Unifying model for two-state and downhill protein folding

D. Mi, W. Q. Meng, and Y. Q. Sun

Phys. Rev. E **83**, 041901 — Published 1 April 2011

DOI: 10.1103/PhysRevE.83.041901

A unifying model for two-sate and downhill protein folding

D. Mi,[1,*] W. Q. Meng,[2] and Y. Q. Sun[2]

[1]*Department of Physics, Dalian Maritime University, Dalian 116026, P. R. China*

[2]*Institute of Environmental Systems Biology,*

Dalian Maritime University, Dalian 116026, P. R. China

Abstract

A protein folding model is proposed at the amino acid level, in which the folding process is divided into two successive stages: the rate-determining step dominated by the "stochastic interactions" of solvent molecules and the rapid phase dominated by the "order interactions" among atoms in polypeptide. The master equation approach is used to investigate the folding kinetics, and an analytical treatment of the master equation yields a simple three-parameter expression for folding time. It is found that both two-sate and downhill protein folding kinetics can be described by a unifying model.

PACS numbers: 87.15.Cc,87.15.hm

* mid@dlmu.edu.cn

I. INTRODUCTION

Protein folding is the physical process by which a random coil polypeptide folds into its unique native structure. As an integral part of the "central dogma" of molecular biology, protein folding is sometimes dubbed the "second half of the genetic code" [1]. In a living cell, a protein is synthesized on a ribosome that makes a protein chain residue by residue. However, up to now, it is difficult to follow the in vivo folding of a nascent protein chain against the background of the huge ribosome. Nearly half a century ago, Anfinsen *et al.* discovered that a globular protein is capable of spontaneous folding in vitro under suitable conditions [2, 3]. Such observations allow one to detach, at least to the first approximation, the study of protein folding physics from the protein folding study in vivo. Hence, experimentalists and theoreticians have mainly focused their efforts on these proteins which can spontaneously fold into their native states in test tube [4, 5].

In vitro experiments showed that many small proteins (less than ~ 110 amino acids) can fold with simple two-state kinetics [6], and can do so amazingly quickly than that by randomly exploring all possible conformations of their unfolded states at the atomic level [7]. Different models have been established to address the speed principle, *e.g.*, framework model [8, 9], hydrophobic collapse model [10, 11], nucleation-condensation mechanism [12, 13], zipping and assembly mechanism [14, 15], *etc.* However, these microscopic theories of folding kinetics generally cannot be responsible for the vast range of folding rates ($10^5 s^{-1} \sim 10^{-1} s^{-1}$) observed in experiments [6, 16], and cannot explain the quantitative relationships between protein folding rates and their topological parameters. The topological parameter here refers to the set of non-covalent contacts. The first famous topological parameter comes from the pioneering work of Plaxco *et al.* [17], where they demonstrated that a topological parameter called "relative contact order" (*RCO*) is statistically linearly correlated to the logarithm of folding rate constant ($\ln k$) of small two-state protein. Since then, a number of other protein topological parameters have been discovered that also have the almost linear relationships with the logarithms of folding rate constant. Such topological parameters include long-range order (*LRO*) [18], total contact distance (*TCD*) [19], absolute contact order (*ACO*) [17, 20], total number of native contacts (*N*) [21], *etc.*

One prediction that emerged from the energy landscape theory of protein folding is that the energy barriers between the folded and the unfolded states are minimized when there is an extreme energetic bias towards the native state, such as at low temperature or in the absence of denaturant [22]. Under such conditions, two-state folders are predicted to be turned into downhill

folding which proceeds through an array of temporary conformations with broad distributions of folding times. Due to the significance of downhill folding in benchmarking molecular dynamics simulations and testing protein folding theories, it has received lot of attention in theory [23, 24], simulation [25–27], and experiment [28, 29] in recent years. The downhill folding is expected to take place very rapidly, approaching the speed limit of folding, which is about $L/100\ \mu s$ (where L is the number of residues) [30].

In this paper, we proposed a generic protein folding model at the amino acid level. The folding process is divided into two successive stages: the first stage is rate-determining step of protein folding, which can be dealt with analytically by solving the master equation. The second one is a rapid free energy downhill stage, which can be regarded as an almost instantaneous process. The model and its conclusions showed that both two-sate and downhill folding kinetics can be described by a unifying model.

II. THEORETICAL MODEL

At the atomic level, there exist vast, complex interactions during protein folding. However, many studies show that the folding rates and mechanisms appear to be largely determined by the topology of the native states [31]. Thus, it is possible to use the coarse-grained model based on the folding rate and native structure to predict some overall features of protein folding.

Suppose a protein has N native contacts in its folded state, we assume that the protein folding process is mainly driven by two different interactions during different folding phases: From the beginning of folding to the formation of the Nth contact, the "stochastic interactions" of solvent molecules dominate. This progress is described by the evolution of the number of native contacts in time, which is a Markov process with transitions between neighboring configurations with n and $(n \pm 1)$ native contacts ($n = 1, 2, ..., N-1$). After that, the phase is dominated by the "order interactions" among atoms in polypeptide, and the conformational changes between forming all N native contacts to final folded state is regarded as an almost instantaneous process. Hence, the first phase is the rate-determining step of protein folding, and is approximately considered as the protein folding time here. Obviously, such folding process is also a funnel one: there are many parallel microscopic routes at the beginning of folding, and fewer and fewer sequential routes with the increase of the number of native contacts, and unique native state at the end of folding.

It should be pointed out here that although both the state with the Nth native contact coming

into existence and native state have the same native contacts, their structures are different at atomic level. During the rapid free energy downhill process from the former state to the latter one, a lot of atoms constituting polypeptide may quickly rearrange their relative positions owing to the interatomic interactions, but all these adjustments in atomic positions are too slight to lead to the change of the formed N native contacts.

In addition, although the second stage is a free energy downhill process, the model does not restrict the overall change trend of free energy in the first stage. If the first stage proceeds uphill in free energy, the folding process corresponds to a two-state folding, and the state with the Nth contact coming into existence corresponds to the "transition state". If the first stage also goes downhill in free energy, just as stage two, the folding process corresponds to one-state folding, which can also be referred to as downhill folding.

The model presented here is the extension of our previous one [32]: it is no longer limited to two-state folding kinetics, but may also include downhill one; a more strict mathematical treatment (the master equation approach) to the model is used to investigate the folding kinetics of small proteins; in addition, instead of the micro parameters (the probability to form or break a native contact and the time of forming or breaking a native contact) in original model, the macro parameters (mean rates of forming and breaking a native contact) would appear in the description of the model.

The above protein folding picture is somewhat similar to the Zwanzig model [33, 34], which is based on the idea of the degree of "correctness" of a protein configuration compared with the native state. Protein folding in Zwanzig model is a random walk process in the space of the degree of "correctness". This is different from ours, in which protein folding process is obviously divided into two different stages. Although a generic reaction coordinate including but not limited to the number of native contacts was considered in Zwanzig model, no any particular definition of "correctness" in physics or chemistry was given obviously. Hence, although the model and its conclusions can qualitatively explain some general folding natures independent of actual protein structures, it could not quantitatively explain the folding kinetics of any specific protein.

III. MATHEMATICAL DERIVATION

The mathematical aspects of the first phase of above model by the master equation approach is now outlined. Let $\rho(n,t)$ be the probability that the polypeptide has formed the nth native contact

at a given time t. This probability will change by gains from the configurations with $(n-1)$ and $(n+1)$ native contacts and losses to the configurations with $(n-1)$ and $(n+1)$ native contacts. Let k_+ and k_- be the rates of forming and breaking a native contact, respectively. Then, $\rho(n,t)$ satisfies the following master equation,

$$\frac{d}{dt}\rho(n,t) = k_+\rho(n-1,t) + k_-\rho(n+1,t) - k_+\rho(n,t) - k_-\rho(n,t). \tag{1}$$

For a protein which has N native contacts in the folded state, the discrete variable n is bounded between 0 and N ($n = 0, 1, 2, ..., N$). The master equation Eq.(1) is valid for $n = 1, 2, ..., N-1$, but meaningless at the boundaries $n = 0$ and N. Two boundary equations should be added as closure conditions. Obviously in our case, $n = 0$ is a reflecting boundary, while $n = N$ is an absorbing boundary. Then, two boundary conditions can be given by

$$\frac{d}{dt}\rho(0,t) = k_-\rho(1,t) - k_+\rho(0,t), \tag{2}$$

$$\frac{d}{dt}\rho(N,t) = k_+\rho(N-1,t). \tag{3}$$

The first-passage time to the absorbing boundary, which is approximately equal to the protein folding time in the present model, is the time that a protein starts from some arbitrary initial the *nth* native contact, to arrive the *Nth* native contact. The mean first-passage time $\tau(n)$ is the average of this time over all ways of getting from n to N. The fundamental equation that determines the mean first-passage time is [35]

$$k_+ \left[\tau(n+1) - \tau(n)\right] + k_- \left[\tau(n-1) - \tau(n)\right] = -1, \tag{4}$$

with the boundary conditions

$$\tau(-1) = \tau(0), \tag{5}$$

$$\tau(N) = 0. \tag{6}$$

It is easy to prove that the solution of Eq.(4) under the above conditions can be expressed as,

$$\tau(n) = \sum_{l=n}^{N} \phi(l) \sum_{m=1}^{l} [k_+(m)\phi(m)]^{-1}, \tag{7}$$

with,

$$\phi(a) = \prod_{b=1}^{a} \frac{k_-(b)}{k_+(b)}. \tag{8}$$

Although the rate constants of forming different contacts are different in general, as an approximation of the actual situation, the rate constant of forming a specific contact is replaced with an average value in this work. So is the rate constant of breaking a specific contact. Then, all the rate constants are independent of position, i.e., $k_-(b) = k_-$, and $k_+(b) = k_+$. Thus, Eq.(8) becomes

$$\phi(a) = K^a. \tag{9}$$

where, $K \equiv k_-/k_+$, is an important parameter in the present model.

It is a natural assumption that there is not any native contact at the beginning of protein folding. Then, take $n = 0$ in Eq.(7) and use Eq.(9), we can obtain the mean folding time,

$$\tau = \frac{K}{(K-1)^2 k_+} K^N - \frac{1}{(K-1)k_+} N - \frac{K}{(K-1)^2 k_+}. \tag{10}$$

Eq.(10) is a simple three-parameter expression for folding time in terms of the mean rates of forming and breaking a native contact k_+, k_-, and the total number of native contacts N. For two-state protein folding, as more and more formations of native contacts correspond to proceeding uphill in free energy, the rate of breaking a contact is higher than the rate of forming it for most contacts. This implies that $k_- > k_+$, or $K > 1$ is generally valid for two-state folding kinetics. In this case, for the actual N values of most small two-state proteins (see next section), comparing with the first term, the last two terms at the right hand of Eq.(10) are too small to considerably contribute to τ. So, we get

$$\tau \approx \frac{K}{(K-1)^2 k_+} K^N. \tag{11}$$

The above equation implies that, for two-state folding kinetics, folding time exponentially increases with the total number of native contacts.

Notice that the mean folding time τ is the reciprocal of the mean folding rate constant k, the logarithm of rate constant of protein folding can be approximately given by

$$\ln k = a - b \cdot N. \tag{12}$$

where, $a = \ln[(K-1)^2 k_+]/K$, $b = \ln K$.

In contrast with two-state folding, as more and more formations of native contacts correspond to proceeding downhill in free energy in downhill folding, the rate of forming a contact is higher than the rate of breaking it for most contacts. This implies that $k_+ > k_-$, or $K < 1$ is generally valid for downhill folding kinetics. In this case, comparing with the second term, the first term and the last term at the right hand of Eq.(10) are too small to considerably contribute to τ. Thus, we have

$$\tau \approx \frac{1}{(1-K)k_+} N. \tag{13}$$

Different with two-state folding case, Eq.(13) implies that folding time linearly increases with the total number of native contacts for downhill folding kinetics. And further, the logarithm of rate constant of downhill folding can be approximately given by

$$\ln k = c - \ln N. \tag{14}$$

where, $k = 1/\tau$, $c = \ln[(1-K)k_+]$.

IV. COMPARISON WITH EXPERIMENTS

Let us now first consider a set of 66 small two-state proteins, where all of their folding rates and native structures were measured by experiments (Among them, 47 proteins come from reference [36], the others come from a protein folding kinetics database [37]). As is well known, protein folding rates are sensitive to a wide variety of environmental conditions, such as temperature, pH, buffer, the concentration of denaturant, etc. To make the protein folding rates reported in Table I as comparable as possible, for the same protein studied in several different conditions, the measurement done at conditions closest to the "standard conditions": $25^\circ C$ at pH 7.0, 50 mM buffer, is selected [37, 38].

By linear regression analysis, we find that the logarithm of folding rate constant $\ln k$ is negative linear correlated with the number of native contacts N

$$\ln k = 10.411 - 0.071 \cdot N, \tag{15}$$

with a correlation coefficient $R = -0.81$. The p-value associated with correlation, $p < 0.0001$, is extremely low, suggesting that the observed correlation is highly unlikely to have arisen by chance in the 66 member test set. Fig.1 shows the linear regressions of $\ln k$ vs. N together with the experimental points. Here, we assume that two residues in the folded protein are in contact if the straight-line distance between their C_α atoms is less than d, and if there are more than l residues between them along the chain. To calculate the total number of native contacts N, we take the cutoffs $d = 7.00$Å and $l = 11$ residues. It is found that cutoffs d from 6.00Å to 8.00Å, l from 4 to 15 residues, do not significantly affect the correlations described in this work. The correlation coefficient remains greater than 0.76.

Different from the native contact in relative contact order, two residues forming a native contact here (cutoffs $d \leq 7.00$Å, $l \geq 11$ residues) are close in space but far in the sequence. In this way, for example, two residues connected by backbone hydrogen bond in alpha helix do not form a native contact in the present model. In fact, this kind of native contact emphasizes the importance of long-range interactions in protein folding.

Comparing the theoretical prediction Eq.(12) and the experimental result Eq.(15), for the two-state folding kinetics, we can determine the two mean rate constants of forming and breaking a contact,

$$k_+ = 6.5 \times 10^6 s^{-1}, k_- = 7.0 \times 10^6 s^{-1}. \tag{16}$$

Eight proteins in table 1, 1E0L, 1ENH, 1L2Y, 1LMB, 1PIN, 1VII, 2A3D and 2PDD, are the so-called ultrafast folding proteins [30]. The mean value of their total numbers of native contacts (nearly 10) is significantly less than that of the other proteins (nearly 70). According to the present folding model, for a two-state folding protein, it is the total number of native contacts determines the height of free energy barrier, and thus the folding rate. The ultrafast folding protein with two-state kinetics has smaller N value, and thus lower folding free energy barrier, and thus faster folding rate.

Assume that a protein has at least one native contact, take $N = 1$ in Eq.(10), and use k_+ value given in Eq.(16), we can estimate the fastest folding time,

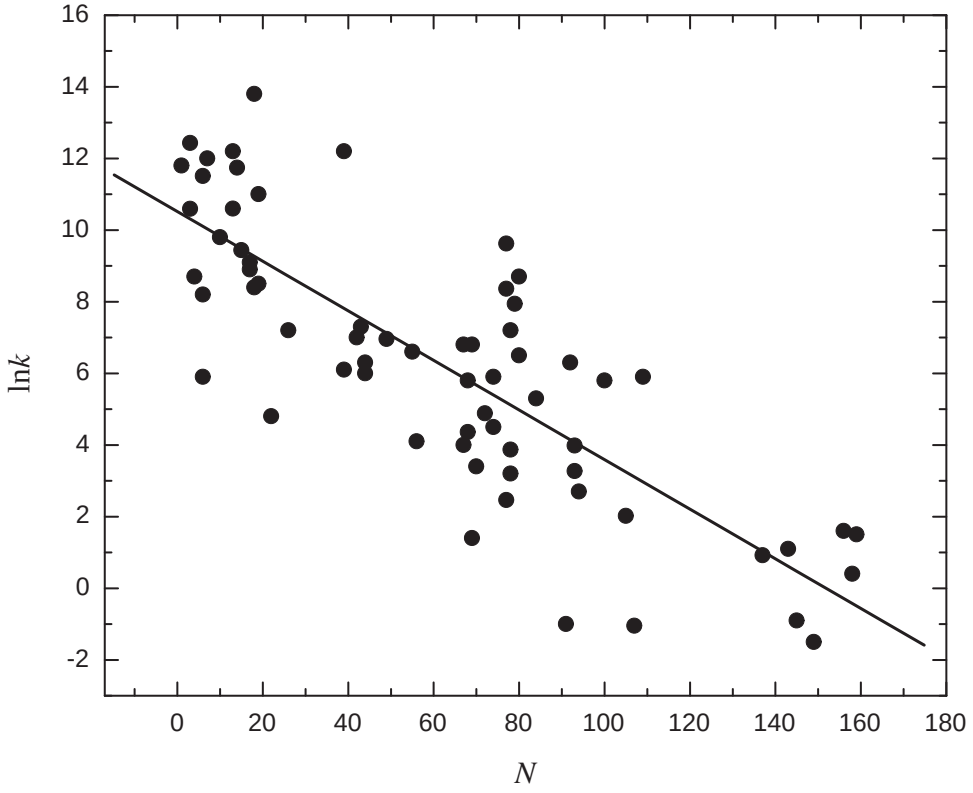

FIG. 1. The linear regression of ln k vs. N together with the experimental points for the 66 small, two-state proteins used in this work.

$$\tau = \frac{1}{k_+} \approx 1.5 \times 10^{-7} s \tag{17}$$

which is close to the folding speed limit that experimental and theoretical approaches predict [30].

Although fast-folding experiments have given evidence for downhill folding in some proteins artificially designed [39, 40], it has been debated whether downhill folding occurs for natural proteins under native conditions [41–43]. Owing to the controversy surrounding the experiments and the lack of accepted experimental data, it is difficult at present to directly verify Eq.(14) by a set of valid downhill folding proteins. However, if the time to form one native contact is the same for downhill and two-state folders, and let the range of the total numbers of native contacts of downhill folders be $1 \sim 100$, then, the range of folding rates from Eq.(14) should be $10^7 \sim 10^5 s^{-1}$,

TABLE I: List of the selected proteins in this article. The columns in this table are as follows: PDB id, Protein Data Bank entry; L, number of residues in the protein used in experimental study; N, total number of native contacts; $\ln k$, the natural logarithm of the experimental folding rates in the water.

Note: Four proteins in the table, 1C8C, 1E0L, 1PRB, and 1T8J are actually the mutants of corresponding wild type proteins, (1C8C)Y34W, (1E0L)W30A, (1PRB)K5L/K39V, and (1T8J)V3Y/F8W, respectively.

PDB id	L	N	$\ln k$	PDB id	L	N	$\ln k$	PDB id	L	N	$\ln k$
1APS	98	149	-1.5	1HRC	104	79	7.94	1RIS	97	109	5.9
1AVZ	57	72	4.88	1IMQ	86	43	7.3	1RLQ	56	68	4.36
1AYI	86	26	7.2	1JMQ	40	18	8.4	1SHF	59	74	4.5
1BA5	49	6	5.9	1JO8	58	77	2.46	1SHG	57	69	1.4
1BDD	60	14	11.74	1JYG	69	17	9.1	1SPR	103	80	8.7
1BRS	89	70	3.4	1L2Y	20	3	12.43	1SRL	56	67	4
1C8C	64	42	7.0	1LMB	80	19	8.5	1T8J	23	1	11.8
1C9O	66	78	7.2	1M9S	76	93	3.98	1TEN	89	143	1.1
1CIS	66	78	3.87	1MJC	69	84	5.3	1U5P	110	19	11.0
1CSP	67	80	6.5	1N88	96	105	2.02	1UBQ	76	74	5.9
1CUN	106	22	4.8	1NTI	86	49	6.96	1URN	96	100	5.8
1DIV1	92	93	3.27	1O6X	81	67	6.8	1VII	36	6	11.51
1DIV2	56	39	6.1	1PBA	81	69	6.8	1WIT	93	158	0.4
1E0L	37	13	10.6	1PGB1	16	7	12.0	1YCC	103	77	9.62
1E0M	37	17	8.9	1PGB2	56	44	6.0	256B	106	39	12.2
1ENH	54	3	10.59	1PIN	34	15	9.44	2A3D	73	13	12.2
1FEX	59	6	8.2	1PKS	76	107	-1.05	2ABD	86	55	6.6
1FKB	107	159	1.5	1PNJ	86	91	-1.0	2ACY	98	137	0.92
1FKF	107	156	1.6	1POH	85	94	2.7	2CI2	64	68	5.8
1FNF	89	145	-0.9	1PRB	47	18	13.8	2PDD	41	10	9.8
1G6P	66	92	6.3	1PSF	69	78	3.2	2PTL	60	56	4.1
1GV2	47	4	8.7	1RFA	78	77	8.36	3GB1	56	44	6.3

which is consistent with the recognized downhill folding rates range [30].

Recently, experimental evidence has emerged that a two-state folding protein can turn into a downhill folding one under some certain solvent conditions. For example, Kim *et al.* have examined the same fast-folding protein under very different solvent conditions, and found that it behaves like a two-state folder in one, a downhill folder in another [44]. In our model, such solvent-tuning two-state to downhill folding corresponds to the parameter K changing from less than 1 to greater than 1.

Finally, it should be pointed out that although the folding rates of an ultrafast two-state folding protein and a downhill folding protein may be very close to each other, they fold through different folding kinetics. For the former, there exists a lower folding free energy barrier, or smaller N value. For the latter, there is no any significant free energy barrier along the reaction coordinate, and its total number of native contacts N may not be very small.

V. SUMMARIZE AND REMARKS

In summary, although it is impossible for a polypeptide to find its native structure by random search among its configuration space at the atomic level, it can find its folded state by random search among its contact number space at the amino acid level. And both two-sate and downhill folding kinetics can be described by a unifying model, depending on the different solvent conditions (temperature, denaturant, PH *etc.*), and therefore the different folding funnel landscapes, and therefore the different parameter values: $K < 1$ or $K > 1$ in the present model.

We acknowledge Dr. W. Wang for helpful discussion. This work was partially supported by NSF China with No. 10875072.

[1] R. Jaenicke, Biochemistry-US **30**, 3147 (1991).

[2] C. B. Anfinsen, E. Haber, M. Sela, and F. H. White, Proc. Natl. Acad. Sci. USA **47**, 1309 (1961).

[3] B. Gutte and R. B. Merrifield, J. Amer. Chem. Soc. **91**, 501 (1969).

[4] K. W. Plaxco, K. T. Simons, I. Ruczinski, and D. Baker, J. Mol. Biol. **39**, 11177 (2000).

[5] D. Barrick, Phys. Biol. **6**, 015001 (2009).

[6] S. E. Jackson, Fold. Des. **3**, R81 (1998).

[7] C. Levinthal, J. Chim. Phys. **65**, 44 (1968).

[8] P. S. Kim and R. L. Baldwin, Annu. Rev. Biochem. **59**, 631 (1990).

[9] J. B. Udgaonkar and R. L. Baldwin, Nature **335**, 694 (1988).

[10] K. A. Dill, Biochemistry **24**, 1501 (1985).

[11] K. A. Dill, Biochemistry **29**, 7133 (1990).

[12] L. S. Itzhaki, D. E. Otzen, and A. R. Fersht, J. Mol. Biol. **254**, 260 (1995).

[13] A. R. Fersht, Curr. Opin. Struct. Biol. **7**, 3 (1997).

[14] K. A. Dill, K. M. Fiebig, and H. S. Chan, Proc. Natl. Acad. Sci. USA. **90**, 1942 (1993).

[15] K. M. Fiebig and K. A. Dill, J. Chem. Phys. **98**, 3475 (1993).

[16] B. Gillespie and K.W. Plaxco, Annu. Rev. Biochem. **73**, 837 (2004).

[17] K. W. Plaxco, K. T. Simons, and D. Baker, J. Mol. Biol. **277**, 985 (1998).

[18] M. M. Gromiha and S. Selvaraj, J. Mol. Biol. **310**, 27 (2001).

[19] H. Y. Zhou and Y. Q. Zhou, Biophys. J. **335**, 458 (2002).

[20] D. N. Ivankov, S. O. Garbuzynskiy, E. Alm, K. W. Plaxco, D. Baker, and A. V. Finkelstein, Protein Sci. **12**, 2057 (2003).

[21] J. N. Onuchic, P. G. Wolynes, Z. Luthey-Schulten and N. D. Socci, Proc. Natl. Acad. Sci. USA **92**, 3626 (1995).

[22] J. D. Bryngelson, J. N. Onuchic, N. D. Socci and P. G. Wolynes, Proteins: Struct. Funct. Genet. **21**, 167 (1995).

[23] P. Bruscolini, A. Pelizzola, and M. Zamparo, J. Chem. Phys. **126**, 215103 (2007).

[24] S. S. Cho, P. Weinkam, and P. G. Wolynes, Proc. Natl. Acad. Sci. USA. **105**, 118 (2008).

[25] G. H. Zuo, J. Wang, and W. Wang, Proteins Struct. Funct. Bioinformatics **63**, 165 (2006).

[26] J. Zhang, W. F. Li, J. Wang, M. Qin, and W. Wang, Proteins: Struct. Funct. Bioinf. **72**, 1038 (2008).

[27] P. L. Freddolino, F. Liu, M. Gruebele, and K. Schulten, Biophys. J. **94**, L75 (2008).

[28] M. Gruebele, C. R. Biol. **328**, 701 (2005).

[29] V. Munoz, Annu. Rev. Biophys. Biomol. Struct. **36**, 395 (2007).

[30] J. Kubelka, J. Hofrichter, and W. A. Eaton, Curr. Opin. Struct. Biol. **14**, 76 (2004).

[31] D. Baker, Nature **405**, 39 (2000).

[32] D. Mi, G. R. Liu, J. S. Wang, and Z. R. Li, J. Theor. Biol. **241**, 152 (2006).

[33] R. Zwanzig, A. Szabo, and B. Bagchi, Proc. Natl. Acad. Sci. USA **89** 20 (1992).

[34] R. Zwanzig, Proc. Natl. Acad. Sci. USA **92** 9801 (1995).

[35] C. W. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences* (Springer,

Berlin, 4th Ed, 2009).

[36] J. S. Guo, D. Mi, and Y. Q. Sun, Physica A **389**, 761 (2010).

[37] N. S. Bogatyreva, A. A. Osypov, and D. N. Ivankov, Nucleic Acids Res. **37**, D342 (2009).

[38] K. L. Maxwell, D. Wildes, A. Zarrine-Afsar, *et al.*, Protein Sci. **14**, 602 (2005).

[39] W. Y. Yang and M. Gruebele, Biophys. J. **87**, 596 (2004).

[40] F. Liu and M. Gruebele, J. Mol. Biol. **370**, 574 (2007).

[41] N. Ferguson, P. J. Schartau, T. D. Sharpe, S. Sato, and A. R. Fersht, J. Mol. Biol. **344**, 295 (2004).

[42] S. Mourad, D. Fushman, and V. Munoz, Nature **442**, 317 (2006).

[43] S. S. Cho, P. Weinkam, and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **105**, 118 (2008).

[44] S. J. Kim, Y. Matsumura, C. Dumont, H. Kihara, and M. Gruebele, Biophys. J. **97**, 295 (2009).