Folding domain B of protein A on a dynamically partitioned free energy landscape

Erik D. Nelson* and Nick V. Grishin

Howard Hughes Medical Institute and University of Texas Southwestern Medical Center, 6001 Forest Park Boulevard, Room ND10.124, Dallas, TX 75235-9050

Edited by José N. Onuchic, University of California at San Diego, La Jolla, CA, and approved December 7, 2007 (received for review June 19, 2007)

The B domain of staphylococcal protein A (BdpA) is a small helical protein that has been studied intensively in kinetics experiments and detailed computer simulations that include explicit water. The simulations indicate that BdpA needs to reorganize in crossing the transition barrier to facilitate folding its C-terminal helix (H3) onto the nucleus formed from helices H1 and H2. This process suggests frustration between two partially ordered forms of the protein, but recent ϕ value measurements indicate that the transition structure is relatively constant over a broad range of temperatures. Here we develop a simplistic model to investigate the folding transition in which properties of the free energy landscape can be quantitatively compared with experimental data. The model is a continuation of the Muñoz–Eaton model to include the intermittency of contacts between structured parts of the protein, and the results compare variations in the landscape with denaturant and temperature to ϕ value measurements and chevron plots of the kinetic rates. The topography of the model landscape (in particular, the feature of frustration) is consistent with detailed simulations even though variations in the ϕ values are close to measured values. The transition barrier is smaller than indicated by the chevron data, but it agrees in order of magnitude with a similar α-carbon type of model. Discrepancies with the chevron plots are investigated from the point of view of solvent effects, and an approach is suggested to account for solvent participation in the model.

folding landscape | frustration | protein folding

The routes a protein travels on the way to its native ensemble are the result of a sensitive exchange between energy and entropy coupling the protein to its solvent environment. Often the basic topography of the folding landscape (1) can be understood just from what would be necessary kinetically for a protein-like molecule to organize itself into a particular native shape (2–4), but the detailed way in which a protein interacts with solvent can qualitatively affect the way it folds (5, 6). The conformation of a protein is coupled to large-scale fluctuations in solvent density (7–9), and for small globular proteins investigated in kinetics experiments, most of the change in solvent exposure (10) can occur before the protein is half folded. The rate at which a protein explores its conformational space is slaved to local solvent motions (11, 12), and together these effects can have a significant influence on the time scale for folding.

Recently methods have become available to simulate proteins with explicit solvent on a time scale long enough to explore folding (13–15). These simulations provide the most precise picture of protein kinetics, but are as a result less wieldy to describe variations that occur across different experiments and solvent conditions (16–18). Typically, much more simplistic α-carbon type models have been used to address these problems, and when guided by experiment they can provide an accurate description of the interplay among protein folding routes (3, 4). A parallel approach established by Muñoz and Eaton (19–21) is to partition the free energy landscape of the protein into a network of small ensembles that can be described by polymer physics methods. This approach has the advantage that solvent effects might be included on a practical time scale, but it is usually less accurate than explicit chain models because the partitioning methods introduce dynamical constraints that conflict with, or oversimplify, the kinetics of the protein (22, 23). Here we develop a method to solve this problem.

In the Muñoz–Eaton model, the configuration space of a protein is in effect partitioned in terms of a local order parameter describing the proximity of chemical bond angles in the residues to their native values. The partitioning is accomplished by labeling the residues in a protein configuration as folded (native like) or unfolded (free, or ensemble like) depending on their order parameter values, and then grouping together configurations that have the same pattern of labels. Configurations with unfolded loops (24, 25) are neglected in the model so that folded segments always represent independent nuclei, and consequently the ensemble free energies are simple to describe. The model can be solved exactly by special techniques (26, 27), and it provides a convenient way to construct basic inferences from experiments (28, 29). Nevertheless, the terms that are being neglected (unfolded loops and intermittency of native contacts between folded segments (30, 31)) are precisely those that coordinate the expulsion of solvent from the protein (2, 13, 32), and it is of interest to describe these terms from the point of view of the original model.

We accomplish this here by continuing the geometric approach of Muñoz and Eaton, including separate order variables to describe the state of contact between the folded segments. The resulting system of constraints resembles the local-order coupling model for explicit chains (33). To test the model, we use it to describe the folding landscape of the B domain of staphylococcal protein A (BdpA), which allows an extensive comparison to thermodynamics and kinetics measurements. BdpA (Fig. 1) is one of the most widely studied proteins of the past decade (34). Folding kinetics and single-molecule FRET measurements indicate that BdpA folds cooperatively over a broad range of solvent conditions (18), yet the system is still small enough to simulate by first principles methods on a computer (13–15). The folding nucleus (16), which forms around the first two helices, includes most of the binding interface from the complex with an IgG Fc fragment (35); H3 is unstructured in the complex, but structured in the transition state in all-atom simulations; however, in crossing the transition barrier, further rearrangement of the helices is required to consolidate contacts with H3 before the protein can fold. These features suggest that BdpA is weakly frustrated by extra "functional" requirements (binding preferences) of *Staphylococcus aureus* (34). But recent measurements to detect temperature changes in the transition structure (17) indicate that whatever frustration exists, it is not strong enough to divide the landscape into kinetically separate routes (29). By fitting our model to thermodynamic stability data (17, 29) we are able to bridge between the kinetics measurements

Author contributions: E.D.N. designed research; E.D.N. performed research; E.D.N. and N.V.G. analyzed data; and E.D.N. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

*To whom correspondence should be addressed. E-mail: enelson@spirit.sdsc.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0705707105/DC1.

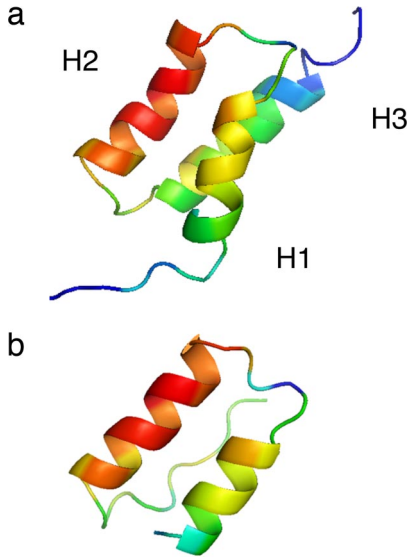

Fig. 1. Native structures of BdpA. (*a*) NMR structure of wild-type BdpA colored according to ϕ values calculated later in this work. ϕ values in the range $0 \leq \phi \leq 0.8$ correspond to blue, green, yellow, orange, and red. (b) Crystal structure from the complex of BdpA with the IgG Fc fragment colored as in *a*.

and the detailed simulations to describe the folding landscape under various solvent conditions.

Model

We consider systems that can be described by a Gō (native only) interaction model in which nonnative contact is repulsive at short distances just to exclude the volume occupied by the protein (3, 4). Let $q_i = 0, 1$ label the local structure of a residue i (folded or unfolded) in a partially unfolded protein as in the Muñoz–Eaton model. Each pattern of labels $\{q_i\}$ describes an ensemble of protein configurations Γ with structurally similar folded segments α_k. To describe the intermittency of contacts between the α_k, we partition Γ into smaller ensembles $\Gamma_n \in \Gamma$ according to whether the degree of native organization between pairs of folded segments α_k and α_l exceeds some threshold value. This is accomplished by providing the folded segments with separate labels (superscripts), $\alpha_k \rightarrow \alpha_k^a$, so that native contact between two segments α_k^a and α_l^b is established only if $a = b$. Leaving to later the question of what actually decides the threshold values for native order, the ensembles Γ_n can each be depicted by a Feynman-like diagram (36) formed by the following procedure [see supporting information (SI) Fig. 5]: First a line is drawn to represent the unfolded protein and labeled circles are placed at points along its length to represent the α_k^a. Circles with the same label are then joined together into nuclei by bending the unfolded lines to form the diagram (note that two circles can have the same label only if contacts exist between them in the native state). Each Γ contains an ensemble Γ_1 corresponding to the fully collapsed (natively connected) configuration of the α_k^a, and each partially collapsed diagram $\Gamma_{n\neq 1}$ corresponds to a topologically different way of "pulling apart" the folded segments in Γ_1. For small proteins like BdpA, there is no need to consider diagrams with more than four folded segments, so the number of diagrams in Γ is always ≤ 15.

To describe the free energy of a diagram the energy is then defined as $E = -\varepsilon \Sigma_{i<j} Q_{ij}(\Gamma_n)$, where ε is the energy per contact, $Q_{ij}(\Gamma_n) = 0$ or Q_{ij}^N, and Q_{ij}^N is the native contact matrix. $Q_{ij}(\Gamma_n) = Q_{ij}^N$ only if residues i and j are located in the same nucleus.

To describe the entropy, first note that besides Γ_1, each Γ contains an ensemble Γ_0 in which the α_k^a are completely pulled apart (all labels are different). The entropy Δs of unfolded residues in Γ_0 can be estimated from the entropy of unfolded proteins (37), and the entropy cost $\Delta S(\Gamma_n)$ to then sequentially

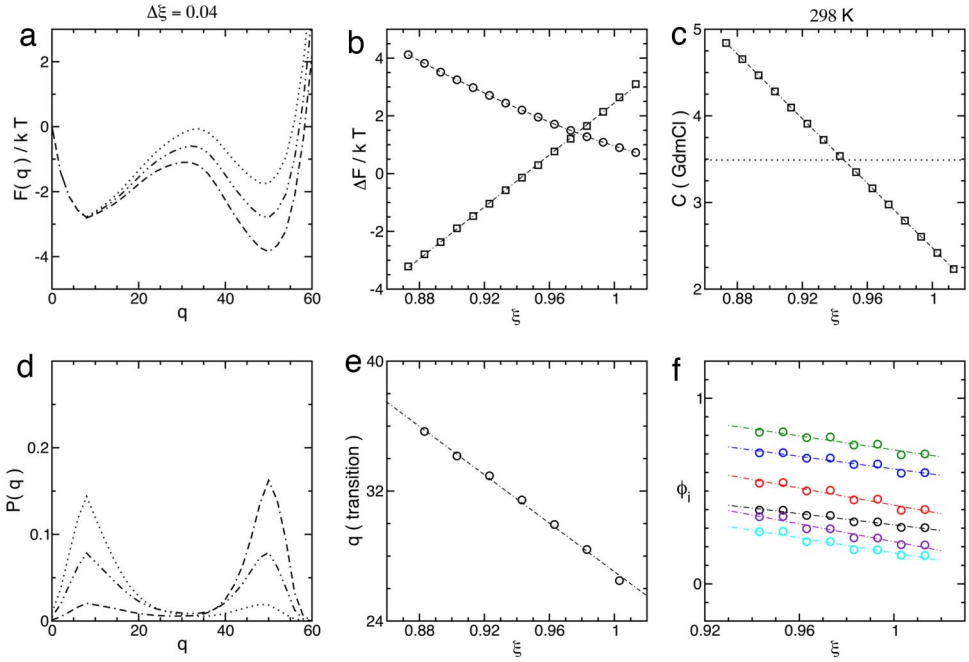

Fig. 2. ξ dependence of landscape parameters. (*a*) Free energy profile across the melting transition, $\xi_m \simeq 0.946$. (*b*) Melting free energy $\Delta F_{\ddagger\text{-}D}$ (□) and transition state barrier $\Delta F_{\ddagger\text{-}D}$ (○). (*c*) Mapping between stability values and denaturant concentration at 25°C. The dotted line indicates the melting concentration $c_m \simeq 3.49$ M GdmCl. Data points occur at integer multiples of $\Delta\xi = 0.01$ ($\Delta c \simeq 0.186$) about $\xi \simeq 0.943$. (*d*) Distribution of states $P(q)$ corresponding to *a*. The stability range corresponds to the FRET data in figure 6 of ref. 18. (*e*) Location of the transition $q^{\ddagger}(\xi)$. The data points were calculated by fitting the barrier to a quadratic. (*f*) ϕ values for the degree of helix formation as in figure 5 of ref. 17 (the color scheme is the same, and the plot describes a similar range of stability). $\phi_i(\xi)$ variations for tertiary structure are slightly larger.

close Γ_0 into any of the diagrams $\Gamma_n \in \Gamma$ will be estimated in terms of the cost for closing a single loop,

$$\Delta s_{\text{loop}} = k \ln[p(\mathbf{x}, \mathbf{x}')\Delta v], \qquad [1]$$

where $p(\mathbf{x}, \mathbf{x}')\Delta v$ is the fraction of unrestricted chain configurations emanating from a point $\mathbf{x}$ that terminate at the point $\mathbf{x}'$ within a small volume Δv about the size of a residue (38) and k is Boltzmann's constant. The entropy of a diagram is then $S = (N - q)\Delta s - \Delta S$, where N is the number of residues, q is the number of folded residues, and ΔS is a sum of terms described by Eq. **1**.

Random loop insertions in proteins are well modeled as Gaussian chains with excluded volume (39), and the persistence length appears close to the size of a backbone unit (40, 41). Therefore, we neglect relatively small excluded volume effects here and use the simplest possible Gaussian approximation,

$$p(\mathbf{x}, \mathbf{x}') = \left[\frac{2\pi l a^2}{3}\right]^{-\frac{3}{2}} e^{-3(\mathbf{x}-\mathbf{x}')^2/2la^2}, \qquad [2]$$

giving

$$\Delta s_{\text{loop}} = -k\frac{3}{2}\left[\ln(l) - \frac{r^2}{la^2}\right] + k\ln\left[\frac{2\pi a^2}{3}\right]^{-\frac{3}{2}}\Delta v, \qquad [3]$$

where l is the number of links, r is the distance between end residues, $a = 3.8$ Å, and for simplicity we let $\Delta v = \frac{4}{3}\pi a^3$, which is about the volume of a large branched residue (38).

Typically, the structure of a diagram makes it obvious how to approximate ΔS in terms of $\Delta s_{\text{loop}}(l, r)$. But we also have to consider situations where, for example, an α_k^a appears along the unfolded part of a loop. This event would lead to a kind of effective loop consisting of two unfolded segments of length l_1 and l_2 joined across a physical distance r' by the folded segment. To describe this diagram, we first interpret the folded segment as if it were an unfolded part x of a loop of total length $l = l_1 + x + l_2$, where

$$x = \left[\frac{r'}{a}\right]^{1+\delta} \qquad [4]$$

and $\delta \in [0, 1]$. The variable part of the entropy cost is then multiplied by

$$\sum(l, x) = 1 - \frac{x+1}{l} \qquad [5]$$

to account for the entropy of segment x. The few more complex diagrams that require this approximation contribute minimally to the results (*SI Text* and SI Fig. 6). The contact matrix and parameters entering into these expressions are described in *Methods*.

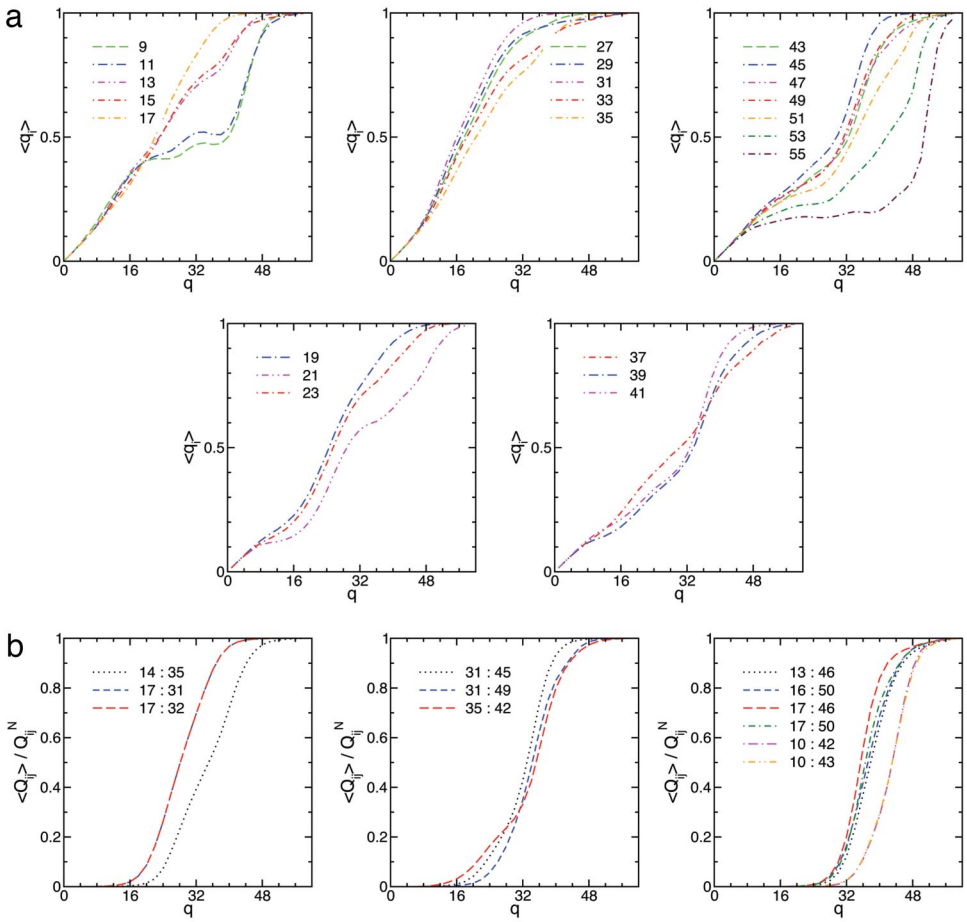

Fig. 3. Degree of local order, $\langle q_i \rangle$, for helix and turn residues (*a*) and degree of interhelical contact order, $\langle Q_{ij} \rangle / Q_{ij}^N$, across helices H1:H2, H3:H2, and H1:H3 near the melting transition (*b*). Keys identify the data for residue index *i* or index pair *i*: *j*. Helix H2 and the C-terminal part of H1 form early, but *a* indicates that some reorganization occurs in H1 and the turns as the transition is crossed at $q^{\ddagger} = 32$. In *b*, H1:H2 contacts form early, then H3:H2 and finally H1:H3 contacts form during the reorganization process described in *a*.

Results

The system can now be depicted as a kind of foam, or network (1) of ensembles with free energy $F(\Gamma_n) = E(\Gamma_n) - TS(\Gamma_n)$, where T is the temperature. To quantitatively compare the model to the data, the melting free energy of the model is mapped to the measured free energy in terms of a stability parameter ξ. To accomplish this, the energy is written as $\varepsilon/T = \xi N\Delta s/Q^N$, where $Q^N = \Sigma_{i<j} Q_{ij}^N$. The free energy profile is $F(q) = -kT\ln Z(q)$ (Fig. 2*a*), where $Z(q) = \Sigma_{\Gamma_n}^q \exp(-F(\Gamma_n)/kT)$ and the sum includes ensembles with q folded residues. The ends of the profile balance when $\xi = 1$.

The profile can be quantified by its critical points in the denatured (D), transition (‡), and native (N) ensembles: The transition barrier is defined as $\Delta F_{\ddagger\text{-D}} = F(q^\ddagger) - F(q^\text{D})$and the melting free energy is $\Delta F_\text{D-N} = kT\ln Z_\text{D}/Z_\text{N}$, where $Z_\text{D} = \Sigma_{q\le q^\ddagger} Z(q)$ and $Z_\text{N} = \Sigma_{q>q^\ddagger} Z(q)$ are the partition functions for denatured and native ensembles (Fig. 2*b*). ξ is mapped to solvent conditions by equating $\Delta F_\text{D-N}(\xi)$ to the melting free energy measured at $T = 25$°C (17). Fig. 2*c* plots this relationship and connects the melting point ξ_m to the melting concentration of denaturant. The predicted melting temperature in 2 M guanidinium chloride (GdmCl) is $T_\text{m} \simeq 50$°C, very close the value $T_\text{m} \simeq 52$°C measured by Sato and Fersht (17), and the melting transition closely resembles single model FRET data (18) (Fig. 2*d*). The folding barrier ($\sim 2kT$ at midpoint) has the same order of magnitude as the barrier for an α-carbon model of the protein with Gō interactions (42).

The actual size of the barrier can be estimated from chevron plots of folding/unfolding rates versus denaturant (18). Assuming two-state (exponential) kinetics, the slope of the folding arm of the chevron predicts a change of $\sim 7kT$ between 0 M GdmCl and melting, which is not inconsistent with estimates from current kinetic models that include solvent effects explicitly (11, 43). In the kinetic model used by Fersht and coworkers, the slope of the chevron, $m_{\ddagger\text{-D}}$, is proportional to the change in solvent-accessible surface area (SASA) between denatured and transition states (18). The fractional change in SASA can be calculated as $m_{\ddagger\text{-D}}/m_\text{D-N} = \partial\Delta G_{\ddagger\text{-D}}/\partial\Delta G_\text{D-N}$, where $m_\text{D-N}$ is the change in SASA between denatured and native states and G is the Gibbs free energy (10). The result is about $m_{\ddagger\text{-D}}/m_\text{D-N} \sim 0.7$–$0.8$ for BdpA. Although our model does not include solvent, we can calculate $m_{\ddagger\text{-D}}/m_\text{D-N}$ from the ratio of slopes $\Delta\Delta F_{\ddagger\text{-D}}/\Delta\Delta F_\text{D-N}$. The result is $m_{\ddagger\text{-D}}/m_\text{D-N} \simeq 0.45$, which is about equal to the fraction of contacts formed at $q^\ddagger$. Although the model and measured transitions occur in roughly the same regions of the profile (range of native order), a smaller fraction of the surface area is lost in the transition. Because the model is mapped to $\Delta F_\text{D-N}$ this value indicates that the slope of $\Delta F_{\ddagger\text{-D}}$ is a bit smaller than in the chevron plots. The difference in scale of activation barriers is very similar to what results when small cooperative (three-particle) terms are added to the Gō model (44). These terms can be related to solvent participation in terms of the n-cluster model of phase separation (9, 45), which, along with the results above, suggests that the coupling between the solvent and the structure of the protein (32, 46, 47) is described inaccurately here. It should be possible to account for this coupling better by defining the strength of pair interactions to depend on the amount of local (in space) native order, and the inclusion of these terms already tends to improve protein ϕ values in explicit chain models (44).

In the Gō model, ϕ values are usually calculated as $\phi_i \simeq \langle Q_i\rangle^\ddagger/Q_i^N$, where $Q_i = \Sigma_j Q_{ij}(\Gamma_n)$ and the angle brackets denote the thermal average at $q^\ddagger$ (48). Fig. 2*f* plots $\phi_i(\xi)$ for comparison with the $\phi_i(T)$ measurements in ref. 17, and Fig. 2*e* plots the location of the transition, $q^\ddagger(\xi)$. Variations in $\phi_i(\xi)$ in Fig. 2*f* roughly agree with the those in $\phi_i(T)$ across the measured temperature range (17), and the uniform increase of curves in Fig. 2*f* with decreasing ξ reflects the slight shift in the transition state toward the native ensemble. Fig. 3 plots the local order, $\langle q_i\rangle$, and the contact order, $\langle Q_{ij}\rangle/Q_{ij}^N$, along the folding profile, and Fig. 4 describes the

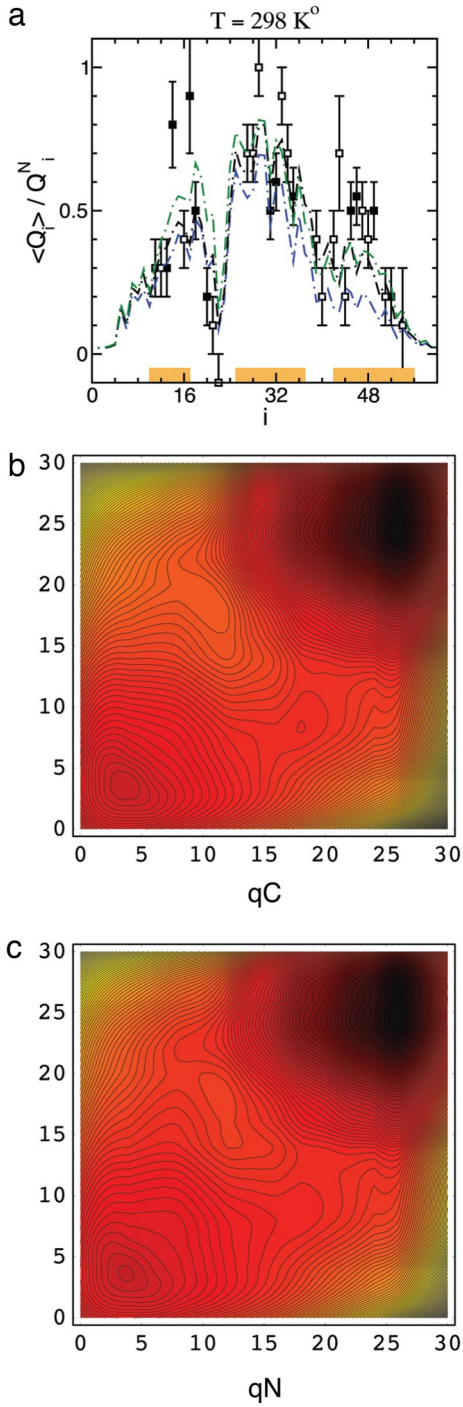

Fig. 4. Two-dimensional projections of the folding network. (*a*) Calculated ϕ values for the wild-type (blue line) and perturbed (black line) contact matrices are shown together with the wild-type melting values (green line). Measured ϕ values for surface (open symbols) and core (filled symbols) residues are from ref. 16. The level of agreement between the perturbed and measured ϕ values is about the same as in ref. 29. (*b* and *c*) Free energy landscapes for two contact models described in *Methods*. The landscapes are a projection of the folding network onto the coordinates $q_N = \Sigma_{i\le N/2} q_i$ and $q_C = \Sigma_{i>N/2} q_i$, where $n = 60$. (*b*) The wild-type NMR structure. (*c*) The same structure weighted by the function $(1 + 5/r_{ij})/2$ for ξ corresponding roughly to 2 M GdmCl at 25°C (levels on the landscapes are separated by $\sim 0.1kT$).

landscape and ϕ values at 2 M GdmCl for different realizations of Q_{ij}^N described in *Methods*. The folding pathway described by Fig. 3 is consistent with the replica exchange (all-atom) simulations of Garcia and Onuchic except for H3 helical content [apparently, this is related to the absence of denaturant (17) and high melting temperature (13) of the simulations]. The model pathway appears to describe early formation of the binding interface H1:H2 followed by reorganization across the transition state to consolidate long-range contacts with H3.

Discussion

Recently, Itoh and Sasai (29) provided a thorough description of BdpA folding by using an exact Muñoz–Eaton model. Here we have followed their approach of fitting $\Delta F_{\text{D-N}}$ to stability measurements to allow a clear comparison of results. Their model predicts that the symmetry of the transition state is broken as ξ is brought through the melting point. However, this process is accompanied by large changes in ϕ values that are not then observed in the $\phi_i(T)$ measurements of Sato and Fersht (17). Actually, all of the models above appear to be describing a similar effect, namely weak frustration between two forms of the protein (22, 34). What separates our model is basically that (*i*) the contact matrix is calculated to reflect the change in exposed surface area (32), and (*ii*) the largest contribution to ϕ_i consists of the diagrams $\Gamma_{n\neq 0}$ that allow unfolded loops.

The discrepancies in $\Delta F_{\ddagger\text{-D}}$ seen here and in the α-carbon model (42) seem to indicate missing solvent/cooperative effects; however, the all-atom model folds across a barrier of only $\sim$1–2kT at midpoint. This situation can be explained somewhat by the conditions of the simulation (13), but it still seems to allow for barrierless folding (28, 49) at high stability, and a calculation of the diffusion constant along the lines of ref. 43 would be helpful here. At the same time, it would be interesting to explore different extensions of the n-cluster model (couplings between Q_{ij} and space-local order) as a means to fit the system to chevron data, and it seems likely that this approach would improve the level of agreement with other measured properties of the free energy landscape.

Methods

To calibrate the model, we applied it to several small proteins where the landscape had been described by (*i*) all-atom simulations or (*ii*) otherwise accurate models where ϕ values were available (13, 14, 50–52). Figs. 2-4 were calculated by using $\delta = 0$, $\Delta s = 1k$ (29), and segments of two residues. The results were insensitive to δ and weakly dependent on the ln(l) part of the entropy cost [adjusting this term to account for excluded volume (36) affects mainly the size of the barrier], but neglecting the extension term r^2/la^2 (25) can lead to more qualitative changes in the landscape (see *SI Text* and SI Fig. 7).

The contact matrix is calculated to reflect the change in exposed surface area on folding (32). A contact is registered in Q_{ij}^N between a pair of atoms in residues i and j if the atoms are within a distance r_c in the NMR structure of the protein and $|i-j|>1$. The contact radii are selected to allow a water molecule between atoms (25): $r_c = 6$ Å for pairs of heavy atoms, 5 Å for heavy and hydrogen atoms, and 4 Å for hydrogens [including hydrogens (53) leads to more cooperative folding]. The unperturbed contact matrix is calculated from the wild-type NMR structure for comparison with ref. 29. In Fig. 4*c* we weight Q_{ij}^N with the function $(1 + 5/r_{ij})/2$ to improve the ϕ values at 2 M GdmCl. This model is similar in all respects the model in Figs. 2–4*b* except for the details of the transition. The better resolved NMR structure of the Y15W mutant (16) leads to more of an imbalance between H1 and H3 ϕ values, but the folded region of the landscape then contains two native-like ensembles similar to what is observed in the all-atom simulations (13). Weighting these contacts by inverse C^α distance did not influence the balance between H1 and H3.

ACKNOWLEDGMENTS. We thank Javier Sancho for helpful comments contributing to this work.

1. Oliveberg M, Wolynes PG (2005) *Q Rev Biophys* 38:245-288.
2. Cheung MS, Garcia AE, Onuchic J (2002) *Proc Natl Acad Sci USA* 99:685-690.
3. Das P, Wilson C, Fossati G, Wittung-Stafshede P, Matthews K, Clementi C (2006) *Proc Natl Acad Sci USA* 102:14569-14574.
4. Chavez L, Gosavi S, Jennings P, Onuchic J (2006) *Proc Natl Acad Sci USA* 103:10254-10258.
5. Gruebele M (2002) *Nat Struct Biol* 9:154-155.
6. Capaldi AP, Kleanthous C, Radford SE (2002) *Nat Struct Biol* 9:209-216.
7. Hummer G (2007) *Proc Natl Acad Sci USA* 104:14883-14884.
8. Miller TF, Vanden-Eijnden E, Chandler D (2007) *Proc Natl Acad Sci USA* 104:14559-14564.
9. ten Wolde PR, Chandler D (2002) *Proc Natl Acad Sci USA* 99:6539-6543.
10. Matouschek A, Fersht A (1993) *Proc Natl Acad Sci USA* 90:7814-7818.
11. Frauenfelder H, Fenimore PW, Chen G, McMahon BH (2006) *Proc Natl Acad Sci USA* 103:15469-15472.
12. Ebbinghaus S, Kim SJ, Heyden M, Yu X, Heugen U, Grubele M, Leitner DM, Havenith M (2007) *Proc Natl Acad Sci USA* 104:20749–20752.
13. Garcia AE, Onuchic JN (2003) *Proc Natl Acad Sci USA* 100:13898-13903.
14. Jayachandaran G, Vishal V, Garcia AE, Pande VS (2006) *J Struct Biol* 157:491-499.
15. Guo Z, Brooks C, Bozcko E (1997) *Proc Natl Acad Sci USA* 94:10161-10166.
16. Sato S, Religa TL, Daggett V, Fersht AR (2004) *Proc Natl Acad Sci USA* 101:6952-6956.
17. Sato S, Fersht AR (2007) *J Mol Biol* 372:254–267.
18. Huang F, Sato S, Sharpe TD, Ying L, Fersht AR (2004) *Proc Natl Acad Sci USA* 104:123-127.
19. Muñoz V, Thompson PA, Hofrichter J, Eaton WA (1997) *Nature* 390:196-199.
20. Muñoz V, Henry ER, Hofrichter J, Eaton WA (1998) *Proc Natl Acad Sci USA* 95:5872-5879.
21. Muñoz V (2001) *Curr Opin Struct Biol* 11:212-216.
22. Nelson ED, Grishin NV (2006) *J Mol Biol* 358:646-653.
23. Bueno M, Ayuso-Tejedor S, Sancho J (2006) *J Mol Biol* 358:813-824.
24. Henry ER, Eaton W (2004) *Chem Phys* 307:163-185.
25. Garbuzynskiy SO, Finkelstein AV, Galzitskaya OV (2004) *J Mol Biol* 336:509-525.
26. Bruscolini P, Pelizzola A (2002) *Phys Rev Lett* 88:258101.
27. Bruscolini P, Pelizzola A, Zamparo M (2007) *J Chem Phys* 126:215103.
28. Garcia-Mira M, Sadqi M, Fischer N, Sanchez-Ruiz JM, Muñoz V (2002) *Science* 298:2191-2195.
29. Itoh K, Sasai M (2006) *Proc Natl Acad Sci USA* 103:7298-7303.
30. Shoemaker BA, Wang J, Wolynes PG (1999) *J Mol Biol* 287:675-694.
31. Qi X, Portman JJ (2007) *Proc Natl Acad Sci USA* 104:10841-10846.
32. Imai T, Harano Y, Kinoshita M, Kovalenko A, Hirata F (2007) *J Chem Phys* 126:225102.
33. Kaya H, Chan HS (2003) *Proteins* 52:524–533.
34. Wolynes PG (2004) *Proc Natl Acad Sci USA* 101:6837-6838.
35. Deisenhofer J (1981) *Biochemistry* 20:2361-2369.
36. Edwards SF, Singh PS (1979) *J Chem Soc Faraday Trans II* 75:1001-1019.
37. d'Aquino JA, Gomez J, Hilser VJ, Lee KH, Amzel LM, Freire E (1996) *Proteins* 25 143-156.
38. Harpaz Y, Gerstein M, Chothia C (1994) *Structure (London)* 2:641-649.
39. Wang L, Rivera EV, Benavides-Garcia MG, Nall BT (2005) *J Mol Biol* 353 719-729.
40. Carrion-Vasquez M, Oberhauser AF, Fowler SB, Marszalek PE, Broedel SE, Clarke J, Fernandez JM (1999) *Proc Natl Acad Sci USA* 96:3694-3699.
41. Zagrovic B, Pande VS (2003) *Nat Struct Biol* 10:955-966.
42. Shea JE, Onuchic JN, Brooks CL (1999) *Proc Natl Acad Sci USA* 96:12512-12516.
43. Yang S, Onuchic JN, Garcia AE, Levine H (2007) *J Mol Biol* 372:756-763.
44. Ejtehadi MR, Avali SP, Plotkin SS (2004) *Proc Natl Acad Sci USA* 101:15088-15093.
45. de Gennes PG (1991) *C R Acad Sci Paris II* 313:1117-1122.
46. Rudriguez-Larrea D, Ibarra-Molero B, Sanchez-Ruiz JM (2006) *Biophys J* 91:L48–L50.
47. Nelson ED, Grishin NV (2004) *Phys Rev E* 70:051096.
48. Lätzer J, Eastwood MP, Wolynes PG (2006) *J Chem Phys* 125:214905.
49. Liu F, Gruebele M (2007) *J Mol Biol* 370:574-584.
50. Lei H, Wu C, Liu H, Duan Y (2007) *Proc Natl Acad Sci USA* 104:4925-4930.
51. Sorenson JM, Head-Gordon T (2002) *J Comp Biol* 9:35-53.
52. Shimada J, Shakhnovich E (2002) *Proc Natl Acad Sci USA* 99:11175-11180.
53. Word JM, Lovell SC, Richardson JS, Richardson DC (1999) *J Mol Biol* 285:1735-1747.