Spatial ranges of driving forces are a key determinant of protein folding cooperativity and rate diversity

Hüseyin Kaya,[1,*] Zeynep Uzunoğlu,[2,†] and Hue Sun Chan[3,‡]

[1]*Department of Biophysics, Faculty of Medicine, University of Gaziantep, 27310 Gaziantep, Turkey*

[2]*Department of Physics, Faculty of Science, Atatürk University, 25240 Erzurum, Turkey*

[3]*Departments of Biochemistry, Molecular Genetics, and Physics, University of Toronto, Toronto, Ontario M5S 1A8, Canada*

(Received 24 May 2013; revised manuscript received 21 August 2013; published 14 October 2013)

The physical basis of two-state-like folding transitions and the tremendous diversity in folding rates is elucidated by directly simulating the folding kinetics of 52 representative proteins. Relative to the results from a common modeling approach, the diversity of the simulated folding rates can be increased from $\sim 10^{2.1}$ to the experimental $\sim 10^{6.0}$ by a modest decrease in the spatial range of the attractive potential. The required theoretical range is consistent with desolvation physics and is notably much more permissive than that needed for two-state-like homopolymer collapse.

DOI: 10.1103/PhysRevE.88.044701 PACS number(s): 87.15.Cc, 87.14.et

Protein folding is a fundamental process at the interface between physics and biology but its energetics is far from being fully understood. In this regard, the collective study of many small single-domain proteins as a paradigmatic system has provided crucial insights [1]. One of the remarkable experimental properties of these proteins with $\gtrsim 40$ and $\lesssim 100$ amino acids is that they can fold at vastly different rates that span ∼6 orders of magnitude or more. The discovery [2] of a striking correlation between native topology and these diverse rates [3] suggested strongly that important principles of protein folding can be revealed by using this experimental behavior as a stringent criterion for evaluating theory [4].

However, providing a rigorus physical basis for the experimental trend turned out to be a nontrivial task. Although folding-rate diversity has been addressed by theories that invoked analytical approximations in lieu of an explicit representation of the protein chains [5], the first attempt to rationalize folding rate diversity by using explicit-chain models resulted in simulated rates that were much less diverse than the experimental rates [6]. Prompted by this mismatch between theory and experiment, subsequent theoretical investigations uncovered that the immense diversity in folding rates is intimately related to the cooperative, two-state-like nature of the folding transitions of single-domain proteins [7,8], and that folding cooperativity can be enhanced by many-body effects [9] such as local-nonlocal coupling [8]. Model studies further indicated that elementary free energy barriers arising from the desolvation process can render the formation of the protein core more akin to an all-or-none process [10], leading to a large diversity in folding rates comparable to that observed experimentally [11,12].

Because desolvation barriers in aqueous potentials of mean force typically entail narrow spatial ranges of attraction, the aforementioned findings led us to propose a simple general organizing principle, namely that a high degree of folding cooperativity and folding-rate diversity are underpinned by generic short-range intrachain interactions. In this Brief Report, we take a first step to assess this hypothesis by asking how folding rates simulated by native-centric Gō-like coarse-grained C_α protein chain models [12–14] are affected by the assumed spatial range of the favorable interactions. The models were constructed according to Ref. [15] with the same parameters for the potential energies and Langevin dynamics except that we now compare different functional forms for the nonbonded native-centric potential.

Recent advances in molecular dynamics simulations, especially through the development of special-purpose machines, have made it possible to directly simulate the folding kinetics of a number of fast-folding proteins using a general transferrable atomic potential [16]. (A summary of folding rates determined by earlier all-atom simulations is provided in Fig. 3 of Ref. [12].) However, even with the tremendous improvements in hardware and software witnessed during the past few years, currently it is not possible to use all-atom molecular dynamics to simulate the slower folding rates among the aforementioned span of ∼6 orders of magnitude in experimental folding rates. Therefore, to make progress in understanding this intriguing phenomenon, we resort to using highly coarse-grained native-centric models. These models have proven to be useful for exploring general principles of folding [12,17], though they do not consider favorable nonnative interactions that can be present in real proteins [18]. In applying these models with their obvious limitations, our goal is to learn how various model interaction schemes affect the trend of folding-rate diversity rather than to predict individual folding rates accurately. After all, consistency in detailed predictions is not always possible even among different all-atom potentials [19].

In the following discussion, the model with the original, well-studied form of the native-centric potential [6,14,15] is referred to as the long-range interaction model (LRIM), wherein the nonbonded energy term for a pair of amino acid residues that are in contact in the native structure in the Protein Data Bank (PDB) is given by a 12–10 Lennard-Jones (LJ) potential,

$$U_{ij}^{\mathrm{LJ}} = \epsilon\left[5\left(r_{ij}^0/r_{ij}\right)^{12} - 6\left(r_{ij}^0/r_{ij}\right)^{10}\right], \qquad (1)$$

*huseyinkaya@gantep.edu.tr
†zeynepu@atauni.edu.tr
‡chan@arrhenius.med.toronto.edu

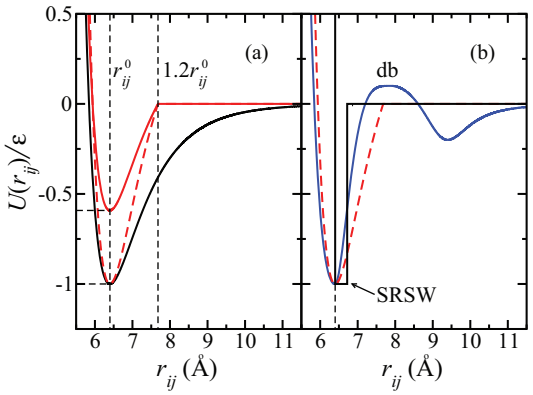

FIG. 1. (Color) Model intrachain attractive potentials. (a) An example LRIM (black) and the corresponding SRIM (solid red) and modified SRIM (dashed red) potentials. (b) Comparing our native-centric modified SRIM potential (dashed red) with that in the desolvation-barrier (db) model (blue) [15] and the nonnative-centric SRSW model of Taylor *et al*. with $\lambda = 1.05$ (black) [21].

where r_{ij} is the distance between amino acids i and j, r_{ij}^0 stands for the corresponding distance in the PDB structure, and ϵ (> 0) sets the energy scale. The total nonbonded native interaction energy $U_{\rm LRIM}^{\rm nb}$ of a conformation is then given by the summation $\sum_{i<j-3}^{\rm native} U_{ij}^{\rm LJ}$ over native pairs $i < j-3$ [15]. A pair of residues are defined here to be in the native contact set if at least two nonhydrogen atoms, one from each residue, are $<$ 4.5 Å apart [20], or if the residues' C_α atoms are $<$ 6.4 Å apart in the PDB structure. The latter criterion followed from our finding of a clear separation in the distribution of $C_\alpha - C_\alpha$ distances at around 6.2 to 6.4 Å among 5 610 different PDB structures.

We compared the behavior of the LRIM model with a new short-range interaction model (SRIM) and a modified SRIM [Fig. 1(a)]. In the SRIM, the LJ terms in the total nonbonded native interaction energy $U_{\rm SRIM}^{\rm nb}$ are shifted up to assume zero value at $r_{ij} = 1.2r_{ij}^0$:

$$U_{\rm SRIM}^{\rm nb} = \begin{cases} \sum_{i<j-3}^{\rm native} \left(U_{ij}^{\rm LJ} + \epsilon \cdot \omega\right) & \text{if} \quad r_{ij} \leqslant 1.2r_{ij}^0 \\ 0 & \text{if} \quad r_{ij} > 1.2r_{ij}^0 \end{cases}, \quad (2)$$

where $\omega = -U_{ij}^{\rm LJ}(r_{ij} = 1.2r_{ij}^0) = 0.40825$ is a constant shifting parameter. In the modified SRIM, the nonbonded term is rescaled to $(U_{ij}^{\rm LJ} + \epsilon \cdot \omega)/(1-\omega)$ to yield a minimum value of $-\epsilon$ as in the LRIM such that the relative strengths of local and nonbonded interactions in the original LRIM are maintained.

Figure 1(b) shows that the attractive range of SRIM or modified SRIM for a typical $r_{ij}^0 \approx 6.4$ Å is comparable to that of the native-centric desolvation-barrier (db) potential [10,11,15] because $(0.2) \cdot 6.4$ Å ≈ 1.3 Å ≈ 1.4 Å $\approx$ radius of a water molecule. This range is much wider, however, than the extremely short attractive range of $(0.05) \cdot 6.4$ Å≈ 0.3 Å (i.e., $\lambda = 1.05$) needed for a short-range square-well (SRSW) homopolymer model with a minimum intermonomer distance $\sigma = 6.4$ Å to undergo two-state coil-globule transition [21].

We applied the LRIM and SRIM models to a set of 52 proteins that cover a broad spectrum of structures consisting of 15 all-α, 18 all-β, and 19 $\alpha + \beta$ proteins, with experimental folding rates covering 5.96 orders of magnitude (Table I). As in previous works [6,14,15,20], we use fractional native contact Q as a progress variable [22]. Using the free-energy profile defined by Q, the transition midpoint for each model protein is determined as the temperature T_m at which the free energies for the denatured and native minima are equal. In an overwhelming majority of cases, the T_m defined in this manner is virtually identical to the temperature of the heat capacity peak. Model folding rates depend on the simulation temperature [15,20]. As in previous comparative studies of native-centric model folding rates [6,11,20,23], each folding rate in this study is simulated near the T_m for the given model protein. Midpoint folding rate is very well correlated with the free-energy barrier at the heat capacity peak, and thus is a good measure of folding cooperativity of a model protein [11,23]. The choice of using simulated midpoint folding rates to compare against experimental folding rates, such as those in Table I, that were largely obtained near room temperature with zero denaturant is necessitated by LRIM's insufficient folding cooperativity under strongly folding conditions [23]. In essence, this approach [6,11,20,23] entails an assumption that the logarithmic range spanned by experimental rates under strongly folding conditions are representative of that spanned by experimental midpoint folding rates, even though midpoint folding rates are not available experimentally for many of the proteins being studied. In the present study, all rates are determined by direct, brute-force simulations of trajectories initiated at $Q < 0.1$ and terminated when $Q > 0.95$. Our directly simulated folding rates are computed without using any bias potential [11] or analytical rescaling [20] (see discussion in the caption for Fig. 3 in Ref. [12]).

Inasmuch as the model and experimental folding rates are related by a universal scaling factor, as we have assumed, the ranges of logarithmic model and experimental rates can be compared without regard to the value of that scaling factor. Figure 2 shows that the LRIM folding rates span only 2.13 orders of magnitude [Fig. 2(b)], falling far short of the experimental range, as has been shown for a more limited set of proteins [11]. In contrast, the SRIM folding rates are much more diverse, spanning 4.80 orders of magnitude [Fig. 2(a)]. Because the relative strengths of local versus nonbonded interactions can affect folding cooperativity [55] and thus folding-rate diversity [7,8], we considered also folding rates in the modified SRIM. The correlation in the inset of Fig. 2(a) between SRIM and modified-SRIM logarithmic folding rates for a subset of 15 proteins suggests that the modified SRIM folding rates for the full set of 52 proteins would span $\approx 1.2107 \times 4.80 = 5.81$ orders of magnitude, which is very similar to the 5.96 orders of magnitude spanned by the experimental folding rates.

As for previous models [7,8], the example in Fig. 3(a) illustrates that the higher folding-rate diversity of the SRIM relative to the LRIM is underpinned by the SRIM's higher degree of folding cooperativity. The overall folding-unfolding barrier along the simulated equilibrium free-energy profile is significantly higher for the SRIM than for the LRIM, and the barrier for the modified SRIM is even higher. This higher folding cooperativity of the modified SRIM relative to the SRIM underlies the modified SRIM's larger diversity in folding rates [Fig. 2(a), inset]. Similar to the db model [11], the

TABLE I. Proteins used in this study. N is number of amino acid residues; $k_f^{\rm sim}$ is simulated folding rate in reciprocal Langevin time steps; $k_f^{\rm expt}$ stands for experimental folding rates in s^{-1} provided in the listed references.

Protein	PDB code	N	$\log_{10} k_{\rm f}^{\rm sim}$ SRIM	$\log_{10} k_{\rm f}^{\rm sim}$ LRIM	$\log_{10} k_{\rm f}^{\rm expt}$	Ref.	Protein	PDB code	N	$\log_{10} k_{\rm f}^{\rm sim}$ SRIM	$\log_{10} k_{\rm f}^{\rm sim}$ LRIM	$\log_{10} k_{\rm f}^{\rm expt}$	Ref.
MerP	1afi	72	−7.50	−6.25	0.26	[24]	Hpr	1poh	85	−7.23	−6.42	1.17	[38]
mAcP	1aps	98	−9.41	−6.97	−0.64	[25]	Photosystem I	1psf	69	−7.09	−5.94	1.39	[39]
Im7	1ayi	86	−5.87	−5.75	3.13	[26]	RafRBD	1rfa	78	−7.31	−6.15	3.63	[26]
Sso7d (Y34W)[a]	1bf4	63	−5.77	−5.43	3.02	[27]	S6	1ris	97	−8.07	−6.08	2.56	[40]
CspB Bc	1c9o	66	−7.54	−6.07	3.14	[28]	Src SH3[a]	1rlq	56	−6.52	−5.78	1.75	[41]
CspB Bs[a]	1csp	67	−7.62	−6.20	2.84	[28]	EC298	1ryk	69	−5.34	−5.57	3.94	[26]
PTL9 C	1div	92	−8.43	−6.46	1.42	[26]	Fyn SH3	1shf	59	−6.88	−5.96	1.95	[42]
NTL9	1div	56	−5.90	−5.84	2.86	[29]	α-spectrin SH3	1shg	57	−7.41	−6.29	0.94	[43]
FBP28 (W30A)[a]	1e0l	37	−5.05	−5.17	4.61	[30]	Src SH2	1spr	103	−6.65	−6.03	3.80	[26]
WW prototype	1e0m	37	−5.03	−4.95	3.84	[30]	BDPA	1ss1	60	−5.19	−5.27	4.98	[44]
Engrailed HD	1enh	54	−5.42	−5.52	4.60	[31]	Tenascin	1ten	89	−8.47	−6.47	0.78	[45]
RAP1 h[a]	1fex	59	−5.15	−5.22	3.56	[31]	TI I27	1tit	89	−8.30	−6.71	1.54	[46]
FKBP12	1fkb	107	−9.51	−6.94	0.63	[32]	Ubiquitin[a]	1ubq	76	−7.30	−6.29	3.18	[26]
9 Fibronectin III	1fnf	90	−9.72	−7.08	−0.40	[33]	U1A	1urn	96	−8.40	−6.69	2.50	[47]
CspB Tm[a]	1g6p	66	−6.60	−6.05	2.75	[28]	E3BD (F166W)	1w4e	45	−5.21	−5.21	4.44	[48]
Cyt C	1hrc	104	−6.22	−5.95	3.80	[24]	BOP (YWLA)[a]	1w4j	51	−5.04	−5.04	5.32	[48]
cMyb	1idy	54	−5.07	−5.00	3.79	[31]	Cyt b562	1yza	106	−5.24	−5.27	3.63	[49]
Im9	1imq	86	−6.02	−5.86	3.18	[26]	bACBP	2abd	86	−5.96	−5.74	3.02	[26]
ABP1 SH3	1jo8	58	−6.72	−6.11	1.07	[26]	ctAcP	2acy	98	−8.69	−6.69	0.36	[50]
Yap[a]	1k9q	40	−5.24	−5.49	3.63	[30]	BBL (H166W)[a]	2bth	45	−5.01	−5.01	5.11	[48]
λ-repressor$_{6-85}$	1lmb	80	−5.66	−5.66	3.69	[34]	CI2[a]	2ci2	64	−6.65	−5.95	1.68	[51]
GW1	1m9s	76	−6.68	−6.02	1.73	[26]	E2[ab]	2pdd	41	−4.98	−5.92	4.25	[52]
CspA	1mjc	69	−7.81	−6.21	2.42	[35]	Protein L[a]	2ptl	62	−6.68	−5.73	1.78	[26]
L23	1n88	96	−8.02	−6.29	1.31	[36]	Urm1	2qjl	99	−9.57	−6.72	1.12	[26]
ADA2h[a]	1o6x	81	−6.66	−5.90	2.95	[26]	Villin 14T	2vik	126	−7.44	−6.25	2.95	[53]
Protein G[a]	1pgb	56	−6.27	−5.62	2.61	[37]	Tendamistat	3ait	74	−6.99	−6.31	1.82	[54]

[a]These proteins are used in the inset of Fig. 2(a).
[b]E3/E1-binding domain of dihydrolipoyl acyltransferase.

SRIM barrier and modified SRIM barrier are shifted to a larger $Q \sim 0.7$ relative to the $Q \sim 0.5$ LRIM barrier [Fig. 3(a)], leading to enhanced populations of local native contacts in the transition state (TS) ensemble [Fig. 3(b)].

Because our model folding rates are simulated at each model protein's respective midpoint temperature T_m, it is pertinent to note that the significantly higher folding-rate diversity of the SRIM relative to the LRIM is not merely a consequence of the different spread of T_m values between these two classes of models. Among the 52 model proteins we study, T_m for LRIM ranges from a minimum $(T_m)_{\rm min} = 0.820$ to a maximum $(T_m)_{\rm max} = 1.159$ in the model unit where the Boltzmann constant $k_{\rm B} = 1$. The corresponding values for SRIM are 0.548 and 0.726. The range of inverse temperature, $\Delta T_m^{-1} \equiv (T_m)_{\rm min}^{-1} - (T_m)_{\rm max}^{-1}$ is thus slightly wider for SRIM $[(\Delta T_m^{-1})_{\rm SRIM} = 0.447]$ than for LRIM $[(\Delta T_m^{-1})_{\rm LRIM} = 0.357]$. However, this difference in ΔT_m^{-1}, viz., $\delta(\Delta T_m^{-1}) \equiv (\Delta T_m^{-1})_{\rm SRIM} - (\Delta T_m^{-1})_{\rm LRIM} = 0.09$, cannot by itself account for the $4.80 - 2.13 = 2.67$ orders of magnitude difference in folding-rate diversity between SRIM and LRIM. In a hypothetical situation in which the folding free-energy barrier heights for all SRIM and LRIM models are identical (all equal to a single $\Delta G^{\ddagger}$ value) at their respective T_ms so that a spread of T_m values is the only possible cause for diversity in midpoint folding rates, the order-of-magnitude difference in folding-rate diversity between SRIM and LRIM would be given by $\log_{10}\{\exp[(\Delta G^{\ddagger})\delta(\Delta T_m^{-1})]\}$. This difference is only $\sim$0.6 order of magnitude even if a very high value $\sim$15 is used for $\Delta G^{\ddagger}$ (in the $k_{\rm B} = 1$ unit) [11]. Moreover, for the subset of 15 proteins that are simulated also using the modified-SRIM $\Delta T_m^{-1} = 0.396$ is smaller than the corresponding SRIM $\Delta T_m^{-1} = 0.485$ for the same subset, but the spread of modified-SRIM folding rates is larger than that of SRIM. Hence, as stated above, the SRIM models' higher diversity in midpoint folding rates is mostly a reflection of their innate tendency to fold more cooperatively.

In summary, the above results show that attractive ranges of pairwise nonbonded interactions only moderately shorter than those in common coarse-grained protein models are sufficient for behaviorial trends similar to those observed experimentally. In addition to highlighting the often under-appreciated fact that appropriate spatial ranges are critical in the design of model protein potentials, our finding has several fundamental implications: First, inasmuch as pairwise attractive ranges are sufficiently narrow, many-body interactions [7–9] are not necessary for (though they can add to) proteinlike cooperativity. Second, an apparent tradeoff exists between sequence specificity and attractive interaction range with regard to folding cooperativity. For our highly specific native-centric models, the attractive ranges need not

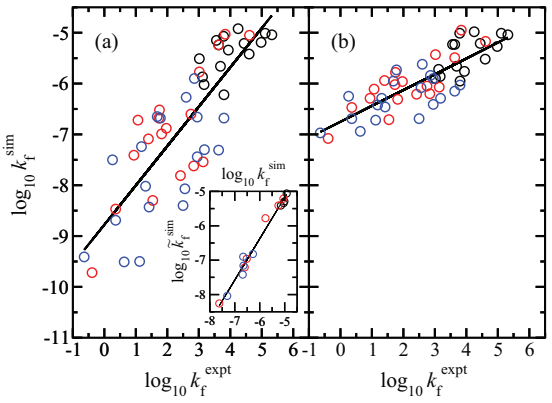

FIG. 2. (Color) Comparison of experimental logarithmic folding rates with their simulated counterparts in the (a) SRIM and (b) LRIM. All k_f^{sim} values were determined using 500 trajectories except 104–180 trajectories were used for the slowest folding 1aps, 1fnf, 2qjl, and 1fkb. The lines are the best linear fits with correlation coefficients $r = 0.83$ (a) and $r = 0.82$ (b). Black, red, and blue data points are for all-α, all-β, and $\alpha + \beta$ proteins, respectively. (Inset) Correlation between the simulated logarithmic folding rates in the SRIM model ($\log_{10} k_f^{sim}$) and modified SRIM model ($\log_{10} \tilde{k}_f^{sim}$) for a representative subset consisting of 4 all-α, 6 all-β, and 5 $\alpha + \beta$ proteins specified in Table I. The plotted best linear fit is given by $\log_{10} \tilde{k}_f^{sim} = 0.82563 + 1.2107 \cdot \log_{10} k_f^{sim}$.

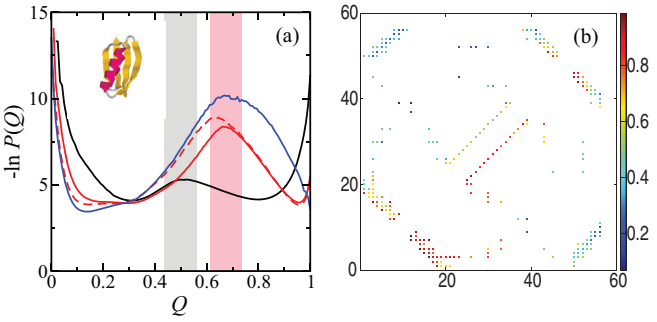

FIG. 3. (Color) (a) PDB structure and free energy profiles of Protein G (1pgb) in the LRIM (black), SRIM (red), modified SRIM (dashed red), and db (blue) models computed near each model's transition midpoint; $P(Q)$ is normalized conformational population. (b) Probability (right scale) maps of native contacts in the putative TSs of SRIM (top-left) and LRIM (bottom-right) which are marked, respectively, by the pink and gray regions in (a).

be very narrow to exhibit proteinlike behaviors. In contrast, for a homopolyer of similar chain length and monomer size, the range of the nonspecific attraction needs to be several times narrower to achieve proteinlike two-state coil-globule transition [21] [Fig. 1(b)]. Third, db [10–12,15] provides a physical basis for the moderately narrow attractive range needed for cooperative protein folding. Our results indicate that the main contributing feature in db potentials to folding cooperativity is the narrow attractive range rather than the positive barrier heights, which can often be quite low as revealed by atomic simulations [11,12]. In this light, how can the extremely narrow attractive ranges needed for two-state homopolymer transition be realized physically is an interesting question. More broadly, it will be instructive to explore how other restrictions on favorable intraprotein interactions, such as the angular restrictions in hydrogen bonding [56], might contribute to folding cooperativity and rate diversity. Recent advances in atomic simulations [16,19] offer exciting prospects in this regard.

H.K. has been supported by the Turkish Academy of Sciences through a Distinguished Young Investigator Program (TÜBA-GEBİP). This work was supported by TÜBİTAK (Grant No. 106T693) and was partly carried out within the framework of the Associateship Scheme of the Abdus Salam International Centre for Theoretical Physics in Trieste, Italy. The numerical calculations were performed in part at TÜBİTAK ULAKBİM High Performance and Grid Computing Center (TRUBA Resources) in Turkey. H.S.C. gratefully acknowledges the support he received from the Canadian Institutes of Health Research (Grant No. MOP-84281) and thanks SciNet of Compute Canada for computational resources.

[1] S. E. Jackson, Fold. Des. **3**, R81 (1998).
[2] K. W. Plaxco, K. T. Simon, and D. Baker, J. Mol. Biol. **277**, 985 (1998).
[3] S. Selvaraj and M. M. Gromiha, J. Mol. Biol. **310**, 27 (2001); H. Zhou and Y. Zhou, Biophys. J. **82**, 458 (2002); C. Micheletti, Proteins **51**, 74 (2003).
[4] H. S. Chan, Nature (London) **392**, 761 (1998); S. Wallin and H. S. Chan, Protein Sci. **14**, 1643 (2005).
[5] E. Alm and D. Baker, Proc. Natl. Acad. Sci. USA **96**, 11305 (1999); V. Muñoz and W. A. Eaton, *ibid.* **96**, 11311 (1999); O. V. Galzitskaya and A. V. Glyakina, Proteins **80**, 2711 (2012).
[6] N. Koga and S. Takada, J. Mol. Biol. **313**, 171 (2001).
[7] A. I. Jewett, V. S. Pande, and K. W. Plaxco, J. Mol. Biol. **326**, 247 (2003).
[8] H. Kaya and H. S. Chan, Proteins **52**, 524 (2003).
[9] S. S. Plotkin, J. Wang, and P. G. Wolynes, J. Chem. Phys. **106**, 2932 (1997); M. P. Eastwood and P. G. Wolynes, *ibid.* **114**, 4702 (2001); M. R. Ejtejadi, S. P. Avall, and S. S. Plotkin, Proc. Natl. Acad. Sci. USA **101**, 15088 (2004); J. J. Portman, Curr. Opin. Struct. Biol. **20**, 11 (2010).
[10] M. S. Cheung, A. E. García, and J. N. Onuchic, Proc. Natl. Acad. Sci. USA **99**, 685 (2002); Y. Levy and J. N. Onuchic, Annu. Rev. Biophys. Biomol. Struct. **35**, 389 (2006).
[11] A. Ferguson, Z. Liu, and H. S. Chan, J. Mol. Biol. **389**, 619 (2009); Corrigendum **401**, 153 (2010).
[12] H. S. Chan, Z. Zhang, S. Wallin, and Z. Liu, Annu. Rev. Phys. Chem. **62**, 301 (2011).
[13] C. Micheletti, J. R. Banavar, A. Maritan, and F. Seno, Phys. Rev. Lett. **82**, 3372 (1999); J.-E. Shea, J. N. Onuchic, and C. L. Brooks, Proc. Natl. Acad. Sci. USA **96**, 12512 (1999).
[14] C. Clementi, H. Nymeyer, and J. N. Onuchic, J. Mol. Biol. **298**, 937 (2000).
[15] H. Kaya and H. S. Chan, J. Mol. Biol. **326**, 911 (2003); Corrigendum **337**, 1069 (2004).

[16] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, Science **334**, 517 (2011); S. Piana, K. Lindorff-Larsen, and D. E. Shaw, Proc. Natl. Acad. Sci. USA **109**, 17845 (2012); **110**, 5915 (2013).
[17] R. B. Best and G. Hummer, Proc. Natl. Acad. Sci. USA **107**, 1088 (2010); R. J. Oliveira, P. C. Whitford, J. Chahine, J. Wang, J. N. Onuchic, and V. B. P. Leite, Biophys. J. **99**, 600 (2010); Z. Zhang and H. S. Chan, Proc. Natl. Acad. Sci. USA **109**, 20919 (2012).
[18] Z. Zhang and H. S. Chan, Proc. Natl. Acad. Sci. USA **107**, 2920 (2010); D. Shental-Bechora, M. T. J. Smith, D. MacKenzie, A. Broom, A. Marcovitz, F. Ghashut, C. Go, F. Bralha, E. M. Meiering, and Y. Levy, *ibid.* **109**, 17839 (2012).
[19] S. Piana, K. Lindorff-Larsen, and D. E. Shaw, Biophys. J. **100**, L47 (2011); K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw, PLoS ONE **7**, e32131 (2012).
[20] L. L. Chavez, J. N. Onuchic, and C. Clementi, J. Am. Chem. Soc. **126**, 8426 (2004).
[21] M. P. Taylor, W. Paul, and K. Binder, Phys. Rev. E **79**, 050801(R) (2009).
[22] A. Săli, E. I. Shakhnovich and M. Karplus, Nature (London) **369**, 248 (1994).
[23] S. Wallin and H. S. Chan, J. Phys. Cond. Matt. **18**, S307 (2006).
[24] K. W. Plaxco, K. T. Simons, I. Ruczinski, and D. Baker, Biochemistry **39**, 11177 (2000).
[25] N. A. van Nuland *et al*., J. Mol. Biol. **283**, 883 (1998).
[26] K. L. Maxwell *et al*., Protein Sci. **14**, 602 (2005).
[27] R. Guerois and L. Serrano, J. Mol. Biol. **304**, 967 (2000).
[28] D. Perl *et al*., Nat. Struct. Biol. **5**, 229 (1998).
[29] B. Kuhlman, D. L. Luisi, P. A. Evans, and D. P. Raleigh, J. Mol. Biol. **284**, 1661 (1998).
[30] N. Ferguson *et al*., Proc. Natl. Acad. Sci. USA **98**, 13002 (2001).
[31] S. Gianni *et al*., Proc. Natl. Acad. Sci. USA **100**, 13286 (2003).
[32] E. R. G. Main, K. F. Fulton, and S. E. Jackson, J. Mol. Biol. **291**, 429 (1999).
[33] K. W. Plaxco, C. Spitzfaden, I. D. Campbell, and C. M. Dobson, J. Mol. Biol. **270**, 763 (1997).
[34] R. E. Burton *et al*., J. Mol. Biol. **263**, 311 (1996).
[35] H. M. Rodriguez, D. M. Vu, and L. Gregoret, Protein Sci. **9**, 1993 (2000).
[36] L. Hedberg and M. Oliveberg, Proc. Natl. Acad. Sci. USA **101**, 7606 (2004).
[37] E. L. McCallister, E. Alm, and D. Baker, Nat. Struct. Biol. **7**, 669 (2000).
[38] N. A. van Nuland *et al*., Biochemistry **37**, 622 (1998).
[39] D. N. Ivankov *et al*., Protein Sci. **12**, 2057 (2003).
[40] D. E. Otzen and M. Oliveberg, Proc. Natl. Acad. Sci. USA **96**, 11746 (1999).
[41] V. P. Grantcharova and D. Baker, Biochemistry **36**, 15685 (1997).
[42] K. W. Plaxco *et al*., Biochemistry **37**, 2529 (1998).
[43] A. R. Viguera *et al*., Biochemistry **33**, 2142 (1994).
[44] S. Sato, T. L. Religa, and A. R. Fersht, J. Mol. Biol. **360**, 850 (2006).
[45] S. J. Hamill, A. E. Meekof, and J. Clarke, Biochemistry **37**, 8071 (1998).
[46] S. B. Fowler and J. Clarke, Structure **9**, 355 (2001).
[47] M. Silow and M. Oliveberg, Biochemistry **36**, 7633 (1997).
[48] N. Ferguson *et al*., J. Mol. Biol. **353**, 427 (2005).
[49] R. Chu, W. Pei, J. Takei, and Y. Bai, Biochemistry **41**, 7998 (2002).
[50] N. Taddei *et al*., Biochemistry **38**, 2135 (1999).
[51] S. E. Jackson and A. R. Fersht, Biochemistry **30**, 10428 (1991).
[52] S. Spector and D. P. Raleigh, J. Mol. Biol. **293**, 763 (1999).
[53] S. E. Choe *et al*., Biochemistry **37**, 14508 (1998).
[54] N. Schonbrunner, K. P. Koller, and T. Kiefhaber, J. Mol. Biol. **268**, 526 (1997).
[55] M. Knott, H. Kaya, and H. S. Chan, Polymer **45**, 623 (2004).
[56] T. Kortemme, A. V. Morozov, and D. Baker, J. Mol. Biol. **326**, 1239 (2003).