Consistent Treatment of Hydrophobicity in Protein Lattice Models Accounts for Cold Denaturation

Erik van Dijk,[1,2,*] Patrick Varilly,[1,†] Tuomas P. J. Knowles,[1,‡] Daan Frenkel,[1,§] and Sanne Abeln[2,‖]
[1]Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom
[2]Centre for Integrative Bioinformatics (IBIVU), Vrije Universiteit, De Boelelaan 1081A, 1081 HV Amsterdam, Netherlands
(Received 24 July 2015; published 16 February 2016)

The hydrophobic effect stabilizes the native structure of proteins by minimizing the unfavorable interactions between hydrophobic residues and water through the formation of a hydrophobic core. Here, we include the entropic and enthalpic contributions of the hydrophobic effect explicitly in an implicit solvent model. This allows us to capture two important effects: a length-scale dependence and a temperature dependence for the solvation of a hydrophobic particle. This consistent treatment of the hydrophobic effect explains cold denaturation and heat capacity measurements of solvated proteins.

DOI: 10.1103/PhysRevLett.116.078101

The stability of the native state of most proteins is typically dominated by interactions between amino acids and through the hydrophobic effect. The direct amino-acid interactions can be attributed to van der Waals and electrostatic forces that are mainly enthalpic in nature. By contrast, hydrophobicity is an interaction emerging from the collective behavior of the solvent and the side chains, and is entropy dominated [1–5] at room temperature for small solutes. The enthalpic amino acid interactions remain relatively constant over the temperature range of interest, while the magnitude of the hydrophobic effect changes with temperature [2,4].

In principle, all-atom simulations could be used to disentangle the role of entropy and enthalpy in protein folding. However, fully atomistic simulations are neither simple, nor cheap—in fact, at present, such simulations are only feasible for studying the folding of relatively small proteins. Moreover, a numerical study of the stability of various protein structures would require simulations over a range of temperatures. Earlier studies [6–11] have shown that a temperature-dependent hydrophobic collapse (rather than folding) can be observed in a strongly coarse-grained model for small, two-dimensional protein chains. Three-dimensional models have shown similar results for homopolymers [11] and peptides [12]. However, these models do not fully capture folding specificity for proteins. In a recent model that does incorporate folding specificity [13], two specific proteins were investigated and a linear correction was added to incorporate the temperature dependence of the hydrophobic effect [14].

In this Letter, we present an extension of the classic protein lattice model first introduced in Ref. [15]. The classic model correctly reproduces the ability of proteins to fold into a unique native structure, and it exhibits denaturation upon heating due to chain entropy alone. Interactions between amino acids are estimated through the frequency of occurrence of close contacts in experimental protein crystal structures [16]. In the Miyazawa and Jernigan (MJ) potential, the interactions are, strictly speaking, free energies that have both entropic and enthalpic components. However, in most coarse-grained simulations, these effective potentials are treated as temperature-independent enthalpies [17–23]. Therefore, they do not model the temperature dependence of the interactions correctly.

In order to model the temperature dependence of the hydrophobic effect, we use an extension of the MJ potential that includes specific solvent-amino acid interaction terms [21]. The derived potential is based on a representative subset of the protein database (PDB) [24]. The hydrophobic effect is volume dominated at small length scales and surface dominated at large length scales. Our model consistently treats this length-scale dependence by dynamically classifying residues into three categories: buried, protein surface, and fully solvated. This categorization allows us to capture the length-scale dependence of the hydrophobic effect according to the Lum-Chandler-Weeks (LCW) theory [1,25]. Our model aims to reproduce the variation in temperature dependence for different length scales of hydrophobic solutes using these implicit solvent terms.

For each residue category, the hydrophobe-water interaction is estimated by a second-order Taylor approximation to the free energy of transfer of hydrophobic particles from an oily environment to water (see Fig. 1). We use this model to investigate three effects that are often associated with the temperature dependence of the hydrophobic effect. First, denaturation upon cooling, or "cold denaturation." Cold denaturation conflicts with the classical view of an entropically favorable state and an enthalpically favorable native state. Second, the structural characteristics of the cold

Published by the American Physical Society under the terms of the Creative Commons Attribution 3.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

denatured state. Third, the temperature dependence of the heat capacity. Using differential scanning calorimetry [26], the heat capacity of the system can be calculated as $C_P = (dQ/dT)_{P,N}$. The heat capacity of the system is commonly used as a well-defined experimental observable to characterize the thermodynamics of the folding transition.

We simulate a protein consisting of 80 residues with Monte Carlo sampling using a classic lattice model to investigate the effect of the entropic contribution of the hydrophobic potential. To model the effective potential for hydrophobe–water interactions, we introduce the following temperature-dependent term for the surface residues (s) and the fully hydrated residues (h)

$$F_{\text{hydr}} = -\alpha_s N_s (T - T_{0,s})^2 - \alpha_h N_h (T - T_{0,h})^2, \qquad (1)$$

describing second order approximations to the theory of the hydrophobic effect [2,5] for both groups. Here, N_s is the number of hydrophobic residues on the surface, N_h is the number of hydrophobic residues that are fully hydrated and T is the temperature in reduced units. The temperature dependence of the fully hydrated (α_h, $T_{0,h}$) and surface (α_s, $T_{0,s}$) residues are set using Ref. [5] [see Fig. 1(a)]. In our lattice model, we define a residue that is fully hydrated as having at least four sides exposed and for a residue that is partially solvated as having at least one, and no more than three sides exposed to the solvent. Fitting the expression in Eq. (1) to the results [2] from the LCW theory yields $\alpha_s = 3.0$ and $T_{0,s} = 0.41$ for the surface term and $\alpha_h = 7.0$ and $T_{0,h} = 0.49$ for the volume solvation term. This assumes that, for the temperature dependence, all amino acids have the same size, while in practice, the volume of amino acids can vary from 75 to 240 Å [27]. To test the sensitivity of our model to this assumption, we performed simulations

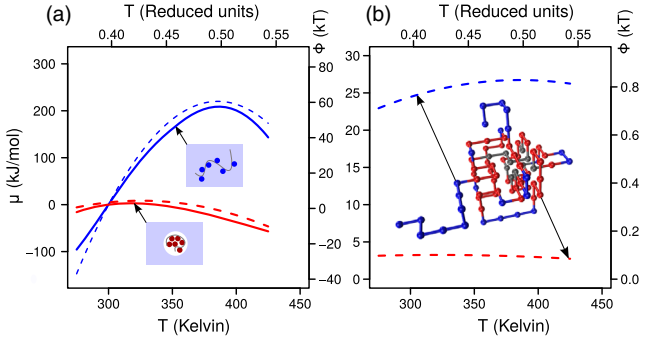

FIG. 1. Comparison between lattice model and LCW theory for a polyphenylalanine hydrophobic chain. (a) The chemical potential for a fully extended chain as a function of temperature (blue lines), and the chemical potential for a compacted chain, which we approximate as a 10 Å sphere (red lines). The dashed lines indicate the approximation made by our lattice model, while the solid lines indicate the theoretical predictions from LCW theory [1,2]. (b) The distinction between surface and fully solvated residues in our model. The blue line shows the potential for the fully solvated residues (corresponding to the residues colored blue), and the red line shows the surface potential (corresponding to the residues colored red).

with three different potentials: a temperature independent potential ($\alpha_s = \alpha_h = 0$), a temperature dependent potential (parameters given above), and a strongly temperature dependent potential, corresponding to amino acids that are 15% larger ($\alpha_s = 4.5$ and $\alpha_h = 11.5$) (derivation shown in Supplemental Material sections "Derivation temperature dependent potential" and "Approximation of hydrophobicity parameters") [28].

First, we probe the folding specificity of this model. The lattice model we use here is sequence dependent. In other words, random sequences will typically not fold into a stable structure, whereas designed sequences do so with a high specificity [17–19,21,29–31].

The number of native contacts (N_{int}) is used as an order parameter for the specificity of protein folding. We define a protein to be folded when $N_{\text{int}} > 75$. The fraction of the simulation spent in this folded state is defined as P_{Fold}. For the purpose of this Letter, umbrella sampling [32] alone is sufficient to sample the configurational space of interest. Figure 2(a) shows that, for all potentials, the protein folds ($P_{\text{Fold}} > 0.5$) at intermediate temperatures and denatures ($P_{\text{Fold}} < 0.5$) at high temperatures. This is consistent with the view of the high-entropy denatured state caused by the chain entropy. For a well-designed protein, the stability of the protein simulated with a temperature-independent potential is a strictly decreasing function of the temperature, since the native state is optimized to be the lowest enthalpy state.

Only the strongly temperature-dependent potential reproduces cold denaturation as well as heat-induced denaturation, see black curve in Fig. 2(a). A very similar folding curve has been observed experimentally for a mutant of cold shock protein Csp [33]. The simulated configurational ensemble of the folded state also includes a small fraction of denatured states ($0.20 < T < 0.37$) as observed in the experiment. Note that Csp, like most proteins, does not show cold denaturation above the freezing point of water. However, statistical investigation has shown that the temperature has a measurable influence on the propensity of hydrophobic amino acids to be buried [34]. This is similar to our observation that proteins become less stable at lower temperatures, but do not denature, for a lower value of the temperature dependence. (Fig. 2, green line).

The structural characteristics of the model were investigated by exploring the free energy landscapes of native contacts (N_{int}) and internal contacts between residues (C_{int}); the latter are used as a measure of compactness. At $T = 0.375$, slightly below the transition temperature ($T = 0.42$), two distinct states can be observed, one where the protein is specifically folded ($N_{\text{int}} > 75$), and one in which the protein is mostly unstructured ($N_{\text{int}} < 25$), with a clear barrier separating the two states [Fig. 2(c)]. Note that the sequence has been designed to fold in this exact structure with 97 native contacts (see Methods in Supplemental Material [28]).

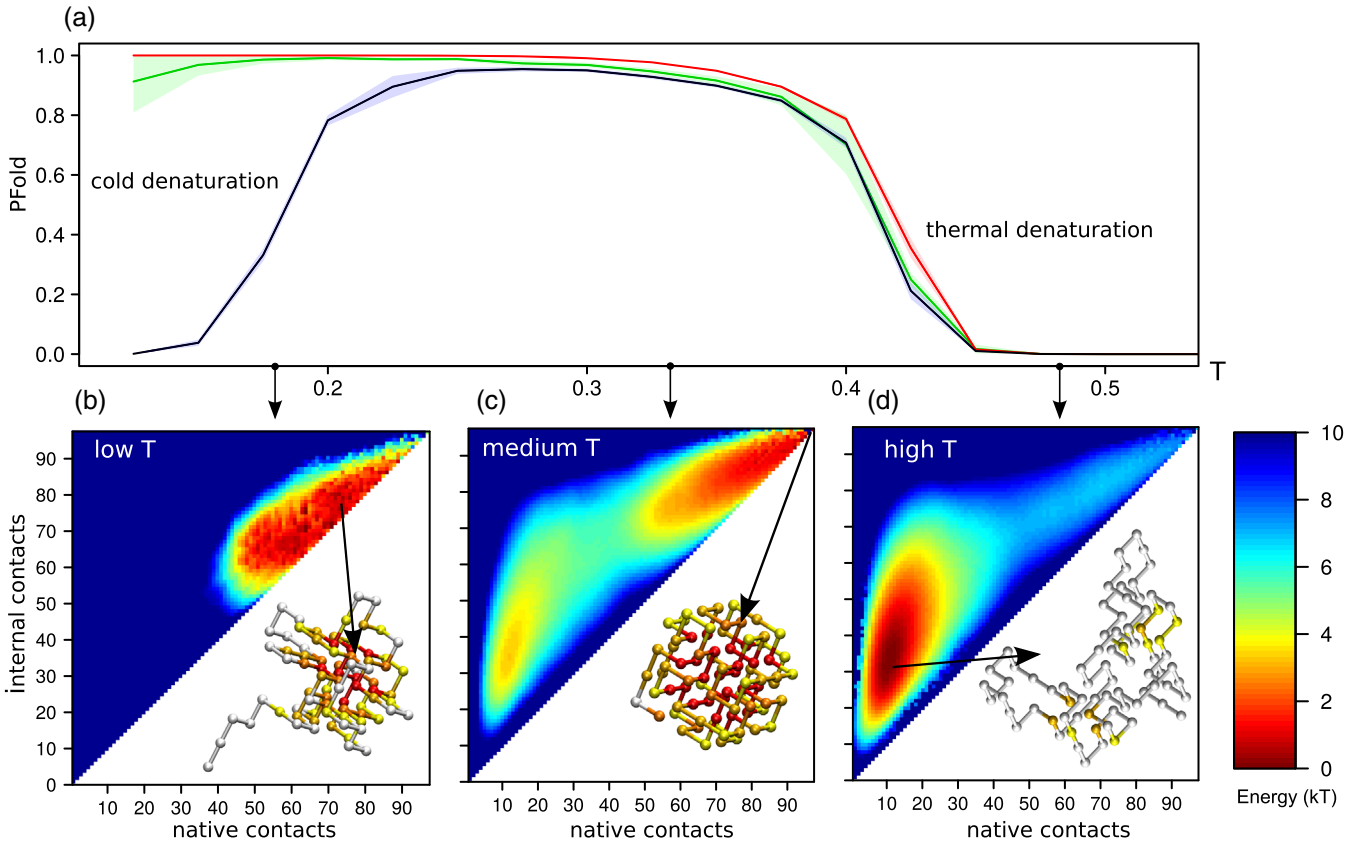

FIG. 2. Temperature-dependent folding stability and structure. The folded state has 97 native contacts. (a) The probability for the model protein to be in the folded state versus temperature, with $\alpha = 0$ (red line), the temperature dependent potential (green line), and a simulation where the temperature dependence is multiplied by 1.5 (black line). The 95% confidence interval is indicated by the shaded area. (b)–(d) Free energy landscapes for the number of native contacts (N_{int}) and all internal contacts (C_{int}) for the simulations with a strong temperature dependence at (b) low temperature ($T = 0.175$), (c) intermediate temperature ($T = 0.375$), and (d) high temperature ($T = 0.475$). For the strong temperature dependent potential, the protein denatures at low temperatures, with many exposed hydrophobic amino acids. However, this denatured structure is a lot more compact than the heat denatured protein, and there are less native contacts present. At intermediate temperatures the protein has the highest stability in its folded configuration (indicated by the arrow) where $N_{\mathrm{int}} = C_{\mathrm{int}} = 97$. At high temperatures, the protein makes only transient contacts.

Comparing Figs. 2(b) and 2(d), it becomes apparent that the cold denatured state has more residual structure than the heat denatured state. The cold denatured configurational ensemble at $T = 0.125$ shows a structure that is compact with approximately two thirds of the native contacts present, similar to experimental NMR observations of pressure-assisted cold denaturation [35], urea-assisted cold denaturation [36], and cold denaturation for a protein that was destabilized by a mutation [37]. Note that, for some disordered proteins, the radius of gyration decreases as the temperature increases [38,39]. This is most likely due to interactions involving charged residues, which play a larger role in disordered proteins. Notably, the λ repressor, which is the most hydrophobic protein in the dataset investigated in Ref. [38], does show a reexpansion at temperatures higher than 319 K [38].

In addition to the structural characteristics, our model allows us to investigate the role of the hydrophobic effect in the thermodynamics of protein folding. We start by investigating the heat capacity of folding. Note that the simulations are performed at constant volume, while the experiments are done at constant pressure. However, the difference is negligible for the system in consideration due to the low compressibility of water [40] (see Supplemental Material [28] for more detail). In order to calculate the heat capacity (C_V in our model) for the temperature-dependent potential, we need to separate the expected enthalpy $\langle E \rangle$ from the entropic part of the hydrophobic potential, F_{hydr} (see Supplemental Material [28]). In a finite system, a phase transition is usually characterized by a sharp peak in the heat capacity that can be observed experimentally [41]. For the temperature-independent potential, we observe only a single peak at the folding transition [Fig. S3(a) [28]]. In contrast, the heat capacity of the temperature dependent potential shows two peaks, one for cold induced denaturation and one for heat induced denaturation [Fig. S3(b) [28]]. Another interesting observation is a linear temperature dependence of the heat capacity in the temperature range where no phase transition is occurring [Fig. 3(a)]. The slope or the linear increase in the baseline of the heat capacity has been investigated in Refs. [42–46]. In the

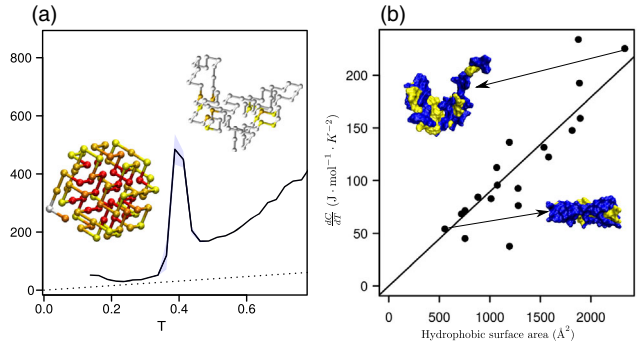

FIG. 3. Heat capacity versus temperature. (a) The heat capacity for the fitted potential as calculated by $C_V = (dE/dT)$, corresponding to the green line in Fig. 2(a), shows a linear increase in the heat capacity with respect to temperature. Our model suggests that the slope of this baseline, which is a lower bound for the heat capacity, should correlate with the amount of hydrophobic surface area. (b) The slope of the heat capacity (dC/dT), shows a strong correlation with the exposed hydrophobic surface area in real protein structure as predicted by our model. Two indicative protein structures, with PDB codes 1J46 and 2ZTA are shown with the hydrophobic and hydrophilic amino acids colored yellow and blue, respectively.

context of the current model, we can understand the linear T dependence of the heat capacity in terms of the exposed hydrophobic groups. Assuming a constant hydrophobe-water contact area and neglecting entropic contributions other than the hydrophobe-water contact area, we can derive a simple lower bound for the heat capacity (see Supplemental Material [28] for derivation)

$$C_V(T) = (2\alpha_s N_s + 2\alpha_h N_h)T, \tag{2}$$

where N_s is the number of hydrophobic amino acids that are at the surface, and N_h the number of hydrophobic amino acids that are fully hydrated. This simple calculation yields a lower bound for the heat capacity in regions where no folding transition occurs, as indicated by the dotted lines in Fig. 3(a). Here, we estimate N_s and N_h for the folded regime from the number of exposed hydrophobes in the native structure ($N_s = 13$, $N_h = 0$). The difference between the lower bound shown in Fig. 3(a) and the simulated results is likely due to the chain entropy in the native ensemble. This is supported by the results from the temperature independent potential [see, also, Fig. S3(a) [28]].

Equation (2) has additional consequences that can be verified experimentally. Initially, as a consistency check, it is easy to see that, from Eq. (2), we can recover the well-known relationship between the change in heat capacity, ΔC_V and the change in hydrophobic surface area upon folding at a given temperature T, e.g., Refs. [47,48]. We can go further, however, and probe the derivative of the heat capacity with respect to the temperature (dC_V/dT), as shown by the dotted lines in Fig. 3(a). Our analysis in Eq. (2) predicts that this slope itself is proportional to the exposed hydrophobic surface area (corresponding to N_h in the model) in a given state.

To test this prediction, we compared the heat capacity slopes for folded proteins tabulated in Ref. [42] with the level of exposed hydrophobic surface area in the corresponding folded structures obtained with [49]. We find a strong correlation ($R^2 = 0.77$) between the slope of the heat capacity and the exposed hydrophobic surface area [see Fig. 3(b)], further supporting the temperature dependence of the hydrophobic effect as a key mechanism underlying the linear increase of the heat capacity.

Previously, a higher slope for DNA-binding proteins has been observed when compared to globular proteins of the same size, see also Supplemental Material Fig. 2, and Refs. [28,42,50,51] was rationalized through the flexibility of DNA-binding domains [42,51]. The results suggest that the increased slope of the heat capacity may be explained solely by a higher exposed hydrophobic surface area. Hence, the strong correlation between the slope of the heat capacity and the exposed hydrophobic surface area for all proteins may be explained by a consistent treatment of hydrophobicity alone (see Fig. S2 [28]). This does not, however, preclude a possible correlation between the flexibility and the amount of exposed surface area.

To conclude, we have presented an extension for a coarse-grained lattice model by including a temperature-dependent hydrophobic term in the interaction potential. The combination of the coarse-grained steric model and the potential ensures appropriate contributions of the solvent-residue and internal contacts. This allows us to separate chain entropy from the solvent entropy. An effective quadratic potential is applied to account for the temperature dependence of the hydrophobic effect. Simulating the model, we observe cold denaturation for realistic parameter settings, suggesting that the hydrophobic effect is the key component in cold denaturation. In addition, we find that the simulated cold denatured state is more compact than the heat denatured state; this is in agreement with experimental observations. Moreover, the model is able to reproduce the characteristic experimental heat capacity curves for protein folding.

Starting from the temperature dependent potential, we derive a simple relation that approximates the heat capacity baseline of the native state of the protein. This relation is tested for a set of real proteins, where we do, indeed, find a correlation between the hydrophobic surface area and the slope of the heat capacity. Hence, our model seems to make accurate predictions for thermodynamic behavior of real proteins. Additionally, the developed relation can potentially be used to calculate an accurate baseline for proteins with a known structure.

The interaction potential is based on a representative set of the protein database, PDB-25 [24], and holds some biases towards rigid and soluble proteins, since they are easier to crystallize (see Supplemental Material [28] for further discussion). The cubic lattice model also has some limitations: secondary structure can not be modeled explicitly, nor is there sufficient molecular detail to predict the true fold of a protein sequence. However, we stress that the use of our

coarse-grained model is not only motivated by considerations of computational cost. Rather, the use of simple, coarse-grained models allows us to reveal the minimal physical ingredients that a model needs to account for cold denaturation.

The temperature-dependent potential developed here is applicable to other (off-lattice) coarse-grained models with implicit solvent-side chain interactions, e.g., Refs. [20,52,53]. The protocol for calculating the heat capacity from a temperature-dependent potential, can be applied to any effective potential with a closed form expression that is continuous and differentiable with respect to $\beta = (1/k_B T)$. Furthermore, the addition to the lattice model itself will enable investigation of temperature dependence of protein aggregation building on previous studies [19,22,54].

The software used to run the simulations are available in Ref. [55].

We would like to thank Alexander Buell, Chris Dobson, and Anton Feenstra for helpful discussions. We would like to thank Robbin Bouwmeester for proofreading the manuscript. S. A. is supported by a Veni grant on the project 'Understanding Toxic Protein Oligomers through Ensemble Characteristics' from the Netherlands Organization for Scientific Research (NWO). T. P. J. K. acknowledges support from ERC Advanced Grant No. 337969. D. F. acknowledges support from ERC Advanced Grant No. 227758 and EPSRC Programme Grant No. EP/I001352/1. P. V. acknowledges EU Marie Curie Grant No. PIIFGA2011300045.

*e.van.dijk@vu.nl
†patvarilly@gmail.com
‡tpjk2@cam.ac.uk
§df246@cam.ac.uk
‖s.abeln@vu.nl

[1] K. Lum, D. Chandler, and J. D. Weeks, J. Phys. Chem. B **103**, 4570 (1999).
[2] D. M. Huang and D. Chandler, Proc. Natl. Acad. Sci. U.S.A. **97**, 8324 (2000).
[3] A. Kim and F. Szoka, Jr., Pharm. Res. **09**, 504 (1992).
[4] B. Widom, P. Bhimalapuram, and K. Koga, Phys. Chem. Chem. Phys. **5**, 3085 (2003).
[5] D. Chandler, Nature (London) **437**, 640 (2005).
[6] B. A. Patel, P. G. Debenedetti, F. H. Stillinger, and P. J. Rossky, Biophys. J. **93**, 4116 (2007).
[7] B. A. Patel, P. G. Debenedetti, and F. H. Stillinger, J. Phys. Chem. A **111**, 12651 (2007).
[8] B. A. Patel, P. G. Debenedetti, F. H. Stillinger, and P. J. Rossky, J. Chem. Phys. **128**, 175102 (2008).
[9] C. L. Dias, T. Ala-Nissila, M. Karttunen, I. Vattulainen, and M. Grant, Phys. Rev. Lett. **100**, 118101 (2008).
[10] S. Romero-Vargas Castrillon, S. Matysiak, F. H. Stillinger, P. J. Rossky, and P. G. Debenedetti, J. Phys. Chem. B **116**, 9540 (2012).
[11] V. Bianco and G. Franzese, Phys. Rev. Lett. **115**, 108101 (2015).
[12] A. Mitsutake, M. Kinoshita, Y. Okamoto, and F. Hirata, J. Phys. Chem. B **108**, 19002 (2004).
[13] A. Davtyan, N. P. Schafer, W. Zheng, C. Clementi, P. G. Wolynes, and G. A. Papoian, J. Phys. Chem. B **116**, 8494 (2012).
[14] B. J. Sirovetz, N. P. Schafer, and P. G. Wolynes, J. Phys. Chem. B **119**, 11416 (2015).
[15] A. Sali, E. Shakhnovich, and M. Karplus, J. Mol. Biol. **235**, 1614 (1994).
[16] S. Miyazawa and R. L. Jernigan, Macromolecules **18**, 534 (1985).
[17] E. I. Shakhnovich, Phys. Rev. Lett. **72**, 3907 (1994).
[18] I. Coluzza, H. G. Muller, and D. Frenkel, Phys. Rev. E **68**, 046703 (2003).
[19] S. Abeln and D. Frenkel, PLoS Comput. Biol. **4**, e1000241 (2008).
[20] I. Coluzza, PLoS One **6**, e20853 (2011).
[21] S. Abeln and D. Frenkel, Biophys. J. **100**, 693 (2011).
[22] S. Abeln, M. Vendruscolo, C. M. Dobson, and D. Frenkel, PLoS One **9**, e85185 (2014).
[23] P. Kukic, A. Kannan, M. J. J. Dijkstra, S. Abeln, C. Camilloni, and M. Vendruscolo, PLoS Comput. Biol. **11**, e1004435 (2015).
[24] S. Griep and U. Hobohm, Nucleic Acids Res. **38**, D318 (2010).
[25] J. D. Weeks, D. Chandler, and H. C. Andersen, J. Chem. Phys. **54**, 5237 (1971).
[26] B. Hallerbach and H.-J. Hinz, Biophys. Chem. **76**, 219 (1999).
[27] A. K. Mishra and J. C. Ahluwalia, J. Phys. Chem. **88**, 86 (1984).
[28] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.116.078101 for more information on the sampling techniques, the protein sequence used in the lattice simulations, the protein identifiers of the protein structures investigated and the associated data, the derivation of the heat capacity and the results of a simplified model.
[29] E. Shakhnovich, G. Farztdinov, A. M. Gutin, and M. Karplus, Phys. Rev. Lett. **67**, 1665 (1991).
[30] E. I. Shakhnovich and A. M. Gutin, Proc. Natl. Acad. Sci. U.S.A. **90**, 7195 (1993).
[31] E. I. Shakhnovich and A. M. Gutin, Protein Eng. **6**, 793 (1993).
[32] A. Grossfield, WHAM: The weighted histogram analysis method, version 2.09, http://membrane.urmc.rochester.edu/content/wham.
[33] T. Szyperski, J. Mills, D. Perl, and J. Balbach, Eur. Biophys. J. **35**, 363 (2006).
[34] E. van Dijk, A. Hoogeveen, and S. Abeln, PLoS Comput. Biol. **11**, e1004277 (2015).
[35] N. Vajpai, L. Nisius, M. Wiktor, and S. Grzesiek, Proc. Natl. Acad. Sci. U.S.A. **110**, E368 (2013).
[36] K.-B. Wong, S. M. V. Freund, and A. R. Fersht, J. Mol. Biol. **259**, 805 (1996).
[37] B. Shan, S. McClendon, C. Rospigliosi, D. Eliezer, and D. P. Raleigh, J. Am. Chem. Soc. **132**, 4669 (2010).
[38] R. Wuttke, H. Hofmann, D. Nettels, M. B. Borgia, J. Mittal, R. B. Best, and B. Schuler, Proc. Natl. Acad. Sci. U.S.A. **111**, 5213 (2014).
[39] P. L. Privalov and G. I. Makhatadze, J. Mol. Biol. **232**, 660 (1993).
[40] D. V. Schroeder, *An Introduction to Thermal Physics* (Addison Wesley, San Francisco, 2000), p. 159.

[41] P. L. Privalov, E. I. Tiktopulo, S. Venyaminov, Y. Griko, G. I. Makhatadze, and N. N. Khechinashvili, J. Mol. Biol. **205**, 737 (1989).
[42] P. L. Privalov and A. I. Dragan, Biophys. Chem. **126**, 16 (2007).
[43] V. Muñoz and J. M. Sanchez-Ruiz, Proc. Natl. Acad. Sci. U.S.A. **101**, 17646 (2004).
[44] P. Farber, H. Darmawan, T. Sprules, and A. Mittermaier, J. Am. Chem. Soc. **132**, 6214 (2010).
[45] P. Bruscolini and A. N. Naganathan, J. Am. Chem. Soc. **133**, 5372 (2011).
[46] C. M. Johnson, Arch. Biochem. Biophys. **531**, 100 (2013).
[47] R. S. Spolar, J. H. Ha, and M. T. Record, Proc. Natl. Acad. Sci. U.S.A. **86**, 8382 (1989).
[48] E. V. Anslyn and D. A. Dougherty, in *Modern Physical Organic Chemistry* (University Science, Sausalito, CA, 2006), Chap. 3, pp. 145–206.
[49] W. Kabsch and C. Sander, Biopolymers **22**, 2577 (1983).
[50] J. Gómez, V. J. Hilser, D. Xie, and E. Freire, Proteins: Struct., Funct., Bioinf. **22**, 404 (1995).
[51] A. N. Naganathan, R. Perez-Jimenez, V. Munoz, and J. M. Sanchez-Ruiz, Phys. Chem. Chem. Phys. **13**, 17064 (2011).
[52] T. X. Hoang, A. Trovato, F. Seno, J. R. Banavar, and A. Maritan, Proc. Natl. Acad. Sci. U.S.A. **101**, 7960 (2004).
[53] S. Auer, F. Meersman, C. M. Dobson, and M. Vendruscolo, PLoS Comput. Biol. **4**, e1000222 (2008).
[54] R. Ni, S. Abeln, M. Schor, M. A. Cohen Stuart, and P. G. Bolhuis, Phys. Rev. Lett. **111**, 058101 (2013).
[55] http://bitbucket.org/edk360/hydrophold.