Capillarity-like growth of protein folding nuclei

Xianghong Qi and John J. Portman*

Department of Physics, Kent State University, Kent, OH 44240

Edited by Peter G. Wolynes, University of California at San Diego, La Jolla, CA, and approved May 21, 2008 (received for review December 6, 2007)

A full structural description of transition state ensembles in protein folding includes the specificity of the ordered residues composing the folding nucleus as well as spatial density. To our knowledge, the spatial properties of the folding nucleus and interface of specific nuclei have yet to receive significant attention. We analyze folding routes predicted by a variational model in terms of a generalized formalism of the capillarity scaling theory that assumes the volume of the folded core of the nucleus grows with chain length as $V_f \approx N^{3\nu}$. For 27 two-state proteins studied, the scaling exponent ν ranges from 0.2 to 0.45 with an average of 0.33. This average value corresponds to packing of rigid objects, although generally the effective monomer size in the folded core is larger than the corresponding volume per particle in the native-state ensemble. That is, on average, the folded core of the nucleus is found to be relatively diffuse. We also study the growth of the folding nucleus and interface along the folding route in terms of the density or packing fraction. The evolution of the folded core and interface regions can be classified into three patterns of growth depending on how the growth of the folded core is balanced by changes in density of the interface. Finally, we quantify the diffuse versus polarized structure of the critical nucleus through direct calculation of the packing fraction of the folded core and interface regions. Our results support the general picture of describing protein folding as the capillarity-like growth of folding nuclei.

diffuse | folding mechanism | interface | nucleation | polarized

The modern theory of protein folding describes the mechanism for folding as an entropic bottleneck arising from the decreasing number of accessible pathways available to a protein as it becomes ordered (1, 2). The collection of partially ordered conformations describing this bottleneck region is known as the transition state ensemble or critical folding nucleus (3). The structure of the transition state ensemble is commonly described through the degree of native-like order of specific residues because this is the kind of structural information that can be directly inferred from site-directed mutagenesis folding experiments. Nevertheless, a complete description of the protein folding mechanism also includes the spatial properties such as size or density of the transition state ensemble. Indeed, shortly after characterizing the transition state ensemble of CI2, Fersht (4) proposed a spatial description of the critical nucleus that supported the kinetic data. The critical nucleus envisioned in this nucleation-condensation mechanism can be thought of as an expanded, partially ordered version of the native state ensemble with concomitant long-range tertiary and local secondary structure. Although diffuse nuclei appear to be the general rule, some nuclei are less diffuse than others (5). Polarized nuclei have highly structured residues that are spatially clustered in the native structure, with the rest of the residues having little definite order (6–9). Such nuclei are similar to the capillarity approximation in homogeneous nucleation, where a stable phase droplet is separated from the metastable phase by a sharp interface. Several theoretical studies have appealed to this analogy to describe the mechanism for folding (10–14). Wolynes (14) describes a nucleus with capillarity-like order in which the interface surrounding a relatively folded core is broadened by wetting of partially ordered residues. In this picture, folding can be described as a wave of order moving across the protein as the edge of the nucleus expands to ultimately consume the entire molecule (14, 15)

The extended partially ordered interface of a capillarity-like ordered nucleus separates space into three regions: a folded core, a partially ordered interface region, and unfolded halo (see Fig. 1). Growth of the nucleus is controlled by fluxes of residues passing through two moving surfaces: One surface separates the folded core and interface, and the other surface separates the interface region and the unfolded halo. As the protein folds, the folded core grows by incorporating residues from the interface region, whereas the evolution of the interfacial region is determined by the net flux of residues entering the interface.

Our analysis is based on folding routes calculated for 28 two-state proteins from a cooperative variational model described in ref. 16. We note this model includes neutral cooperativity due to repulsive excluded volume interactions. This form of cooperativity has been shown to broaden the range of barrier heights, allowing direct comparison between calculated and measured folding rates (16). Although this cooperativity tends to sharpen the interface between folded and unfolded regions, the interface from this model is not so sharp that each residue is either completely folded or completely unfolded, as assumed in some other analytic models (17–19). In fact, an unbiased analysis of the spatial properties of the folding nucleus fundamentally depends the model's ability to describe partial order.

The capillarity approximation of folding nuclei is based on classical nucleation theory of first order phase transition kinetics (17, 11). Within the capillarity approximation, the free energy of a nucleus with volume V_f and surface area A_f can be written as a sum of two terms

$$F = -\Delta f V_f + \gamma A_f, \qquad [1]$$

where Δf denotes the bulk free energy difference per unit volume between the unfolded and folded ensembles, and γ is the surface tension between the folded and unfolded regions.

A folded core with native-like density has a volume per monomer independent of its size. Relaxing this assumption, we take the number of residues in the folded core, N_f, to scale with its volume, V_f, according to

$$V_f = b^3 N_f^{3\nu}. \qquad [2]$$

Here, ν is the scaling exponent associated with the length scale of the folded core $R \approx b_0 N_f^{\nu}$, and b^3 is a geometry-dependent elementary volume proportional to the monomer volume, b_0^3. The free energy of a folded nucleus with N_f residues then has the form: (14)

Author contributions: X.Q. and J.J.P. designed research, performed research, analyzed data, and 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: jportman@kent.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0711527105/DCSupplemental.

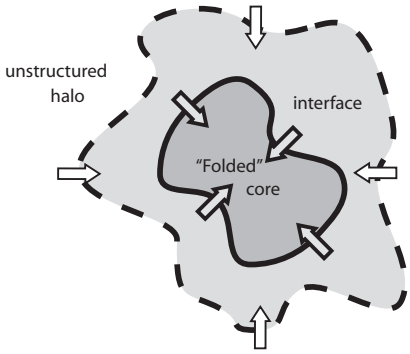

Fig. 1. Illustration of folding nucleus: folded core, interfacial region, and unfolded halo. Growth of the nucleus can be characterized by fluxes entering the folded core and interfacial regions.

$$F(N_f) = -\Delta f b^3 N_f^{3\nu} + \gamma b^2 N_f^{2\nu}. \quad [3]$$

At the folding transition temperature, T_f, finite size depression of the surface energy suggests that $\gamma \approx \Delta f b N^{\nu}$, where N is the number of monomers in the protein. The maximum of the free energy occurs at $N_f^{\dagger} = (2/3)^{1/\nu}N$, with an associated free-energy barrier that scales as $\Delta F^{\dagger} \approx N^{2\nu}$. If we assume that the folded core has native-like packing, $\nu = 1/3$, and b^3 is the native-like volume per monomer, so that $N_f^{\dagger} = (2/3)^3 N$, and $\Delta F^{\dagger} \approx N^{2/3}$ (13, 14).

Simulations and alternative theoretical considerations also suggest that barrier height (logarithm of the folding time) scales sublinearly on chain length, $\Delta F^{\dagger} \approx N^{p}$, with $0 \le p \le 1$ (13, 14, 20, 21). Direct analysis of folding rate data to determine the scaling exponent p encounters the difficulty that the range of N is too small to distinguish between different values of p (22–26). So, although it may be reasonable to expect that the scaling of the barrier height with chain length is universal for sufficiently large proteins, the size of typical two-state proteins (≈100 aa) may well be too small to be governed by this generic behavior. In this case, both specificity and size of these smaller proteins should generally determine the properties of the critical nuclei. In this article, we assume that Eq. **2** is valid to describe the growth of the nucleus in all of the two-state proteins, but the exponent ν and volume b^3 are allowed to be protein-specific.

Results

Characterizing the Folded Core and the Interface. In the variational model considered in this article, partially ordered configurations are described by a variational Hamiltonian, $\mathcal{H}_0$, corresponding to a stiff polymer chain inhomogeneously constrained to the native structure. Because this model is described in detail in ref. 13, we focus here on how to define folded core, interface, and unfolded regions along the calculated folding route. This is not as straightforward as one might expect because the concept directly couples specificity of the nucleus with the spatial density.

We characterize the degree of structure of each residue by the extent of localization about the native structure $\{\boldsymbol{r}^N\}$, $\rho_i = \langle \exp(-\alpha^N(\boldsymbol{r}_i - \boldsymbol{r}_i^N)^2)\rangle_0$, with $\alpha^N = 0.1$. Here, the subscript denotes the average with respect to the Boltzmann weight with $\mathcal{H}_0$. Denoting the native density at the globule and native state by $\rho_i(G)$ and $\rho_i(N)$, respectively, we consider the normalized density

$$\tilde{\rho}_i = \frac{\rho_i - \rho_i(G)}{\rho_i(N) - \rho_i(G)} \quad [4]$$

as a set of order parameters characterizing the folding of each residue. Progress along the folding route can be monitored by the global structural parameter $Q = 1/N\Sigma\,\tilde{\rho}_i$. We use the normalized native densities to define a fiducial set of folded residues, $\{F\}$, with $\tilde{\rho}_i > 0.6$, as shown in Fig. 2*A*. Next, we define the spatial region of the folded core through the relative contribution of the density of the folded residues in $\{F\}$, $n_f(\boldsymbol{r}) = \Sigma_{\{F\}}\langle\delta(\boldsymbol{r} - \boldsymbol{r}_i)\rangle_0$, to the total density, $n(\boldsymbol{r}) = \Sigma_{i=1}^{N}\langle\delta(\boldsymbol{r} - \boldsymbol{r}_i)\rangle_0$. The spatial extent of the folded core and interfacial regions in this analysis is determined by an indicator function

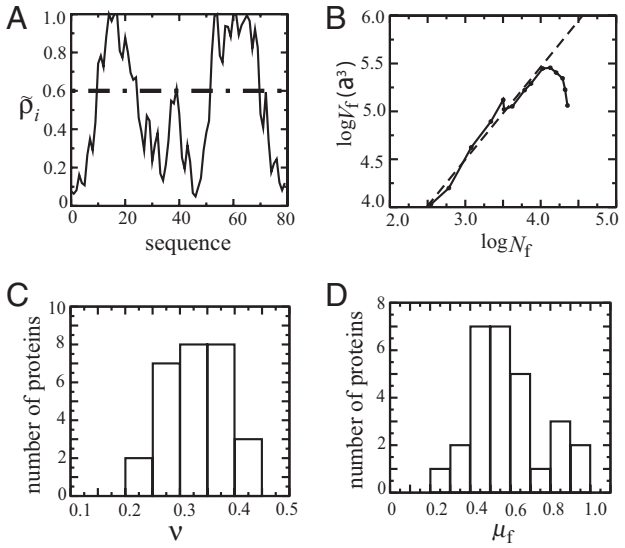

Fig. 2. Scaling of the folded core with number of monomers. *A* and *B* correspond to λ⁻repressor (1lmb). (*A*) Residues with native density $\tilde{\rho}_i > 0.6$ (indicated by the dashed line) define a fiducial set of folded residues. (*B*) Linear fit of log V_f vs. log N_f (dashed line) gives the exponent of $V_f \approx N_f^{3\nu}$. In this example, the fitting equation is $y = 5.6 + 0.97x$, so that $\nu = 0.32$, and $b^3 = 5.6a^3$; a = 3.8 Å is the average distance between the α carbons. (*C*) Histogram of scaling exponent ν for 28 proteins. (*D*) Histogram of the packing fraction of the folded core of the critical nucleus at T_f.

$$\tilde{n}(\boldsymbol{r}) = \frac{n_f(\boldsymbol{r})}{n(\boldsymbol{r})}, \quad [5]$$

where $0 \le \tilde{n}(\boldsymbol{r}) \le 1$. We define the folded core region, $\mathcal{V}_f$, as the points $\{\boldsymbol{r}_f\}$ for which the density of the fiducial folded residues contributes at least 50% to the total density ($\tilde{n}(\boldsymbol{r}) \ge 0.5$). The number of residues in this folded core region can be found by numerically integrating the density over the core region, $N_f = \int_{\mathcal{V}_f} n(\boldsymbol{r})d\boldsymbol{r}$. The volume of the core region is given by $V_f = \int_{\mathcal{V}_f} d\boldsymbol{r}$.

Similarly, the interfacial region, $\mathcal{V}_{\rm int}$, is defined as the points $\{\boldsymbol{r}_{\rm int}\}$ for which $0.1 \le \tilde{n}(\boldsymbol{r}_{\rm int}) < 0.5$. The number of interfacial residues and volume of the interface is given by $N_{\rm int} = \int_{\mathcal{V}_{\rm int}} n(\boldsymbol{r})d\boldsymbol{r}$, and $V_{\rm int} = \int_{\mathcal{V}_{\rm int}} d\mathrm{r}$, respectively.

The number of residues and the volume can be used to define a mean packing fraction of the folded core and partially ordered interface by

$$\mu_f = \frac{N_f}{V_f} v_0 \quad \text{and} \quad \mu_{\rm int} = \frac{N_{\rm int}}{V_{\rm int}} v_0, \quad [6]$$

respectively. Here, v_0 is the calculated volume per particle of the native structure at the folding transition temperature, T_f. The growth of the nucleus can be characterized by the way the packing fractions μ_f and $\mu_{\rm int}$ change along the folding route.

Growth of Folding Nucleus Along the Folding Route. As illustrated in Fig. 2 *A* and *B*), the changes in N_f and V_f along a folding route can be fit to Eq. **2** to give an estimate of the scaling exponent ν for each protein. Fig. 2*C* shows the distribution of predicted ν from the folding routes obtained from the variational model for

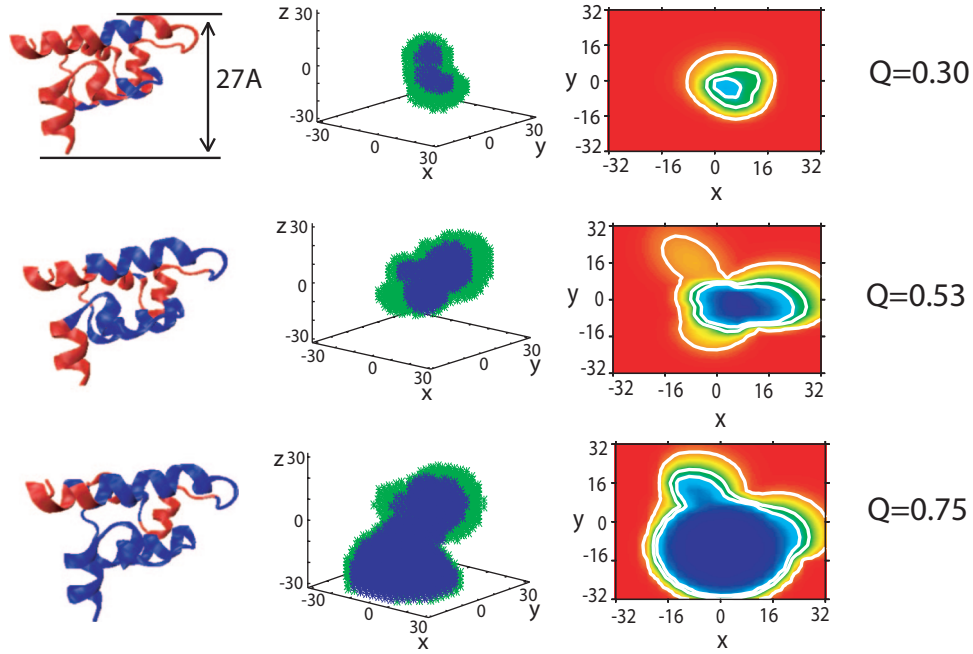

Fig. 3. Illustration of growth of folding nucleus and interface along the folding route (increasing Q for the λ-repressor protein (1lmb). Column 1 shows the three-dimensional folded structure with the fiducial set of folded residues colored blue and the unfolded residues colored red. In column 2, the folded core (colored blue) is surrounded by the interfacial region (colored green). Column 3 is a projection of the indicator function $\tilde{n}(\boldsymbol{r})$ that defines the folded and interfacial regions in space. The values correspond to $\max_z \tilde{n}(x,y,z)$, ranging from 1 (blue) to 0 (red) in steps of 0.01. Contour lines correspond to 0.1, 0.5, 0.7. Column 4 gives the corresponding Q value for each row. The critical nucleus corresponds to $Q = 0.53$. The units for three plots are in angstroms. This protein belongs to pattern C (balanced growth). The three-dimensional structure was produced by VMD (29).

28 two-state proteins discussed in ref. 13. The predicted scaling exponent ν ranges between 0.2 and 0.4, with an average of $\nu = 0.33$. The mean exponent is very close to the scaling associated with close-packed rigid objects, $\nu = 1/3$. For comparison, recent detailed statistical models indicate that the scaling exponent for the unfolded state of a protein is about $\nu = 0.59$ (27), whereas a wide variety of protein folded structures suggest that proteins with <300 aa have compact folded structures ($\nu = 0.3$), and larger proteins are less dense ($\nu = 0.4$) (28).

The mean packing fraction of the core scales with the number of monomers as:

$$\mu_f = \frac{v_0}{b^3} N_f^{1-3\nu}. \qquad [7]$$

For the close packing value $\nu = 1/3$, μ_f is independent of the number of monomers. When $\nu > 1/3$, the core becomes less compact as monomers are added to the core. This is the familiar scaling from loosely packed or fractal objects. When $\nu < 1/3$, the core density increases as more monomers are incorporated into the core. This can be understood as the consolidation of structure in the folded core as folding progresses.

Although the spatial structure of the critical folding nucleus (transition-state ensemble), is discussed in more detail below, it is instructive to consider the value of the mean packing fraction of the core here. Fig. 2*D* shows the distribution of packing fractions of the folded core evaluated at the maximum free-energy barrier between folded and unfolded states at T_f. The packing fraction has a wide range, from 0.2 to 1.0. Although some of the transition-state nuclei have compact cores, the average packing fraction is only 0.59. This means that although the growth of a typical folded core corresponds to rigidly packed objects, a typical transition state ensemble has a folded core with twice the volume as the volume of same number of monomers in the native-state conformation ($b^3 \approx 2v_0$). That is, the monomers composing the nucleus are typically much less localized than in the native state.

Fig. 3 shows a representative example of the growth of the folded core and the interface region. Early in the folding, the folded nucleus is small and compact, surrounded by a partially folded interface. This small nucleus is partially ordered, occupying about twice the volume of the corresponding residues in the native state. Structural fluctuations giving nuclei corresponding to $Q < Q^\dagger$ are unstable with respect to the unfolded state due to relatively large surface free-energy cost associated with small nuclei, whereas structural fluctuations with $Q > Q^\dagger$ tends to evolve to the folded state. As the nucleus grows, the volume of the nucleus evolves as interfacial regions are incorporated into the core, whereas unfolded residues become part of the partially ordered interface.

Growth Patterns of the Nucleus. The structural growth of the folding nucleus can be understood as the competition between growth of the folded core and the evolution of the interface. The flux of residues entering core through the interface region controls the growth of the core, and the net flux of residues entering interface region from the unfolded halo controls the growth of the interface (see Fig. 1). The evolution of the nucleus along the folding route can be monitored through changes in the number of residues, volume, and packing fraction as a function the global order parameter Q. The scaling relation given in Eq. **2**, for example, is a parametric equation of $V_f(Q)$ and $N_f(Q)$. As noted below in Eq. **8**, the density of the folded core can increase, decrease, or stay the same as residues are incorporated into the core depending on the value of the scaling exponent ν. Here, we consider evolution of both the folded core and the interface between folded and unfolded regions along the folding route to describe the growth of the nucleus. In particular, we focus on the changes in the packing fraction $\mu_f(Q)$ and $\mu_{\text{int}}(Q)$ along the folding route. That is, we consider the signs of

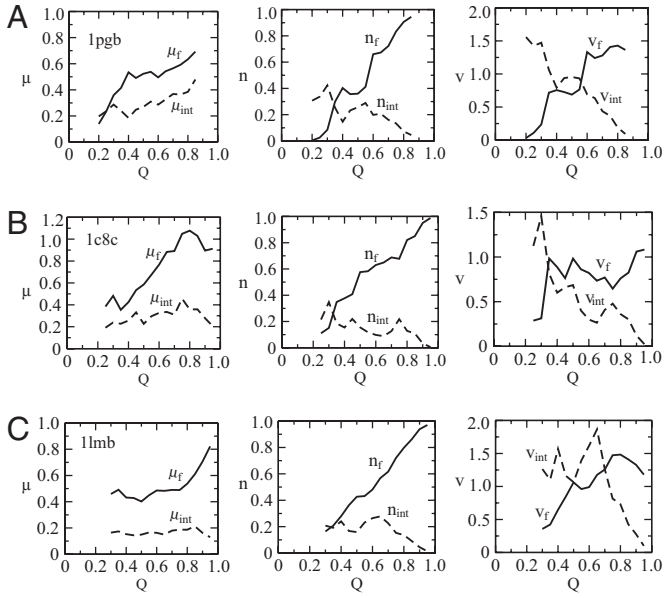

Fig. 4. Examples of three folding patterns. Patterns A–C correspond to (*A–C*), respectively. Column 1 shows the mean packing fraction in folded core (μ_f, solid line) and interface region (μ_{int}, dashed line) as function of *Q*; column 2 shows the number of particles in the folded core (normalized by the chain length, n_f, solid line), and interface region (n_{int}, dashed line) as a function of *Q*; column 3 shows the volume of folded core (normalized by the volume at native state, v_f, solid line) and interface region (v_{int}, dashed line) as a function of *Q*.

$$\dot{\mu}_f(Q) = \frac{d\mu_f}{dQ} \quad \text{and} \quad \dot{\mu}_{\text{int}}(Q) = \frac{d\mu_{\text{int}}}{dQ} \qquad [8]$$

to identify different modes of growth. From the two-state proteins used in this study, we can identify three distinct scenarios as illustrated in Fig. 4. [Similar plots for all of the protein studies are given in supporting information (SI) Figs. S1–S5 and *SI Text*.]

Pattern A (consolidation of core and interface). As shown in Fig. 4*A*, proteins described by this growth pattern have $\dot{\mu}_f(Q) > 0$ and $\dot{\mu}_{\text{int}}(Q) > 0$. For these proteins, the number of folded residues increases faster than volume increases so that the folded core region becomes increasingly more tightly packed along folding route. The consolidation of the core is closely related to the evolution of the interface region, shown as dashed lines in Fig. 4. The growth of the interface region determined by the density is similar, although this time, both the number of residues and their occupying volume are decreasing along the folding route. Both the core and interface become more compact until the interface is almost completely depleted near the folded configuration.

Pattern B (consolidation of the core only). As shown in Fig. 4*B*, proteins described by this growth pattern have $\dot{\mu}_f(Q) > 0$ and $\dot{\mu}_{\text{int}}(Q) \approx 0$. This means that the folded core consolidates along the folding route as in pattern A, but the interface region evolves with relatively constant density. In the illustration shown in Fig. 4*B*, the N_{int} and V_{int} remain relatively constant throughout much of the growth. Then the interface can be seen as a channel transferring residues from the unfolded halo to the folded core with no net flux of residues through the interfacial surface. For some other proteins in this class, a decrease in the number of interfacial residues is balanced by a proportional decrease in the interfacial volume (see Figs. S1–S5 and *SI Text*).

Pattern C (balanced growth). As shown in Fig. 4*C*, proteins described by this growth pattern have $\dot{\mu}_f(Q) \approx 0$ and $\dot{\mu}_{\text{int}}(Q) \approx 0$. For this set of proteins, the cores do not consolidate until the end of folding. Rather, the evolution of the number of residues and the volume keeps the density of both regions relatively constant. (The same comments about balanced evolution of the interface relevant to pattern B apply here as well.)

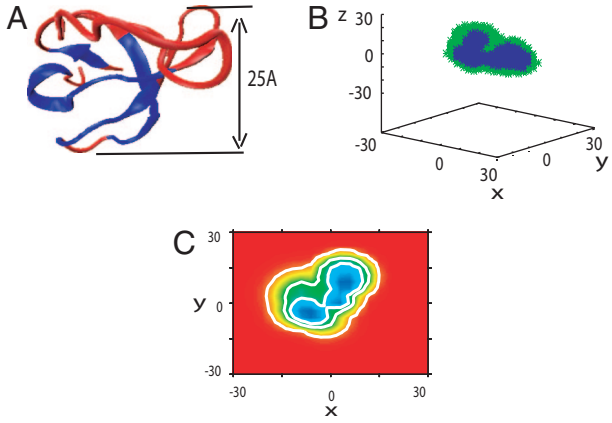

Fig. 5. An example of a polarized critical nucleus ($Q = 0.45$) for Src-SH3 (1srl). Plots *A–C* correspond to the middle row of the diffuse nucleus shown in Fig. 3.

Of the 27 proteins studied, three proteins are difficult to classify by this scheme (see Figs. S1–S5 and *SI Text*). The packing fractions of folded cores for these proteins show clearly sharp variations between low and high values. Some of the ruggedness of these curves is due to the rigid cutoff values defining the fiducial set of folded and interface residues as well as the three spatial regions of the nucleus. Whether this accounts for the anomalous calculated behavior for these three anomalous proteins is not yet clear. The growth mode of the nucleus for the 27 proteins considered in this article (1pgb16 is too small to have a compact folded core) can be roughly classified as follows: pattern A: 1pgb, 1a0n, 1pks, 1pin, 1psf, 1shg, 2ptl; pattern B: 1c8c, 1coa, 1enh, 1fkb, 1hdn, 1vii, 1wit, 2pdd; pattern C: 1aps, 1csp, 1fnf9, 1imq, 1mef, 1o6x, 1srl, 1ten, 1lmb; Exceptions: 1div, 1urn, 2abd.

Polarized Versus Diffuse Critical Nucleus. A folding mechanism is typically characterized by the structure of the critical nucleus. The spatial structure of the transition state ensemble, inferred from ϕ-value analysis, has often been qualitatively summarized as either diffuse or polarized (30). Intermediate ϕ-values spread across a large portion of the protein sequence indicate a diffuse nucleus. In contrast, polarized transition states are inferred when only one part of the structure has relatively high ϕ-values and the rest of the residues have low ϕ-values. In addition to a bimodal distribution of ϕ-values, the ordered residues in a polarized transition-state ensemble are located in one region in the native configuration. Polarized and diffuse critical nuclei are sometimes called localized and delocalized transition-state ensembles, respectively (31). Of course, the critical nucleus of a given protein is expected to have structural properties somewhere between the two ideal limits. The second row of Fig. 3 gives an example a diffuse critical nucleus (1lmb). For comparison, Fig. 5 shows the corresponding plots for a protein with a polarized critical nucleus (1srl). Comparing Figs. 3 and 5, it is clear that the interface of 1lmb is much broader than the interface region of 1srl. Furthermore, the folded core of 1lmb is much more diffuse than the folded core of 1srl.

Characterizing a capillarity-like ordered nucleus as either diffuse or polarized is a statement of the sharpness of the interface as well as the compactness of the core. For convenience, we monitor both regions by the normalized volume per monomer (inverse packing fraction): $1/\mu_f$ and $1/\mu_{\text{int}}$. The results for the two-state proteins considered in this work are shown in

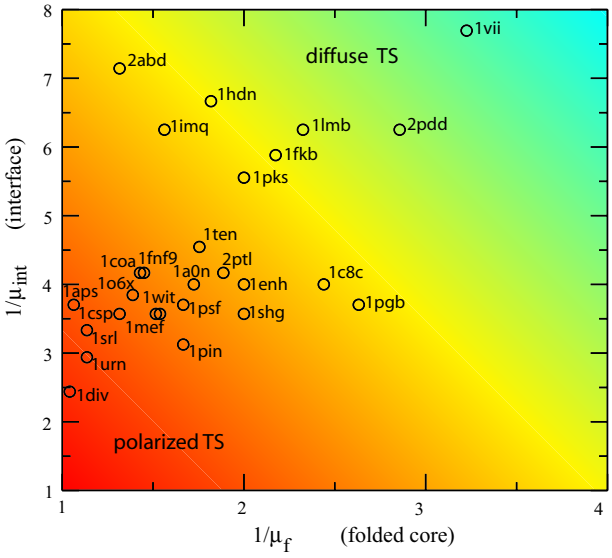

Fig. 6. Inverse packing fraction of the interface and folded core. for 27 two-state proteins. The gradual change in color shows the continuous change from polarized nuclei (red) to diffuse nuclei (cyan).

Fig. 6. Nuclei with small values of $1/\mu_f$ and $1/\mu_{\rm int}$ are more polarized, possessing relatively compact cores and sharp interfaces (similar to those envisioned in the strict capillarity approximation). Diffuse nuclei, on the other hand, have extended regions of partial order that correspond to larger values of $1/\mu_f$ and/or $1/\mu_{\rm int}$. We note that relatively polarized nuclei can have cores that are still loosely packed compared with the native-state density (e.g., 1pgb). Furthermore, relatively diffuse nuclei can have tightly packed cores but extended interfaces (e.g., 2abd, 1imq, 1fkb). Our analysis suggests polarized critical nuclei consistent with classification inferred by experimental ϕ-value analysis [such as 1csp (9), 1srl (7), 1shg (6), 1pin (8), 2ptl (32), and 1pgb (33)]. Our model also is consistent with several proteins classified as having diffuse critical nuclei [such as 1lmb (34), 2abd (35), 1imq (36), and 1fkb (37)]. This favorable comparison for a wide variety of proteins is a reassuring assessment of the model. At the same time, we realize that this comparison is necessarily qualitative and subjective.

With this caveat in mind, there are several proteins for which the model appears at odds with the characterization of the critical nucleus inferred from ϕ-values [such as CI2 (4), 1aps (38), U1A (39), and 1pgb (33)]. In these exceptional cases, ϕ-value distributions indicate that the critical nucleus is rather diffuse, but our model predicts more polarized nuclei. This tendency may indicate that the model is too cooperative for these proteins, because high cooperativity is expected to have sharp interfaces and more polarized transition-state ensembles.

Discussion

In this article, we directly characterize folding in terms of the capillarity-like growth of the folding nucleus. The nature of the partially folded interfacial region between the folded core and unfolded halo is the central focus of characterizing the growth modes of the nucleus. We find that the growth of the nucleus can be classified into three different patterns: the core and interface both condense along the folding route (pattern A); the core condenses at the expense of the interfacial region (pattern B); and the growth of the core is balanced by the monomers entering the interfacial region from the unfolded halo (pattern C). The picture of the core as close packing of rigid monomers appears to be valid on average, although the size of the effective monomers is larger than one would expect for a native-like, compact core. This analysis clarifies that diffuse nuclei inferred by the distribution of intermediate ϕ-values, for example, can arise from either a diffuse folded core, a broad interfacial region, or both. The predictions from our calculations can be tested from the analysis of the evolution of ϕ-values as a function of the movement of the transition state ensemble ($\beta^\dagger$) pioneered by Oliveberg and coworkers (39, 40). The variational model considered here includes a uniform "neutral," excluded volume type cooperativity developed to account of general trends in the absolute folding rates of two-state proteins (13). The exceptional qualitative discrepancies of the polarized versus diffuse characterization of the critical nucleus (such as CI2, 1aps, 1pgb, and U1A) permit an opportunity to assess the form and strength of the cooperativity of this model. The spatial density of the critical nucleus can be used as an independent criterion to check the value of the cooperativity obtained by simultaneously fitting ϕ-values and barrier height by the parameterization of the cooperativity for each protein. There are some indications that one should consider variations in the strength of the cooperativity for different proteins (although, admittedly, this is very closely tied to the specific form of the cooperativity in the model). For example, Ejtehadi and Plotkin (41) recently found that the strength of cooperativity from three-body interactions can be tuned for each protein to bring simulations of ϕ-values into better agreement with experimental measurement. The generally good qualitative agreement between our calculations and experimental inferences about the spatial extent of folding nuclei suggest that tuning the excluded volume strength for each protein would not greatly improve the results presented here for the majority of the proteins studied.

ACKNOWLEDGMENTS. This work was supported in part by a grant from the Ohio Board of Regents Research Challenge program.

1. Leopold PE, Montal M, Onuchic JN (1992) Protein folding funnels: A kinetic approach to the sequence–structure relationship. *Proc Natl Acad Sci USA* 89:8721–8725.
2. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways and the energy landscape of protein folding: A synthesis. *Proteins Struct Funct Genet* 21:167–195.
3. Onuchic JN, Wolynes PG, Luthey-Schulten Z, Socci ND (1995) Toward an outline of the topography of a realistic protein folding funnel. *Proc Natl Acad Sci USA* 92:3626–3630.
4. Itzhaki LS, Otzen DE, Fersht AR (1995) The structure of the transition state for folding of chymotrypsin inhibitor 2 analyzed by protein engineering methods: Evidence for a nucleation–condensation mechanism for protein folding. *J Mol Biol* 254:260–288.
5. Alm E, Morozov AV, Kortemme T, Baker D (2002) Simple physical models connect theory and experiment in protein folding kinetics. *J Mol Biol* 322:463–476.
6. Martinez JC, Serrano L (1999) The folding transition state between sh3 domain is conformationally restricted and evolutionarily conserved. *Nat Struct Biol* 6:1010–1016.
7. Riddle DS, *et al.* (1999) Experiment and theory highlight role of native state topology in sh3 folding. *Nat Struct Biol* 6:1016–1024.
8. Jager M, Nguyen H, Crane JC, Kelly JW, Gruebele M (2001) The folding mechanism of a beta-sheet: The ww domain. *J Mol Biol* 311:373–393.
9. Garcia-Mira MM, Bochringer D, Schmid FX (2004) The folding transition state of the cold shock proteins is strongly polarized. *J Mol Biol* 339:555–569.
10. Bryngelson JD, Wolynes PG (1990) A simple statistical field theory of heterpolymer collapse with applications to protein folding. *Biopolymers* 30:177–188.
11. Guo Z, Thirumalai D (1995) Kinetics of protein folding: Nucleation mechanism, time scales, and pathways. *Biopolmers* 36:83–102.
12. Abkevich VI, Gutin AM, Shakhnovich EI (1994) Specific nucleus as the transition state for protein folding: Evidence from the lattice model. 33:10026–10036.
13. Finkelstein AV, Badretdinov AY (1997) Rate of protein folding near the point of thermodynamic equilibrium between the coil and the most stable chain fold. *Fold Des* 2:115–121.
14. Wolynes PG (1997) Folding funnels and energy landscapes of larger proteins within the capillarity approximation. *Proc Natl Acad Sci USA* 94:6170–6175.
15. Oliveberg M, Wolynes PG (2005) The experimental survey of protein-folding energy landscapes. *Q Rev Biophys* 38:245–288.
16. Qi X, Portman JJ (2007) Excluded volume, local structural cooperativity, and the polymer physics of protein folding rates. *Proc Natl Acad Sci USA* 104:10841–10846.
17. Galzitskaya OV, Finkelstein AV (1999) A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. *Proc Natl Acad Sci USA* 96:112999–113004.
18. Alm E, Baker D (1999) Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc Natl Acad Sci USA* 96:11305–11310.

19. Muñoz V, Eaton WA (1999) A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc Natl Acad Sci USA* 96:11311–11316.
20. Thirumalai D (1995) From minimal models to real proteins: Time scales for protein folding kinetics. *J Phys I* 5:1457–1467.
21. Koga N, Takada S (2001) Role of native topology and chain length in protein folding. *J Mol Biol* 313:171–180.
22. Ivankov D, *et al.* (2003) Contact order revised: Influence of protein size on the folding rate. *Protein Sci* 12:2057–2062.
23. Galzitskaya OV, Garbuzynskiy SO, Ivankov DN, Finkelstein AV (2003) Chain length is the main determinant of the folding rate for proteins with three-state folding kinetics. *Proteins Struct Funct Genet* 51:162–166.
24. Ivankov DN, Finkelstein AV (2004) Prediction of protein folding rates from the amino acid sequence-predicted secondary structure. *Proc Natl Acad Sci USA* 101:8942–8944.
25. Li MS, Klimov DK, Thirumalai D (2004) Thermal denaturation and folding rates of single domain proteins: Size matters. *Polymer* 45:573–579.
26. Naganathan AN, Muñoz V (2005) Scaling of folding times with protein size. *J Am Chem Soc* 127:480–481.
27. Jha AK, Colubri A, Freed KF, Sosnick TR (2005) Statistical coil model of the unfolded state: Resolving the reconciliation problem. *Proc Natl Acad Sci USA* 102:13099–13104.
28. Xu X, Leitner D (2003) Anomalous diffusion of vibrational energy in proteins. *J Chem Phys* 119:12673–12679.
29. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graphics* 14:33–38.
30. Grantcharova V, Alm E, Baker D, Horwich AL (2001) Mechanism of protein folding. *Curr Opin Struct Biol* 11:70–82.
31. Geierhaas CD, Best RB, Paci E, Vendruscolo M, Clarke J (2006) Structural comparison of the two alternative transition states for folding of ti i27. *Biophys J* 91:263–275.
32. Kim DE, Fisher C, Baker D (2000) A breakdown of symmetry in the folding transition state of protein l. *J Mol Biol* 298:971–984.
33. McCallister EL, Alm E, Baker D (2000) Critical role of beta-hairpin formation in protein g folding. *Nat Struct Biol* 7:669–673.
34. Burton RE, Huang GS, Daugherty MA, Calderone TL, Oas TG (1997) The energy landscape of a fast-folding protein mapped by Ala–Gly substitutions. *Nat Struct Biol* 4:305–310.
35. Kragelund BB, *et al.* (1999) The formation of a native-like structure containing eight conserved hydrophobic residues is rate limiting in two-state protein folding of acbp. *Nat Struct Biol* 6:594–601.
36. Friel CT, Capaldi AP, Radford SE (2003) Structural analysis of the rate-limiting transition states in the folding of im7 and im9: Similarities and differences in the folding of homologous proteins. *J Mol Biol* 326:293–305.
37. Fulton KF, Main ER, Daggett V, Jackson SE (1999) Mapping the interaction present in the transition state for unfolding/folding of fkbp12. *J Mol Biol* 291:445–461.
38. Chiti F, *et al.* (1999) Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nat Struct Biol* 6:1005–1009.
39. Ternstrom T, Mayor U, Akke M, Oliveberg M (1999) From snapshot to movie: ϕ Analysis of protein folding transition states taken one step further. *Proc Natl Acad Sci USA* 96:14854–14859.
40. Shen T, Hofmann CP, Oliveberg M, Wolynes PG (2005) Scanning malleable transition state ensembles: Comparing theory and experiment for folding protein u1a. *Biochemistry* 44:6433–6439.
41. Ejtehadi MR, Avall SP, Plotkin SS (2004) Three-body interactions improve the prediction of rate and mechanism in protein folding models. *Proc Natl Acad Sci USA* 101:15088–15093.