Model for the Nucleation Mechanism of Protein Folding

Y. S. Djikaev† and Eli Ruckenstein*

Department of Chemical and Biological Engineering, State University of New York at Buffalo, Buffalo, New York 14260

Received: August 23, 2006; In Final Form: October 26, 2006

A nucleation-like pathway of protein folding involves the formation of a cluster containing native residues that grows by including residues from the unfolded part of the protein. This pathway is examined by using a heteropolymer as a protein model. The model heteropolymer consists of hydrophobic and hydrophilic beads with fixed bond lengths and bond angles. The total energy of the heteropolymer is determined by the pairwise repulsive/attractive interactions between nonlinked beads and by the contribution from the dihedral angles involved. The parameters of these interactions can be rigorously defined, unlike the ill-defined surface tension of a cluster of protein residues that constitutes the basis of a previous nucleation model. The main idea underlying the new model consists of averaging the dihedral potential of a selected residue over all possible configurations of all neighboring residues along the protein chain. The resulting average dihedral potential depends on the distance between the selected residue and the cluster center. Its combination with the average pairwise potential of the selected residue and with a confining potential caused by the bonds between the residues leads to an overall potential around the cluster that has a double-well shape. Residues in the inner (closer to the cluster) well are considered as belonging to the folded cluster, whereas those in the outer well are treated as belonging to the unfolded part of the protein. Transitions of residues from the inner well into the outer one and vice versa are considered as elementary emission and absorption events, respectively. The double-well character of the potential well around the cluster allows one to determine the rates of both emission and absorption of residues by the cluster using a first passage time analysis. Once these rates are found as functions of the cluster size, one can develop a self-consistent kinetic theory for the nucleation mechanism of folding of a protein. The model allows one to evaluate the size of the nucleus and the protein folding time. The latter is evaluated as the sum of the times necessary for the first nucleation event to occur and for the nucleus to grow to the maximum size (of the folded protein). Depending on the diffusion coefficients of the native residues in the range from 10^{-6} to 10^{-8} cm^2/s, numerical calculations for a protein of 2500 residues suggest that the folding time ranges from several seconds to several hundreds of seconds.

1. Introduction

Proteins play an overwhelmingly dominant role in life. If a specific job has to be performed in a living organism, then it is almost always a protein that does it. Life depends on thousands of different proteins whose structures are fashioned so that the individual protein molecules combine, with exquisite precision, with other molecules. For a protein molecule to carry out a specific biological function, it has to adopt a well-defined (native) three-dimensional structure.[1,2] The formation of this structure (of a biologically active globular protein) constitutes the core of the so-called "protein folding problem".[3] Many thermodynamic and kinetic aspects of the process remain obscure, and its mechanism is elusive.[4−9]

Experiment and simulation suggest that there are multiple pathways for protein folding.[5−24] It is believed that initially an unfolded protein transforms very quickly into a compact (but not native) configuration with a few, insignificant number of tertiary contacts. The transition from such a compact configuration to the native one has been suggested to occur via two distinct mechanisms. One of them can be referred to as a "transition state mechanism" whereby the tertiary contacts of the native structure are formed as the protein passes through a sequence of intermediate states thus gradually achieving its unique spatial configuration.[7−24] The protein in the intermediate states has a nativelike overall topology but is stabilized by incorrect hydrophobic contacts. These states correspond to misfolded forms of the native protein. The transition from these intermediate, misfolded states to the correctly folded, native structure is a slow process (compared to the formation of the compact configuration) that occurs on a relatively large time scale (because it involves a large-scale rearrangement of the molecule).[15,16] Alternatively, the transition from the compact "amorphous" configuration to the native state occurs immediately following the formation of a number of tertiary contacts.[15,16] This mechanism is similar to nucleation; i.e., once a critical number of (native) tertiary contacts is established the native structure is formed without passing through any detectable intermediates.[15]

So far most of the work on protein folding has been carried out by using either Monte Carlo (MC) or molecular dynamics (MD) simulations. A rigorous theoretical treatment of protein folding by means of statistical mechanics is hardly practicable because of the extreme complexity of the system, although some approximate treatments have been already reported.[26,27] Moreover, a theoretical model for the nucleation mechanism of the process has so far remained underdeveloped.[15,28,29] The model,

* Author to whom correspondence should be addressed. Phone: (716) 645-2911 ext. 2214. Fax: (716) 645-3822. E-mail: feaeliru@acsu.buffalo.edu.
† E-mail: idjikaev@eng.buffalo.edu.

10.1021/jp065464j CCC: $37.00
Published on Web 01/05/2007

a thermodynamic one, considers the formation of a cluster of protein residues and calculates its free energy change much like the classical nucleation theory (CNT) does. The cluster is characterized by the number of residues ν (with the mole fraction of the hydrophobic ones assumed to be known). As usual in CNT, the critical cluster (nucleus) corresponds to the maximum of the free energy of formation as a function of ν. Such an approach involves the use of the concept of surface tension for a cluster consisting of protein residues. This quantity is an intrinsically ill-defined physical quantity and can be considered only as an adjustable parameter; clearly, no direct experimental measurement thereof is possible. After the formation of a nucleus (a critical size cluster of residues), the protein quickly reaches its native state.

In this paper we present a new, microscopic model for the nucleation mechanism of protein folding. The new model is based on "molecular" interactions both long-range (i.e., pairwise repulsion/attraction) and configurational (bond and dihedral angles) in which protein residues are involved. These parameters can be rigorously defined, and it should be possible (although not straightforward) to determine them theoretically, computationally, or experimentally. The ill-defined surface tension of a cluster of protein residues (within a protein in a compact but non-native configuration) does not enter into the new model, which is thus more reasonable than the old one. The main idea underlying the new model consists of averaging the dihedral potential of a selected residue over all possible configurations of neighboring residues. The resulting average dihedral potential depends on the distance between the residue and the cluster center. Its combination with the average long-range potential between the cluster and the residue and with a confining potential due to the bonds between the residues generates a pair of potential wells around the cluster with a barrier between them. The residues in the inner well are considered to belong to the cluster (part of the protein with correct tertiary contacts) while those in the outer well are treated as belonging to the mother phase (amorphous part of the protein with incorrect tertiary contacts). Transitions of residues from the inner well into the outer one and vice versa are considered as elementary emission and absorption events, respectively. The rates of emission and absorption of residues by the cluster are determined by using a first passage time analysis.[30−35] Once these rates are found as functions of the cluster size, one can develop a self-consistent kinetic theory for the nucleation mechanism of folding of a protein. For example, the size of the critical cluster (nucleus) is then found as the one for which these rates are equal. The time necessary for the protein to fold can be evaluated as the sum of the times necessary for the appearance of the first nucleus and the time necessary for the nucleus to grow to the maximum size (of the folded protein in the native state). Both can be obtained if the emission and absorption rates are known as functions of the cluster size.

The paper is structured as follows. In section 2 we describe a random heteropolymer chain whereby a protein molecule is often modeled[7,15] and outline a CNT-based model for the nucleation mechanism for protein folding. A new, microscopic model is proposed in section 3, and the results of numerical calculations are presented in section 4. A brief discussion and conclusions are summarized in section 5.

2. Heteropolymer Chain as a Protein Model and a CNT-Based Model for the Nucleation Mechanism of Protein Folding

2.1. Heteropolymer as a Protein model. A heteropolymer was introduced as a simple model of a protein in the MD and MC simulations of protein folding dynamics.[7,15] The polypeptide chain of a protein was modeled as a heteropolymer consisting of N connected beads that can be thought of as representing the α-carbons of various amino acids. The heteropolymer may consist of hydrophobic (b), hydrophilic (l), and neutral (n) beads. Two adjacent beads are connected by a covalent bond of fixed length η. This model (and its variants), completed with the appropriate potentials described below, has been shown[7,15,26,28] to be able to capture the essential characteristics of protein folding even though it contains only some of the features of a real polypeptide chain. For example, this model ignores the side groups although they are known to be crucial for intramolecular hydrogen bonding.[1] Besides, the presence of a solvent (water) in a real physical system has been usually accounted for too simplistically, although protein dynamics were reported to become more realistic in those MD simulations where the solvent molecules were explicitly taken into account.[36] Despite these limitations, various modifications of the heteropolymer model[6,26,28,37−40] shed light on some important details regarding the transition of a protein from its unfolded state to the native one.[7,15]

The total energy of the heteropolymer (polypeptide chain) can contain three different types of contributions. First, the contribution from repulsive/attractive forces between pairs of nonadjacent beads (these can be, for example, of Lennard-Jones or other types). The next contribution can arise from the harmonic forces due to the oscillations of bond angles. Finally, there is a contribution from the dihedral angle potential due to the rotation around the peptide bonds. There are various ways to model these three types of contributions.[7,15,37−39]

A pair interaction between two nonadjacent beads i and j at a distance d away from each other can be considered to be of the form[7,15]

$$\phi_{ij}(d) = \begin{cases} 4\epsilon_{\mathrm{b}}[(\eta/d)^{12} - (\eta/d)^{6}] & (i, j = \mathrm{b}) \\ 4\epsilon_{\mathrm{l}}[(\eta/d)^{12} + (\eta/d)^{6}] & (i = \mathrm{l}, j = \mathrm{b}, \mathrm{l}) \\ 4\epsilon_{\mathrm{n}}(\eta/d)^{12} & (i = \mathrm{n}, j = \mathrm{b}, \mathrm{l}, \mathrm{n}) \end{cases} \quad (1)$$

where η is the bond length (fixed) and ϵ_{b}, ϵ_{l}, and ϵ_{n} are energy parameters.

The angle β between two successive bonds (in the heteropolymer) can be regarded to be subjected to the harmonic potential

$$\phi_{\beta} = \frac{k_{\beta}}{2}(\beta - \beta_0)^2 \quad (2)$$

where the spring constant k_{β} is sufficiently large for the deviation of the bond angles from the average value β_0 to be small. As argued in refs 7 and 15, the bond angle forces play a minor role in protein folding/unfolding; hence all bond angles can be set to be equal to β_0.

The dihedral angle potential arises due to the rotation of three successive peptide bonds connecting four successive beads and is related to the dihedral angle δ via

$$\phi_{\delta} = \epsilon'_{\delta}(1 + \cos\delta) + \epsilon''_{\delta}(1 + \cos 3\delta) \quad (3)$$

where ϵ'_{δ} and ϵ''_{δ} are independent energy parameters that depend on the nature and sequence of the four beads involved in the dihedral angle δ. This potential has three minima, one in the trans configuration at $\delta = 0$ and two others in the gauche configurations at $\delta = \pm\arccos\sqrt{(3\epsilon''_{\delta} - \epsilon'_{\delta})/12\epsilon''_{\delta}}$ (the former one being the lowest).

The above structure of the potential functions of a heteropolymer was suggested in ref 6. Discrete analogues (for a protein on a diamond lattice) of eqs 1 and 3 completed with a "cooperativity potential" were also proposed,[37−39] and random energy models were used.[28] It was shown[7,15] by MD simulations, which employed low friction Langevin dynamics, that a proper balance between the above three contributions (eqs 1−3) to the total energy of the heteropolymer ensures that the heteropolymer folds into a well-defined β-barrel structure. The balancing between these contributions was achieved by adjusting the energy parameters ϵ_b, ϵ_l, ϵ_n, ϵ'_δ, and ϵ''_δ for each of the types of beads. It was also found[7,15] that the balance between the dihedral angle potential, which tends to stretch the molecule into a state with all bonds in a trans configuration, and the attractive hydrophobic potential is crucial to induce folding into a β-barrel-like structure upon cooling. If attractive forces are excessively dominant, then they make the heteropolymer fold into a globulelike structure, while an overwhelming dihedral angle potential forces the chain to remain in an unfolded (elongated) state (even at low temperatures) with bonds mainly in the trans configuration.

The possibility that a nucleation-like mechanism constitutes a viable pathway for protein folding was first suggested in refs 15 and 28. The formalism of the nucleation theory was used to evaluate the size of a critical cluster (nucleus) of native protein residues, the formation of which leads to a rapid transition of the whole protein to its native state. Let us denote the total number of residues in the protein by N_0 and consider the formation of a cluster having a correct tertiary structure in an unfolded protein. The free energy of formation of such a cluster of ν native residues (i.e., residues which are in the same state as they are in the native protein) can be written in the framework of CNT as

$$W = -\nu\Delta\mu + \sigma 4\pi\lambda^2\nu^{2/3} \tag{4}$$

where $\Delta\mu \equiv \mu_u - \mu_f$ is the difference between the free energies per residue in the unfolded and folded (native) states, respectively (marked with the subscripts "u" and "f"), σ is the "surface" tension of the boundary between the cluster (having a native structure) and the unfolded part of the protein, $\lambda = (3\upsilon/4\pi)^{1/3}$, and υ is the volume of a protein residue in its native state.

It was argued[15] that the initial stage of protein folding is driven by the hydrophobic attractive forces so that the volume term (i.e., the first one) in eq 4 is determined by the number of hydrophobic contacts in the cluster and hence can be written as $-(1/2)\epsilon_b\chi\nu(\chi\nu - 1)$, where χ is the mole fraction of hydrophobic residues in the cluster (assumed to be the same as in the whole protein). As a result, the number of residues in the critical cluster is given by $\nu_c = (8\pi\sigma\lambda^2/3\chi^2\epsilon_b)^{3/4}$, which for typical values of λ, σ, and ϵ_b was estimated[15] to be on the order of 10. In ref 28, $\Delta\mu$, which appears in the volume term of eq 4, was evaluated to be on the order of $0.1k_BT$ (k_B being the Boltzmann constant and T the absolute temperature). The surface tension was argued to arise because the amino acid residues located at the cluster surface interact more strongly with the cluster interior than with the unfolded part of the protein. Since the interaction energies in protein folding are on the order of k_BT, the surface tension σ could be obtained from $\sigma 4\pi\lambda^2 \approx k_BT$, and the number of residues in the critical cluster was evaluated[28] to be on the order of 100 (for $N_0 = 150$). Despite the significant difference between them, both estimates corroborate the idea that the nucleation mechanism constitutes a viable pathway for protein folding.[18−22,41−45]

3. A Kinetic Approach to the Nucleation Mechanism of Protein Folding

The above nucleation model of protein folding involves thermodynamics within the framework of CNT. However, it was argued that the concept of surface tension may not be adequate for too small clusters (such as those of interest in nucleation),[30−32] not to mention the assumption (of CNT) that it is equal to the surface tension of a planar interface. Although CNT produces reasonable agreement with experiment for unary nucleation, its application to multicomponent nucleation leads to several inconsistencies and large discrepancies with experiment[46−49] that are blamed on the use of the concept of surface tension. In the case of protein folding this problem is even more complex because σ in eqs 4 is an ill-defined quantity that cannot be determined experimentally due to the nonexistence of bulk "folded" and "unfolded" proteins as real physical phases, not to mention a flat interface between them.

To avoid the use of macroscopic thermodynamics in the kinetic theory of unary nucleation, an alternative approach was proposed by Ruckenstein and co-workers[30−32] on the basis of a mean first passage time analysis. Unlike CNT, the new theory[30−32] is built upon molecular interactions and does not make use of the surface tension for the tiny clusters involved in nucleation. Instead, the theory[30−32] exploits the fact that one can derive and solve the kinetic equation of nucleation (hence find the nucleation rate) if the emission and absorption rates of a cluster are determined as functions of its size. For the rate of absorption of molecules by the cluster, the new theory uses (as CNT does) a standard gas-kinetic expression,[50] but (and this is the main idea of the new approach[30−32]) the rate of emission of molecules by the cluster is determined via a mean first passage time analysis. This time is calculated by solving a single-molecule master equation for the probability distribution function of a surface layer molecule moving in a potential well around the cluster. The master equation is a Fokker−Planck equation in the phase space that can be reduced to the Smoluchowski equation owing to the hierarchy of the characteristic time scales involved in the evolution of the single-molecule distribution function with respect to coordinates and momenta.[30−32] Recently, a further development of that kinetic theory was proposed by combining it[33,34] with the density functional theory (DFT) and extending it to binary[34] and heterogeneous[35] systems.

Although the emission rate of the cluster in refs 30−35 was found via a first passage time analysis, the absorption rate there was calculated using an expression derived in the framework of the gas-kinetic theory,[50] which assumes a Maxwellian distribution of the velocities of the mother phase molecules. While this assumption is unquestionably valid for vapor-to-liquid nucleation in dilute (if not ideal) gases, it becomes increasingly inaccurate as the density of the mother phase increases and molecular interactions therein become non-negligible. Clearly, this assumption (hence the absorption rate based thereupon) is inadequate in the treatment of cluster formation during protein folding. Indeed, the amino acid residues of the protein are successively linked by bonds of virtually fixed length and fixed angle between each pair.

In this section we will present a new, kinetic model for the nucleation mechanism of protein folding based on a first passage analysis that will be used to determine not only the rate of emission (of native residues from the cluster) but also (unlike refs 30−35) the rate of absorption (of non-native residues by the cluster). The general formalism of our model is a mean first

passage time analysis, but a major difference, compared to refs 30−35, is the double-well potential generated in this case around the cluster.

3.1. Potential Well around a Cluster of Native Residues. A heteropolymer chain as a protein model, proposed in refs 7 and 15 and described above, consists of three types of beads: neutral, hydrophobic, and hydrophilic. The neutral beads (n) play an important role in that model. Their interaction with each other is purely repulsive, and the dihedral angle forces are considered to be weaker for the bonds involving them so that the bending is enhanced where they are present. Indeed, MD simulations[7,15] show that such a heteropolymer acquires a β-barrel shape in the lowest energy configuration, with neutral residues appearing mostly in the bend regions. For the sake of simplicity, we will consider a heteropolymer consisting of only hydrophobic (b) and hydrophilic (l) beads rather than three kinds of beads. However, this simplification requires us to rebalance the ϵ's in eq 1 and particularly in eq 3 to allow the formation of loops and turns in a folded heteropolymer chain.

Consequently, we will adopt a two-component analogue of a heteropolymer chain with the pair interaction, bond angle, and dihedral angle potentials given by eqs 1−3 as a model for a protein. Besides, one must introduce a confining potential that excludes the volume free of protein residues. Therefore, the formation of a cluster consisting of native residues during protein folding can be regarded as binary nucleation. We shall therefore present a nucleation mechanism model of protein folding in terms of binary nucleation by using a first passage time analysis[30−35] that accounts for the particular shape of the present potential.

Consider a binary cluster of spherical shape (with sharp boundaries and radius R) immersed in a binary fluid mixture. In the original papers regarding nucleation[30−32] and recent developments,[33−35] a molecule of component i (i = b, l) located in the surface layer of the cluster was considered to perform a thermal chaotic motion in a spherically symmetric potential well $\phi_i(r)$ resulting from the pair interactions of this molecule with those in the cluster.

Assuming pairwise additivity of the interactions between beads, $\phi_i(r)$ is provided by

$$\phi_i(r) = \sum_j \int_V \mathrm{d}\mathbf{r}'\, \rho_j(r')\phi_{ij}(|\mathbf{r}' - \mathbf{r}|) \tag{5}$$

Here $\mathbf{r}$ is the coordinate of the surface bead i, $\rho_j(r')$ (j = 1, 2) is the number density of beads of component j at point $\mathbf{r}'$ (spherical symmetry is assumed with the cluster center chosen as the origin of the coordinate system), and $\phi_{ij}(|\mathbf{r}' - \mathbf{r}|)$ is the interaction potential between two beads of components i and j at points $\mathbf{r}$ and $\mathbf{r}'$, respectively. The integration in eq 5 has to be carried out over the whole volume of the system, but the contribution from the unfolded part can be assumed to be small and accounted for by particular choices of ϵ_b and ϵ_l.

For nucleation in proteins the potential well $\psi_i(r)$ for a residue of type i around the cluster is determined not only by the potential $\phi_i(r)$ but also by three other constituents, $\phi_\beta(r)$, $\bar{\phi}_\delta(r)$, and ϕ_co, which represent the bond angle, dihedral angle, and confining potentials, respectively

$$\psi_i(r) = \phi_i(r) + \phi_\beta(r) + \bar{\phi}_\delta(r) + \phi_\mathrm{co}$$

Without affecting the generality of the model, one can significantly simplify the algebra and eventual numerical calculations by assuming that all bond angles are fixed and equal to $\beta_0 = 105°$. Under this assumption the contribution to the potential energy of the protein arising from the bond angle potential is constant and does not depend on the distance r between the selected bead and the center of the cluster. Therefore, the term $\phi_\beta(r)$ on the right-hand side of the above equation can be disregarded (or, equivalently, be chosen as a reference level for the potential energy), i.e.,

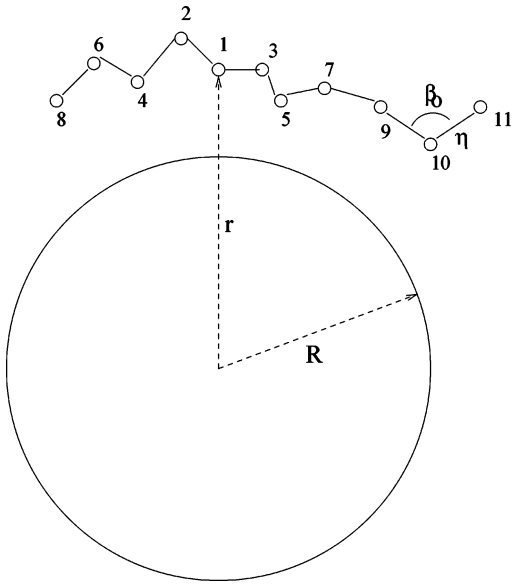

Figure 1. Scheme of a piece of a heteropolymer chain around a spherical cluster of radius R. Bead 1 is in the plane of the figure, whereas other beads may all lie in different planes, but all bond angles are equal to 105°, and their lengths are equal to η. The distance between the selected bead 1 and the center of the cluster is r.

$$\psi_i(r) = \phi_i(r) + \bar{\phi}_\delta(r) + \phi_\mathrm{co} \tag{6}$$

Because of the bonds between the residues, none of them can be located outside a volume that depends on the densities of the folded and unfolded regions. Assuming that volume to be spherical, its radius r_co is given by the expression $r_\mathrm{co} = [R^3 + 3(N_0 - \nu)/4\pi\rho_\mathrm{u}]^{1/3}$ with $\nu = (4\pi/3)R^3\rho_\mathrm{f}$, ρ_u and ρ_f being the unfolded and folded densities, respectively. Consequently, the confining potential has the form

$$\phi_\mathrm{co} = \begin{cases} 0 & (r < r_\mathrm{co}) \\ \infty & (r \geq r_\mathrm{co}) \end{cases} \tag{7}$$

The term $\bar{\phi}_\delta(r)$ in $\psi_i(r)$ is due to the dihedral angle potential of the whole protein. Consider a bead 1 (of type b or *l)* at a distance $\tilde{d}$ from the surface of the cluster, so that $r = R + \tilde{d}$ (Figure 1). The total dihedral angle potential ϕ'_δ of the whole protein chain for a given configuration of beads 1, 2, 3, ..., N_0 is a function of r, $\mathbf{r}_2$, $\mathbf{r}_3$, $\mathbf{r}_4$, $\mathbf{r}_5$, $\mathbf{r}_6$, and $\mathbf{r}_7$ and can be written in the form (for simplicity, the arguments $\mathbf{r}_2$, $\mathbf{r}_3$, $\mathbf{r}_4$, $\mathbf{r}_5$, $\mathbf{r}_6$, and $\mathbf{r}_7$ of ϕ'_δ are not explicitly noted)

$$\phi'_\delta(r) = \phi_\delta(\delta^{642}_{421}(r)) + \phi_\delta(\delta^{421}_{213}(r)) + \phi_\delta(\delta^{213}_{135}(r)) + \phi_\delta(\delta^{135}_{357}(r)) + \text{constant}$$

where δ^{ijk}_{jkl} is a dihedral angle between two planes, one of which is determined by beads i, j, and k and the other by beads j, k, and l. The last term on the right-hand side of the above equation, representing the contributions from the dihedral angles involving beads 8, 9, ..., N_0, does not depend on r because these dihedral angles remain unaffected as the position of bead 1 changes. Hence this contribution (constant) is omitted hereafter

without being specified since it can be regarded as affecting only the reference level for $\psi_i(r)$

$$\phi'_\delta(r) = \phi_\delta(\delta^{642}_{421}(r)) + \phi_\delta(\delta^{421}_{213}(r)) + \phi_\delta(\delta^{213}_{135}(r)) + \phi_\delta(\delta^{135}_{357}(r)) \quad (8)$$

For a given location of bead 1, various configurations of beads 2, 3, ..., N_0 (subject to the constraints on the bond length and bond angle as well as to the constraint of excluded cluster volume) lead to various sets of dihedral angles. However, variations in the locations of beads 8, 9, ..., N_0 lead to variations in the dihedral potential that are independent of r. Thus, the dihedral term $\bar{\phi}_\delta(r)$ (less a term independent of r and omitted hereafter) on the right-hand side of eq 6 can be obtained by averaging eq 8 with the Boltzmann factor $\exp[-\phi'_\delta(r)/k_BT]$ (and appropriate normalization constant) over all possible configurations of beads 2, 3, ..., 7 and assigning the result to a selected bead with fixed coordinates, i.e., bead 1

$$\bar{\phi}_\delta(r) = f^{-1} \int_{\Omega_{18}} d\mathbf{r}_2\, d\mathbf{r}_3\, d\mathbf{r}_4\, d\mathbf{r}_5\, d\mathbf{r}_6\, d\mathbf{r}_7\, \phi'_\delta(r) \times \exp[-\phi'_\delta(r)/k_BT] \quad (9)$$

where f is a normalization constant

$$f^{-1} \int_{\Omega_{18}} d\mathbf{r}_2\, d\mathbf{r}_3\, d\mathbf{r}_4\, d\mathbf{r}_5\, d\mathbf{r}_6\, d\mathbf{r}_7 \exp[-\phi'_\delta(r)/k_BT] \quad (10)$$

and Ω_{18} is the integration region in an 18-dimensional space. The 18-fold integrals in eqs 9 and 10 can be reduced to 7-fold integrals by taking into account the constraints of fixed bond length and fixed bond angle (see Appendix A)

$$\bar{\phi}_\delta(r) = f^{-1} \sum_{i,j,k,m,n=-}^{+} \int_0^{2\pi} d\varphi_2 \int_{L_2} dr_2 \int_{L_3^i} dx_3 \int_{L_4^j} dx_4 \int_{L_5^k} dx_5 \times \int_{L_6^m} dx_6 \int_{L_7^n} dx_7\, {r_2}^2 \sin\Theta_2(r, r_2) \tilde{\phi}_{ijkmn}(r, \varphi_2, r_2, x_3, ..., x_7) \times \exp[-\tilde{\phi}_{ijkmn}(\varphi_2, r_2, x_3, ..., x_7)/k_BT] \quad (11)$$

$$f = \sum_{i,j,k,m,n=-}^{+} \int_0^{2\pi} d\varphi_2 \int_{L_2} dr_2 \int_{L_3^i} dx_3 \int_{L_4^j} dx_4 \int_{L_5^k} dx_5 \times \int_{L_6^m} dx_6 \int_{L_7^n} dx_7\, {r_2}^2 \sin\Theta_2(r, r_2) \times \exp[-\tilde{\phi}_{ijkmn}(r, \varphi_2, r_2, x_3, ..., x_7)/k_BT] \quad (12)$$

Each of the summation indices in eqs 11 and 12 takes on two values, denoted (see Appendix A) + and −, so that there are 5^2 terms in the sum differing by the integrand as well as by the integration ranges (except for L_2 which is independent of i, j, k, m, and n). The definitions of the coordinates φ_2, r_2, x_3, x_4, x_5, x_6, and x_7, the explicit forms of the integration ranges L_2, L_3^i, L_4^j, L_5^k, L_6^m, and L_7^n, and the rules for transforming the function $\phi'_\delta(r)$ into the function $\tilde{\phi}_{ijkmn}(\varphi_2, r_2, x_3, ..., x_7)$ are provided in Appendix A.

Figure 2 presents typical shapes of the constituents $\phi_i(r)$ and $\bar{\phi}_\delta(r)$ of the potential well as functions of the distance from the cluster center as well as the overall potential well $\psi_i(r)$ itself. (For details regarding the numerical calculations see section 4.) The contribution $\phi_i(r)$, arising from the pairwise interactions, has a familiar form[30−35] reminiscent of the underlying Lennard-Jones potential, while the contribution from the average dihedral potential has a rather remarkable behavior. Starting with its maximum value at the cluster surface, it monotonically decreases with increasing r until it becomes constant for $r \geq \tilde{r}$ (section 4, eq 32). Thus, except very short distances from the cluster surface where $\phi_i(r)$ sharply decreases from ∞ to its global minimum, the potential $\psi_i(r)$ is shaped by two competing terms, $\phi_i(r)$ and $\bar{\phi}_\delta(r)$, which increase and decrease, respectively, with increasing r and by the confining potential ϕ_{co}. As a result, the overall potential $\psi_i(r)$ has a double-well shape: The inner well is separated by a potential barrier from the outer well. This shape of $\psi_i(r)$ is of crucial importance to our nucleation mechanism model of protein folding because it allows one to use the mean first passage time analysis for the determination of the rates of absorption and emission of beads by the cluster.

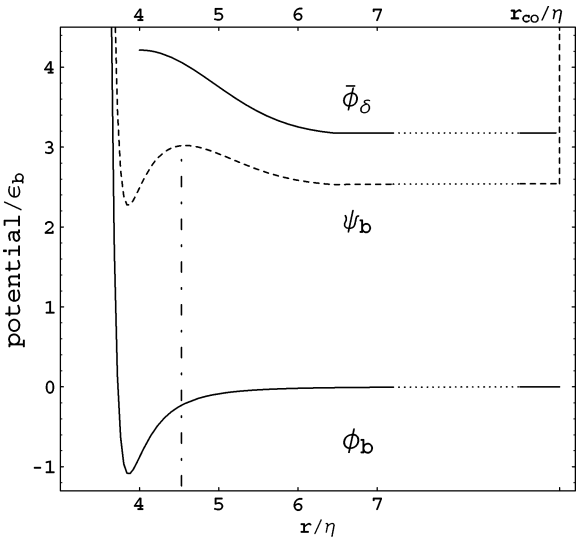

Figure 2. Typical shapes of the potentials $\phi_b(r)$ (lower solid curve), $\bar{\phi}_\delta(r)$ (upper solid curve), and $\psi_b(r)$ (dashed curve) for a hydrophobic bead around the cluster as functions of the distance r from the center of the cluster of radius $R = 3\eta$. The vertical dashed line at $r \simeq 4.53\eta$ indicates the location of the maximum of the barrier between the ipw and the opw that determines the widths λ_b^{iw} and λ_b^{ow} of both potential wells. The outer boundary of the opw ($r_{co} \simeq 12.99\eta$) was assumed to coincide with the outer boundary of the volume wherein the whole protein is encompassed.

3.2. Determination of the Emission and Absorption Rates. Formulating our model in the spirit of mean first passage time analysis,[30−35] a bead (residue) is considered as belonging to a cluster as long as it remains in the inner potential well (hereinafter referred to as "ipw") and as dissociated from the cluster when it passes over the barrier between the ipw and the outer potential well (hereinafter referred to as "opw"). The rate of emission, W^-, is determined by the mean time necessary for the passage of the bead from the ipw over the barrier into the opw. Likewise, a bead is considered as belonging to the unfolded part of the heteropolymer (protein) as long as it remains in the opw and as absorbed by the cluster when it passes over the barrier between the opw and the ipw. The rate of emission, W^+, is determined by the mean time necessary for the passage of the bead from the opw over the barrier into the ipw.

The mean first passage time of a bead escaping from a potential well is calculated on the basis of a kinetic equation governing the chaotic motion of the bead in that potential well. The chaotic motion of the bead is assumed to be governed by the Fokker−Planck equation for the single-particle distribution function with respect to its coordinates and momenta, i.e., in the phase space.[51−53] Under favorable conditions, the Fokker−Planck equation reduces to the Smoluchowski equation, which involves diffusion in an external field.[52,53] Solving that

equation,[30−35] one can obtain the following expressions for W^- and W^+, the emission and absorption rates (see Appendix B for details), respectively

$$W^- = n^{iw} D^{iw} \omega^{iw} \tag{13}$$

$$W^+ = n^{ow} D^{ow} \omega^{ow} \tag{14}$$

Here the superscripts "iw" and "ow" mark the quantities for the inner and outer potential wells, respectively; n is the number of beads in the well, D is the diffusion coefficient of the bead, and ω^{iw} and ω^{ow} are defined as

$$\omega^{iw} \equiv 1/(D^{iw}\bar{\tau}^{iw}) \quad \omega^{ow} \equiv 1/(D^{ow}\bar{\tau}^{ow}) \tag{15}$$

where $\bar{\tau}^{iw}$ and $\bar{\tau}^{iw}$ are the mean first passage times for a bead in the ipw to cross over the barrier into the opw and vice versa, respectively. Note that in the original work[30−32] on the mean first passage time analysis in the nucleation theory and its recent development[33−35] this method was used only for determining W^- but not W^+ . In the present model, the double-well shape of the overall potential $\psi(r)$ around the cluster allows one to use the first passage time analysis to determine W^+ as well. Clearly, the quantities W^- ,W^+ , ω^{iw}, ω^{iw}, $\bar{\tau}^{iw}$, and $\bar{\tau}^{iw}$ are functions of the cluster size and composition (for their explicit forms see Appendix B). However, since the overall composition of the protein (heteropolymer) is fixed, one can assume that the cluster that forms during folding has a constant composition equal to the overall protein composition, which reduces the theory to a unary nucleation theory. (Strictly speaking, one can develop a theoretical model for the nucleation mechanism of protein folding without this assumption, which would lead to a binary nucleation theory, but this would drastically complicate the problem computationally.) Under these assumptions the aforementioned quantities are functions of only the size of the cluster (say, its radius R or the total number of beads ν therein).

3.3. Equilibrium Distribution and Steady-State Nucleation Rate. During protein folding, clusters of various sizes may emerge and exist simultaneously with different probabilities. Let us denote the distribution of clusters with respect to the number of beads in a cluster at time t by $g(\nu,t)$. Once the emission and absorption rates $W^- = W^-(\nu)$ and $W^+ = W^+(\nu)$ are known as functions of cluster size, one can find the equilibrium distribution of clusters $g_e(\nu)$ and solve the kinetic equation of nucleation to find the steady-state nucleation rate.

According to the principle of detailed balance

$$W^+(\nu-1)g_e(\nu-1) = W^-(\nu)g_e(\nu) \tag{16}$$

which can be rewritten as

$$\frac{g_e(\nu)}{g_e(\nu-1)} = \frac{W^+(\nu-1)}{W^-(\nu)} \tag{17}$$

By applying eq 17 to $(\nu - i)$ with $i = 2, 3, ..., \nu - 1$ and multiplying the right-hand and left-hand sides of all equalities, one obtains

$$\frac{g_e(\nu)}{g_e(1)} = \prod_{i=1}^{\nu-1} \frac{W^+(\nu-i)}{W^-(\nu-i+1)} \tag{18}$$

The equilibrium distribution of clusters $\nu = 1$ is just the number density of residues in a compact (but unfolded) protein, i.e., $g_e(1) = \rho_u$, so that eq 18 can be rewritten as

$$g_e(\nu) = \rho_u \frac{W^+(1)}{W^+(\nu)} \prod_{i=1}^{\nu-1} \frac{W^+(\nu-i+1)}{W^-(\nu-i+1)} \tag{19}$$

Let us introduce the function $G(\nu) = -k_B T \ln[g_e(\nu)/\rho_u]$. Clearly, $G(\nu)$ in the present theory plays a role similar to the free energy of cluster formation in CNT[54−56] and can be expected to have a shape similar to the latter with a global maximum attained at some critical ν_c (hereinafter the subscript "c" marks quantities at the critical point). In the vicinity of the critical size the function $G(\nu)$ can be accurately represented by its bilinear form $G(\nu) = G_c + (1/2)G_c''(\nu - \nu_c)^2$, where G_c'' is the second derivative of $G(\nu)$ at $\nu = \nu_c$. The essence of our model, as an alternative to the CA-based theory, consists of constructing the equilibrium distribution of clusters, $g_e(\nu)$, and the function $G(\nu)$ without employing classical thermodynamics.

The kinetic equation of nucleation in the vicinity of the critical point can be derived from the discrete equation governing the temporal evolution of the distribution $g(\nu, t)$

$$\frac{\partial g(\nu,t)}{\partial t} = W^+(\nu-1)g(\nu-1,t) - W^+(\nu)g(\nu,t) + W^-(\nu+1)g(\nu+1,t) - W^-(\nu)g(\nu,t) \tag{20}$$

This equation assumes that the evolution of clusters occurs through the absorption and emission of a bead. For the sake of simplicity, we neglect the multimer absorption and emission. Assuming $\nu \gg 1$, one can rewrite eq 20 in a continuous form by using Taylor series expansions for its right-hand side and retaining only the lowest order terms. The resulting partial differential equation has the Fokker−Planck form

$$\frac{\partial g(\nu,t)}{\partial t} = -\frac{\partial}{\partial \nu} J(\nu,t) \tag{21}$$

where

$$J(\nu,t) = -A(\nu)g(\nu,t) - \frac{\partial}{\partial \nu}(B(\nu)g(\nu,t)) \tag{22}$$

Here, the coefficients $A(\nu)$ and $B(\nu)$ depend on the emission and absorption rates W^- and W^+ of the cluster

$$A(\nu) = W^+(\nu) - W^-(\nu) \quad B(\nu) = -\frac{1}{2}[W^+(\nu) + W^-(\nu)]$$

However, according to the principle of detailed balance, $W^-(\nu + 1)/W^+(\nu) = \exp[-G(\nu) + G(\nu + 1)] \simeq \exp[\partial G(\nu)/\partial \nu]$. The first derivative $\partial G(\nu)/\partial \nu$ is equal to zero at the critical point $\nu = \nu_c$ and is close to zero in its vicinity where, in addition, $W^+(\nu + 1) \simeq W^+(\nu)$. Therefore, $W^-(\nu)/W^+(\nu) \simeq 1 + o(\partial G(\nu)/\partial \nu)$, where $o(x)$ denotes a quantity on the same order of magnitude as x. Neglecting the terms of order $o(\partial G(\nu)/\partial \nu)$, one can assume that $W^+(\nu) + W^-(\nu) \simeq 2W^+(\nu)$ and $W^+(\nu) \simeq W^+(\nu_c) \equiv W_c^+$; hence $B \simeq -W_c^+$. Because at equilibrium $J = 0$, eq 22 leads to

$$A(\nu)g_e(\nu) + B(\nu)\frac{\partial g_e(\nu)}{\partial \nu} = 0$$

whence one obtains

$$A(\nu) = -B(\nu)\frac{1}{g_e(\nu)}\frac{\partial g_e(\nu)}{\partial \nu} \simeq W_c^+ \frac{\partial G(\nu)}{\partial \nu}$$

and J in eq 22 acquires the form

$$J(\nu,t) = -W_c^+\left(\frac{\partial}{\partial\nu} + \frac{\partial G(\nu)}{\partial\nu}\right)g(\nu,t) = -W_c^+ g_e(\nu)\frac{\partial}{\partial\nu}\frac{g(\nu,t)}{g_e(\nu)} \tag{23}$$

The substitution of this equation into eq 21 leads to the equation

$$\frac{\partial g(\nu,t)}{\partial t} = W_c^+\frac{\partial}{\partial\nu}\left[\frac{\partial}{\partial\nu} + \frac{\partial G}{\partial\nu}\right]g(\nu,t) \tag{24}$$

The steady-state solution of eq 24 in the vicinity of ν_c, subject to the conventional boundary conditions

$$\frac{g(\nu, t)}{g_e(\nu)} \to 1 \quad (\nu \to 0)$$

$$\frac{g(\nu, t)}{g_e(\nu)} \to 0 \quad (\nu \to \infty) \tag{25}$$

where $g_e(\nu)$ is the equilibrium distribution, provides the steady-state nucleation rate[54−56] $J(\nu,t) = J_s$. Indeed, the rightmost equality in eq 23 in the steady state takes the form

$$J_s = -W_c^+ g_e(\nu)\frac{\partial}{\partial\nu}\frac{g(\nu)}{g_e(\nu)} \tag{26}$$

Using the boundary conditions (eq 25) and the bilinear form of $G(\nu) = G_c + (1/2)G_c''(\nu - \nu_c)^2$ in the vicinity of the critical point, one can easily integrate the above equation and obtain for J_s the expression

$$J_s = \frac{W_c^+}{\sqrt{\pi}\Delta\nu_c}\rho_u\, e^{-G_c/k_BT} \tag{27}$$

where $\Delta\nu_c = |\partial^2 G/\partial\nu^2|_c^{-1/2}$.

3.4. Evaluation of the Protein Folding Time. Knowing the emission and absorption rates as functions of ν as well as the nucleation rate J_s, one can estimate the time t_f necessary for the protein to fold via nucleation. Roughly speaking, protein folding (via nucleation) consists of two stages. During the first stage, a critical cluster of native residues is formed (nucleation proper). During this stage, i.e., for $\nu < \nu_c$, the emission rate W^- is larger than W^+, but the cluster still can attain the critical size by means of fluctuations. At the second stage the nucleus grows via regular absorption of native residues, which dominates their emission, $W^- < W^+$ for $\nu > \nu_c$. Thus, the folding time is given by

$$t_f \simeq t_n + t_g \tag{28}$$

where t_n is the time necessary for one critical cluster to nucleate within a compact (but still unfolded) protein and t_g is the time necessary for the nucleus to grow up to the maximum size, i.e., attain the size of the folded protein.

The time t_n of the first nucleation event can be estimated as

$$t_n \simeq 1/J_s V_0 \tag{29}$$

where V_0 is the volume of the unfolded protein in a compact state. The growth time t_g can be found by solving the differential equation

$$\frac{d\nu}{dt} = W^+(\nu) - W^-(\nu) \tag{30}$$

subject to the initial conditions $\nu = \nu_c$ at $t = 0$ and $\nu = N_0$ at $t = t_g$. The solution of eq 30 is given by the integral

$$t_g \simeq \int_{\nu_c}^{N_0}\frac{d\nu}{W^+(\nu) - W^-(\nu)} \tag{31}$$

4. Numerical Evaluations

In this section we will present some numerical results of the application of our model to the folding of a model protein, namely, a heteropolymer consisting of a total of 2500 hydrophobic and hydrophilic residues, with the mole fraction of hydrophobic residues $\chi_0 = 0.75$. The interactions between a pair of nonlinked beads were modeled by Lennard-Jones (LJ) type potentials (eq 1), whereas the potential due to the dihedral angle δ was modeled according to eq 3. The presence of water molecules was not taken into account explicitly but was assumed to be included into the model via the potential parameters.

All numerical calculations were carried out for the following values of the interaction parameters

$$\eta = 5.39 \times 10^{-8}\ \text{cm}$$

$$\epsilon_l = (2/700)\epsilon_b$$

$$\epsilon_\delta' = \epsilon_\delta'' = 0.3\epsilon_b$$

$$\epsilon_b/kT = 1$$

The typical density of a folded protein was evaluated using the data from refs 57 and 58 and was set to $\rho_f\eta^3 = 1.05$, whereas the typical density of an unfolded protein in the compact configuration was set to be $\rho_u = 0.25\rho_f$. (Note that comparable values for ρ_f and ρ_d were also suggested in ref 28.) Taking into account the results of ref 59, the diffusion coefficients in the ipw and the opw were assumed to be related as $D^{iw}\rho_f = D^{ow}\rho_u$. Because of the lack of reliable data on the diffusion coefficient of a residue in a protein chain, D^{iw} was assumed to vary between 10^{-6} and 10^{-8} cm ^{2}/s.

Figure 2 presents typical shapes of the potentials $\phi_b(r)$ (lower solid curve), $\bar{\phi}_\delta(r)$ (upper solid curve), and $\psi_b(r)$ (dashed curve) for a hydrophobic bead around the cluster as functions of the distance r from the center of a cluster of radius $R = 3\eta$. The potential $\phi_b(r)$ is due to the pairwise interactions of the Lennard-Jones type and has a shape reminiscent thereof. Previous applications[30−35] of a mean first passage time analysis to nucleation employed this kind of potential well around a cluster.

The average dihedral potential (assigned to a selected bead) $\bar{\phi}_\delta(r)$ has a maximum value at the cluster surface and decreases monotonically with increasing r until it becomes constant for $r \geq \tilde{r}$, with $\tilde{r} - R$ being the maximum distance between beads 1 and 6 (or beads 1 and 7), which depends on R, η, and β_0

$$\tilde{r} = R + \eta\sqrt{3 - 2\cos\beta_0 + 2\sqrt{2(1 - \cos\beta_0)}\sin\frac{\beta_0}{2}} \tag{32}$$

Such a behavior of $\bar{\phi}_\delta(r)$ can be thought to be a consequence of an increase in the entropy of the heteropolymer chain as the selected bead approaches the cluster surface for $r < \tilde{r}$. It occurs because the configurational space available to the neighboring beads (2−7) becomes increasingly restricted. This piece of the heteropolymer chain (beads 1−7) would not feel the presence of the cluster any more at $r > \tilde{r}$ if there were no pairwise interactions.

As a result of the combination of $\phi_b(r)$, $\bar{\phi}_\delta(r)$, and ϕ_{co}, the overall potential $\psi_b(r)$ has a double-well shape: The inner well is separated by a potential barrier from the outer well. The geometric characteristics of the wells (widths, depths, etc.) and the height and location of the barrier between them depend on the interaction parameters ϵ_b, ϵ_l, and $\epsilon_\delta = \epsilon'_\delta = \epsilon''_\delta$. The larger the ratio $\epsilon_\delta/\epsilon_b$, the higher the barrier between the wells, the wider the ipw, and the narrower the opw. Note that the barrier has different heights for the ipw and opw beads. For a given N_0 the location r_{co} of the outer boundary of the opw is determined by the size of the cluster and densities ρ_f and ρ_u (with the above choices thereof $r_{co} \simeq 12.99\eta$ for a cluster $R = 3\eta$). The existence of an opw allows one to consider the absorption of a bead by the cluster as an escape of the bead from the opw by crossing over the barrier into the ipw. One can therefore use a mean first passage time analysis for the determination not only of the emission rate but also of the rate of absorption of beads by the cluster. Since the use of the traditional expression for the absorption rate (based on the gas-kinetic theory)[46] is rather inadequate for the cluster growth within the protein, the double-well shape of $\psi_i(r)$ is of crucial importance in our model for the nucleation mechanism of protein folding.

In Figure 3 the average dihedral potential (assigned to a selected bead) $\bar{\phi}_\delta(r)$ is plotted as a function of r for three clusters of sizes (a) $R = 3\eta$, (b) $R = 6\eta$, and (c) $R = 9\eta$. The points represent the actual numerical results obtained by using Monte Carlo integration with 1×10^6 to 2×10^6 points in calculating the 7-fold integrals in eqs 11 and 12. The vertical dashed lines correspond to $\tilde{r}$ (which was defined above), and $\bar{\phi}_\delta(r)$ is expected to be constant for $r > \tilde{r}$. The solid lines are analytical fits by an expression of the form $a + b \exp[-c(r - d)^2]$. Within the accuracy of our calculations the parameters a, b, and c of this fit do not change with R, while the parameter d is roughly $R + \eta$. Clearly, with an increased accuracy we may eventually find some dependence of a, b, and c on R. Undoubtedly, the calculation of $\bar{\phi}_\delta(r)$ constitutes the most time-consuming step in applying our model to real proteins. It took approximately 24 h to obtain one value of $\bar{\phi}_\delta(r)$ (one point in Figure 3) by using Mathematica on a Dell computer with a 3 GHz Pentium 4 processor and 512 Mb of memory.

Figure 4 presents W^- and W^+, the emission and absorption rates, respectively, as functions of the cluster size R. The location of the intersection of these functions determines the size of the critical cluster, R_c. The emission rate is greater than the absorption rate, $W^-(r) > W^+(r)$, for clusters with $R < R_c$, whereas for clusters larger than the nucleus absorption dominates emission, $W^-(r) < W^+(r)$ for $R > R_c$. Note that both W^- and W^+ increase with increasing R, but W^- increases roughly linearly with R whereas W^+ rapidly increases by several orders of magnitude after the cluster becomes supercritical. This occurs because the width of the opw quickly decreases as the cluster grows, whereas the outer height of the barrier between the ipw and the opw virtually does not change. For this reason it becomes increasingly easier for a bead located in the opw to cross the barrier and to fall into the ipw.

The behavior of W^- and W^+ also explains our estimates for the characteristic times of the first nucleation event t_n, growth time t_g, and total folding time t_f provided by eqs 28, 29, and 31. Although these values depend strongly on R_c and on D^{iw} (R_c itself does not depend on D^{iw} but only on the ratio D^{ow}/D^{iw}), we have always $t_n \gg t_g$; i.e., the time of protein folding is mainly determined by the time necessary for the first nucleation event to occur. Physically, this happens because the increase of the cluster size from $\nu = 1$ to $\nu = \nu_c$ takes place only via

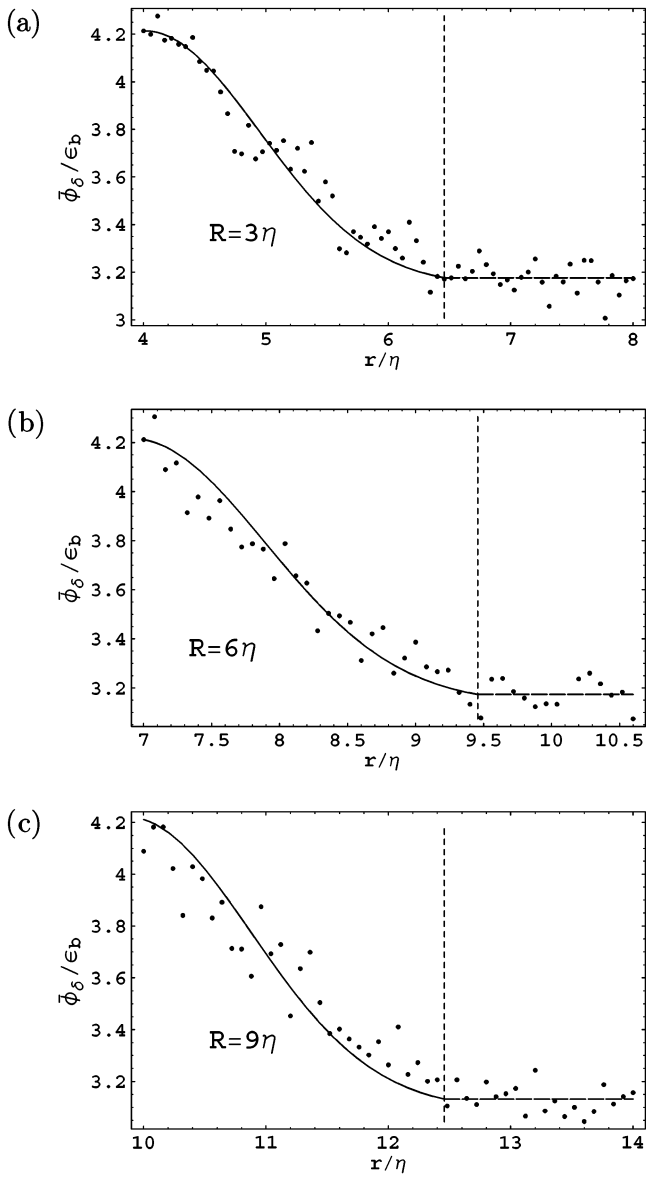

Figure 3. Average dihedral potential (assigned to a selected bead) $\bar{\phi}_\delta(r)$ as a function of r for three clusters of sizes (a) $R = 3\eta$, (b) $R = 6\eta$, and (c) $R = 9\eta$. The points represent the actual numerical results obtained by using Monte Carlo integration in eqs 11 and 12. The vertical dashed lines correspond to $\tilde{r}$. The solid lines are analytical fits by the expression $a + b \exp[-c(r - d)^2]$.

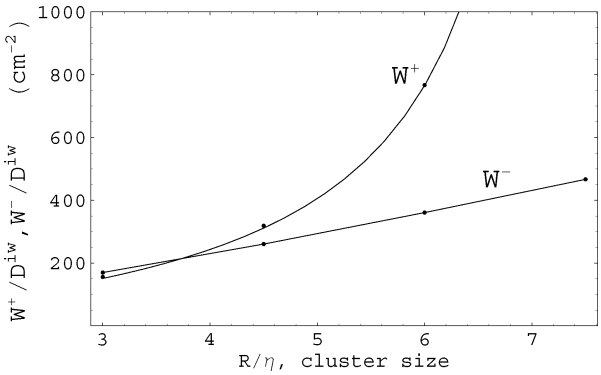

Figure 4. W^- and W^+, the emission and absorption rates, respectively, as functions of cluster size R. The location of the intersection of these functions determines the size of the critical cluster, R_c.

fluctuations that have to overcome the natural tendency of a small cluster to decay ($W^+ < W^-$ for $\nu < \nu_c$). For

supercritical clusters, W^+ overwhelms W^- so quickly that the fluctuations are unable to impede the natural tendency of the cluster to grow. (Strictly speaking, this is true only for $\nu > \nu_c + \Delta\nu_c$, but for rough estimates eq 31 is acceptable.) With the above choice of the system parameters and with D_{iw} in the range from 10^{-6} to 10^{-8} cm^2/s, for $N_0 = 2500$ our model predicts $R_c \simeq 3.7\eta$, $\nu_c \simeq 220$, and characteristic times for protein folding in the range from several seconds to several hundreds of seconds, which is in a good agreement with experimental data.[5]

5. Conclusions

So far most of the work on protein folding has been carried out by using either Monte Carlo or molecular dynamics simulations. A rigorous theoretical treatment of protein folding by means of statistical mechanics is hardly practicable because of the extreme complexity of the system. Simulations have suggested that there are multiple pathways for a protein to fold and one of them has been identified as reminiscent of nucleation. However, a theoretical model involving the nucleation mechanism was based on classical nucleation theory (CNT), hence on classical thermodynamics.[15,28] In that approach the size of the critical cluster (nucleus) was provided by the location of the maximum of the free energy of cluster formation as a function of the cluster size. In such a model the free energy of cluster formation depends on the surface tension of a cluster of protein residues. This quantity is an ill-defined physical quantity and can be considered only as an adjustable parameter. According to the nucleation mechanism, after the formation of a nucleus (critical size cluster of residues), the protein quickly reaches its native state.

In this paper we have presented a new, microscopic model for the nucleation mechanism of protein folding. A protein is considered as a heteropolymer consisting of two types of beads (hydrophobic and hydrophilic) linked with bonds of fixed length. All bond angles are also assumed to be fixed and equal to 105°. All nonadjacent beads are assumed to interact via a Lennard-Jones-like potential. Besides these interactions, the total energy of the heteropolymer contains a contribution from dihedral angles of all triads of successive links. Unlike the old model, the present one is developed without recurring to the CNT approach but is based on the above "molecular" interactions, both pairwise and configurational. The parameters of these potentials can be rigorously defined, unlike the ill-defined surface tension of a cluster of protein residues within a protein.

The main idea underlying the new model consists of averaging the dihedral potential in which a selected residue is involved over all possible configurations of neighboring residues. The resulting average dihedral potential depends on the distance between the residue and the cluster center. It has a maximum at the cluster surface and monotonically decreases with increasing distance therefrom. Its combination with the average potential due to pairwise interactions between the selected residue and those in the cluster and with a confining potential provides a double potential well around the cluster with a barrier between the two wells. Residues in the inner well are considered to belong to the cluster (part of the protein with correct tertiary contacts) while those in the outer well are treated as belonging to the mother phase (amorphous part of the protein with incorrect tertiary contacts). Transitions of residues from the inner well into the outer one and vice versa are considered as elementary emission and absorption events, respectively. The rates of these processes are determined by using a mean first passage time analysis. Once these rates are found as functions of the cluster size, one can develop a self-consistent kinetic theory for the nucleation mechanism of protein folding. For example, the size of the critical cluster (nucleus) is then found as the one for which these rates are equal. The time necessary for the protein to fold can be evaluated as the sum of the times necessary for the appearance of the first nucleus and the time necessary for the nucleus to grow to the maximum size (of the folded protein in the native state).

For numerical illustration we have considered a model protein consisting of 2500 beads with the mole fraction of hydrophobic beads equal to 0.75. The composition of the cluster during its formation and growth was assumed to be constant and equal to the composition of the whole protein. This allows one to consider the model as a single-component one. The size of the critical cluster and the folding time predicted by the model depend very much on the parameters of the system, such as the interaction parameters, densities of the protein in the unfolded (but compact) and folded states, diffusion coefficients therein, etc. With appropriate choices of interaction parameters and densities, the size of the critical cluster predicted by our model is approximately 220 residues, the free energy of nucleus formation being approximately $20k_BT$. These results suggest that the quantity equivalent to the "surface tension" in the old model of nucleation in a protein should be smaller than the previous estimates of the latter[15,28] by approximately 1 order of magnitude. Considering the diffusion coefficient of native residues in the range from 10^{-6} to 10^{-8} cm^2/s, the characteristic time of protein folding was estimated to be in the range from several seconds to several hundreds of seconds. These values are consistent with the experimental data[5] on typical folding times of proteins as well as with estimates provided by other theoretical models[28] and simulations.[7,15]

A further development of our model will require the removal of several simplifying assumptions that we have employed in the present paper. For example, it would be more appropriate to model a protein as a three-component heteropolymer (including not only hydrophobic and hydrophilic residues, but also neutral ones). This will result in longer numerical calculations of the average dihedral potential because the dihedral potential involving neutral beads is expected to be weaker and requires separate calculations. Next, the cluster composition during its formation and growth can quite significantly depend on the cluster size, particularly in the vicinity of the critical size; hence the assumption that it is constant can lead to inaccuracies in the results. By including neutral beads in the model and allowing the cluster composition to differ from that of the protein will result in a binary or even ternary nucleation mechanism of protein folding. However, besides some increase in the computational efforts there is no conceptual difficulty in developing the model in those directions.

Appendix A. Derivation of Eq 11 from Eq 9

Consider beads 1−7 of the unfolded part of the heteropolymer (Figure 1) around a cluster of radius R and denote the distance between bead 1 and the cluster center by r. It is convenient to choose a Cartesian system of coordinates with the origin in the cluster center in such a way that the coordinates of bead 1 are $x_1 = 0$, $y_1 = 0$, $z_1 = r$. The Cartesian coordinates of other beads will be denoted by x_i, y_i, z_i ($i = 2, ..., 7$). The Cartesian coordinates of bead 2 are related to its spherical ones r_2, Θ_2, φ_2

$$x_2 = r_2 \sin\Theta_2 \cos\varphi \quad y_2 = r_2 \sin\Theta_2 \sin\varphi \quad z_2 = r_2 \cos\Theta_2 \tag{A1}$$

At a given r (since the location of bead 1 is fixed), the polar angle Θ_2 of bead 2 is uniquely determined by r_2 due to the

constant bond length constraint

$$\Theta_2(r, r_2) = \arccos[(r^2 + r^2{}_2 - \eta^2)/(2r_2 r)] \quad (0 \le \Theta_2 \le \pi) \tag{A2}$$

whereas the azimuthal angle $0 \le \phi_2 \le 2\pi$. The distance r_2 varies in the range $r_{2\min} \le r_2 \le r + \eta$, where $r_{2\min} = \max(R, r - \eta)$, which determines the integration range L_2 in eqs 11 and 12. Thus the integration with respect to r_2 in eqs 9 and 10 reduces to the integration with respect to φ_2 and r_2 with fixed Θ_2 (given by eq A2) in eqs 11 and 12.

For given locations of beads 1 and 2, the possible locations of beads 3 and 4 lie on circles of radius $r_1 = \eta \sin\beta_0$ with their location and orientation completely determined by the coordinates of beads 1 and 2. This is due to the constraints that all bond angles are equal to β_0 and all bond lengths are equal to η. Due to the same constraints, if the locations of beads 2 and 4 are given, then the possible locations of bead 6 lie on a circle of radius r_1, whereof the location and orientation are completely determined by the coordinates of beads 2 and 4. Further, for given locations of beads 1 and 3, the possible locations of bead 5 lie on a circle of radius r_1 with the position and orientation completely determined by the coordinates of beads 1 and 3. Finally, for the given locations of beads 3 and 5, the possible locations of bead 7 are on a circle of radius r_1, with the location and orientation completely determined by the coordinates of beads 3 and 5.

Let us consider bead s with unknown coordinates and two other beads, c and n (closest to and next to the closest to bead s), with fixed (known) coordinates x_c,y_c,z_c and x_n,y_n,z_n. For example, if $s = 7$, then $c = 5$ and $n = 3$; if $s = 4$, then $c = 2$ and $n = 1$. Bead s lies on a circle of radius r_1 with the coordinates of the center

$$x_0 \equiv x_0(x_n,y_n,z_n, x_c,y_c,z_c) = x_n + (x_c - x_n)\frac{\eta(1 + |\cos\beta_0|)}{\sqrt{(x_c - x_n)^2 + (y_c - y_n)^2 + (z_c - z_n)^2}} \tag{A3}$$

$$y_0 \equiv y_0(x_n,y_n,z_n, x_c,y_c,z_c) = y_n + (y_c - y_n)\frac{\eta(1 + |\cos\beta_0|)}{\sqrt{(x_c - x_n)^2 + (y_c - y_n)^2 + (z_c - z_n)^2}} \tag{A4}$$

$$z_0 \equiv z_0(x_n,y_n,z_n, x_c,y_c,z_c) = z_n + (z_c - z_n)\frac{\eta(1 + |\cos\beta_0|)}{\sqrt{(x_c - x_n)^2 + (y_c - y_n)^2 + (z_c - z_n)^2}} \tag{A5}$$

The coordinate x_s of bead s can change in the range

$$x_c - x_b \le x_s \le x_c + x_b \tag{A6}$$

where

$$x_b \equiv x_b(x_n,y_n,z_n, x_c,y_c,z_c) = \eta \sin\beta_0 \sqrt{\frac{((y_c - y_0)^2 + (z_c - z_0)^2)}{(x_c - x_n)^2 + (y_c - y_n)^2 + (z_c - z_n)^2}} \tag{A7}$$

For a given x_s, the coordinate y_s of bead s can have only one of two values

$$y_s^{\pm} = y_s^{\pm}(x_s, x_n,y_n,z_n, x_c,y_c,z_c) = y_0 + \frac{-(x_c - x_0)(y_c - y_0)(x_s - x_0)}{(y_c - y_0)^2 + (z_c - z_0)^2} \pm \frac{|z_c - z_0|\sqrt{[(y_c - y_0)^2 + (z_c - z_0)^2]\eta^2 \sin^2\beta_0 - [(x_c - x_n)^2 + (y_c - y_n)^2 + (z_c - z_n)^2](x_s - x_0)^2}}{(y_c - y_0)^2 + (z_c - z_0)^2} \tag{A8}$$

For given x_s and $y_s^{\pm}$, the coordinate z_s of bead s can have only a single value

$$z_s^{\pm} \equiv z_s(x_s, y_s^{\pm}, x_n,y_n,z_n, x_c,y_c,z_c) = z_0 - \frac{(x_c - x_0)(x_s - x_0) + (y_c - y_0)(y_s^{\pm} - y_0)}{z_c - z_0} \tag{A9}$$

Since the volume occupied by the cluster is unavailable to beads 2, 3, ..., 7, an additional constraint is imposed on x_s, y_s, and z_s, namely, $x_s^2 + y_s^2 + z_s^2 > R^2$. Subject to this constraint, the double inequality in eq A6 and eqs A7−A9 completely determine the integration ranges in eqs 11 and 12. To transform the function $\phi'_\delta(r) \equiv \phi'_\delta(r,\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4,\mathbf{r}_5,\mathbf{r}_6,\mathbf{r}_7)$ into the function $\tilde{\phi}_{ijkmn}(r,\varphi_2,r_2,x_3,...,x_7)$, the variables y_s ($s = 3, ..., 7$) and z_s ($s = 3, ..., 7$) in the former must be replaced by $y_s^{\pm}$ and $z_s^{\pm}$ in all possible combinations, each of which gives rise to a term in the sums in eqs 11 and 12. (Note that $z_s^+ = z_s(x_s, y_s^+, ...)$ and $z_s^- = z_s(x_s, y_s^-, ...)$.) The polar angle of bead 2 must be replaced by $\Theta_2(r,r_2)$.

For example, consider the term with $i = +$, $j = -$, $k = +$, $m = +$, $n = -$ in the sum in eqs 11 and 12. In the corresponding integrands, the function $\tilde{\phi}_{+-++-}(\varphi_2, r_2, x_3, ..., x_7)$ is obtained from $\phi'_\delta(r, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4, \mathbf{r}_5, \mathbf{r}_6, \mathbf{r}_7)$ as follows:

y_3 must be replaced by $y_3^+ \equiv y_3^+(x_3, x_2, y_2, z_2, x_1, y_1, z_1)$ (defined by eq A8) and z_3 by $z_3^+ \equiv z_3(x_3, y_3^+, x_2, y_2, z_2, x_1, y_1, z_1)$ (defined by eq A9), y_4 must be replaced by $y_4^- \equiv y_4^-(x_4, x_1, y_1, z_1, x_2, y_2, z_2)$ and z_4 by $z_4^- \equiv z_4(x_4, y_4^-, x_1, y_1, z_1, x_2, y_2, z_2)$, y_5 must be replaced by $y_5^+ \equiv y_5^+(x_5, x_1, y_1, z_1, x_3, y_3^+, z_3^+)$ and z_5 by $z_5^+ \equiv z_5(x_5, y_5^+, x_1, y_1, z_1, x_3, y_3^+, z_3^+)$, y_6 must be replaced by $y_6^+ \equiv y_6^+(x_6, x_2, y_2, z_2, x_4, y_4^-, z_4^-)$ and z_6 by $z_6^+ \equiv z_6(x_6, y_6^+, x_2, y_2, z_2, x_4, y_4^-, z_4^-)$, y_7 must be replaced by $y_7^- \equiv y_7^-(x_7, x_3, y_3^+, z_3^+, x_5, y_5^+, z_5^+)$ and z_7 by $z_7^- \equiv z_7(x_7, y_7^-, x_3, y_3^+, z_3^+, x_5, y_5^+, z_5^+)$. The coordinates x_2, y_2, and z_2 are functions of r_2 and φ_2 provided by eqs A1 and A2.

Appendix B. Derivation of Expressions for the Emission and Absorption Rates by a Mean First Passage Time Analysis

Let us use the mean first passage time analysis[30−35] to calculate the emission and absorption rates of a cluster during protein folding. The mean first passage time of a bead escaping from a potential well is calculated on the basis of a kinetic equation governing the chaotic motion of the bead in that potential well. The chaotic motion of the bead is assumed to be governed by the Fokker−Planck equation for the single-particle distribution function with respect to its coordinates and momenta, i.e., in the phase space.[51−53] Prior to the passage event, the evolution of a bead in both the ipw and the opw occurs in a dense enough medium (cluster of folded residues or unfolded but compact part of the protein), where the relaxation time for its velocity distribution function is very short and negligible compared to the characteristic time scale of the passage process.

Under these conditions, the Fokker−Planck equation reduces to the Smoluchowski equation, which involves diffusion in an external field.[52,53] In the case of spherical symmetry it can be written in the form[30−32]

$$\frac{\partial p_i(r,t|r_0)}{\partial t} = D_i r^{-2} \frac{\partial}{\partial r}\left(r^2 \, \mathrm{e}^{-\Psi_i(r)} \frac{\partial}{\partial r} \mathrm{e}^{\Psi_i(r)} p_i(r,t|r_0) \right) \quad \text{(B1)}$$

where $p_i(r,t|r_0)$ is the probability of observing a bead of species i (i = b, l) between r and $r + \mathrm{d}r$ at time t given that initially it was at a radial distance r_0, D_i is its diffusion coefficient in the well, and $\Psi_i(r) = \psi_i(r)/kT$.

The mean passage time depends on the initial position (distance from the center of the cluster) r_0 of the bead. It is convenient to use the backward Smoluchowski equation,[30−32] which expresses the dependence of the transition probability $p_i(r,t|r_0)$ on r_0

$$\frac{\partial p_i(r,t|r_0)}{\partial t} = D_i {r_0}^{-2} \, \mathrm{e}^{\Psi_i(r_0)} \frac{\partial}{\partial r_0}\left({r_0}^2 \, \mathrm{e}^{-\Psi_i(r_0)} \frac{\partial}{\partial r_0} p_i(r,t|r_0) \right) \quad \text{(B2)}$$

Let us first consider the ipw and find the emission rate of beads therefrom. The probability that a bead i, initially at a distance r_0 within the surface layer, will remain in this region after time t is given by the so-called survival probability[30−32]

$$q_i(t|r_0) = \int_R^{R+\lambda_i^{\mathrm{w}}} \mathrm{d}r \, r^2 p_i(r, t|r_0) \quad (R < r_0 < R + \lambda_i^{\mathrm{w}}) \quad \text{(B3)}$$

where R is the radius of the cluster and λ_i^{w} is the width of the potential well determined by the location of the barrier between the ipw and the opw (Figure 2). (Previously,[30−35] λ_i^{w} was defined so that a bead i could be assumed to be dissociated from the cluster at the distance $R + \lambda_i^{\mathrm{w}}$ from its center. The width λ_i^{w} had to be chosen such that $\lambda_i^{\mathrm{w}}/R \ll 1$ and $\Psi_i(R + \lambda_i^{\mathrm{w}})/\Psi_i^{\mathrm{w}} \ll 1$, where Ψ_i^{w} was the depth of the well.) The probability for the dissociation time to be between 0 and t is equal to $1 - q_i(t|r_0)$, and the probability density for the dissociation time is given by $-\partial q_i/\partial t$. The first passage time is provided by[51−53]

$$\tau_i(r_0) = -\int_0^\infty t \frac{\partial q_i(t|r_0)}{\partial t} \mathrm{d}t = \int_0^\infty q_i(t|r_0) \, \mathrm{d}t \quad \text{(B4)}$$

The equation for the first passage time is obtained by integrating the backward Smoluchowski eq B2 with respect to r and t over the entire range and using the boundary conditions $q_i(0|r_0) = 1$ and $q_i(t|r_0) \to 0$ as $t \to \infty$ for any r_0. This yields[30−32]

$$-D_i {r_0}^{-2} \, \mathrm{e}^{\Psi(r_0)} \frac{\partial}{\partial r_0}\left({r_0}^2 \, \mathrm{e}^{-\Psi_i(r_0)} \frac{\partial}{\partial r_0} \tau_i(r_0) \right) = 1 \quad \text{(B5)}$$

One can solve eq B5 by assuming a reflecting inner boundary of the well ($\mathrm{d}\tau_i/\mathrm{d}r_0 = 0$ at $r_0 = R$) and the radiation boundary condition at the outer boundary ($\tau_i(r_0) = 0$ at $r_0 = R + \lambda_i^{\mathrm{w}}$). One thus obtains for the first passage time

$$\tau_i(r_0) = \frac{1}{D_i} \int_{r_0}^{R+\lambda_i^{\mathrm{w}}} \mathrm{d}y \, y^{-2} \, \mathrm{e}^{\Psi_i(y)} \int_R^y \mathrm{d}x \, x^2 \, \mathrm{e}^{-\Psi_i(x)} \quad \text{(B6)}$$

The average dissociation time, $\bar{\tau}_i$, or the mean first passage time, is obtained by averaging $\tau_i(r_0)$ with the Boltzmann factor over all possible initial positions r_0

$$\bar{\tau}_i = \frac{1}{Z} \int_R^{R+\lambda_i^{\mathrm{w}}} \mathrm{d}r_0 \, {r_0}^2 \, \mathrm{e}^{-\Psi_i(r_0)} \, \tau_i(r_0) \quad \text{(B7)}$$

with

$$Z = \int_R^{R+\lambda_i^{\mathrm{w}}} \mathrm{d}r_0 \, {r_0}^2 \, \mathrm{e}^{-\Psi_i(r_0)} \quad \text{(B8)}$$

Defining the quantity ω_i as

$$\omega_i \equiv 1/(D_i \bar{\tau}_i) \quad \text{(B9)}$$

the emission rate, W_i^-, can be expressed as

$$W_i^- = \frac{N_i^{\mathrm{w}}}{\bar{\tau}_i} = N_i^{\mathrm{w}} D_i w_i \quad \text{(B10)}$$

where N_i^{w} denotes the number of molecules in the well. Clearly, $\bar{\tau}_i$, ω_i, and W_i^- are functions of R and $\chi = \nu_{\mathrm{b}}/(\nu_{\mathrm{b}} + \nu_{\mathrm{l}})$, where ν_i (i = b, l) is the number of beads of type i in the cluster.

Considering the opw, a similar procedure can be used to find the rate of absorption of beads by the cluster, W_i^+.

Acknowledgment. This work was supported by the National Science Foundation through Grant No. CTS-0000548.

References and Notes

(1) Creighton, T. E. *Proteins: Structure and Molecular Properties*; W. H. Freeman: San Francisco, 1984.
(2) Stryer, L. *Biochemistry*, 3rd ed.; W.H. Freeman: New York, 1988.
(3) Ghelis, C.; Yan, J. *Protein Folding*; Academic Press: New York, 1982.
(4) Anfinsen, C. B. *Science* **1973**, *181*, 223−230.
(5) Nölting, B. *Protein Folding Kinetics: Biophysical Methods*, 2nd ed.; Springer-Verlag: Berlin, 2006.
(6) Honeycutt, J. D.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 3526−3529.
(7) Honeycutt, J. D.; Thirumalai, D. *Biopolymers* **1992**, *32*, 695−709.
(8) Weissman, J. S.; Kim, P. S. *Science* **1991**, *253*, 1386−1393.
(9) Creighton, T. E. *Nature (London)* **1992**, *356*, 194−195.
(10) Kim, P. S.; Baldwin, R. I. *Annu. Rev. Biochem.* **1982**, *51*, 459.
(11) Kim, P. S.; Baldwin, R. I. *Annu. Rev. Biochem.* **1990**, *59*, 631.
(12) Creighton, T. E. *Biochem. J.* **1990**, *240*, 1.
(13) Creighton, T. E. *Prog. Biophys. Mol. Biol.* **1978**, *33*, 231.
(14) Guo, Z.; Thirumalai, D.; Honeycutt, J. D. *J. Chem. Phys.* **1992**, *97*, 525−535.
(15) Guo, Z.; Thirumalai, D. *Biopolymers* **1995**, *36*, 83−102.
(16) Thirumalai, D.; Guo, Z. *Biopolymers* **1995**, *35*, 137−140.
(17) Nölting, B.; Golbik, R.; Neira, J .L.; Soler-Gonzalez, A. S; Schreiber, G.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *84*, 826.
(18) Nölting, B. *J. Theor. Biol.* **1998**, *194*, 419.
(19) Nölting, B. *J. Theor. Biol.* **1999**, *197*, 113.
(20) Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10869.
(21) Fersht, A. R. *Cur. Opin. Struct. Biol.* **1997**, *7*, 3.
(22) Shakhnovich, E. I. *Cur. Opin. Struct. Biol.* **1997**, *7*, 29.
(23) Shakhnovich, E. I.; Abkevich, V.; Ptitsyn, O. *Nature* **1996**, *379*, 96.
(24) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. *J. Chem. Phys.* **1994**, *101*, 6052.
(25) Finkelstein, A. V.; Badretdinov, A. Y. *Folding Des.* **1997**, *2*, 115.
(26) (a) Bryngelson, J. D.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7524. (b) Bryngelson, J. D.; Wolynes, P. G. *J. Phys. Chem.* **1998**, *93*, 6902.
(27) (a) Shakhnovich, E. I.; Gutin, A. M. *Nature* **1990**, *346*, 773. (b) Shakhnovich, E. I.; Gutin, A. M. *J. Phys. A: Math. Gen.* **1989**, *22*, 1647.
(28) Bryngelson, J. D.; Wolynes, P. G. *Biopolymers* **1990**, *30*, 177.
(29) (a) Dill, K. A. *Biochemistry* **1985**, *24*, 1501. (b) Dill, K. A. *Biochemistry* **1990**, *29*, 7133.
(30) Narsimhan, G.; Ruckenstein, E. *J. Colloid Interface Sci.* **1989**, *128*, 549.
(31) Ruckenstein, E.; Nowakowski, B. *J. Colloid Interface Sci.* **1990**, *137*, 583.
(32) Nowakowski, B.; Ruckenstein, E. *J. Colloid Interface Sci.* **1990**, *139*, 500.
(33) Djikaev, Y. S.; Ruckenstein, E. *J. Chem. Phys.* **2005**, *123*, 214503.
(34) Djikaev, Y. S.; Ruckenstein, E. *J. Chem. Phys.* **2006**, *124*, 124521.
(35) Djikaev, Y. S.; Ruckenstein, E. *J. Chem. Phys.* **2006**, *124*, 194709.
(36) Levitt, M.; Sharon, R. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 8557.

(37) Skolnick, J.; Kolinski, A.; Yaris, R. *Biopolymers* **1989**, *28*, 1059.
(38) Kolinski, A.; Skolnick, J.; Yaris, R. *Biopolymers* **1987**, *26*, 937.
(39) (a) Sikorski, A.; Skolnick, J. *Biopolymers* **1989**, *28*, 1097. (b) Sikorski, A.; Skolnick, J. *J. Mol. Biol.* **1990**, *215*, 183.
(40) Skolnick, J.; Kolinski, A. *Science* **1990**, *250*, 1121.
(41) Levinthal, C. In *Mossbauer Spectroscopy in Biological Systems*; Debrunner, P., Tsibris, J. C. M., Mönck, E., Eds.; University of Illinois Press: Urbana, IL, 1968.
(42) Wetlaufer, D. *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 697.
(43) Tsong, T. Y.; Baldwin, R.; McPhie, P. *J. Mol. Biol.* **1972**, *63*, 453.
(44) Moult, J.; Unger, R. *Biochemistry* **1991**, *30*, 3816.
(45) Dill, K.; Fiebig, K.; Chan, H. S. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 1942.
(46) Flageollet, C.; Dihn Cao, M.; Mirabel, P. *J. Chem. Phys.* **1980**, *72*, 544.
(47) (a) Wilemski, G. *J. Phys. Chem.* **1987**, *91*, 2492. (b) Wilemski, G. *J. Chem. Phys.* **1984**, *80*, 1370.
(48) Laaksonen, A. *J. Chem. Phys.* **1992**, *97*, 1983.
(49) Djikaev, Y. S.; Napari, I.; Laaksonen, A. *J. Chem. Phys.* **2004**, *97*, 9752.
(50) Abraham, F. F. *Homogeneous Nucleation Theory*; Academic Press: New York, 1974.
(51) Chandrasekhar, S. *Rev. Mod. Phys.* **1949**, *15*, 1.
(52) Gardiner, C. W. *Handbook of Stochastic Methods*; Springer Series in Synergetics 13; Springer: New York, 1983.
(53) Agmon, N. *J. Chem. Phys.* **1984**, *81*, 3644.
(54) Lothe, J.; Pound, G. M. J. In *Nucleation*; Zettlemoyer, A. C., Ed.; Marcel-Dekker: New York, 1969.
(55) *Nucleation Theory and Applications*; Schmelzer, J. W. P., Röpke, G., Priezzhev, V. B., Eds.; Joint Institute for Nuclear Research: Dubna, Russia, 1999.
(56) Kashchiev, D. *Nucleation: Basic Theory with Applications*; Butterworth-Heinemann: Oxford, U. K., 2000.
(57) Harpaz, Y.; Gerstein, M.; Chothia, C. *Structure* **1994**, *2*, 641−649.
(58) Huang, D. M.; Chandler, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *15*, 8324−8327.
(59) (a) Hirshfelder, J. O.; Curtiss, C. F.; Bird, R. B. *Molecular Theory of Gases and Liquids*; Wiley: New York, 1964. (b) Wakeham, W. A. *J. Phys. B: At. Mol. Phys.* **1973**, *6*, 372.