suggests that, in this case, the transition state ensemble may be composed of a significantly more diverse range of structures than that found for other proteins.

Looking for trends

In general, proteins that fold with two-state kinetics are small, typically < 100 residues in length (FKBP12 is the largest reported example, at 25°C, with 107 residues). Apart from chain length, there are no other general trends; examples are known with very different structures, stabilities

and folding rates (Figure 1 and Table 1). Stabilities are found in the range 2–8 kcal mol^{-1}, whereas both unfolding and refolding rates are found to vary by more than a factor of 10^5 (this corresponds to a difference of 7 kcal mol^{-1} in energy terms at 25°C).

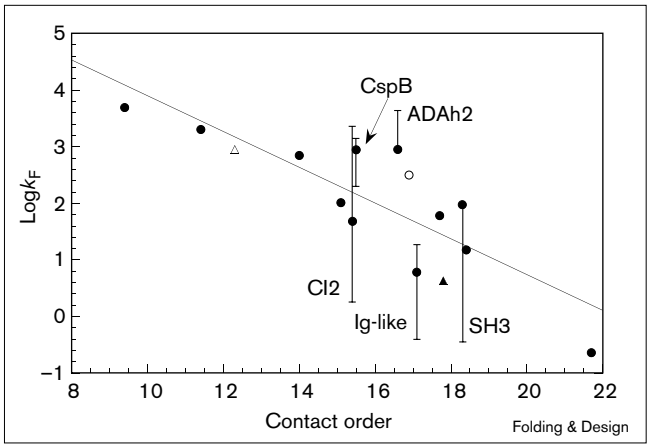
Secondary structure

It has been proposed that structure determines both the rate at which a protein folds and the position of the transition state on the reaction coordinate. In particular, it has been suggested that α-helical proteins should fold fast whereas all-β proteins should fold slowly as a result of the relative importance of local versus non-local interactions in determining structure and stability. Indeed, a correlation is observed for small α-helical proteins that fold rapidly with rates of more than 400 s^{-1} (see Table 1). Not all α-helical proteins fold fast, however; several relatively small α-helical proteins fold slowly through populated intermediate states, for example, apomyoglobin [48] and p16 (L. Itzhaki, personal communication). In addition, these simple ideas do not apply to α/β and β proteins. For example, α/β proteins have been found to fold with rate constants of the order of 900 s^{-1} and 0.2 s^{-1}. Likewise, some all-β proteins fold fast (rates > 1000 s^{-1}) whereas others are slow (0.4 s^{-1}). In addition, there is no clear correlation between the position of the transition state on the reaction coordinate (β_T) and the structure. The position of the transition state (β_T) is in the range 0.6–0.8 for α/β proteins, 0.5–0.9 for β proteins, and 0.4–0.8 for α proteins. It is clear that correlations with structure, if they exist, are more complex than simply based on the amount of α or β elements of secondary structure.

Contact order: local versus non-local interactions

In their search for correlations between the structure, equilibrium properties and the folding of a series of structurally unrelated proteins that fold with simple kinetics, Plaxco *et al.* [38] recently introduced the concept of a contact order to describe the topological complexity of a given protein fold. The contact order is a measure of the number of contacts (< 6 Å) a residue makes with other residues that are local in sequence, relative to the number of contacts with residues distant in sequence. Thus, highly helical proteins, which have a large number of local contacts, have low contact orders, whereas α/β and β structures have higher contact orders, reflecting the importance of long-range interactions to protein stability. Thus, the contact order is related to secondary structure, but is also influenced substantially by the tertiary fold. For the dataset of 12 proteins studied by Plaxco *et al.* [38] two particular correlations emerged. One was between the contact order and the natural logarithm of the refolding rate (correlation coefficient R = 0.8); a slightly weaker one connected the contact order and the position of the transition state (R = 0.68). Thus, these correlations suggest that the lower the contact order the faster the folding rate and the less native-like (i.e., early on the folding pathway) the transition state.

Figure 2



Correlation between folding rate in water and contact order [38] for the 12 structurally unrelated proteins used in the study of Plaxco *et al.* [38]. The filled circles correspond to monomeric λ repressor [5,6], bovine ACBP [7,8], cytochrome *c* [9], *Bacillus subtilis* CspB [12], SH3 domain (fyn; [19]), tenascin [21], chymotrypsin inhibitor 2 (CI2; [4]), activation domain procarboxypeptidase (ADAh2; [22]), ubiquitin [24,25], histidine-containing phosphocarrier protein (Hpr; [28]), Ig-binding domain of protein L [26], muscle acylphosphatase (AcP; [29]). The solid line is the best fit for these data (correlation coefficient = 0.8). Superimposed on these data are the recently published refolding rates for FKBP12 (S.E.J., E.R.G. Main and K.F. Fulton, unpublished observations; open circle), U1A spliceosomal protein ([27]; filled triangle) and villin ([30]; open triangle). For CspB, tenascin, and the SH3 domain from fyn, error bars indicate the range of refolding rates that have been measured for structurally homologous proteins. For CI2 the error bars indicate the range of refolding rates measured for single point substitutions, and for ADAh2 the error bars indicate the refolding rate obtained for a helix-stabilised mutant [88].

Figure 2 shows the correlation between $\log k_F$ and the contact order for the 12 proteins in the data set of Plaxco *et al.* [38]. Values for three other, structurally unrelated proteins, FKBP12, U1A spliceosomal protein and the N-terminal domain of villin, are also shown and fit reasonably well with the correlation. The value of $\log k_F$ for tendamistat, however, is significantly higher than predicted by contact order alone. This may result from the two disulphide bonds, which would tend to restrict conformational space in the unfolded state and which may result in faster folding. The structure of a final example, the single-chain arc repressor, can be inferred from that of the dimeric protein. This has a high helical content and hence suggests a low contact order in the monomer, which is therefore predicted to fold rapidly. Indeed, the single chain arc repressor has one of the fastest rates of folding observed [23].

Despite this apparently strong correlation it is clear from Table 1 that homologous proteins with similar topologies, and therefore contact orders, can fold with very different rates. In these cases, it is difficult to know which protein is the most representative of its fold. In order to illustrate

the variation in refolding rates for structurally related proteins error bars have been included in Figure 2 to show the slowest and fastest refolding rates measured for a particular structural family. These results clearly show that there is a significant spread of values, suggesting that topology is not the sole determinant of the rate of folding. In addition, protein engineering studies have shown that there can be significant differences in the rate of folding on mutation, even for single substitutions. Figure 2 shows error bars representing the fastest folding mutant of CI2, Arg48→Phe, which folds with a rate of 2270 s^{-1} in water [49], and the slowest folding mutant, Leu49→Ala, which folds with a rate of 1.8 s^{-1} in water [50]. Double and triple mutants can have even larger effects, again without affecting the contact order. It has also been shown that changes to the hydrophobic core of the four-helical bundle protein ROP, which do not significantly affect the structure of the native state, affect the rates of folding by up to three orders of magnitude [51]. In addition, in some cases, such as CspB and the fyn SH3 domain, there is a significant difference in the contact order calculated from different structures obtained by X-ray crystallography and NMR (see Table 1). Thus, the concept of contact order may have some value in predicting folding characteristics of proteins, but evidently other factors must be at least as significant.

Comparison of proteins that fold with two-state and three-state kinetics

What factors determine whether stable intermediates are populated on the folding pathway? Since it was established that intermediates are not prerequisites for the fast efficient folding of proteins, the role of intermediate states on folding pathways has been under discussion. In some cases, it has been suggested that intermediates observed during folding reactions are off-pathway species. Given the difficulty in distinguishing kinetically between off-pathway intermediates in rapid equilibrium with the unfolded state and obligate on-pathway intermediates, this is still the subject of debate [52]. It is becoming apparent, however, that in some cases intermediates represent misfolded species, which act as kinetic traps. It has also been suggested that the presence of intermediate states on folding pathways may slow the folding process [53].

What structural, thermodynamic or kinetic features determine whether a protein folds through a stable intermediate state? Table 2 summarises the kinetic and thermodynamic data for folding for some of the proteins that have been shown to fold to their native states via a populated intermediate ([54–63]; F. Rousseau, J.W.H. Schymkowitz, M. Sánchez del Pino and L.S. Itzhaki, unpublished observations). This is not a comprehensive list of proteins with multi-state kinetics, but includes those proteins for which there is data available on the denaturant dependence of the unfolding and refolding rates. For a more comprehensive discussion of intermediate states in protein folding see recent reviews [64–67].

Neither the rates of folding, 0.5–900 s^{-1}, nor the position of the rate-limiting transition state, β_T = 0.6–0.9, nor the topological complexity of the native fold (as measured by the contact order), are significantly different for the proteins listed in Table 2 from those which fold with two-state kinetics, Table 1. The main differences appear to be chain length and stability.

The proteins can be grouped into two classes depending on chain length — those < 110 residues in length and those longer. Comparing proteins with < 110 residues with proteins in Table 1 it is clear that one of the major differences is stability. In general, the proteins that fold with three-state kinetics are more stable than proteins observed to fold with two-state kinetics. Reduced cytochrome *c*, CD2 and spliceosomal U1A protein appear to be exceptions to this rule, although there may be reasons for this. Reduced horse and yeast cytochrome *c* are significantly more stable than their corresponding oxidised forms; but the kinetics of refolding for these proteins have only been measured at concentrations of GdnHCl > 1 M [10]. Thus, kinetic intermediates at lower concentrations cannot be ruled out. Likewise, ^{10}FN-III, which has the highest stability of the β-sandwich proteins shown in Table 1, folds through an intermediate state (S.J. Hamill, E. Cota, C. Chothia and J. Clarke, unpublished observations). This is consistent with studies that have demonstrated that one can switch from three-state to two-state kinetics by destabilising the native state (which may also result in destabilisation of the intermediate relative to the unfolded state), and from two-state to three-state kinetics by stabilising the protein with sodium sulphate [25]. In some cases, the addition of sodium sulphate does not switch the kinetics from two-state to three-state, suggesting that intermediate states, if present, are high in energy. This has been shown for FKBP12 (S.E.J., E.R.G. Main and K.F. Fulton, unpublished observations) and protein L [26].

Kinetic traps. There is growing evidence, for a number of proteins that have previously been shown to fold with multi-state kinetics, that the intermediate states populated in these studies result in kinetic traps. In some cases, conditions have been found that result in rapid folding with two-state kinetics. For example, at pH > 6 the folding of cytochrome *c* is slow as a result of the formation of a misfolded structure involving a non-native haem-ligand interaction. A histidine, or the N-terminal amino group, can coordinate the haem group resulting in a misfolded species. The non-native ligand must dissociate from the haem group before refolding to the final native state can take place [68–70]. This off-pathway, misfolded, intermediate can be eliminated, and the folding simplified, by refolding in the presence of 0.2 M imidazole (this coordinates the

haem group preventing the non-native histidine-haem interaction forming; [9]). Alternatively, at pH < 4.9, the histidines are protonated and cannot coordinate to the haem. In the former case, the protein refolds rapidly with simple two-state kinetics [9]; in the latter case, folding is rapid but a 'roll-over' is still observed in the plots of $\ln k_F$ versus denaturant concentration indicating that at [GdnHCl] < 1.2 M the rate is less than expected from a simple two-state model, as a result of either the formation of an intermediate or a competing reaction [71,72].

Studies of lysozyme have revealed the existence of multiple folding pathways [73–75]. Under strongly native conditions the majority of molecules (> 80%) fold on a slow pathway with a well-populated intermediate state. This intermediate state and the transition between it and the native state have been extensively characterised and it has been shown that this intermediate state represents a kinetic trap to folding — the polypeptide chain has to rearrange before proceeding to the native state. A much smaller fraction of molecules (< 20%) fold rapidly along a 'fast track'. Kiefhaber [57] has studied this fast track using "interrupted refolding experiments", and results suggest that the fast pathway may correspond to that for a two-state transition with no intermediate states populated. Other experiments, however, have suggested the existence of a highly native-like intermediate on this pathway too [75].

Peptidyl-prolyl isomerisation is frequently a rate-limiting step in the folding of proteins that have one or more prolines in a *cis* conformation in the native state [76]. Disulphide-intact ribonuclease A, for example, has two *cis* prolines in the native structure and the folding of the majority of molecules is limited by slow proline isomerisation. Studies have now shown that there is a small (~6%) fraction of the molecules for which folding is not limited by proline isomerisation, and, in this case, folding is very rapid — 20 s^{-1} at 1.5 M GdnHCl, 15°C [77].

In many cases, proteins that are observed to fold slowly may do so as a result of a misfolded species, on or off pathway. In these cases, there may be a 'fast track' to folding which is only populated under certain experimental conditions. The folding is not, as previously thought, intrinsically slow.

Comparison of the folding of monomeric and dimeric proteins

Thermodynamic and kinetic data for the folding of three dimeric proteins are summarised in Table 3. The folding of the Arc represssor [78–81] and GCN-4 [82] is concurrent with dimerisation, whereas the rate-determining step in the folding of the ROP dimer is unimolecular [51]. Thus, there is a rate-limiting structural rearrangement after the fast association of two chains. Dimeric proteins, like small monomeric proteins, can fold with simple two-state kinetics or via populated intermediate states. The range of folding rates (at 10 μM protein) and β_T are within the range found for monomeric proteins. Thus, it appears that there are no intrinsic differences between the folding of monomers and dimers.

Protein engineering studies

Protein engineering techniques have been used to dissect the interactions and structure present in the transition state for folding for a number of proteins ([50,83–86]; K.F. Fulton, E.R.G. Main and S.E.J., unpublished observations). The effect of a mutation on the energetics of the native state ($\Delta\Delta G_{U-F}$) and the transition state relative to the unfolded state ($\Delta\Delta G_{U-\ddagger}$), can be measured using a combination of equilibrium and kinetic experiments [46]. The two are compared by defining a Φ value as $\Delta\Delta G_{U-\ddagger}/\Delta\Delta G_{U-F}$ [46]. Φ values can be taken as a measure of the extent of structure formation in the transition state:

Table 3

Kinetic and thermodynamic parameters for the folding of dimeric proteins.

Protein	Reference	PDB code	Chain length	Structure	Denaturant	$\Delta G_{U-F}^{H_2O}$ (kcal mol^{-1} $dimer^{-1}$)	m (kcal $mol^{-1}M^{-1}$)	$[D]_{50\%}$ (M)	Temp. (°C)	[a]$k_F^{H_2O}$ (M^{-1} s^{-1})	$k_F^{H_2O}$ (10 μM protein s^{-1})	$m^{\ddagger-U}$ (kcal mol^{-1} M^{-1})	$k_U^{H_2O}$ (s^{-1})	$m^{\ddagger-F}$ (kcal mol^{-1} M^{-1})	[b]β_T
Arc repressor															
wild type	[78–80]	1ARQ	2 × 53	$(\beta\alpha_2)_2$	Urea	9.6	1.4		20	9.1×10^6	91		0.2		0.75
wild type	[81]	1ARR			GdnHCl	10.1	3.0		25	8.4×10^6	84	1.83	0.2	0.88	0.68
PL8					GdnHCl	12.7	3.1		25	7.4×10^6	74	2.10	0.002	1.14	0.65
MYL mutant					Urea	14.3	[c]2.6		25	3.0×10^6	3000				0.4
[d]ROP	[51]	1ROP													
wild type		1RPR	2 × 63	$(\alpha_2)_2$	GdnHCl	7.7	2.4	3.3	25	[e]0.013	0.013		9.8×10^{-7}	0.47	0.8
Ala_2Leu_2 (1+8)						6.3	2.9	2.3	25	2.1	2.1	1.8×10^{-4}			
Ala_2Leu_2 −6						8.1	2.7	2.7	25	4.0	4.0	4.9×10^{-2}			
GCN-4	[82]	2ZTA 3DGC 1YSA	2 × 33	$(\alpha)_2$	GdnHCl	10.5	1.8		5	4.2×10^5	4.2	0.97	3.3×10^{-3}	0.88	0.52

[a]At low protein concentration the refolding of dimers often follows first-order kinetics with respect to the protein concentration such that $k_{obs} = k_F$[protein], as expected for a bimolecular reaction. For Arc repressor and GCN-4 the refolding rate constants are given as the first-order rate constant. For ROP the rate-limiting step is unimolecular and the rate corresponds to the rate observed and is independent of protein concentration. [b]Calculated from $1 - m^{\ddagger-F}/m$. [c]Average value. [d]Values for k_F^{H2O} and k_U^{H2O} are calculated from Table 1 of [51] using an average $m^{\ddagger-F}$ of 0.8 M^{-1} and an average value for $m^{\ddagger-U}$ of 1.5 M^{-1} and extrapolating the data to 0 M GdnHCl. [e]For wild-type ROP the refolding rates are strongly dependent on ionic strength – values given are in 100 mM Tris-HCl (pH 8), 0.2 mM EDTA and 0.1 M KCl.

Φ values of 0 indicate that there is little structure in the region surrounding the point of mutation in the transition state, Φ values of 1 indicate that the region is highly structured, and intermediate Φ values indicate partial structure formation.

By far the best characterised protein using this approach is CI2, for which > 150 mutations have been made at sites throughout the protein [84]. In general, the Φ values are low, in the range 0.2–0.5 in the hydrophobic core, and slightly higher in the single α helix, whereas residues at the N-cap of the α helix have some of the highest Φ values of around 0.6–0.8. These residues interact with two residues in the β sheet to form a core, which is the most highly structured region of the protein in the transition state. This has led Fersht and coworkers [53,84] to propose a nucleation-condensation mechanism for folding. Once a critical number of interactions have been made in the transition state — the formation of a 'folding nucleus' — the remaining structure condenses rapidly around this. Thus, formation of the folding nucleus is the rate-limiting step in folding. Protein engineering and folding experiments on a much larger protein, CheY, have also identified a folding nucleus consistent with this mechanism [86].

The transition state for folding of the SH3 domain from α-spectrin has also been characterised using protein engineering techniques [16,85]. For the wild-type protein, eight mutants have been analysed and the Φ values found to be in the range 0.2–0.6. More importantly, this group extended their protein engineering studies on the wild-type protein and looked at two circular permutants that have the same native structure, the same compact transition state and the same simple two-state kinetics [85]. Comparing the Φ values obtained for wild type and two circular permutants they were able to show that, although the gross properties of the system were very similar, there were significant differences in the structures of the transition states. They concluded that folding nuclei need not be specific and that no compulsory initiation event was necessary for the folding of this five-stranded orthogonal β-sheet structure. This study also shows the limitations on interpreting gross overall properties and the need for detailed information on the structure and interactions in a transition state.

Protein engineering studies have also been performed on the monomeric form of the λ repressor [47]. Seven Ala→Gly substitutions have been made, which destabilise the protein, and the effect of these substitutions on the folding and unfolding rates, and the position of the transition state has been measured. This protein is unusual as mutations significantly change the compactness of the transition state measured by β_T. Values vary between 0.4 and 0.9. These results have been interpreted in terms of a highly plastic transition state — suggesting that the transition state ensemble may be a diverse set of structures.

A Φ value analysis on 34 mutants of FKBP12 has shown that the transition state for folding is only weakly structured with Φ values in the hydrophobic core in the range 0–0.6 (K.F. Fulton, E.R.G. Main and S.E.J., unpublished observations). In comparison to CI2 and barnase, Φ values for residues in the α helix of FKBP12 are all close to zero, suggesting that the helix is largely unstructured in the transition state. The data is consistent with the nucleation-condensation mechanism.

In comparison with the proteins discussed above, many Φ values from protein engineering experiments on the ribonuclease barnase were found to be close to 0 or 1 [83]. Barnase folds through a populated intermediate state, which is thought to be an obligate intermediate and not a misfolded, off-pathway species [34,36,37]. These results have led to the proposal that barnase folds according to the framework model in which a large amount of secondary structure is formed in subdomains in the intermediate state and the rate-limiting step involves these subdomains coalescing [87].

Conclusions

Many proteins, and domains of larger proteins, have been shown to fold with two-state kinetics. From the number of studies on these simple two-state systems some trends are beginning to emerge. It is evident, however, that there are many factors influencing the rate of folding, and position of the transition state on the folding pathway. Contact order, a measure of the relative number of local interactions compared to non-local interactions, may be one factor. It is clear though, that in many cases, other factors are equally important. Further studies are needed to determine some of these factors, and the balance between them.

Although the folding rate, and position of the transition state on the folding pathway, tell us something about the nature of the folding reaction, these parameters do not provide many structural details on how a protein folds. Protein engineering experiments have provided more detailed energetic and structural information, which has resulted in new folding models. With the number of protein engineering studies currently in progress in many laboratories it will be interesting to see what insights a future comparative study will yield.

Acknowledgements

I thank Jane Clarke, Sarah Perrett and John Armstrong for helpful discussions.

References

1. Levinthal, C. (1968). Are there pathways for protein folding? *J. Chim. Phys.* **85**, 44-45.
2. Kim, P.S. & Baldwin, R.L. (1982). Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* **51**, 459-489.
3. Kim, P.S. & Baldwin, R.L. (1990). Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* **59**, 631-660.
4. Jackson, S.E. & Fersht, A.R. (1991). Folding of chymotrypsin inhibitor-2. 1. Evidence for a two-state transition. *Biochemistry* **30**, 10428-10435.

5. Huang, G.S. & Oas, T.G. (1995). Structure and stability of monomeric λ repressor: NMR evidence for two-state folding. *Biochemistry* **34**, 3884-3892.
6. 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.
7. Kragelund, B.B., Robinson, C.V., Knudsen, J., Dobson, C.M. & Poulsen, F.M. (1995). Folding of a four-helix bundle: studies of acyl-coenzyme A binding protein. *Biochemistry* **34**, 7217-7224.
8. Kragelund, B.B., *et al*., & 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.
9. Chan, C.K., *et al*., & Hofrichter, J. (1997). Submillisecond protein folding kinetics studied by ultrarapid mixing. *Proc. Natl Acad. Sci. USA* **94**, 1779-1784.
10. Mines, G.A., Pascher, T., Lee, S.C., Winkler, J.R. & Gray, H.B. (1996). Cytochrome-*c* folding triggered by electron-transfer. *Chem. Biol.* **3**, 491-497.
11. Schonbrunner, N., Koller, K.-P. & Kiefhaber, T. (1997). Folding of the disulfide-bonded β-sheet protein tendamistat: rapid two-state folding without hydrophobic collapse. *J. Mol. Biol.* **268**, 526-538.
12. Schindler, T., Herrler, M., Marahiel, M.A. & Schmid, F.X. (1995). Extremely rapid folding in the absence of intermediates. *Nat. Struct. Biol.* **2**, 663-673.
13. Perl, D., *et al*., & Schmid, F.X. (1998). Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nat. Struct. Biol.* **5**, 229-235.
14. Reid, K.L., Rodriguez, H.M., Hillier, B.J. & Gregoret, L.M. (1998). Stability and folding properties of a model β-sheet protein, *Escherichia coli* CspA. *Protein Sci.* **7**, 470-479.
15. Viguera, A., Martinez, J., Filimonov, V., Mateo, P. & Serrano, L. (1994). Thermodynamic and kinetic-analysis of the SH3 domain of spectrin shows a 2-state folding transition. *Biochemistry* **33**, 2142-2150.
16. Viguera, A.R., Serrano, L. & Wilmanns, M. (1996). Different folding transition-states may result in the same native structure. *Nat. Struct. Biol.* **3**, 874-880.
17. Grantcharova, V.P. & Baker, D. (1997). Folding dynamics of the src SH3 domain. *Biochemistry* **36**, 15685-15692.
18. Guijarro, J.I., Morton, C.J., Plaxco, K.W., Campbell, I.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.
19. Plaxco, K.W., *et al*., & Dobson, C.M. (1998). The folding kinetics and thermodynamics of the Fyn-SH3 domain. *Biochemistry* **37**, 2529-2537.
20. Plaxco, K.W., Spitzfaden, C., Campbell, I.D. & Dobson, C.M. (1997). A comparison of the folding kinetics and thermodyanmics of two homologous fibronectin type III modules. *J. Mol. Biol.* **270**, 763-770.
21. 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.
22. Villegas, V., *et al*., & Serrano, L. (1995). Evidence for a two-state transition in the folding process of the activation domain of human procarboxypeptidase A2. *Biochemistry* **34**, 15105-15110.
23. Robinson, C.R. & Sauer, R.T. (1996). Equilibrium stability and submillisecond refolding of a designed single-chain Arc repressor. *Biochemistry* **35**, 13878-13884.
24. Khorasanizadeh, S., Peters, I.D., Butt, T.R. & Roder, H. (1993). Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry* **32**, 7054-7063.
25. 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. *Nat. Struct. Biol.* **3**, 193-205.
26. Scalley, M.L., *et al*., & Baker, D. (1997). Kinetics of folding of the IgG binding domain of peptostreptococcal protein L. *Biochemistry* **36**, 3373-3382.
27. Silow, M. & Oliveberg, M. (1997). High-energy channeling in protein folding. *Biochemistry* **36**, 7633-7636.
28. van Nuland, N.A.J., *et al*., & Dobson, C.M. (1998). Slow co-operative folding of a small globular protein HPr. *Biochemistry* **37**, 622-637.
29. van Nuland, N.A.J., *et al*., & Dobson, C.M. (1998). Slow folding of muscle acylphosphatase in the absence of intermediates. *J. Mol. Biol*., in press.
30. Choe, S.E., Matsudaira, P.T., Osterhout, J., Wagner, G. & Shakhnovich, E.I. (1998) Folding kinetics of villin 14T, a protein domain with a central β-sheet and two hydrophobic cores. *Biochemistry*, in press.
31. 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.
32. Matouschek, A., & Fersht, A.R. (1993). Application of physical organic chemistry to engineered mutants of proteins - Hammond postulate behavior in the transition state of protein folding. *Proc. Natl Acad. Sci. USA* **90**, 7814-7818.
33. Matouschek, A., Otzen, D.E., Itzhaki, L.S., Jackson, S.E. & Fersht, A.R. (1995). Movement of the position of the transition state in protein folding. *Biochemistry* **34**, 13656-13662.
34. Dalby, P.A., Oliveberg, M. & Fersht, A.R. (1998). Movement of the intermediate and rate-determining transition state of barnase on the energy landscape with changing temperature. *Biochemistry* **37**, 4674-4679.
35. Pace, C.N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266-279.
36. Dalby, P.A., Clarke, J., Johnson, C.M. & Fersht, A.R. (1998). Folding intermediates of wild-type and mutants of barnase. II. Correlation of changes in equilibrium amide exchange kinetics with the population of the folding intermediate. *J. Mol. Biol.* **276**, 647-656.
37. Dalby, P.A., Oliveberg, M. & Fersht, A.R. (1998). Folding intermediates of wild-type and mutants of barnase. I. Use of Φ-value analysis and m-values to probe the cooperative nature of the folding pre-equilibrium. *J. Mol. Biol.* **276**, 625-646.
38. Plaxco, K.W., Simons, K.T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985-994.
39. Dobson, C.M., Sali, A. & Karplus, M. (1998). Protein folding: a perspective from theory and experiment. *Angew. Chem. Int. Ed. Engl.* **37**, 868-893.
40. Munoz, V. & Serrano, L. (1996). Local versus nonlocal interactions in protein folding and stability – an experimentalist's point of view. *Fold. Des.* **1**, R71-R77.
41. Plaxco, K.W. & Dobson, C.M. (1996). Time-resolved biophysical methods in the study of protein-folding. *Curr. Opin. Struct. Biol.* **6**, 630-636.
42. Miranker, A., Robinson, C.V., Radford, S.E., Aplin, R.T. & Dobson, C.M. (1993). Detection of transient protein-folding populations by mass-spectrometry. *Science* **262**, 896-900.
43. Hooke, S.D., *et al*., & Dobson, C.M. (1995). Cooperative elements in protein-folding monitored by electrospray-ionization mass-spectrometry. *J. Am. Chem. Soc.* **117**, 7548-7549.
44. Miranker, A., Robinson, C.V., Radford, S.E. & Dobson, C.M. (1996). Investigation of protein-folding by mass-spectrometry. *FASEB J.* **10**, 93-101.
45. Chiti, F., *et al*., & Dobson, C.M. (1998). Structural characterisation of the transition state for folding of muscle acylphophatase. *J. Mol. Biol*., in press.
46. Fersht, A.R., Matouschek, A. & Serrano, L. (1992). The folding of an enzyme. 1. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224**, 771-782.
47. Burton, R.E., Huang, G.S., Daugherty, M.A., Calderone, T.L. & Oas, T.G. (1997). The energy landscape of a fast-folding protein mapped by Ala→Gly substitutions. *Nat. Struct. Biol.* **4**, 305-310.
48. Eliezer, D., Yao, J., Dyson, H.J. & Wright, P.E. (1998). Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat. Struct. Biol.* **5**, 148-155.
49. 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*, in press.
50. Jackson, S.E., elMasry, N. & Fersht, A.R. (1993). Structure of the hydrophobic core in the transition state for folding of chymotrypsin inhibitor 2: a critical test of the protein engineering method of analysis. *Biochemistry* **32**, 11270-11278.
51. Munson, M., Anderson, K.S. & Regan, L. (1997). Speeding up protein folding: mutations that increase the rate at which ROP folds and unfolds by over four orders of magnitude. *Fold. Des.* **2**, 77-87.
52. Baldwin, R.L. (1996). On-pathway versus off-pathway folding intermediates. *Fold. Des.* **1**, R1-R8.
53. Fersht, A.R. (1995). Optimization of rates of protein folding: the nucleation-collapse mechanism for the folding of chymotrypsin inhibitor 2 (CI2) and its consequences. *Proc. Natl Acad. Sci. USA* **92**, 10869-10873.
54. Schreiber, G. & Fersht, A.R. (1993). The refolding of cis-peptidylprolyl and trans-peptidylprolyl isomers of barstar. *Biochemistry* **32**, 11195-11203.

55. Parker, M., Dempsey, C.E., Lorch, M. & Clarke, A.R. (1997). Acquisition of native β-strand topolgy during the rapid collapse phase of protein folding. *Biochemistry* **36**, 13396-13405.
56. Matouschek, A., Kellis, J.T., Jr, Serrano, L., Bycroft, M. & Fersht, A.R. (1990). Transient folding intermediates characterized by protein engineering. *Nature* **346**, 440-445.
57. Kiefhaber, T. (1995). Kinetic traps in lysozyme folding. *Proc. Natl Acad. Sci. USA* **92**, 9029-9033.
58. Parker, M.J., Spencer, J. & Clarke, A.R. (1995). An integrated kinetic-analysis of intermediates and transition-states in protein-folding reactions. *J. Mol. Biol.* **253**, 771-786.
59. Munoz, V., Lopez, E.M., Jager, M. & Serrano, L. (1994). Kinetic characterization of the chemotactic protein from *Escherichia coli*, CheY - kinetic-analysis of the inverse hydrophobic effect. *Biochemistry* **33**, 5858-5866.
60. Golbik, R., Zahn, R., Harding, S.E. & Fersht, A.R. (1998). Thermodynamic stability and folding of GroEL minichaperones. *J. Mol. Biol.* **276**, 505-515.
61. Raschke, T.M. & Marqusee, S. (1997). The kinetic folding intermediate of ribonuclease h resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nat. Struct. Biol.* **4**, 298-304.
62. Yamasaki, K., Ogagahara, K., Yutani, K., Oobatake, M. & Kanaya, S. (1995). Folding pathway of *Escherichia coli* ribonuclease HI: a circular dichroism, fluorescence and NMR study. *Biochemistry* **34**, 16552-16562.
63. Parker, M.J., Sessions, R.B., Badcoe, I.G. & Clarke, A.R. (1996). The development of tertiary interactions during the folding of a large protein. *Fold. Des.* **1**, 145-156.
64. Creighton, T.E., Darby, N.J. & Kemmink, J. (1996). The roles of partly folded intermediates in protein folding. *FASEB. J.* **10**, 110-118.
65. Clarke, A.R. & Waltho, J.P. (1997). Protein folding pathways and intermediates. *Curr. Opin. Biotechnol.* **8**, 400-410.
66. Roder, H. & Colon, W. (1997). Kinetic role of early intermediates in protein folding. *Curr. Opin. Struct. Biol.* **7**, 15-28.
67. Fink, A.L. (1995). Compact intermediate states in protein-folding. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 495-522.
68. Colon, W., Wakem, L.P., Sherman, F. & Roder, H. (1997). Identification of the predominant non-native histidine ligand in unfolded cytochrome *c*. *Biochemistry* **36**, 12535-12541.
69. Elove, G.A., Bhuyan, A.K. & Roder, H. (1994). Kinetic mechanism of cytochrome-c folding - involvement of the heme and its ligands. *Biochemistry* **33**, 6925-6935.
70. Hammack, B., Godbole, S. & Bowler, B.E. (1998). Cytochrome *c* folding traps are not due solely to histidine-heme ligation: direct demonstration of a role for N-terminal amino group-heme ligation. *J. Mol. Biol.* **275**, 719-724.
71. Sosnick, T.R., Mayne, L. & Englander, S.W. (1996). Molecular collapse - the rate-limiting step in 2-state cytochrome-c folding. *Proteins* **24**, 413-426.
72. Sosnick, T.R., Mayne, L., Hiller, R. & Englander, S.W. (1994). The barriers in protein-folding. *Nat. Struct. Biol.* **1**, 149-156.
73. Matagne, A., Radford, S.E. & Dobson, C.M. (1997). Fast and slow tracks in lysozyme folding: insight into the role of domains in the folding process. *J. Mol. Biol.* **267**, 1068-1074.
74. Radford, S.E., Dobson, C.M. & Evans, P.A. (1992). The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* **358**, 302-307.
75. Lu, H., Buck, M., Radford, S.E. & Dobson, C.M. (1997). Acceleration of the folding of hen lysozyme by trifluoroethanol. *J. Mol. Biol.* **265**, 112-117.
76. Schmid, F.X. (1991). Catalysis and assistance of protein folding. *Curr. Opin. Struct. Biol.* **1**, 36-41.
77. Houry, W.A., Rothwarf, D.M. & Scheraga, H.A. (1994). A very fast phase in the refolding of disulfide-intact ribonuclease-A - implications for the refolding and unfolding pathways. *Biochemistry* **33**, 2516-2530.
78. Milla, M. & Sauer, R. (1994). P22 arc repressor - folding kinetics of a single-domain, dimeric protein. *Biochemistry* **33**, 1125-1133.
79. Schildbach, J.F., Milla, M.E., Jeffrey, P.D., Raumann, B.E. & Sauer, R.T. (1995). Crystal-structure, folding, and operator binding of the hyperstable arc repressor mutant PLI8. *Biochemistry* **34**, 1405-1412.
80. Waldburger, C.D., Jonsson, T. & Sauer, R.T. (1996). Barriers to protein-folding - formation of buried polar interactions is a slow step in acquisition of structure. *Proc. Natl Acad. Sci. USA* **93**, 2629-2634.
81. Hendsch, Z.S., Jonsson, T., Sauer, R.T. & Tidor, B. (1996). Protein stabilization by removal of unsatisfied polar groups - computational approaches and experimental tests. *Biochemistry* **35**, 7621-7625.
82. Zitzewitz, J.A., Bilsel, O., Luo, J.B., Jones, B.E. & Matthews, C.R. (1995). Probing the folding mechanism of a leucine-zipper peptide by stopped-flow circular-dichroism spectroscopy. *Biochemistry* **34**, 12812-12819.
83. Serrano, L., Matouschek, A. & Fersht, A.R. (1992). The folding of an enzyme. 3. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. *J. Mol. Biol.* **224**, 805-818.
84. Itzhaki, L.S., Otzen, D.E. & Fersht, A.R. (1995). The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-collapse mechanism for protein folding. *J. Mol. Biol.* **254**, 260-288
85. Viguera, A.R., Blanco, F.J. & Serrano, L. (1995). The order of secondary structure elements does not determine the structure of a protein but does affect its folding kinetics. *J. Mol. Biol.* **247**, 670-681.
86. Lopez-Hernandez, E. & Serrano, L. (1996). Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein, CI-2. *Fold. Des.* **1**, 43-55.
87. Fersht, A.R. (1995). Mapping the structures of transition-states and intermediates in folding - delineation of pathways at high-resolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **348**, 11-15.
88. Viguera, A.R., Villegas, V., Aviles, F.X. & Serrano, L. (1997). Favourable native-like helical local interactions can accelerate protein folding. *Fold. Des.* **2**, 23-33.