Theoretical Study of a Landscape of Protein Folding−Unfolding Pathways. Folding Rates at Midtransition

ARTICLE *in* BIOCHEMISTRY · SEPTEMBER 2001

Impact Factor: 3.02 · DOI: 10.1021/bi010625p · Source: PubMed

2 AUTHORS:

Dmitry Ivankov
Centre for Genomic Regulation
29 PUBLICATIONS **755** CITATIONS

SEE PROFILE

Alexei V Finkelstein
Russian Academy of Sciences
149 PUBLICATIONS **5,556** CITATIONS

SEE PROFILE

Theoretical Study of a Landscape of Protein Folding−Unfolding Pathways. Folding Rates at Midtransition†

Dmitry N. Ivankov and Alexei V. Finkelstein*

Institute of Protein Research, Russian Academy of Sciences, 142290, Pushchino, Moscow Region, Russia

Received March 28, 2001; Revised Manuscript Received June 4, 2001

ABSTRACT: This paper presents a new method for calculating the folding−unfolding rates of globular proteins. The method is based on solution of kinetic equations for a network of folding−unfolding pathways of the proteins. The rates are calculated in the point of thermodynamic equilibrium between the native and completely unfolded states. The method has been applied to all the proteins listed by Jackson [Jackson, S. E. (1998) *Folding Des. 3,* R81−R91] and some peptides. Although the studied protein chains differ by more than 1 order of magnitude in size and exhibit two- as well as three-state kinetics in water, and their folding rates cover more than 11 orders of magnitude, the theoretical estimates are reasonable close to the experimentally measured folding rates in midtransition (the correlation coefficient being as high as 0.78). This means that the presented theory (having no adjustable parameters at all) is consistent with the experimental observations.

Many proteins fold and unfold without any visible intermediates, by a simple two-state transition (*1*). Many other proteins exhibit a more complicated three-state transition, if all the range of conditions is taken into account; close to thermodynamic equilibrium between the native and unfolded states, however, they also demonstrate no visible intermediates and a simple two-state transition (*1, 2*). Some basic correlation between the protein size and its folding rate has been suggested during recent years (*3*−*6*), but this correlation is not as high (*2*). Recent empirical and theoretical studies of proteins that undergo a simple two-state transition have instead indicated that topology largely determines the folding rates (*7*−*10*). It has been shown also that theoretical investigations of folding/unfolding pathways are able to outline the residues of key importance for protein folding (the "folding nucleus"), provided the 3D[1] structures of native proteins are given (*11, 12*).

Theoretically, folding and unfolding are most simple in the vicinity of the point of thermodynamic equilibrium between the native and unfolded states (*2, 5, 6*). This paper presents a new method for calculating the folding−unfolding rates of globular proteins in this point, where the folding and unfolding rates coincide. The method is based on solution of kinetic equations for a network (similar to that studied in ref *12*) of the folding−unfolding pathways of the protein three-dimensional structure. The method has been applied to all the set of proteins listed in the review (*1*) and to some peptides. The studied polypeptides differ by more than 1 order of magnitude in size and exhibit two- as well as three-state kinetics in water, and their folding rates cover more than 11 orders of magnitude. Nevertheless, the theoretical estimates are reasonable close to the experimentally measured folding rates in midtransition (the correlation coefficient being as high as 0.78). This means that the presented theory (having no adjustable parameters) is consistent with the experimental observations.

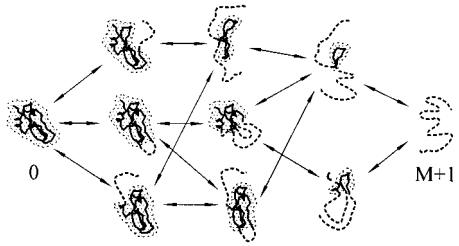

FIGURE 1: Unfolding (and folding) semifolded intermediates (only a small part of them shown) and a network of unfolding pathways. Each arrow corresponds to an elementary step, i.e., to the transition of one "chain link" (of several residues) from the globular, nativelike part of the intermediate (bold line against the background of a dotted cloud denoting the globule) to the coil (dotted line).

MATERIALS AND METHODS

Network of the Folding−Unfolding Pathways. Following the earlier developed theory (*5, 12*), we consider a simplified stepwise unfolding (Figure 1), at each step of which a "chain link" (i.e., a chain fragment of one or a few residues) is removed from the native 3D structure. A semifolded protein with some given residues fixed in their native positions and the other residues being disordered can be described as a "microstate", since a disordered link corresponds to an ensemble of many conformations. The removed links are assumed to form the random coil; they lose all their nonbonded interactions and gain the coil entropy (except that spent to close the disordered loops protruding from the remaining globule; see Figure 1 and eq 9 below). We assume

† This work was supported by the Russian Foundation for Basic Research (Grants 01-04-48322 and 01-04-48329), by the CRDF Grant RB2-2022, and by an International Research Scholar's Award to A.V.F. from the Howard Hughes Medical Institute.

* To whom correspondence should be addressed. Tel./fax: +7095-924-0493. E-mail: afinkel@vega.protres.ru.

[1] Abbreviations: 3D, three-dimensional; NMR, nuclear magnetic resonance.

10.1021/bi010625p CCC: $20.00
Published on Web 07/24/2001

that the links remaining in the globule keep their native positions and that the unfolded regions do not fold into another, non-native globule. Thus, we actually neglect non-native interactions, which makes our model similar to that of Gō (*13*). The next simplification is that we concentrate on the stability (actually, on the instability) of partly unfolded intermediates, rather than on a detailed description of the chain motions. Further, to facilitate the computations, we divide the *N*-residue chain not into *N* residues but into *L* "links" of a few (*N*/*L*) residues each. Then the network of the folding−unfolding pathways (Figure 1) includes 2^L microstates; the taken $L \leq 18$ ensures a sufficiently rapid computation.

Kinetic Equations. Let us consider an ensemble of protein molecules. The number n_i of the molecules having microstate i ($i = 0, 1, \ldots, M, M+1$, where $M = 2^L - 2$, "0" is the native state, "$M+1$" is the coil state, and "1", ..., "M" are the semifolded intermediates; see Figure 1) changes with time t according to the usual kinetic equation

$$\frac{dn_i}{dt} = -\sum_{j=0}^{M+1} k_{ij} n_i + \sum_{j=0}^{M+1} k_{ji} n_j \qquad (1)$$

where k_{ij} is the rate of transition from the ith to the jth state. These equations can be solved in a quasi-stationary approximation (*14*), i.e., under the assumption that

$$\frac{dn_i}{dt} = 0 \qquad (2)$$

for all the intermediates "1", ..., "M".

This approximation is correct when all the semifolded intermediates ("1", ..., "M") have a high free energy and therefore a low statistical weight (i.e., they are present in a very low quantity) as compared with the initial and final states (*14*). Thus, it is valid for the "all-or-none" (in thermodynamics) folding and unfolding, which is typical of proteins (*15*) close to thermodynamic midtransition (in kinetics, it is called the "two-state transition").

Taking into account eq 1, system 2 can be represented as

$$\frac{d\mathbf{n}}{dt} = -\mathbf{A}n + \mathbf{B}n_0 + \mathbf{C}n_{M+1} = 0 \qquad (3)$$

where $\mathbf{A}$ is an $M \times M$ matrix with the elements $A_{ij} =$

$$\begin{cases} \sum_{p=0}^{M+1} k_{ip} & i = j \\ -k_{ji} & i \neq j \end{cases}$$

where $1 \leq i, j \leq M$, $\mathbf{B}$ is a vector with the components $B_i = k_{0i}$ ($1 \leq i \leq M$), $\mathbf{C}$ is a vector with the components $C_i = k_{M+1,i}$ ($1 \leq i \leq M$), n_0 is the population of the initial, native state "0", n_{M+1} is the population of the final, coil state "$M+1$", and $\mathbf{n}$ is a vector with the components n_i ($i = 1$, ..., M). The changes in the populations of the final and initial states are

$$\frac{dn_{M+1}}{dt} = -\sum_{p=1}^{M} k_{M+1,p} n_{M+1} + \mathbf{C}^*\mathbf{n} = -\frac{dn_0}{dt} \qquad (4)$$

where $\mathbf{C}^*$ is a vector with the components $C_i^* = k_{i,M+1}$ ($1 \leq i \leq M$).

The vector

$$\mathbf{n} = \mathbf{A}^{-1}\mathbf{B}n_0 + \mathbf{A}^{-1}\mathbf{C}n_{M+1} \qquad (5)$$

is a solution of system 3, and from eq 4, the decrease in the population of the initial state is

$$\frac{dn_0}{dt} = -[\mathbf{C}^*\mathbf{A}^{-1}\mathbf{B}]n_0 + [\sum_{p=1}^{M} k_{M+1,p} - \mathbf{C}^*\mathbf{A}^{-1}\mathbf{C}]n_{M+1} \qquad (6)$$

The vector $\mathbf{X} = \mathbf{A}^{-1}\mathbf{B}$ is obtained from the solution of equation $\mathbf{Ax} = \mathbf{B}$ by Zeidel's iteration method (*16*). Thus, $K_u = \mathbf{C}^*\mathbf{X}$ is the rate constant for the unfolding, i.e., the rate of transition from "0" (the native state) to "$M+1$" (the coil) through the network of the unfolding pathways. Then the time of protein unfolding is $1/K_u$. At the same time, $K_f = \sum_{p=1}^{M} k_{M+1,p} - \mathbf{C}^*\mathbf{A}^{-1}\mathbf{C}$ is the rate of folding. At the equilibrium between the native and unfolded state (i.e., when their free energies F_0 and F_{M+1} are equal), $K_u = K_f$.

For the elementary transition rates, we use the approximation given by eq 7, which is usual (*17*) for the Metropolis scheme of kinetic simulations based on the Monte Carlo method (*18*):

$$k_{ij} = k_0 \times \begin{cases} 0 & \text{if transition } i \rightarrow j \text{ is physically impossible} \\ 1 & \text{if transition } i \rightarrow j \text{ is possible and } F_i \geq F_j \\ \exp\left(-\frac{F_j - F_i}{RT}\right) & \text{if transition } i \rightarrow j \text{ is possible and } F_i < F_j \end{cases} \qquad (7)$$

(Transition is "possible" when j is obtained from i by removing one link from the nativelike part of i or by adding one link to this nativelike part.) F_i is the free energy of microstate i, R is the gas constant, T is the temperature, and k_0 is the rate constant for a downhill (in free energy) step.

Such estimates of the elementary transition rates are certainly rough, but they can be used when the rate of the overall process is determined by the instability of the transition state(s) (as in the problem we solve now) rather than by a detailed description of the chain motions.

Since we use "chain links" of $\lambda = N/L$ residues (where N is the number of residues in the chain and L is the number of the links), k_0 must depend on λ: the unfolding of a chain link of 10 residues is certainly slower than unfolding of one residue. It is reasonable to assume that propagation of the unfolded region by a downhill sequential unfolding of λ residues takes, on the average, approximately λ times more time than a downhill unfolding of one residue. Thus, the elementary transition time is proportional to the number λ of residues in the chain link. Since unfolding of a residue may be considered as a two-step unfolding (unfolding of the backbone and of the side-chain), we estimate k_0 as

$$k_0 = 1/(2\tau_0 N/L) \qquad (8)$$

where $\tau_0 = 10$ ns is the experimentally measured time of unfolding of one α-helical residue (*19*) (where only the backbone unfolds). However, the theory-to-experiment correlation virtually does not change (see below) when we use

Table 1: List of Proteins and Polypeptides[a]

no.	protein	ref	PDB	length	$\ln(t)_w$	$\ln(t)_{mt}$	$\ln(t)_t$
1	α-helix of 21 residues[b]	21	*b*	21	−15.5[b]	−15.5[b]	−13.7
2	β-hairpin from Ig binding protein (41−56)	22	1PGB	16	−12.0	−12.0	−14.5
3	monomeric λ-repressor (λ_{6-85}), WT	23	1LMB	80	−8.5	−6.5	−1.8
4	monomeric λ-repressor (λ_{6-85}), G46A;G48A	23	*c*	80	−11.4	−4.9	*c*
5	ACBP bovine	24	2ABD	86	−6.6	1.2	−4.9
6	ACBP rat	25	*c*	86	−6.0	1.5	*c*
7	ACBP yeast	25	*c*	86	−8.3	0.2	*c*
8	cytochrome *c*	26	1HRC	104	−7.9	−2.3	3.4
	cytochrome *c* [horse, oxidized (Fe^{III})]	27			−6.0	*d*	
9	cytochrome *c* [yeast, oxidized (Fe^{II})]	27	1YCC	103	−9.6	*d*	0.8
10	tendamistat	28	2AIT	74	−4.2	5.0	1.7
11	CspB (*Bacillus subtilis*)	29	1CSP	67	−7.0	−3.2	−6.7
		30			−6.5	−2.7	
12	CspB (*S. caidolyticus*)	30	*c*	66	−7.2	−0.1	*c*
13	CspB (*Thermotoga maritima*)	30	*c*	68	−6.3	2.6	*c*
14	CspA	31	1MJC	69	−5.3	−1.7	−3.4
15	SH3 domain (α-spectrin)	32	1SHG	57	−2.1	1.8	−0.8
		33			−1.4	3.0	
16	SH3 domain (src)	34	1SRL	56	−4.0	0.3	−1.0
17	SH3 domain (PI3 kinase)	35	1PKS	76	1.1	4.8	4.9
18	SH3 domain (fyn)	36	1SHF	59	−4.5	3.2	−2.9
19	9FN-III	37	1FNF	90	0.9	*d*	8.2
20	twitchin	*e*	1WIT	93	−0.4	5.8	8.5
21	tenascin (short form)	38	1TEN	89	−1.1	4.2	7.4
22	tenascin (long form)	38	*c*	92	−1.8	6.7	*c*
23	10FN-III	38	1FNF	94	−5.0	*d*	7.2
24	Ci2	39	2CI2	64	−3.9	3.7	−1.3
25	activation domain procarboxypeptidase A2 (ADAh2)	40	1AYE	78	−6.8	−1.4	0.8
26	arc repressor (dimer single chain)[f]	41	*c*	121	−9.3	−2.5	*c*
27	IgG binding domain of streptococcal protein L	42	2PTL	62	−4.1	1.9	2.2
28	spliceosomal protein U1A	43	1URN	96	−5.8	2.7	5.3
29	Hpr (histidine-containing phosphocarrier protein)	44	1HDN	85	−2.7	3.0	4.7
30	FKBP12	45	1FKB	107	−1.5	5.4	7.1
31	muscle-AcP	46	1APS	98	1.5	7.5	7.1
32	villin 14T	47	2VIK	126	−6.8	0.2	5.9
33	ubiquitin, WT*[g] V26A	48	*c*	76	−4.6	0.4	*c*
34	ubiquitin, WT*[g]	48	1UBQ	76	−5.9	1.6	−6.3
	ubiquitin, WT* + 0.4 M Na_2SO_4[g]	48			−6.8	1.0	
35	barstar	49	1BTA	89	−3.4	1.4	1.6
36	CD2, pH 7.0	50	1HNG	97	−1.8	3.4	3.8
	CD2, pH 4.5	50			−2.6	*d*	
37	barnase	51	1BNI	108	−2.6	4.3	7.9
38	suc 1	*e*	*c*	113	−4.2	*d*	*c*
39	lysozyme (hen egg white)	52	1HEL	129	−1.3	7.4	4.4
40	CheY	53	3CHY	128	−1.0	1.6	−1.2
41	p16	*e*	*c*	148	−3.5	*d*	*c*
42	GroEL apical domain (191−345)	54	1DK7	146	−0.8	2.1	8.8
43	ribonuclease H (*Escherichia coli*), pH 5.5	55	2RN2	155	0.5	7.6	6.5
		56			−1.4	3.5	
44	N-terminal domain from PGK	52	1PHP	175	−2.3	1.2	9.1
45	C-terminal domain from PGK	57	1PHP	208	3.5	8.5	9.1

[a] List of proteins (refs *23*−*57*) is taken from ref *1*. If some protein was investigated at different temperatures, the experiment at the temperature closest to 25 °C is presented here. The midtransition point was defined by crossing of the extrapolated branches of the chevron plot. If several 3D structures (i.e., several PDB entries) are available for some protein, the X-ray structure is used in our calculation; otherwise, the averaged NMR structure; otherwise the NMR structure, model 1. The proteins numbered from 3 to 33 exhibit the two-state folding within the whole range of experimental conditions; from 34 to 45 exhibit the three-state folding when the native state is much more stable than the denatured one. [b] There is no PDB entry for the Ala-rich 21-residue α-helix studied; the ideal $(Ala)_{21}$ α-helix was used in our calculation of the transition time. Both $\ln(t)_w$ and $\ln(t)_{mt}$ values refer to 25 °C. [c] There is no PDB entry; consequently, no calculation of the transition time has been done. [d] The lowest part of the chevron (the midtransition) is not defined; consequently, there is no experimental estimate of the folding/unfolding rate in the midtransition. [e] Only a personal communication or unpublished observations are cited in ref *1*. [f] Two Arc repressor monomers (of 53 residues, PDB entries 1ARR and 1PAR) have been connected by a glycine-rich linker in the folding experiment; there is no PDB entry for such a construction. [g] WT* is F45W mutant; the PDB entry 1UBQ was used in our calculations of the transition time. The experimental data from the Table 2 of ref *1* are used since it has been shown that ubiquitin folds via three-state kinetics in water (*49*).

a link size-independent $k_0 = 1/(10\ \text{ns}) = 10^8\ \text{s}^{-1}$, as it should be when the overall rate of a process is determined by instability of the transition state(s) rather than by a detailed description of the chain motions.

Free Energy Estimate. The free energy of microstate *i* (having ν_i disordered residues and the given $N - \nu_i$ residues keeping their native positions and conformations) is taken (*12*) as

$$F_i = \epsilon \sum_{(\alpha<\beta-1)\in \text{nat}.i}\sum \delta_{\alpha\beta} - T[\nu_i\sigma_1 + \sum_{\text{loops}\in i} S_{\text{loop}}] \qquad (9)$$

The first double sum is taken over all nonneighbor residues α, β keeping their native positions in *i*. The second sum is taken over all the closed disordered loops protruding from the globular, nativelike part of *i* (e.g., the intermediate given in the middle of the lower row of Figure 1 contains one

closed loop and one tail, and the next one in this row has both the N- and C-terminal tails and one closed loop). $\delta_{\alpha\beta}$ is the number of atom−atom contacts (at the distance of <5 Å) between residues α and β in the native 3D structure, ϵ is the energy of one atom−atom contact, T is the temperature, and σ_1 is the entropy difference between the coil and native states of a residue (according to ref *15*, we take $\sigma_1 = 2.3R$, R being the gas constant). The native state (with $\nu_0 = 0$) and the coil (with $\nu_{M+1} = N$) have equal free energies at the midtransition; i.e., ϵ and T are connected here by equation $\epsilon/T = -N\sigma_1/\sum\sum_{1\le\alpha<\beta-1\le N} \delta_{\alpha\beta}$. Thus, the ϵ/T value is individual for each protein and refers to its midtransition (cf. ref *10*). Note that the rate constants (eq 7) depend just on the ϵ/T ratio rather than on ϵ or T separately. The entropy spent to close a disordered loop between fixed residues α and β is estimated (*6*, *12*) as

$$S_{\text{loop}} = -\frac{5}{2}R\ln|\alpha-\beta| - \frac{3}{2}R\frac{r_{\alpha\beta}^2 - a^2}{2aA|\alpha-\beta|} \qquad (10)$$

Here $r_{\alpha\beta}$ is the distance between the C_α atoms of residues α and β, $a = 3.8$ Å is the distance between the neighbor C_α atoms in the chain, and A is the persistent length for a polypeptide (according to ref *20*, we take $A = 20$ Å).

RESULTS AND DISCUSSION

To compare the developed theory with experiment, we took (Table 1) all the proteins from the list of Jackson (*1*) and two short peptides [α-helix (*21*) and β-hairpin (*22*)] studied by Eaton's group (*21*, *22*). For all the proteins we took [or computed, from ref *1* or from the original papers (*23*−*57*)] the rates of their folding/unfolding in the point of thermodynamic equilibrium between the native and the denatured forms. The theoretical unfolding rates were computed, as described above, from unfolding the 3D protein structures, taken from the PDB (*58*) according to their codes listed by Jackson. The computed unfolding times refer to $L = 18$ except for the β-hairpin which consists of 16 residues (thus, $L = 16$ is used here). This number of chain links was taken to obtain the most detailed model of unfolding, on one hand, and to have a reasonable time of calculation, on the other. However, the results for $L = 15$ and $L = 12$ are nearly the same as for $L = 18$ (data not shown). The elementary transition rate k_0 is estimated from eq 8. However, the correlation with experiment virtually does not change when the simplest estimate, $k_0 = 10^8$ s^{-1}, is used for all proteins and peptides (data not shown).

Figure 2 presents a comparison of theoretical and experimental data. One can see that the points in Figure 2 gather around the diagonal of the plot, which means a rather good approximation of the experimental folding times by their theoretical estimates. The coefficient of correlation is as high as 0.78 (0.57, if the short peptides are excluded from the comparison). This correlation is significant for the following reasons: First, we did not use any adjustable parameters in our theory; second, the theory is based on very simple estimates of kinetic constants and free energies and on a straightforward analysis of the folding/unfolding pathways.

The presented theory is general. It works nearly equally for small peptides, middle-size two-state folding proteins, and larger three-state folders (the coefficient of correlation is 0.78 for all the set of peptides and proteins listed in Table 1, 0.59 for two-state folders, and 0.44 for three-state folders). In contrast, the well-known "order parameter" (CO) of Plaxco and Baker (*7*) was developed for the two-state folding proteins only and correlates with their folding times at the level as high as 0.8−0.9 (*7*, *59*). But the CO fails to correlate with folding times of short peptides (*10*) and of the three-state folders: for the latter, the correlation is even negative, −0.55, according to the data taken from ref *1*.

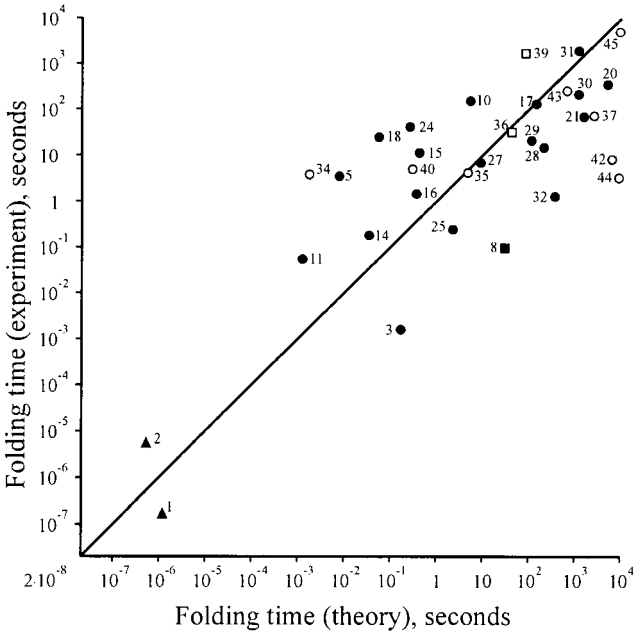

FIGURE 2: Plot of the experimental vs the theoretically estimated folding/unfolding times in the points of midtransition for the polypeptides and proteins listed in Table 1: triangles, α-helix and β-hairpin; squares, proteins with S−S bonds; open symbols, proteins with three-state folding kinetics at a low denaturant concentration. The number near the point corresponds to the protein number in the Table 1. If there are multiple experimental data for some protein (see Table 1), the time corresponding to the mean value of ln(t) is shown.

However, it should be noted that the coincidence between theory and experiment is still far from being perfect: at extreme, some experimental values differ from the theoretical estimates by 3 orders of magnitude. This requires the farther development of a more accurate theory and, in particular, a more accurate estimate of the free energies of the semifolded intermediates and the kinetic constants.

ACKNOWLEDGMENT

We are grateful to O.V. Galzitskaya, S.E. Jackson, K.V. Plaxco, and D. Baker for fruitful discussions.

REFERENCES

1. Jackson, S. E. (1998) *Folding Des. 3*, R81−R91.
2. Galzitskaya, O. V., Ivankov, D. N., and Finkelstein, A. V. (2001) *FEBS Lett. 489*, 113−118.
3. Thirumalai, D. (1995) *J. Phys.* (*Orsay, France*) *15*, 1457−1469.
4. Gutin, A. M., Abkevich, V. I., and Shakhnovich, E. I. (1996) *Phys. Rev. Lett. 77*, 5433−5436.
5. Finkelstein, A. V., and Badretdinov, A. Ya. (1997) *Folding Des. 2*, 115−121.
6. Finkelstein, A. V., and Badretdinov, A. Ya. (1997) *Mol. Biol. (Moscow) 31*, 391−398.

7. Plaxco, K. V., Simons, K. T., and Baker, D. (1998) *J. Mol. Biol. 277*, 985−994.
8. Fersht, A. R. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 1525−1529.
9. Baker, D. (2000) *Nature 405*, 39−42.
10. Muñoz, V., and Eaton, W. A. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 11311−11316.
11. Alm, E., and Baker, D. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 11305−11310.
12. Galzitskaya, O. V., and Finkelstein, A. V. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 11299−11304.
13. Ueda, Y., Taketomi, H., and Gō, N. (1975) *Int. J. Pept. Protein Res. 7*, 445−459.
14. Emmanuel, N. M., and Knorre, D. G. (1984) *The course of chemical kinetics*, pp 282−290, Vysshaya Shkola, Moscow.
15. Privalov, P. L. (1979) *Adv. Protein Chem. 33*, 167−241.
16. Bahvalov, N. S., Zhidkov, N. P., and Kobelkov, G. M. (1987) *Numerical Methods*, pp 287−292, Nauka, Moscow.
17. Abkevich, V. I., Gutin, A. M., and Shakhnovich, E. I. (1994) *Biochemistry 33*, 10026−10036.
18. Hilhorst, H. J., and Deutch, J. M. (1975) *J. Chem. Phys. 63*, 5153−5161.
19. Zana, R. (1975) *Biopolymers 14*, 2425−2428.
20. Flory, P. J. (1969) *Statistical Mechanics of Chain Molecules*, pp 274−281, Interscience, New York.
21. Thompson, P. A., Eaton, W. A., and Hofrichter, J. (1997) *Biochemistry 36*, 9200−9210.
22. Muñoz, V., Thompson, P. A., Hofrichter, J., and Eaton, W. A. (1997) *Nature 390*, 196−199.
23. Burton, R. E., Huang, G. S., Daugherty, M. A., Fullbright, P. W., and Oas, T. G. (1996) *J. Mol. Biol. 263*, 311−322.
24. Kragelund, B. B., Robinson, C. V., Knudsen, J., Dobson, C. M., and Poulsen, F. M. (1995) *Biochemistry 34*, 7217−7224.
25. Kragelund, B. B., Højrup, P., Jensen, M. S., Schjerling, C. K., Juul, E., Knudsen, J., and Poulsen, F. M. (1996) *J. Mol. Biol. 256*, 187−200.
26. Chan, C.-K., Hu, Y., Takahashi, S., Rousseau, D. L., Eaton, W. A., and Hofrichter, J. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 1779−1784.
27. Mines, G. A., Pascher, T., Lee, S. C., Winkler, J. R., and Gray, H. B. (1996) *Chem. Biol. 3*, 491−497.
28. Schönbrunner, N., Koller, K.-P., and Kiefhaber, T. (1997) *J. Mol. Biol. 268*, 526−538.
29. Schindler, T., Herrler, M., Marahiel, M. A., and Schmid, F. X. (1995) *Nat. Struct. Biol. 2*, 663−673.
30. Perl, D., Welker, C., Schindler, T., Schröder, K., Marahiel, M. A., Jaenicke, R., and Schmid, F. X. (1998) *Nat. Struct. Biol. 5*, 229−235.
31. Reid, K. L., Rodriguez, H. M., Hillier, B. J., and Gregoret, L. M. (1998) *Protein Sci. 7*, 470−479.
32. Viguera, A. R., Martínez, J. C., Filimonov, V. V., Mateo, P. L., and Serrano, L. (1994) *Biochemistry 33*, 2142−2150.
33. Viguera, A. R., Serrano, L., and Wilmanns, M. (1996) *Nat. Struct. Biol. 3*, 874−880.
34. Grantcharova, V. P., and Baker, D. (1997) *Biochemistry 36*, 15685−15692.
35. Guijarro, J. I., Morton, C. J., Plaxco, K. W., Campbell, I. D., and Dobson, C. M. (1998) *J. Mol. Biol. 276*, 657−667.
36. Plaxco, K. W., Guijarro, J. I., Morton, C. J., Pitkeathly, M., Campbell, I. D., and Dobson, C. M. (1998) *Biochemistry 37*, 2529−2537.
37. Plaxco, K. W., Spitzfaden, C., Campbell, I. D., and Dobson, C. M. (1997) *J. Mol. Biol. 270*, 763−770.
38. Clarke, J., Hamill, S. J., and Johnson, C. M. (1997) *J. Mol. Biol. 270*, 771−778.
39. Jackson, S. E., and Fersht, A. R. (1991) *Biochemistry 30*, 10428−10435.
40. Villegas, V., Azuaga, A., Catasús, Ll., Reverter, D., Mateo, P. L., Avilés, F. X., and Serrano, L. (1995) *Biochemistry 34*, 15105−15110.
41. Robinson, C. R., and Sauer, R. T. (1996) *Biochemistry 35*, 13878−13884.
42. Scalley, M. L., Yi, Q., Gu, H., McCormack, A., Yates, J. R., III, and Baker, D. (1997) *Biochemistry 36*, 3373−3382.
43. Silow, M., and Oliveberg, M. (1997) *Biochemistry 36*, 7633−7637.
44. Van Nuland, N. A. J., Meijberg, W., Warner, J., Forge, V., Scheek, R. M., Robillard, G. T., and Dobson, C. M. (1998) *Biochemistry 37*, 622−637.
45. Main, E. R. G., Fulton, K. F., and Jackson, S. E. (1999) *J. Mol. Biol. 291*, 429−444.
46. Van Nuland, N. A. J., Chiti, F., Taddei, N., Raugei, G., Ramponi, G., and Dobson, C. M. (1998) *J. Mol. Biol. 283*, 883−891.
47. Choe, S. E., Matsudaira, P. T., Osterhout, J., Wagner, G., and Shakhnovich, E. I. (1998) *Biochemistry 37*, 14508−14518.
48. Khorasanizadeh, S., Peters, I. D., and Roder, H. (1996) *Nat. Struct. Biol. 3*, 193−205.
49. Schreiber, G., and Fersht, A. R. (1993) *Biochemistry 32*, 11195−11203.
50. Parker, M. J., Dempsey, C. E., Lorch, M., and Clarke, A. R. (1997) *Biochemistry 36*, 13396−13405.
51. Matouschek, A., Kellis, J. T., Jr., Serrano, L., Bycroft, M., and Fersht, A. R. (1990) *Nature 346*, 440−445.
52. Parker, M. J., Spencer, J., and Clarke, A. R. (1995) *J. Mol. Biol. 253*, 771−786.
53. Muñoz, V., Lopez, E. M., Jager, M., and Serrano, L. (1994) *Biochemistry 33*, 5858−5866.
54. Golbik, R., Zahn, R., Harding, S. E., and Fersht, A. R. (1998) *J. Mol. Biol. 276*, 505−515.
55. Raschke, T. M., and Marqusee, S. (1997) *Nat. Struct. Biol. 4*, 298−304.
56. Yamasaki, K., Ogasahara, K., Yutani, K., Oobatake, M., and Kanaya, S. (1995) *Biochemistry 34*, 16552−16562.
57. Parker, M. J., Sessions, R. B., Badcoe, I. G., and Clarke, A. R. (1995) *Folding Des. 1*, 145−156.
58. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rogers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *Eur. J. Biochem. 80*, 319−324.
59. Plaxco, K. V., Simons, K. T., Ruczinski, I., and Baker, D. (2000) *Biochemistry 39*, 11177−11183.

BI010625P