Exploring the Interplay between Topology and Secondary Structural Formation in the Protein Folding Problem

Margaret S. Cheung,[†] John M. Finke,[†] Benjamin Callahan,[‡] and José N. Onuchic*,[†]

The Center for Theoretical Biological Physics and Department of Physics, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0319, and Department of Physics, Iowa State University, Ames, Iowa 50011-3160

Received: February 20, 2003; In Final Form: July 16, 2003

Simple models with a single bead representation (C_α models) have been successful in providing a qualitative understanding of the folding mechanism of small globular proteins. Can we go beyond this qualitative understanding and make more detailed quantitative connections to experiments? To achieve this goal, a tractable framework of protein representations whose complexity falls between C_α and all-atom representations is needed to address different energetic competing factors during folding events. Such a model conserves the low computational expense inherent in minimalist models while enhancing the understanding of side-chain packing not existent in simple C_α models. In this work, we present a minimalist representation of protein structures that are used to investigate the competition between native side-chain contacts and nonspecific backbone hydrogen bonds. Our results suggest that native tertiary contacts and dihedrals force the nonspecific hydrogen bonds to adopt native configurations and retain a funneled landscape. In addition, the use of an angular component in the hydrogen bond interaction prevents non-native conformations.

I. Introduction

The protein folding problem has been a challenge for both experimentalists and theoreticians. Several explanations of the underlying mechanisms for folding dynamics have been proposed.[1−10] Because most small and intermediate size globular proteins have sufficiently small energetic frustration,[2] their energy landscapes are funnel-like,[11] and protein topology strongly influences what are the structured regions in the folding transition state ensemble. These conclusions are in good agreement with mutagenetic analyses[12−14] and the distribution of kinetics rates.[15] These experimental observations have been successfully explained by energy landscape theory[16] and the folding funnel concept.[11] Computational studies using topology-based off-lattice molecular models with low resolution (C_α models)[17−26] capture general structural features of the folding landscape. Several recent theoretical studies using this model have successfully provided an overall agreement with experimental results, addressing issues as experimental ϕ values,[27] effects of circular permutation,[28] the distribution of kinetic rates,[29] and the desolvation of the hydrophobic core.[30] C_α models, however, may be too simple to address higher resolution details such as the steric effects of side-chains and backbone hydrogen bonding.

Side-chains play nontrivial roles in determining protein thermostability,[31−33] protein−protein interactions,[34] designability of de novo proteins,[33,35] and protein recognition.[36,37] To address their importance in the protein folding problem, we have developed a protein representation that introduces a single bead (C_β) to represent each side-chain for nonglycine amino acids. This model increases the molecular complexity without compromising computational costs. Reduced representations of proteins, with two to six beads for each amino acid, have been used by many research groups.[38−44] Particularly, simplified models using two-beads per residue ($C_\alpha-C_\beta$ or backbone-side-chain model) were first introduced by Wolynes and co-workers[39] and Klimov and Thirumalai.[40] In addition, although many details are included in fully atomistic models, an interpretation of simulation data toward an understanding of basic mechanisms can be a very complex. Minimalist models simplify many of the molecular complexities inherent in protein simulations, thereby emphasizing the essential physics we are interested in uncovering. As the level of complexity of these models increases, determining the energetic interactions between these beads becomes nontrivial as the involvement of many-body effects is pertinent in the energy function (e.g., chiral atomic orientations, or hydrogen bonding).[39,43,45]

Backbone hydrogen bond formation is relevant to protein folding, protein stability, and ligand binding.[46,47] The role of backbone hydrogen bonding in protein folding, however, has not been fully investigated theoretically and computationally. In this study, side-chain beads interact with pairwise contact energies, whereas main chain beads follow a many-body interaction involving both distance and angular[48] terms, designed to capture the essential physics of hydrogen bonding.

In this study, we chose chymotrypsin inibitor 2 (CI2) as a protein model because it has been extensively studied experimentally and computationally. In addition, this 65-residue protein contains three different types of secondary structures (α helix, parallel, and antiparallel β sheets), making it a system that includes all three kinds of secondary structures. Initially a Gō-like Hamiltonian, allowing up to two native backbone hydrogen bonds per C_α bead is investigated. Subsequently, backbone hydrogen bonds are modulated with an angular-dependent term to restrict hydrogen bond orientations. This treatment forces competition among C_α pairs. Only those pairs that have appropriate angular alignment are favored. Combined with native side-chain interactions, angular-dependent backbone hydrogen bonding avoids non-native traps and reinforces the funnel-like nature of the folding landscape.

* To whom correspondence should be addressed.
[†] University of California.
[‡] Iowa State University.

10.1021/jp034441r CCC: $25.00
Published on Web 09/12/2003

TABLE 1 [a]

	models	E_s	E_{chiral}	$E_{native}^{C\alpha^{pairs}}$	$E_{native}^{(C\alpha^{pairs})^c}$	$E_{non-native}^{C\alpha^{pairs}}$	$E_{non-native}^{(C\alpha^{pairs})^c}$
(a)	basic	v	x	LJ	LJ	Rep	Rep
(b)	Chiral-B0	v	v	LJ	LJ	Rep	Rep
(c)	Chiral-B0-ns	v	v	LJ	LJ	LJ	Rep
(d)	Chiral-B1	v	v	HB	LJ	Rep	Rep
(e)	Chiral-B10	v	v	HB	LJ	Rep	Rep
(f)	Chiral-B1-ns	v	v	HB	LJ	HB	Rep
(g)	Chiral-B10-ns	v	v	HB	LJ	HB	Rep

[a] E_s is the structural Hamiltonian that includes the bond length potential, bond angle potential, dihedral angle potential. E_{chiral} is the chiral potential. $E_{native}^{C\alpha^{pairs}}$ is the energy function for native C_α pair contacts. $E_{native}^{(C\alpha^{pairs})^c}$ is the energy function for native contacts that are not native C_α pairs. $E_{non-native}^{C\alpha^{pairs}}$ is the energy function for non-native C_α pair contacts. $E_{non-native}^{(C\alpha^{pairs})^c}$ is the energy function for non-native contacts that are not C_α pairs. "v" represents that the energy function is included, whereas "x" represents otherwise. "LJ/HB/Rep" represents that the Lennard-Jones/angular-dependent/repulsive potential is used for the corresponding contact pairs.

II. Model and Methods

A. Chain Representation. To investigate the interplay between backbone formation and tertiary contact formation, we use an off-lattice model that consists of only two beads per amino acid to describe the structure of CI2. The first bead occupies the position of the α carbon (C_α) of each amino acid to approximate the backbone structure, and the second (C_β) occupies the center of mass of the side-chain for all amino acids (except glycine). It is referred as the "$C_\alpha-C_\beta$" model in this study, although the "C_β" bead is not at the genuine β carbon position of an amino acid.

B. Hamiltonian. In this study, we attempt to discriminate the differing physical interactions between side-chain and backbone beads. The C_α carbons of real amino acids favor L-isomers, a chiral restraint unresolved by minimalist C_α models. Side-chain interactions are pairwise and are designed to provide a minimally frustrated energy landscape; that is, the folding dynamics of the side-chains is mainly driven by a topology-based potential represented with a Gō-like[49] energy function.

For backbone interactions, we propose an energy function that addresses the angular behavior of backbone hydrogen bonding (HB) by incorporating many-body structural alignments into the contact formation. We start from a simple and intuitive $C_\alpha-C_\beta$ Gō-like Hamiltonian, where only "native" backbone pairs are assigned. We continue by adding new terms, such as chiral restraints and an angular-dependent hydrogen bonding term. Finally, we adjust the backbone hydrogen bonds to become less Gō-like by allowing any distant C_α pairs to interact (we make them to capture the "nonspecific" feature in the backbone hydrogen bond formation). The following models are investigated (details are in Table 1): (a) a "Basic" model lacking constraints on chirality and hydrogen-bond angular dependence; (b) a "Chiral-B0" model that imposes constraints on chirality but it has no angular dependence in the hydrogen bond formation; (c) a "Chiral-B0-nsHB" model that is identical to model b, but which also allows nonspecific C_α, C_α interactions; (d) a "Chiral-B1" model that imposes constraints on chirality and includes a weak angular dependence in the backbone hydrogen bond formation; (e) a "Chiral-B10" model that is identical to model d, but with a stronger angular dependence in the backbone hydrogen bond formation; (f) a "Chiral-B1-nsHB" model that is identical to model d, but which also allows nonspecific interactions between the backbone hydrogen bonds (no sequence information); (g) a "Chiral-B10-nsHB" that is identical to model e, but which also allows nonspecific interactions between the backbone hydrogen bonds.

The structural Hamiltonian E_s is composed of the following terms for each model:

$$E_s = \sum E_{bond} + \sum E_{angle} + \sum E_{dihedral} \quad (1)$$

(a). Basic

$$E = E_s + \sum_{native} E_{nonbond}^{LJ} + \sum_{non\text{-}native} E_{nonbond}^{Rep} \quad (2)$$

(b). Chiral-B0

$$E = E(\text{Basic}) + \sum E_{chiral} \quad (3)$$

(c). Chiral-B0-nsHB

$$E = E_s + \sum E_{chiral} + \sum_{native} E_{nonbond}^{LJ} + \sum_{non\text{-}native}^{(C_\alpha pairs)} E_{nonbond}^{LJ} + \sum_{non\text{-}native}^{(C_\alpha pairs)^c} E_{nonbond}^{Rep} \quad (4)$$

(d)/(e). Chiral-B1/Chiral-B10

$$E = E_s + \sum E_{chiral} + \sum_{native}^{(C_\alpha pairs)} E_{HB} + \sum_{native}^{(C_\alpha pairs)^c} E_{nonbond}^{LJ} + \sum_{non\text{-}native} E_{nonbond}^{Rep} \quad (5)$$

(f)/(g). Chiral-B1-nsHB/Chiral-B10-nsHB

$$E = E_s + \sum E_{chiral} + \sum^{(C_\alpha pairs)} E_{HB} + \sum_{native}^{(C_\alpha pairs)^c} E_{nonbond}^{LJ} + \sum_{non\text{-}native}^{(C_\alpha pairs)^c} E_{nonbond}^{Rep} \quad (6)$$

E_{bond}, E_{angle}, $E_{dihedral}$, and E_{chiral} are bond length, bond angle, dihedral angle, and chiral potentials, respectively. $E_{nonbond}^{LJ}$ and $E_{nonbond}^{Rep}$ are nonbonded potentials, which use either a type of Lennard-Jones (LJ) or a hard-core (Rep) energy function for contacts that can be either "native" or "non-native" pairs. The notation of "(C_α pairs)" at the top of sums refers to the set of distant nonbonded C_α pairs, whereas "(C_α pairs)^c" refers to all other nonbonded contacts that are not C_α pairs. E_{HB} is the angular-dependent potential for the backbone hydrogen bonds. An index of summation can go over a set of native C_α pairs or over all of the C_α pairs, depending on the choice of model.

1. Nonbonded Interactions using a Lennard-Jones Potential ($E_{nonbond}^{LJ}$). A Lennard-Jones potential is used for nonbonded interactions of "C_β, C_β" or "C_α, C_α" pairs between residues separated by at least three amino acids along the chain

$$E_{nonbond}^{LJ} = \sum_{|i-j|>3} \epsilon\left[5\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10}\right] \quad (7)$$

For the native pairs, σ_{ij} is the distance in the native configuration. For models which allow non-native C_α, C_α contacts, σ_{ij} is set to 4.65 Å. ϵ is the solvent averaged energy.

2. Nonbonded Interactions using a Hard-Core Potential ($E_{nonbond}^{Rep}$). A hard-core potential is assigned to represent excluded volume effects. Sizes of the excluded volume are

determined by the radii of interacting beads: the radius of C_α bead is 1.9 Å; the radius of C_β bead is the distance between this C_β and its bonding C_α, whose sizes vary by the types of side-chains. As a consequence, excluded volume effects can be modulated by sizes of beads, reflected in the potential below

$$E_{\text{nonbond}}^{\text{Rep}} = \sum_{i,j} \epsilon \left(\frac{f\sigma_{ij}}{r_{ij}}\right)^{12} \tag{8}$$

where i and j are indices of beads. $\sigma_{ij} = r_i^0 + r_j^0$, where r_i^0 and r_j^0 are the radii of interacting beads. To avoid local clashes, nearby repulsive contacts are not allowed and only medium/long-range indices are allowed for repulsive contacts: $|i - j| \geq 2$ for non-native C_α^i and C_β^j pairs; $|i - j| \geq 4$ for non-native C_α^i and C_α^j pairs; $|i - j| \geq 2$ for non-native C_β^i and C_β^j pairs. $f = 0.7$ is the scaling factor to avoid clashes that may destabilize the native state.

3. Angular-Dependent Potential for Backbone Hydrogen Bonds E_{HB}. To model the directional behavior of backbone hydrogen bond (HB) formation, an energy function must be capable of capturing three important features: (1) Hydrogen bonding must allow the three primary types of secondary structures. (2) Only a limited number of bonding partners should be allowed for a given C_α bead because few hydrogen bonds are observed per residue in proteins. (3) Because, in principle, hydrogen bond interactions do not depend on the type of amino acid, the model must also accommodate "nonspecific" hydrogen bonds between backbone C_α's.

We propose an energy function for backbone hydrogen bond formation that is capable of satisfying these criteria. This energy function is composed of a product of two terms: a Lennard-Jones(LJ) radial function and an angular-dependent function. This energy function takes the structural alignment (i.e., many-body effects) into account during contact formation. The allowed orientations of a pairwise HB interaction is energetically limited, substantially reducing the number of conformations between any pair of interacting C_α's

$$E_{\text{HB}} = \sum_{\text{HB};|i\text{-}j| > 3} U(r_{ij}) A(\rho_{i-1,\,i,\,j,\,j+1}) \tag{9}$$

$U(r_{ij})$ is the LJ energy function that has the same form as eq 7, where r_{ij} is the interacting distance between C_α^i and C_α^j in Figure 2; σ_{ij} is set to σ_{ij}^0 for native hydrogen bonds and σ_{ij} is set to 4.65 Å for the non-native ones. Rules for defining native hydrogen bonds are set by Vieth et al.[38] For Gō models, each C_α can form at most two native hydrogen bonds. $A(\rho_{i-1,\,i,\,j,\,j+1})$ is a function incorporating hydrogen bond alignments, described by a parameter ρ. It is defined by four C_α atoms (two C_α's on each side of the backbone) as shown in Figure 2

$$A(\rho) = \frac{1}{1 + B\left[(1 + \cos\rho)(1 - \cos\rho)\left(1 - \frac{\cos\rho}{\cos\rho_\alpha}\right)\right]^2} \tag{10}$$

$A(\rho)$ follows a sigmoid saturation curve. It approaches 1 if $\rho = 0$, $\rho = \pi$, or $\rho = \pm\rho_\alpha$; otherwise it is 0. In other words, it favors interactions when the local orientation of the backbone are in configurations favoring parallel β sheets, antiparallel β sheets, or left and right-handed α helices. The angular dependence is modulated by a factor B in the denominator that tunes the steepness of the function. Two values of B are chosen to investigate the relative influence of backbone hydrogen bond stability: $B = 1$ for low angular dependence and $B = 10$ for high angular dependence.

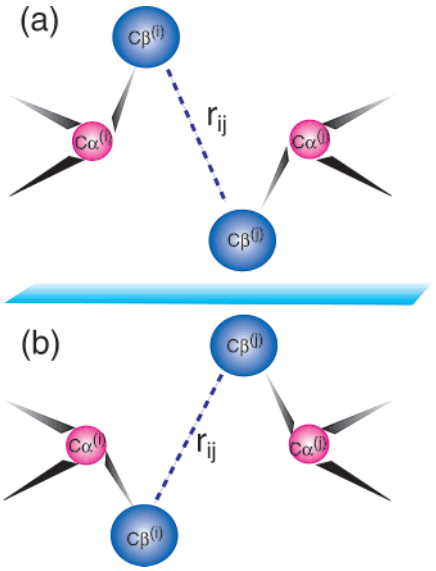

Figure 1. Schematic representation showing that interactions of residues whose side-chains (i.e., C_β^i, C_β^j) allow chiral isomers are energetically indistinguishable.

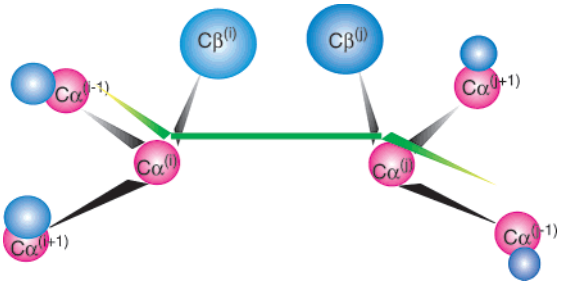

Figure 2. Schematic representation of the "pseudo dihedral" angle (defined by four C_α beads shown in green) that describes the alignment of paired backbone C_α's. This angle (whose vectors are defined by two neighboring C_α atoms at each side of the backbone and by a nonbonded C_α pair across through space) serves as a parameter to modulate the angular dependent term in the energy function that describes backbone hydrogen bonds.

4. Chirality. An introduction of C_β side-chain beads can result in unphysical chirality (D) in certain residues. To remove this problem, we incorporate an energy term that assigns the correct symmetry such that L-isoforms are always favored. This function was previously introduced by Takada et al.[39]

$$E_{\text{chiral}} = \sum_{\text{chiral}} \frac{1}{2} k_c (c - c_0)^2 \tag{11}$$

where c, the triple scalar product, is defined by $r_{C_\alpha^i C_\beta^i} \cdot (r_{C_\alpha^i C_\alpha^{i-1}} \times r_{C_\alpha^i C_\alpha^{i+1}})$. c_0 is the measurement obtained from the native structure. $k_c = 20\epsilon$.

5. Other Structural Information. A flexible chain is used in this study and is defined by the following terms:

$$E_{\text{bond}} = \sum_{\text{bonds}} \frac{1}{2} k_b (r - r_0)^2 \tag{12}$$

$$E_{\text{angle}} = \sum_{\text{angles}} \frac{1}{2} \kappa_\theta (\theta - \theta_0)^2 \tag{13}$$

$$E_{\text{dihedral}} = \sum_{\text{dihedrals}}^{C_\alpha - C_\alpha - C_\alpha - C_\alpha} \kappa_\phi^{(n)} [1 - \cos(n(\phi - \phi_0))] \tag{14}$$

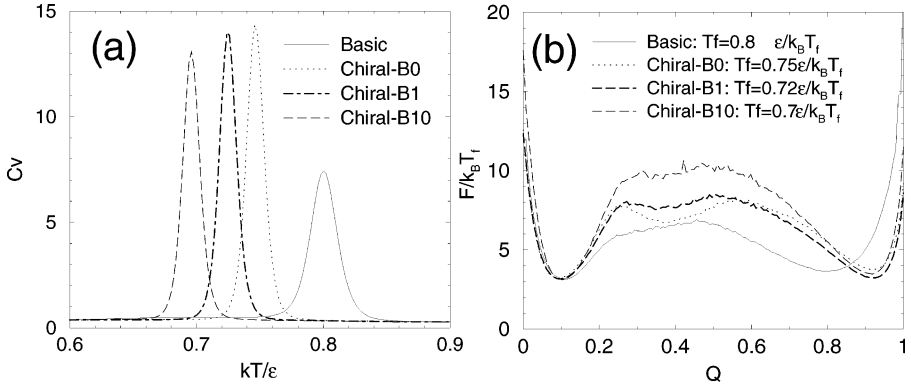

Figure 3. Differences in thermodynamic signatures of models with/without chiral constraints and effects of the angular-dependence in the Hamiltonian. Panel (a) shows the specific heat plotted as a function of temperature, and the peak of the profiles denotes the folding temperature (T_f). Panel (b) shows the free energy diagrams as a function of Q (Q is the fraction of native contact formation). The T_f of the Chiral-B0 model is lower than the Basic model, indicating a reduction of configurational entropy in the native state. Because both native energies are the same, the enthalpy change given in panel a reflects a change in the cooperativity of folding; the higher peak in the profile of Chiral-B0 (with chiral constraints) indicates a stronger cooperativity of the folding transition. The folded minimum of Chiral-B0 is much closer to $Q = 1$ implies that chiral constraints help to remove side-chain disordering in the native basin. Although chiral constraints help remove disordering of side-chains, it overly stabilizes the mini-core region of CI2 at the transition states (TS). This over-populated states caused a "dip" about $Q = 0.4$ at the top of the free energy profile.

r is the distance between two adjacent beads. θ is the angle defined by three consecutive beads (C_α's or C_β's). ϕ is the dihedral angle defined over four consecutive C_α's. r_0, θ_0, and ϕ_0 are determined from the native structure. $k_b = 100\epsilon$, $\kappa_\theta = 20\epsilon$, $\kappa_\phi^{(1)} = \epsilon$, $\kappa_\phi^{(3)} = 0.5\epsilon$.

C. Molecular Dynamics. We employ a standard molecular dynamics method carried out with the AMBER6[50] program as an integrator. The dynamics of the system is studied at constant temperature with the Berendsen algorithm with a coupling time 0.01 τ_s.[51] Temperature is measured in units of (ϵ/k_B), where k_B is the Boltzmann constant. The integration time step is 0.0001 τ_s, and data is collected every 0.1 τ_s, where time is in units of reduced time, $\tau_s = (m\sigma^2/\epsilon)^{1/2}$. To ensure appropriate sampling, dynamics is simulated for sufficiently long runs (over 1×10^8 integration time steps). SHAKE is used to keep the bond length constant in the simulations. WHAM[52,53] is used to calculate the thermodynamic quantities of the system such as the free energy profile, specific heats, or ensemble averages.

III. Results and Discussion

A. Chiral Compounds. The construction of a minimal protein representation that includes side-chain degrees of freedom needs a chirality constraint since most amino acids in proteins are L-stereoisomers. In our "$C_\alpha - C_\beta$" model, we distinguish two nonsuperimposable mirror images (shown in Figure 1) by specifying only L isomers via a harmonic constraint (eq 11).

Effects of chiral constraints are analyzed by comparing the thermodynamic signatures between the Chiral-B0 model (with chiral constraints) and the Basic model (without chiral constraints). The folding temperature (T_f) of the Chiral-B0 model is lower, indicating a reduction in the number of allowed states (configurational entropy) in the folded minimum (Figure 3a). Because both models have the same native energy, a change in specific heat would imply cooperativity in folding. A sharper peak in the specific heat of Chiral-B0 model denotes a higher cooperativity of folding. Also, the folded minimum of the free energy profile of the Chiral-B0 model is closer to $Q = 1$ (Figure 3b), implying that chiral constraints help to reduce the level of side-chain disorder in the native basin. Finally, in the free energy profile of the Chiral-B0 model, there is a "dip" at the top of the transition states (TS). We found that the populated configurations of this dip have significant contacts in the mini-core region of CI2. This dip is attributed to the chiral constraint because side-chain disorder at the mini-core region is reduced and that contact formation becomes selected favorable in this region of CI2. The heterogeneity of the contact probabilities increases and can be reflected by the route measure, R(Q).[54,61] This route measure in Figure 7 implies that there are fewer folding routes in the Chiral-B0 model (with a higher value of R($Q = 0.5$) $\approx$ 0.6) than the Basic model (with a lower value of R($Q = 0.5$) $\approx$ 0.3) at the TS.

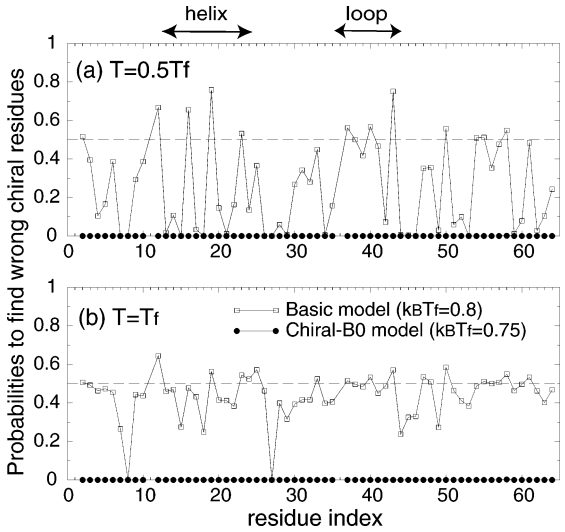

Figure 4. Probabilities to find wrong chiral residues under chiral constraints. A residue is assumed to sample a wrong chiral position if there is a sign change in the triple scalar product of any given configuration relative to the native state. The top panel shows probabilities to find wrong chiral residues over an entire ensemble of both Basic (without chiral constraints) and Chiral-B0 models (with chiral constraints) at $0.5T_f$, whereas the probabilities at the lower panel are calculated at T_f. In both panels, there are more residues in the wrong chiral position for the Basic model, especially those located at the helix and loop regions. Furthermore, at T_f, the probabilities of the Basic model approach 0.5 (a visual guideline is provided by a horizontal dashed line), an indication of a racemic mixture in which both isoforms become energetically indistinguishable.

Next, we attempt to quantify chiral fluctuations in the folding dynamics. A residue is considered to sample a wrong chiral conformation if there is a change in the sign of the triple scalar product (eq 11) relative to its native state one. The probability

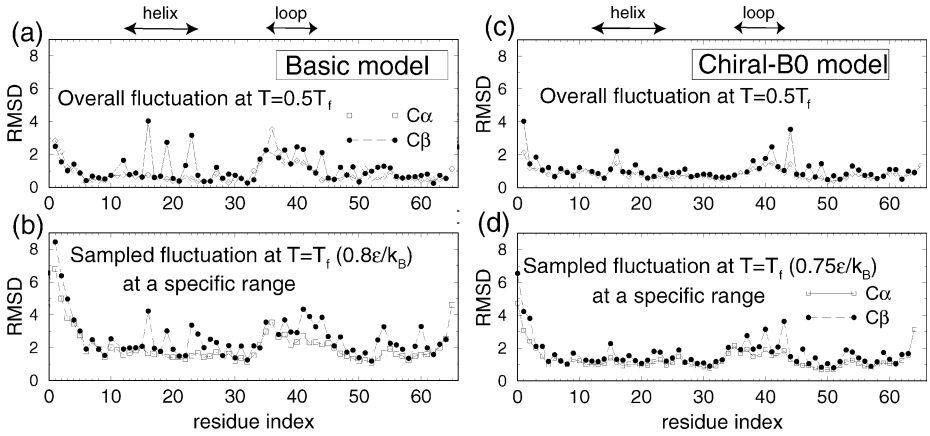

Figure 5. Overall fluctuations of the C_α and C_β beads with/without chiral constraints. In panel a, the Basic model (without chiral constraints) has larger fluctuations in C_β beads than the Chiral-B0 model (with chiral constraints) in panel c, especially those located at the helix and the loop regions. At T_f ($k_BT_f/\epsilon = 0.8$ for the Basic model, $k_BT_f/\epsilon = 0.75$ for the Chiral-B0 model), sampled fluctuations at the folded minimum for the Basic and Chiral-B0 model are shown in panels b and d, respectively. In panel d, as chiral constraints are added to the Hamiltonian, the fluctuations for C_α beads decrease by 30%, and those for C_β beads decrease by 50%. Together with the results shown in Figure 4, it is clear that the additions C_β beads cause artificial chiral traps, which can be removed by chiral terms in the model Hamiltonian.

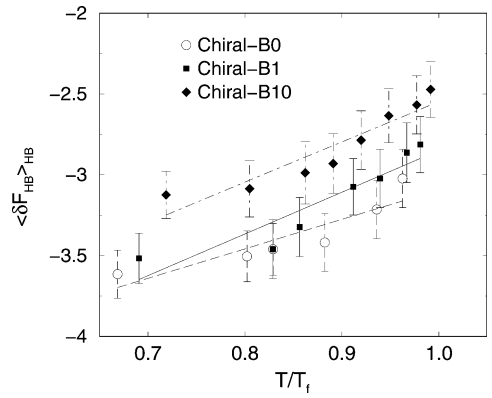

Figure 6. Averaged free energy costs ($\langle \delta F_{HB} \rangle_{HB}$) to order backbone hydrogen bonds (HB) in the folded ensemble are shown as a function of the normalized temperature (by the folding temperature T_f). $\langle \delta F_{HB} \rangle_{HB} = \langle -k_BT \ln P_{Formed}/P_{Notformed} \rangle_{HB}$ is the free energy difference of the formed (P_{Formed}) and not formed ($P_{Notformed}$) probabilities for each HB in the folded ensemble and then averaged over all native HBs. The bars in the profile are standard deviations that reflect the distribution of stabilities of native backbone HBs. It shows a linear dependence (positive slope), indicating an entropy loss to order backbone hydrogen bonds as the angular dependence in the contact formation increases. The slope of profiles when plotted against normalized temperature (by T_f) is almost the same for Chiral-B1 and Chiral-B10 models, indicating that their changes in T_f correspond to the entropy loss attributed to ordering backbones into correct orientations. In addition to the entropy loss, offsets in $\langle \delta F_{HB} \rangle_{HB}$ indicate that interactions are destabilized energetically. Clearly, the Chiral-B10 model (with the strongest angular dependence in the contact formation) that has the largest losses in both entropy and energy shows the smallest stability among all models.

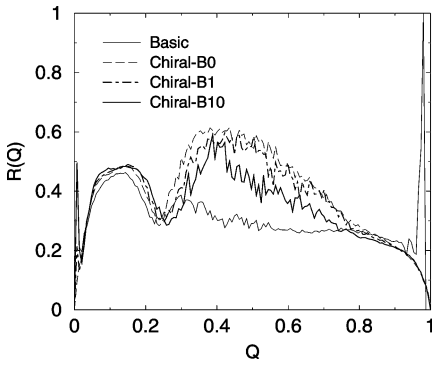

Figure 7. Heterogeneity in the distribution of the contact probabilities for the different models is affected by chiral constraints and the angular-dependent backbone formation. The folding route, $R(Q)$, is the parameter to measure the dispersion of contact formation as these protein models evolve from the unfolded ensemble to the native basin. As $R(Q)$ approaches 1(0), the dispersion is high(low) and the system chooses fewer (more) routes at a given Q. The effects of the chiral constraint helps to reduce the level of disordering of side-chains and therefore affects dynamics at two stages: (1) At the folded states where the packing of beads play an important role and the elimination of wrong chiral conformations allow routes to access $Q = 1$ more efficiently. Indeed, the Basic model (without chiral constraints) shows a large value of $R(Q)$ near $Q = 1$, indicating that this system is impeded from approaching $Q = 1$. (2) At the transition state ensemble (TSE), the chiral constraint overly stabilizes the mini-core region of the CI2. It results in increasing the dispersion of the contact formation at TSE; $R(Q; \text{TSE})$ is larger in the Chiral-B0 model than in the Basic model. The angular dependence in the hydrogen bond interaction strongly affects the stability of pair interactions. This effect reduces the over-populated states in the mini-core region of CI2 at the TSE in the Chiral-B0. The Chiral-B10 model that has the strongest angular-dependence in the contact formation, which leads to a low $R(Q; \text{TSE})$ value compared to the Chiral-B0 and Chiral-B1 models.

to find C_α's in an incorrect chiral arrangement is shown in Figure 4. At low temperatures (Figure 4a), the Basic model has more residues in the wrong C_α chiral arrangements, especially those located at the helix and loop regions. As the temperature increases to the folding temperature(Figure 4(b)), a probability of 0.5 is reached in the Basic model when both L and D isoforms become energetically indistinguishable, which is an indication of a racemic mixture.

Furthermore, to determine how the lack of a chiral constraint affects the ordering of the C_β and C_α beads in the folded states, we calculate the RMSD of each bead at different temperatures in Figure 5. At low temperatures ($0.5T_f$), the Basic model (Figure 5a) shows that there are larger fluctuations of C_β's at the helix and the loop regions compared to Chiral-B0 model (Figure 5c). Effects of non-native chiral conformations have also been discussed by Hardin et al.[55] in their kinetic studies of protein folding. These incorrect chiral conformations are interpreted as artificial "traps" which exist because of the lack of a chiral constraint.

As temperature increases and approaches T_f, ensemble fluctuations of each bead are analyzed at the folded minimum of the free energy profile ($Q \approx 0.8$ for Basic model in Figure 5b; $Q \approx 0.9$ for Chiral-B0 model in Figure 5d). In this

temperature range, not only C_β's, but also C_α beads have larger fluctuations for the Basic model (Figure 5b). After the correct L symmetry is constrained, fluctuations of C_α beads decrease by 30% and fluctuations of C_β beads decrease by 50% (Figure 5d). Together with the results in Figure 4, we conclude that the introduction of C_β beads causes artificial chiral traps and that an energy function which assigns the correct handed-symmetry can correct this problem. Therefore, chiral constraints must be considered to produce physically relevant models when multiple-bead amino acid representations are used. An analysis based on models that do not impose chiral constraints should be interpreted with care.[44]

B. Free Energy Cost to form Backbone Hydrogen Bonds. Interactions of backbone hydrogen bonds (HB) are necessary to include in order to study secondary structure formation. Unlike side-chains, which allow a large number of bonding pairs per residue, backbone C_α's should only allow a few pairs. A cooperative model of backbone dynamics can be achieved by including many-body effects[43] into C_α pairing interactions. This cooperativity can be achieved by either (1) providing an additional energetic gain if nearby pairs of hydrogen bonds are also formed[38,39] or (2) by placing an angular dependence term on contact formation in the energy function.[41,56] This study uses the latter approach, which requires only the relative positions of a few backbone atoms to incorporate. Several groups use specific orientations (e.g., the vector of C═O···H—N to be perpendicular to the backbone) to favor hydrogen bond interactions. This approach provides a good description for secondary structures such as α helices and antiparallel β sheets. However, it is not inclusive of parallel β sheets. To remove this limitation, an energy function that promotes equally spaced C_α atoms in both forms of β sheets has been proposed.[57] However, this model lacks a formulation that can be applied to molecular dynamics simulations.

We provide a general expression for an angular dependence term that allows transitions between the three major kinds of secondary structures to accommodate molecular dynamics simulations. This angular term destabilizes the energy of C_α,C_α pairs whose spatial orientations do not fit to the definition of common secondary structures, and thereby defines a restricted range of conformational space for each C_α,C_α pair. It also imposes an implicit excluded volume effect on HB formation because it lowers the chance for one C_α bead to participate in multiple HB pairs. In this subsection, we explore models that only include pairs of native backbone hydrogen bonds. In the next subsection, more general models which allow nonspecific backbone interactions (beyond native HB pairs) are discussed.

To quantify the free energy costs to "dock" HBs into correct orientations, Figure 6 shows the average free energy cost ($\langle \delta F_{HB} \rangle_{HB}$) to order backbone hydrogen bonds (HB) in the folded ensemble as a function of temperature below T_f. $\langle \delta F_{HB} \rangle_{HB} = \langle -k_B T \ln P_{Formed}/P_{Notformed} \rangle_{HB}$ is the free energy difference of the formed (P_{Formed}) and not formed ($P_{Notformed}$) for each native HB pairs in the folded ensemble. This quantity is averaged over all pairs of native backbone HBs. Offsets of $\langle \delta F_{HB} \rangle_{HB}$ indicate that contact energies are destablized when the angular dependence is incorporated in the contact formation. Moreover, when the free energies of the different models are fitted linearly, they show positive slopes that denote a loss of configurational entropy to order secondary structures. If the angular dependence in the contact formation increases, these slopes should become steeper because it would be more difficult to orientate the backbone into the lower-energy conformation.

When δF_{HB} is plotted against temperatures normalized to the T_f of each model (as shown in Figure 6), the different models can be fitted approximately to the same slope. This observation indicates that the loss of configurational entropy tracks linearly with the change in folding temperatures of each model. This relationship can be understood by using a simple two-state, statistical model, where T/T_f is proportional to the configurational entropy in the system. Given two states, the energy difference between the folded state and the unfolded state is $\delta\epsilon$. δn is the ratio of number of configurations in the unfolded state and the folded state. The temperature to have equally populated states is $\delta\epsilon/k_B \ln \delta n$. Then, around T_f, the free energy is a linear function of $-\delta\epsilon T/T_f$, which in our case is proportional to configurational entropy in this system.

The angular dependent term for HB formation preferentially reduces the stability of the contacts in the mini-core region of CI2 which, in the absence of the angular term, have poor backbone alignment in the transition state ensemble. As a consequence, the distribution of contact probabilities at the TS becomes more homogeneous, reflected by a lowered route measure ($R(Q = 0.5) \approx 0.4$) for the Chiral-B10 model in Figure 7. In addition, both free energy profiles of Chiral-B1 and Chiral-B10 in Figure 3b do not show a "dip" at the top of the barrier, indicating that the transition state ensemble is no longer populated by the same conformations observed with the Chiral-B0 model. Together with chiral constraints, we provide a $C_\alpha - C_\beta$ representation which models side-chain interactions, hydrogen bond formation, and which behaves similarly to CI2 in folding experiments.

C. Interplay of Topology-Driven Interactions and Secondary Structure Formation. Although side-chain interactions are highly sequence-dependent, backbone hydrogen bonding (HB) does not, in principle, depend significantly on the amino acid side-chains (i.e., nonspecific interactions). Thus, pairing of the HBs can be nonspecific and may promote non-native conformations. Here, the angular dependence is introduced into the nonspecific C_α, C_α contact formation. Backbone interactions are still constrained to specific orientations by the angular term so that the number of possible pairing HBs orientations per amino acid is effectively reduced. Consequently, the angular term acts to prevent non-native, collapsed conformations, even though non-native HBs pairs are allowed.

In this subsection, we generalize backbone interactions by allowing non-native hydrogen bonds, whereas those of side-chains are still kept Gō-like. This approach is useful to explore the interplay between nonspecific backbone interactions and specific side-chain interactions. Interestingly, when the angular dependence is incorporated in nonspecific backbone HB formation, there are dramatic differences in the ensemble-averaged properties. We first examined a case where there is no angular dependence and all pairing HB interactions are indistinguishable (i.e., model Chiral-B0-nsHB). In Figure 8, at a region where temperature is above T_f, the radius of gyration (R_g) of the Chiral-B0-nsHB model is the smallest one observed compared to other models, which is an indication of the existence of the most compact unfolded states. As the angular dependence is systematically introduced, the R_g of the unfolded states increases. It implies that those unfolded states become more extended because less non-native HBs are likely to be formed. Indeed, when we plot the free energy profiles of both Chiral-B10-nsHB and Chiral-B0-nsHB models as a function of both Q and HB, we found that the size of unfolded basin dramatically falls from 70 HBs (Chiral-B0-nsHB) to 35 HBs (Chiral-B10-nsHB) (data are not shown).

If the angular dependence is incorporated into the contact

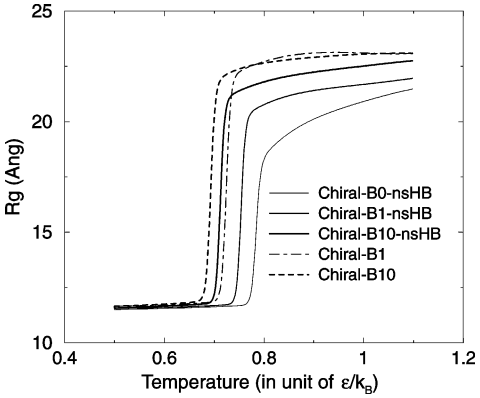

Figure 8. Effects of angular-dependent interactions in nonspecific backbone hydrogen bond (HB) formation are observed from plots of the radius of gyration (R_g) as a function of temperature. The dashed profiles are Gō-like models; the solid profiles are models that allow nonspecific backbone HB's. Because the models Chiral-B0-nsHB, Chiral-B1-nsHB, and Chiral-B10-nsHB have different levels of angular-dependence in the backbone formation, their R_g's vary in the unfolded ensemble at temperatures above T_f. As expected, the Chiral-B0-nsHB has the smallest R_g, indicating the presence of more collapsed conformations. As the angular dependence in the backbone formation increases (Chiral-B1-nsHB, Chiral-B10-nsHB), the Rg of the unfolded ensemble increases. It implies that the misfolded conformations are more extended because the angular term helps the backbone interactions to avoid non-native hydrogen bonding. The models with nonspecific hydrogen bonds approach the standard Gō-like behavior because non-native interactions are mostly avoided.

formation, there are free energy costs to order backbone hydrogen bonds into secondary structure elements. Backbone HBs become difficult to form when the angular dependence is introduced to the Hamiltonian. At high temperatures, backbone interactions with angular dependence can be more destabilized than those with no angular dependence. When the temperature decreases, because of differences in stabilities of the backbone HBs imposed by the angular-dependent energy function, C_α beads start to compete for pairing backbone HBs. During this temperature evolution, a bias from the side-chain energies is crucial in finding the native backbone HBs from the many nonspecific backbone hydrogen bonds.

In Figure 8, we discuss the role of this angular dependence in the interplay between secondary structural formation and the side-chain interactions driven by topology. Profiles of nonspecific hydrogen bond models (solid lines) converge to those of Gō-like hydrogen bond models (dashed lines) as the angular dependence in the backbone HB formation increases. Such convergence of the "Gō model + hydrogen bonding" to a standard "Gō model" is underpinned by the free energy costs imposed by angular-dependent backbone interactions. As the angular free energy costs increase, the possibility to order non-native secondary structures in the unfolded states greatly deminishes. With a large angular term, this system converges to a Gō-like model where non-native hydrogen bond interactions act repulsive. Our results are comparable to the analysis of the "Gō + Hydrogen bond" model proposed by Hardin et al,[55] despite the fact that a different energy function for the backbone HB is used.

IV. Conclusion

In minimalist C_α models, the many degrees of freedom of a single amino acid are renormalized into a few simple interactions which are able to capture the essential physics of protein folding. Although the global folding behavior of C_α models qualitatively agrees with experiments, all-atom simulations reveal that a more quantitative and a higher resolution protein representation is required to explain the details of the folding mechanism. This observation raises a critical question: Which level of complexity of a protein representation is needed to provide quantitative agreement with experiments? On one hand, we know that a pure C_α model has its limitations such as a lacking description of side-chain packing. On the other hand, we wish to preserve a minimal complexity that captures the essential physics of folding, thereby reducing the computational effort and facilitating analysis. In this study, we use a minimal representation of a protein model that goes beyond the C_α model and allows the study of side-chain interactions and backbone hydrogen bonding.

Because of the aforementioned reason, we choose a "$C_\alpha-C_\beta$" model. The introduction of C_β beads to represent side-chains, however, creates artificial chiral traps. To solve this problem, we used an energy function to assign the correct handed-symmetry. This chirally assigned protein representation improves the RMSD of conformations in the folded basin. Also, the basin of the free energy of the folded states shifts closer to $Q = 1$. As expected, these folded structures show an increased level of side-chain ordering.

An angular dependence is included in the contact formation between C_α pairs to account for the directional behavior of the backbone hydrogen bonds. Our study shows that the interplay between the secondary structure and tertiary native contact formation maintains cooperativity in the folding transition. The stability of side-chain interactions can be perturbed sensibly by angular alignments of the local backbone conformation. For example, in the folding of CI2, "native" backbone hydrogen bonds at the mini-core region help exclude some side-chain conformations that do not have the proper angular orientation. The stability of the native side-chain interactions can be altered consequently by this effect; the wrongly over-populated mini-core encountered in the transition state ensemble(TSE) of previous simulations becomes destabilized, and the new model behaves more protein-like.

In our study, we show that there are free energy costs to order backbone hydrogen bonds. This property can be observed by changing the stability of the backbone HB through a systematic variation of angular dependence in the contact formation. A change of T_f accounts for both the energy and configurational entropy costs needed to "dock" the local construct of the backbone into appropriate secondary structures. Because of the added free energy cost needed to order a backbone, our results suggest that the formation of side-chain contacts must occur concurrently to aid the folding process. This cooperativity between side-chain interactions and backbone hydrogen bonding has also been highlighted in recent studies by others.[58]

Although parts of the energy function can be non-Gō-like, such as C_α, C_α interactions, the energy function still leads to the correct folding as long as other contributions in the structural Hamiltonian and C_β, C_β interactions remain Gō-like. As the protein (e.g., CI2) folds toward the native state, native interactions between side-chain contacts help to introduce bias toward native C_α pairs. Also, the angular-dependent hydrogen bonding helps to eliminate non-native conformations throughout the folding process. Without native side-chain interactions, native hydrogen bonds become very unlikely. In future studies, it will be interesting to investigate whether non-native side-chain interactions (present in more physically relevant potentials) contribute with mis-matched backbone hydrogen bonding to produce mis-folded configurations suggested by some experiments.[59]

Acknowledgment. M.S.C. thanks Dr. Peter Wolynes for critical comments and physical insights during her thesis defense which includes this study, Drs. Chinlin Guo, Herbert Levine, Dmitri Klimov, Andrzej Kolinski, Anders Irbäck, and Chigusa Kobayashi for helpful discussions on hydrogen bond modeling, and Ms. Leslie Chavez for revising the manuscript. Work in San Diego has been sponsored by the NSF (MCB-0084797 with additional support from PHY-0216576 and 0225630). Support has been also provided by the NIH (GM064936-01) and La Jolla Interfaces in Science Interdisciplinary Training Program and the Burroughs Welcome Fund (J.M.F.), and the NSF (PHY-0097854) Research Experience for Undergraduates (B.C.). Additional support for computation has been provided by W. M. Keck Foundation.

References and Notes

(1) Onuchic, J. N.; H. Nymeyer, H.; García, A. E.; Chahine, J.; Socci, N. D. *Adv. Protein Chem.* **2000**, *53*, 87−153.
(2) Bryngelson, J. D.; Wolynes, P. G. *Proc. Natl Acad. Sci. U.S.A.* **1987**, *84*, 7524−7528.
(3) Scheraga, H. A. *Protein Sci*. **1992**, *1*, 691.
(4) Dill, K. A.; Chan, H. S. *Nature Struct. Biol*. **1997**, *4*, 10−19.
(5) Eaton, W. A.; Munoz, V.; Thompson, P.; Chan, C. K.; Hofrichter, J. *Curr. Opin. Struct. Biol.* **1997**, *7*, 10−14.
(6) Telford, J. R.; Wittung-Stafshede, P.; Gray, H. B.; Winkler, J. R. *Acc. Chem. Res.* **1998**, *31*, 755−763.
(7) Gruebele, M. *Annu. Rev. Phys. Chem.* **1999**, *50*, 485−516.
(8) Shea, J.-E.; Brooks, C. L.; Onuchic, J. N. *Annu. Rev. Phys. Chem.* **2001**, *52*, 499−535.
(9) Dokholyan, N. V.; Li, L.; Ding, F.; Shakhnovich, E. I. *Proc. Natl Acad. Sci. U.S.A.* **2002**, *99*, 8637−8641.
(10) Myers, J. K.; Oas, T. G. *Annu. Rev. Biochem*. **2002**, *71*, 783−815.
(11) Leopold, P. E.; Montal, M.; Onuchic, J. N. *Proc. Natl Acad. Sci. U.S.A.* **1992**, *89*, 8721−8725.
(12) Jackson, S. E.; el Masry, N.; Fersht, A. R. *Biochemistry* **1993**, *32*, 2, 11270−11278.
(13) Martinez, J. C.; Serrano, L. *Nature Struct. Biol.* **1999**, *6*, 1010−1016.
(14) Riddle, D. S.; Grantcharova, V. P.; Santiago, J. V.; Alm, E.; Ruczinski, I.; Baker, D. *Nature Struct. Biol*. **1999**, *6*, 1016−1024.
(15) Plaxco, K. W.; Simons, K. T.; Baker, D. *J. Mol. Biol.* **1998**, *277*, 985−994.
(16) Onuchic, J. N.; Wolynes, P. G.; Luthey-Schulten, Z. A.; Socci, N. D. *Proc. Natl Acad. Sci. U.S.A.* **1995**, *92*, 3626−3630.
(17) Levitt, M.; Warshel, A. *Nature* **1975**, *253*, 694−698.
(18) Friedrichs, M. S.; Goldstein, R. A.; Wolynes, P. G. *J. Mol. Biol.* **1991**, *222* (4), 1013−1034.
(19) Irbäck, A.; Potthast, F. *J. Chem. Phys.* **1995**, *103*, 10298−10305.
(20) Guo, Z.; Thirumalai, D.; Honeycutt, J. D. *J. Chem. Phys.* **1992**, *97* (1), 525−535.
(21) Berry, R. S.; Elmaci, N.; Rose, J. P.; Vekhter, B. *Proc. Natl Acad. Sci. U.S.A.* **1997**, *94*, 9520−9524.
(22) Guo, Z.; Brooks, C. L. *Biopolymers* **1997**, *42*, 745−757.
(23) Nymeyer, H.; García, A. E.; Onuchic, J. N. *Proc. Natl Acad. Sci. U.S.A.* **1998**, *95*, 5921−5928.
(24) Shea, J.-E.; Onuchic, J. N.; Brooks, C. L. *J. Chem. Phys.* **2000**, *113*, 7663−7671.
(25) Berriz, G. F.; Shakhnovich, E. I. *J. Mol. Biol.* **2001**, *310*, 673−685.
(26) Karanicolas, J.; Brooks, C. L. *Protein Sci*. **2002**, *11*, 2351−2361.
(27) Clementi, C.; Nymeyer, H.; Onuchic, J. N. *J. Mol. Biol.* **2000**, *298*, 937−953.
(28) Clementi, C.; Jennings, P. A.; Onuchic, J. N. *J. Mol. Biol.* **2001**, *311*, 879−890.
(29) Koga, N.; Takada, S. *J. Mol. Biol.* **2001**, *313*, 171−180.
(30) Cheung, M. S.; García, A. E.; Onuchic, J. N. *Proc. Natl Acad. Sci. U.S.A.* **2002**, *99*, 685−690.
(31) Lim, W. A.; Sauer, R. T. *Nature* **1989**, *339*, 31−36.
(32) Sandberg, W. S.; Terwilliger, T. C. *Science* **1989**, *245*, 54−57.
(33) Dahiyat, B. I.; Mayo, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10172−10177.
(34) Gonzalez, L., Jr.; Plecs, J. J.; Alber, T. *Nature Struct. Biol.* **1996**, *3*, 510−515.
(35) Walsh, S. T. R.; Sukharev, V. I.; Betz, S. T.; Vekshin, N. L.; DeGrado, W. F. *J. Mol. Biol.* **2001**, *305*, 361−373.
(36) Mittermaier, A.; Varani, L.; Muhandiram, D. R.; Kay, L. E.; Varani, G. *J. Mol. Biol.* **1999**, *294*, 967−979.
(37) Lee, A. L.; Kinnear, S. A.; Wand, A. J. *Nature Struct. Biol.* **2000**, *7*, 72−77.
(38) Vieth, M.; Kolinski, A.; Brooks, C. L.; Skolnick, J. *J. Mol. Biol.* **1995**, *251*, 448−467.
(39) Takada, S.; Luthey-Schulten, Z.; Wolynes, P. G. *J. Chem. Phys.* **1999**, *110*, 11616−11629.
(40) Klimov, D. K.; Thirumalai, D. *Proc. Natl. Acad. U.S.A.* **2000**, *97*, 2544−2549.
(41) Irbäck, A.; Sjunnesson, F.; Wallin, S. *Proc. Natl Acad. Sci. U.S.A.* **2000**, *97*, 13614−13618.
(42) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Chem. Phys.* **2001**, *115*, 2323−2347.
(43) Guo, C.; Cheung, M. S.; Levine, H.; Kessler, D. A. *J. Chem. Phys.* **2002**, *116*, 4353−4365.
(44) Ding, F.; Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. *Biophys. J.* **2002**, *83*, 3525−3532.
(45) Shimizu, S.; Chan, H. S. *J. Chem. Phys.* **2001**, *115*, 1414−1421.
(46) Cordier, F.; Wang, C.; Grzesiek, S.; Nicholson, L. K. *J. Mol. Biol.* **2000**, *304*, 497−505.
(47) Krantz, B. A.; Srivastrava, A. K.; Nauli, S.; Baker, D.; Sauer, R. T.; Sosnick, T. R. *Nature Struct. Biol.* **2002**, *9*, 458−463.
(48) Leach, A. R. In *Molecular modelling: principles and applications*; Longman: Harlow, U.K., 1996.
(49) Go, N. *J. Stat. Phys.* **1983**, *30* (2), 413−423.
(50) Amber6. Case, D. A.; Pearlman, D.; Caldwell, J. W.; T. E. Cheatam, I.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; et al. ALC Press: Lake Oswego, OR, 1999.
(51) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81* (8), 3684−3690.
(52) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1988**, *61*, 2635−2638.
(53) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13* (8), 1011−1021.
(54) Plotkin, S. S.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 6509−6514.
(55) Hardin, C.; Luthey-Schulten, Z.; Wolynes, P. G. *Proteins* **1999**, *34*, 281−294.
(56) Klimov, D. K.; Betancourt, M. R.; Thirumalai, D. *Fold. Des.* **1998**, *3*, 481−496.
(57) Ilkowski, B.; Skolnick, J.; Kolinski, A. *Macromol. Theory. Simul.* **2000**, *9*, 523−533.
(58) Fernández, A.; Scheraga, H. A. *Proc. Natl Acad. Sci. U.S.A.* **2003**, *100* (1), 113−118.
(59) Kuwata, K.; Shastry, R.; Cheng, H.; Hoshino, M.; Batt, C. A.; Goto, Y.; Roder, H. *Nature Struct. Biol.* **2001**, *8*, 151−155.
(60) ρ_α is obtained by calculating the pseudo dihedral angle of a canonical helical turn. This pseudo dihedral angle for an α helix (i.e., $\rho_{i-1,i,i+4,i+5}$) is 0.465999 (rad).
(61) The route number, $R(Q) = \langle\sigma\rangle/Q(1 - Q)$, where σ is the distribution (variance) of probabilities of the contact formation for each native contact, and $\langle\sigma\rangle$ is the measurement averaged over all native contact pairs. $R(Q)$ measures the distribution of contact formation and is used to indicate the amount of folding routes funneled to the native state; $R(Q)$ approaches 1 (0), if the system has less (more) routes as it moves toward the native state.