JCTC Journal of Chemical Theory and Computation

Coarse-Grained Model of Coil-to-Helix Kinetics Demonstrates the Importance of Multiple Nucleation Sites in Helix Folding

Alan E. van Giessen and John E. Straub*

Department of Chemistry, Boston University, Boston, Massachusetts 02215

Received December 26, 2005

Abstract: An extension of a coarse-grained, implicit-solvent peptide model wherein each amino acid residue is represented by four interaction sites is presented and discussed. The model is used to study the coil-to-helix transition of five peptide sequences, ranging from all hydrophobic to all hydrophilic, for a 10-residue peptide. The thermodynamics of the folding transition are analyzed and discussed for each sequence, and the stability of the α-helix is correlated with the hydrophobic content of the sequence. In addition, for each sequence, the folding kinetics of the transition from random coil to full α-helix are analyzed, and the mean folding time is determined. Folding times vary from 59 ns for the most hydrophobic sequence to 132 ns for the most hydrophilic sequence. These folding times compare very well with those measured in experments. All sequences show single-exponential kinetics. A plot of the mean folding time versus the reciprocal of the Zimm−Bragg parameter σ—a measure of the free energy cost of nucleating a helix—is shown to be nonlinear, in contrast to the predictions of many theories of the coil-to-helix transition. It is proposed that the origin of this nonlinearity is due to multiple helix nucleation sites, indicating that even for short peptides such as those studied here, multiple folding pathways play an important role in the transition from random coil to native state.

1. Introduction

One of the fundamental problems in the study of protein folding is the coil-to-helix transition of proteins and peptides which form α-helices. That α-helices are common structural motifs in many biologically relevant proteins only underscores the importance of understanding the coil-to-helix transition. Many theoretical models for this transition have been developed, beginning with the pioneering work of Schellman,[1] Zimm and Bragg,[2] and Lifson and Roig[3] more than 40 years ago. For a recent survey of advances in the theory of the coil-to-helix transition, see the review by Doig.[4]

It is the model of Zimm and Bragg[2] which first defined the parameters by which the coil-to-helix transition is frequently described. In their model, which is isomorphic with the one-dimensional Ising model, the Hamiltonian is

$$H = -J\sum_{ij} s_i s_j - H\sum_{is_i} \tag{1}$$

where the spin−spin coupling constant J and the external field H are expressed in units of kT. Zimm and Bragg introduced the parameters s and σ, defined as

$$\sigma s = \exp(H) \qquad s = \exp(J + H) \tag{2}$$

which determine the free energy of helix propagation and nucleation, respectively. At the folding transition temperature, $s = 1$. Takano and co-workers[5] have shown recently how well the Zimm−Bragg model describes the thermodynamics of the coil-to-helix transition when compared to an all-atom molecular dynamics simulation.

The thermodynamics of the coil-to-helix transition is well understood, due largely to the use of the Zimm−Bragg model. The kinetics of the transition, however, are still poorly understood, despite much progress, both experimental and

* Corresponding author e-mail: straub@bu.edu.

10.1021/ct0503318 CCC: $33.50
Published on Web 04/22/2006

theoretical, over the past decade. Details of the folding rate and of the folding mechanism, or mechanisms, are still unclear. Early estimates for the folding rate were on the order of 1 μs.[6] More accurate temperature-jump experiments by Williams et al.[7] revealed a folding time of 160 ns for a 21 residue alanine-based peptide. Thompson et al.[8] and Lednev et al.[9] measured a slightly longer folding time of 220 ns and 240 ns, respectively. Using a stopped-flow CD measurement, Clarke[10] and co-workers measured a considerably longer folding time of miliseconds for a 16-residue peptide. Another set of T-jump experiments on a 153-residue globular protein by Woodruff and co-workers[11−13] showed a relaxation rate in the range of 10−160 ns. In the past few years, Gai and co-workers have performed both stopped-flow experiments[14,15] and temperature jump experiments[16−18] that have established a time scale on the order of 200 ns.

Theoretical predictions for the folding time of an α-helix tend to be considerably shorter than the 200 ns time scale established by experiment, though some are in agreement. An early MD simulation by Daggett and Levitt[19] suggested the time scale for helix propagation to be 100 ps. Coarse-grained simulations by Thirumalai and co-workers indicated the folding time at the folding temperature to be roughly 20 ns.[20] A different coarse-grained model developed by Takada and co-workers,[21] which will be discussed in more detail below, gives a folding time of approximately 15 ns. Margulis et al.[22] measured a folding time of 1 ns for a molecular dynamics simulation of an alanine pentapeptide in explicit solvent. One calculation that does agree with experiment is the nucleation-elongation theory of Doshi and Muñoz.[23] They establish a folding time of 150−300 ns.

Many models of the coil-to-helix transition have sought to relate the folding time, τ, to the Zimm−Bragg parameters s and σ. In a seminal work, Schwartz[24] estimated the folding time at the midpoint of the coil-to-helix transition (when $s=1$) to be

$$\tau = \frac{1}{4\sigma k_F} \tag{3}$$

where k_F is the rate of adding an additional helical residue at the helix end. Brooks[25] proposed a model based on a sequential formation of helical residues and demonstrated that the mean time for the folding/unfolding process scaled as $\tau \sim 1/\sigma$. In all these models, it is found that the mean folding time, τ, is inversely proportional to the Zimm−Bragg parameter σ. In addition, we note that all these models include the assumption that the helix propagates from a single nucleation site.

One recent theoretical model that relaxes this assumption is that due to Buchete and Straub.[26] This model, referred to as the active helix Ising model, is also based on the Zimm−Bragg model and allows one to solve the mean-first passage time equation. Buchete and Straub numerically determined the mean first passage time for a range of $1/\sigma$ values at fixed s and found significant nonlinear behavior. They observed linear behavior for small values of σ and strongly nonlinear behavior for large values of σ ($\sigma \geq 0.005$). This nonlinear behavior is increasingly important for longer polypeptide chains and for smaller values of the propagation constant s.

In contrast to almost all theoretical predictions, Gai[17] found evidence that the folding time for the coil-to-helix transition does *not* scale linearly with $1/\sigma$. The only model which predicts this behavior is the active helix Ising model of Buchete and Straub.[26] As noted above, one of the key assumptions in many models, but not that of Buchete and Straub, is that a helix is formed from a single nucleation site. It is suggested[26] that the origin of the nonlinearity lies in the ability of a peptide, even one as short as 10 residues in length, to have multiple nucleation sites. In this work, we propose to study the kinetics of the coil-to-helix transition via computer simulation for several model peptides in order to investigate the dependence of the folding time on the Zimm−Bragg parameter σ and to determine the relevance of multiple nucleation sites.

A natural choice for this investigation is the use of molecular dynamics (MD) simulations. The most accurate approaches employ all-atom MD simulations using an explicit molecular representation of the solvent. At the present time, for studies of the thermodynamics and kinetics of large-scale conformational transitions, and for phenomena that occur on time scales of hundreds of nanoseconds, such approaches are computationally too demanding in applications involving all but small peptides and proteins. Consequently, there is an ongoing effort to develop coarse-grained models of proteins using a reduced number of degrees of freedom. The most appealing approach is to include solvent effects implicitly in the interaction potentials and to replace the atoms in each amino acid residue by a small number of interaction sites, thereby drastically reducing the number of particles and interactions necessary for the calculation.

In this work, we present an extension of one such reduced model, originally developed by Takada et al.[21] In section 2, the peptide model is presented, while in section 3, the Langevin dynamics used to propagate the motion of the peptide forward in time is described. In section 4, the thermodynamcis of the coil-to-helix transition is discussed as well as the kinetics of the folding transition. Finally, section 5 presents some conclusions.

2. Peptide Model

The coarse-grained model used for the peptide in this work is a refinement of that proposed by Takada et al.[21] Similar structural models have been used by Hall[27] to study peptide aggregation. The structural model consists of four particles or "united atoms" per amino acid residue, shown schematically in Figure 1. Three of these united atoms represent the peptide backbone: one represents the amide nitrogen and its hydrogen, another the α-carbon and its hydrogen, and the third the carbonyl carbon and its oxygen. This high level of backbone representation is essential for reproducing correct secondary structure in the folded peptide.[28] The fourth united atom represents the amino acid side chain.

The model as presented here includes two types of side-chain interaction sites: hydrophobic and hydrophilic. While it is possible to introduce more detailed interaction potentials that more closely mimic the chemical identity of all the amino acids, for the present purpose it is sufficient to limit ourselves to this "two-letter" amino acid model. It has long

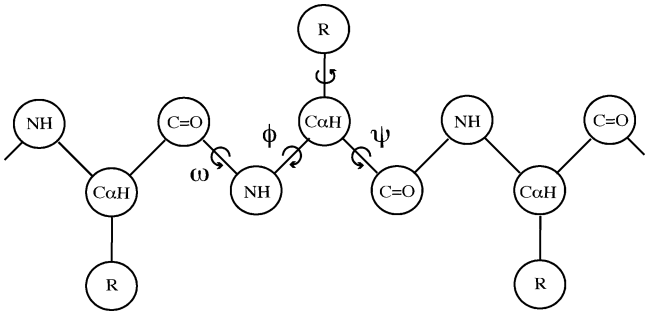

Figure 1. Schematic representation of the peptide model showing the dihedral angles ϕ, ψ, and ω. All bond lengths are fixed.

Table 1. Amino Acid Sequences

label	sequence[a]	T_f	τ (ns)
A	*PPPPPPPPPP*	411.3	59.4
B	*HPPPHPPPHP*	366.6	78.6
C	*PHPHPHPHPH*	349.4	118.7
D	*PHHHPHHHPH*	344.0	120.2
E	*HHHHHHHHHH*	323.8	131.7

[a] "P" denotes hydrophobic and "H" denotes hydrophilic amino acid residues.

Table 2. Structural Parameters

van der Waals diameters	σ (Å)	σ_{local} (Å)
C_α	3.30	2.64
C'	3.56	2.94
N	2.94	2.36
C_β	4.50	4.50

bond lengths	r (Å)
$C_\alpha-C'$	1.52
$C_\alpha-N$	1.45
$C'-N$	1.33
$C_\alpha-C_\beta$	1.80

bond angles	degrees	k_θ
$N-C_\alpha-C'$	111.6	200.0
$C_\alpha-C'-N$	117.5	200.0
$C'-N-C_\alpha$	120.0	200.0
$C'-C_\alpha-C_\beta$	110.0	200.0
$N-C_\alpha-C_\beta$	110.0	200.0

been realized that such a coarse-grained approximation captures many fundamental aspects of protein folding and can adequately be used to study the effect of amino acid sequence on equilibrium and dynamic properties.

In this work, we study five different 10-residue amino acid chains in order to determine the effect of amino acid sequence on the equilibrium properties of the coil-to-helix transition and on the mean folding time for helix formation. We use the letter *P* to denote a hydrophobic residue and the letter *H* to denote a hydrophilic reside. The five sequences are given in Table 1. Sequence A consists of 10 hydrophobic residues, while sequence E consists of 10 hydrophilic residues. Sequences B, C, and D are mixtures of hydrophobic and hydrophilic residues, with, in order, increasing hydrophilic content.

Table 3. Energetic Parameters

torsion potential	kcal/mol	
$v_{2,\phi}$	0.00	$-\pi < \phi < 0$
$v_{2,\phi}$	0.20	$0 < \phi < \pi$
$v_{3,\phi}$	0.45	$-\pi < \phi < -\pi/3$
$v_{3,\phi}$	4.00	$-\pi/3 < \phi < \pi/3$
$v_{3,\phi}$	0.45	$\pi/3 < \phi < \pi$
$v_{2,\psi}$	0.00	$-\pi < \psi < 0$
$v_{2,\psi}$	0.00	$0 < \psi < \pi$
$v_{3,\psi}$	1.50	$-\pi < \psi < -\pi/3$
$v_{3,\psi}$	4.00	$-\pi/3 < \psi < \pi/3$
$v_{3,\psi}$	0.45	$\pi/3 < \psi < \pi$
v_ω	40.0	

chiral potential	kcal/mol	degrees
k_χ	100.0	
$\chi_{0,i}$		52.52

van der Waals potential	kcal/mol
ϵ	0.060
ϵ_{local}	0.033

The interaction potentials can be divided into two types: local and nonlocal

$$V = V_{local} + V_{nonlocal} \tag{4}$$

The local interaction potentials consist of the bond angle, dihedral angle, 1−4 van der Waals, and an improper dihedral potential to maintain the chirality of the side chain

$$V_{local} = V_{BA} + V_\phi + V_\psi + V_\omega + V_{vdW-local} + V_\chi \tag{5}$$

The values for the various structural and energetic parameters are given in Tables 2 and 3. The bond angle potential is harmonic about the equilibrium bond angle and is given by

$$V_\theta = \sum \frac{1}{2} k_\theta (\theta_i - \theta_{0,i}) \tag{6}$$

The dihedral angle energy is given by

$$V_{TOR} = V_\phi + V_\psi + V_\omega \tag{7}$$

with

$$V_\phi = \sum_i \frac{1}{2}[v_{2,\phi}(1 - \cos 2\phi_i) + v_{3,\phi}(1 + \cos 3\phi_i)] \tag{8}$$

$$V_\psi = \sum_i \frac{1}{2}[v_{2,\psi}(1 - \cos 2\psi_i) + v_{3,\psi}(1 + \cos 3\psi_i)] \tag{9}$$

$$V_\omega = \sum_i \frac{1}{2} v_\omega (1 + \cos \omega_i) \tag{10}$$

The values for $v_{2,\phi}$, $v_{3,\phi}$, $v_{2,\psi}$, $v_{3,\psi}$, and v_ω were carefully chosen to produce, in conjunction with the van der Waals potential, a Ramachandran plot with realistic energy barriers for the alanine dipeptide,[29] shown in Figure 2. The barrier height between the α-helix and β-sheet regions is 4.2 kcal/mol. These values are also given in Table 3. The box in Figure 2 defines the α-helical region and is centered on the

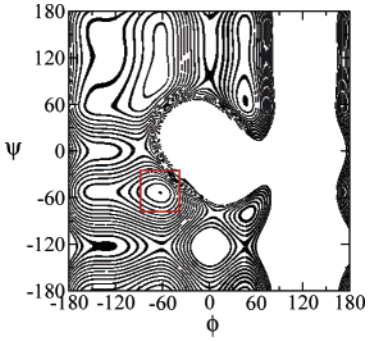

Figure 2. Ramachandran plot for the alanine dipeptide. Contour lines are 0.5 kcal/mol apart. The global minimum is at $\phi = -63$, $\psi = -54$. The "box" defines the α-helical region (see text for details).

average values of $\langle\phi\rangle_\alpha = -63°$ and $\langle\psi\rangle_\alpha = -54°$, where the subscript α indicates the averages are over helical configurations. The box in Figure 2 defines the α-helical region and marks out the area 25 degrees on either side of the minimum. Finally, the chirality is preserved via a harmonic potential for the improper dihedral angle formed by the vectors connecting the α-carbon and the nitrogen, the carbonyl carbon and the nitrogen, and the β-carbon and the α-carbon. The potential has the form

$$V_\chi = \sum \frac{1}{2} k_\chi (\chi_i - \chi_{0,i}) \tag{11}$$

Values for k_χ and $\chi_{0,i}$ are given in Table 3.

The nonlocal interaction has three contributions: the van der Waals interaction, hydrogen bonding, and the hydrophobic effect

$$V_{\text{nonlocal}} = V_{\text{vdW}} + V_{\text{HB}} + V_{\text{HP}} \tag{12}$$

The novel feature of the model developed by Takada et al.[21] is the dependence of the hydrogen bonding scheme and the hydrophobic interaction on the local peptide density, here referred to as ρ_I for residue I. The strength of both interactions is scaled by a function $0 \le S(\rho) \le 1$, which depends on the local density. For example, in the hydrogen bond potential, S is small for low peptide densities and equal to 1 for high densities. In this way, the competition for hydrogen bond formation between the (implicit) solvent and other hydrogen bond donors on the peptide is mimicked.

In full, the hydrogen bond potential is

$$V_{\text{HB}} = \epsilon_{\text{HB}} \sum_{ij(I \ge J+3)} S_{\text{HB},IJ} u_{\text{HB}}^{(\text{a,r})}(r_{ij}) + \frac{1}{2} \epsilon_{\text{HB}} \sum_I S_{\text{HB,c}}(\rho_I) \tag{13}$$

where i is the ith atom and is located in residue I. The strength of the interaction is given by ϵ_{HB}. The scaling function $S_{\text{HB},IJ}$ is given by $S_{\text{HB},IJ} = [S_{\text{HB}}(\rho_I) + S_{\text{HB}}(\rho_J)]/2$. The distance-dependent interaction between i and j is

$$u_{\text{HB}}^{(\text{a})}(r_{ij}) = 5\left(\frac{\sigma_{\text{HB}}}{r_{ij} - r_{\text{HB}}}\right)^{12} - 6\left(\frac{\sigma_{\text{HB}}}{r_{ij} - r_{\text{HB}}}\right)^{10} \tag{14}$$

and

$$u_{\text{HB}}^{(\text{r})}(r_{ij}) = 3\left(\frac{\sigma_{\text{HB}}}{r_{ij} - r_{\text{HB}}}\right)^{10} \tag{15}$$

The superscript (a) stands for attractive and (r) for repulsive. The attractive and repulsive forces in the hydrogen bond potential were introduced by Takada et al.[21] to model the anisotropy of the hydrogen bond. The attractive potential, eq 14, is used for the interaction between a carbonyl carbon and an amide nitrogen, while the repulsive potential, eq 15, is used for the interaction between an amide nitrogen with another amide nitrogen, between an amide nitrogen and an α-carbon, between a carbonyl carbon and another carbonyl carbon, and between a carbonyl carbon and an α-carbon. For a more indepth description of the hydrogen bond potential, readers are referred to the original work of Takada et al.[21]

The scaling function $S_{\text{HB}}(\rho_I)$, where $\rho_I = \sum_K u_{\text{HP}}(r_{\text{IK}})$, is defined as

$$S_{\text{HB}}(x, x_{\min}, x_{\max}) = \begin{bmatrix} 0 & \text{if } x < x_{\min} \\ \frac{1}{2}\left(1 + \cos\left(\pi \frac{x_{\max} - \boldsymbol{x}}{x_{\max} - x_{\min}}\right)\right) & \text{if } x_{\min} \le x \le x_{\max} \\ 1 & \text{if } x > x_{\max} \end{bmatrix} \tag{16}$$

The second term in eq 13 is a penalty term that accounts for buried non-hydrogen-bonded pairs. The function $u_{\text{HP}}(r)$ is defined below.

There are two contributions to the hydrophobic interaction, one from the side chains and one from the α-carbons. The interaction is given by

$$V_{\text{HP}} = \sum_I \delta_I \epsilon_{\text{HP},I}^{(\alpha)} S_{\text{HP}}(\rho_I) + \sum_\mu \delta_\mu \epsilon_{\text{HP},\mu}^{(\beta)} S_{\text{HP}}(\rho_\mu) \tag{17}$$

where I represents α-carbons and μ represents side-chain interaction sites. The parameter $\delta_{I,\mu}$ is equal to 1 for hydrophobic residues and -1 for hydrophilic residues. Similar to that used in the hydrogen bonding scheme, the scaling function $S_{\text{HP}}(\rho_I)$, where $\rho_I = \sum_K u_{\text{HP}}(r_{IK})$, is defined as

$$S_{\text{HP}}(x, x_{\max}) = \begin{bmatrix} \cos\left(\frac{\pi}{2} \frac{x_{\max} - x}{x_{\max}}\right) & \text{if } x \le x_{\max} \\ 1 & \text{if } x > x_{\max} \end{bmatrix} \tag{18}$$

The switching function u_{HP} used in both the hydrogen bonding scheme and the hydrophobic interaction is defined by

$$u_{\text{HP}} = \begin{bmatrix} 1 & \text{if } r < \sigma_{\text{HP1}} \\ \frac{1}{2}\left(1 + \cos\left(\pi \frac{r - \sigma_{\text{HP1}}}{\sigma_{\text{HP2}} - \sigma_{\text{HP1}}}\right)\right) & \text{if } \sigma_{\text{HP1}} < r < \sigma_{\text{HP2}} \\ 0 & \text{if } r > \sigma_{\text{HP2}} \end{bmatrix} \tag{19}$$

Finally, the van der Waals interaction is given by

$$V_{\text{vdW}} = \sum_{i,j>i} \phi_{ij}(r) \tag{20}$$

where

$$\phi_{ij}(r) = 4\epsilon\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad (21)$$

The cross-diameter σ_{ij} is given by $\sigma_{ij} = [\sigma_i + \sigma_j]/2$. For interactions between particles connected by three covalent bonds (1−4 pairs), the interaction strength ϵ and the diameter σ_{ij} are replaced by their reduced or "local" counterparts ϵ_{local} and $\sigma_{ij,\text{local}}$. The reduced parameters are introduced because these short-range interactions are better modeled by using the atomic parameters instead of the united atom parameters. Values for these parameters are given in Tables 2 and 3.

3. Langevin Dynamics

The motion of the model peptide is described using Langevin dynamics. The physical interaction between the solute molecule and solvent is mimicked by a random force, Γ, and solvent viscosity is modeled by a damping term with a coefficient ζ. The equation of motion for a generalized coordinate x_i is

$$m\ddot{x}_i = F_i + \Gamma_i - \zeta_i\dot{x}_i \quad (22)$$

The force on particle i due to the molecular configuration is represented by F_i. The friction constant ζ_i is related to the viscosity of water, η, by Stoke's Law

$$\zeta_i = 6\pi a_i\eta \quad (23)$$

where a_i is the effective radius of each particle and is equivalent to the sum of the van der Waals radius and the radius of a water molecule, 1.4 Å. As is usual in Langevin dynamics, the random force Γ_i has a mean of zero and a variance of

$$\langle\Gamma_i(t)\Gamma_i(t')\rangle = 2\zeta_i k_B T\delta(t - t') \quad (24)$$

Equation 22 is solved using the Velocity Verlet algorithm. The position at time $t + h$ is given by[30]

$$x_i(t + h) = x_i(t) + h\dot{x}_i(t) + \frac{h^2}{2m}[F_i + \Gamma_i - \zeta_i\dot{x}_i(t)] \quad (25)$$

Similarly, the velocity at time $t + h$ is given by

$$\dot{x}_i(t + h) = (1 - \alpha)(1 - \alpha + \alpha^2)\dot{x}_i(t) + \frac{h}{2m}(1 - \alpha + \alpha^2)[F_i(t) + \Gamma_i(t) + F_i(t + h) + \Gamma_i(t + h)] \quad (26)$$

where we have used the shorthand $\alpha = h\zeta_i/2m$. Bond lengths are held fixed via the RATTLE algorithm.

To improve the sampling of phase space, the Replica Exchange Method[31−34] is used. In this method, several *noninteracting* replicas are simulated in parallel, each at a different temperature. At regular intervals, a Monte Carlo exchange step is attempted between two replicas, say i and j, at neighboring temperatures, T_i and T_j. The transition probability of this replica exchange is given by

$$W(X \rightarrow X') = \begin{bmatrix} 1 & \text{if } \Delta \le 0 \\ \exp\{-\Delta\} & \text{if } \Delta > 0 \end{bmatrix} \quad (27)$$

where

$$\Delta = (\beta_i - \beta_j)[E_j - E_i] \quad (28)$$

Here, E_i is the potential energy for replica i at temperature $\beta_i = 1/kT_i$. The temperatures are chosen to be equally spaced on a logarithmic temperature scale. Exchanges are attempted every 10 ps, and the replica exchange acceptance ratios vary from 15% to 40%.

4. Results and Analysis

4.1. Thermodynamics. The thermodynamic and structural properties of all five peptides were studied using replica exchange MD. The fluctuations in the total energy and in the molecular configuration were measured. The first, given by the heat capacity C_v, is commonly used to determine the location of the collapse transition T_θ. The heat capacity is defined by

$$C_v(T) = \frac{\partial E_{\text{total}}}{\partial T} = \frac{\langle E^2\rangle - \langle E\rangle^2}{k_B T^2} \quad (29)$$

where the first equality is from thermodynamics, and the second is from statistical mechanics. Figure 3 shows the heat capacity for all five peptides as a function of the temperature. The peak in C_v is the collapse transition temperature. The data shown in Figure 3 were determined via the second equality and were subjected to the weighted histogram analysis method (WHAM). The heat capacity was also determined via the first equality and found to be in agreement with the statistical mechanical definition.

The peak in the conformational fluctuations of the peptide is used to determine the folding temperature, T_f, below which the polypeptide is predominantly in the native configuration. A measure of how much a given conformation differs from the native state is given by the parameter χ, called the "overlap function". There is no unique way of defining such a parameter, though all reasonable definitions lead to similar results. We follow Vietshans et al.[30] in defining χ as

$$\chi = \frac{1}{N_\alpha{}^2 - 5N_\alpha + 6}\sum_{i=1}^{N_\alpha-3}\sum_{j=i+3}^{N_\alpha}\Theta(\epsilon - |r_{ij} - r_{ij}^N|) \quad (30)$$

Here, N_α corresponds to the number of α-carbons, r_{ij} is the distance between α-carbons i and j, and r_{ij}^N is the same distance in the native state. Θ is the Heaviside function and is equal to 1 when its argument is positive and is equal to zero otherwise. Specifically, the Heaviside function is 1 when the difference between the pair distance r_{ij} and distance between the same pair in the native state, r_{ij}^N, is less than some tolerance ϵ. Thus, only "nativelike" pair-distances contribute to the sum in eq 30. The parameter ϵ is set to 0.5 Å. Note that χ is equal to 1 in the native state. We define the native state as a helix with ϕ and ψ angles of 63° and 54°, respectively. The fluctuations in χ are measured by

$$\Delta\chi = \langle\chi^2\rangle - \langle\chi\rangle^2 \quad (31)$$

The behavior of $\Delta\chi$ is shown in Figure 4. The temperature of the peak in $\Delta\chi$ for each sequence is the same as that for the heat capacity. Values for the folding temperature are given in Table 1. Note that the location of the peaks for both

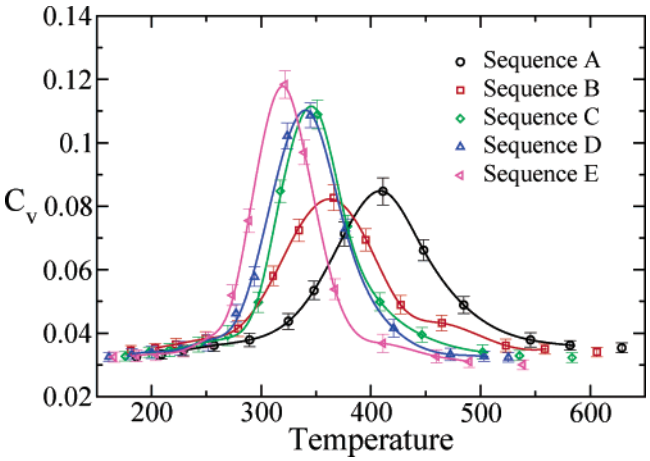

Figure 3. A plot of the heat capacity C_v versus the temperature. The peak in C_v is the collapse temperature, T_θ. The circles are sequence A (black), the squares are sequence B (red), the diamonds are sequence C (green), the up triangles are sequence D (blue), and the left triangles are sequence E (magenta). The colors and symbols are consistent throughout.

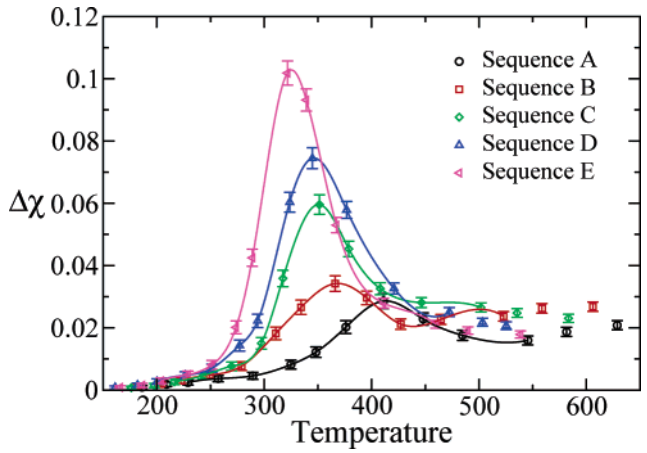

Figure 4. The fluctuations $\Delta\chi$ plotted versus the temperature. The peak in $\Delta\chi$ is the folding temperature, T_f. In this model, $T_f = T_\theta$. Note that the magnitude of the fluctuations increases, and the width of the transition region decreases as the hydrophilic content increases. The symbols are the same as in Figure 3.

C_v and $\Delta\chi$ shift to higher temperatures with increasing hydrophobicity. The more hydrophilic peptides are more easily solvated and consequently have a less stable α-helix and a folding transition at a lower temperature. The relative heights of the peaks are also consistent with the peaks in the heat capacity: the more hydrophobic sequences have smaller fluctuations at the folding transition than the more hydrophilic sequences. This is to be expected, as the more hydrophilic the peptide, the more readily it will be solvated. Note that while sequences C and D have similar transition temperatures, the width of the transition region for sequence C is narrower than that for sequence D.

Comparison with a plot of the radius of gyration, R_g, as a function of temperature, shown in Figure 5, clearly shows that R_g changes dramatically within the transition region centered on T_θ. Above T_θ, the radius of gyration increases with increasing temperature with the rate of increase greater

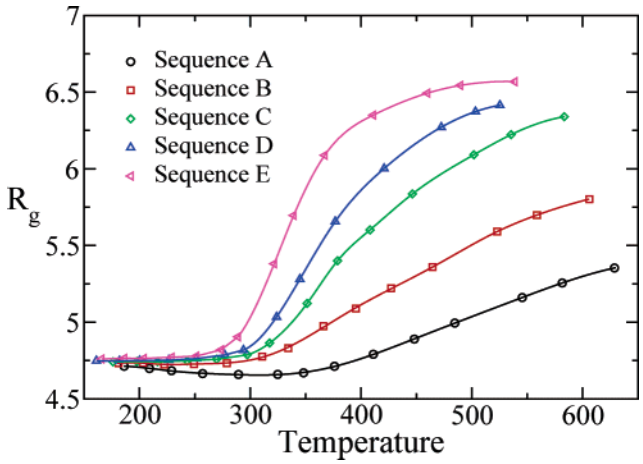

Figure 5. The radius of gyration R_g as a function of temperature. Note that R_g for sequence A has a mimumim. The symbols are the same as in Figure 3.

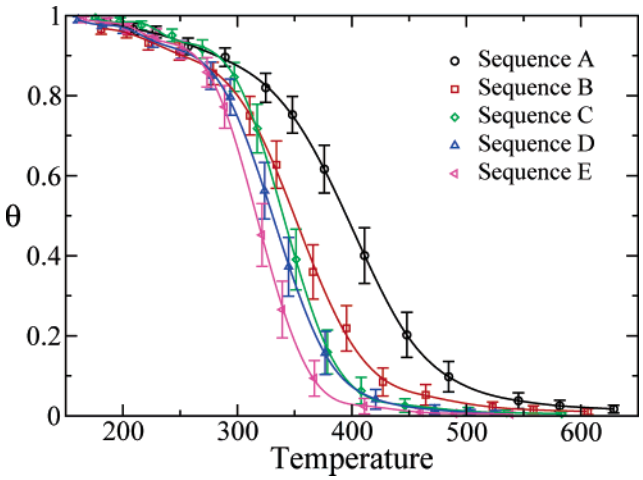

Figure 6. A plot of the fractional helicity θ versus temperature. The width of the transition region decreases with increasing hydrophilic content. The symbols and colors are the same as in Figure 3.

the more hydrophilic the peptide sequence. Below T_θ, the peptides all have a constant radius of gyration equal to that of a full α-helix, while above T_θ, the radius of gyration increases with increasing T. One interesting feature of the radius of gyration for sequence A is that it passes through a minimum at 305 K. This is due to the strong hydrophobic attraction of the P side chains which cause the α-helix to compress slightly. At very low temperatures, this compression is canceled by the bond- and dihedral angle potentials, which become increasingly important as the temperature decreases.

Figure 6 shows the behavior of the average helicity θ as a function of temperature. The helicity is defined as

$$\theta = \frac{N_H}{N_{res} - 4} \tag{32}$$

where N_H is the number of helical hydrogen bonds and N_{res} is the number of residues in the polypeptide chain. For each peptide sequence, $\theta = 1$ at low temperatures, indicating a full α-helix, and $\theta \approx 0$ at high temperatures, where each

peptide is fully unfolded and has no α-helical hydrogen bonds. The width of the transition region decreases with increasing hydrophobicity. Due to the short length of the peptides, sequence effects play a role in the behavior of θ. For example, sequence B has the lowest value for θ at low temperatures when it would be expected, as a more hydrophilic sequence, to have a value comparable to sequence A. This is due to fraying of the helix at the N-terminus.

One of the most useful models to describe the coil-to-helix transition is due to Zimm and Bragg.[2,5] Two parameters are central to their analysis: s, which is related to the free energy of helix propagation, and σ, which is related to the free energy of helix nucleation.

The fractional helicity for a peptide of N residues is given by

$$\theta = \frac{1}{N}\frac{\partial \ln Q}{\partial \ln s} \tag{33}$$

where Q is the partition function and can be written as a sum over eigenfunctions of the transition matrix

$$Q = \sum_n \lambda_n^N \tag{34}$$

In the case of the Zimm−Bragg model, $n = 2$, the partition function can be written as

$$Q = \lambda_+^N + \lambda_-^N \tag{35}$$

where

$$\lambda_\pm = \frac{1}{2}(1+s) \pm \frac{1}{2}\sqrt{(1-s)^2 + 4\sigma s} \tag{36}$$

To determine s and σ, eq 33 was fit to the fractional helicity determined via simulation and shown in Figure 6. Following the analysis of Takano et al.,[5] we have used the Levenburg-Marquardt nonlinear least-squares algorithm[35] to fit eq 33.

The behaviors of s and σ as a function of temperature are shown in Figures 7 and 8, respectively. In agreement with Ohkubo and Brooks,[36] we find that the large-N approximation for s and σ used in previous work[37−39] is not suitable for chain lengths of 10 residues. The behavior of s is qualitatively similar to that seen by Ohkubo and Brooks.[36] For a given T, the more hydrophobic sequence generally has a higher value of s. The exception is for sequence B, since s was determined from the fractional helicity, and the fractional helicity was low at low temperatures due to fraying of the helix ends. At the lowest temperatures sampled, s is between 1.5 and 2.0 for all sequences. As the temperature increases, s decreases, passing through 1.0 at a lower temperature than the folding temperature as determined by the peak in $\Delta\chi$. At the highest temperatures sampled, s is approximately 0.5, which is higher than expected, though the more hydrophilic the sequence, the smaller the value of s.

The behavior of σ is very interesting. At low and high temperatures, σ is appoximately 0.005 for the most hydrophobic sequence and 0.001 for the most hydrophilic sequence. However, at the folding temperature, σ has a maximum. The most hydrophobic sequence has a maximum

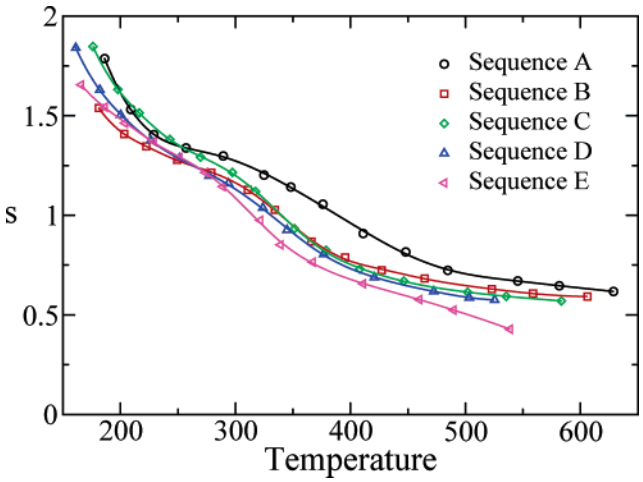

Figure 7. The Zimm−Bragg paramater s plotted as a function of temperature. A measure of the folding temperature is where $s = 1$. In this case, $T(s = 1)$ is generally lower than the folding temperature as determined by the peak in the heat capacity. The symbols and colors are the same as in Figure 3.

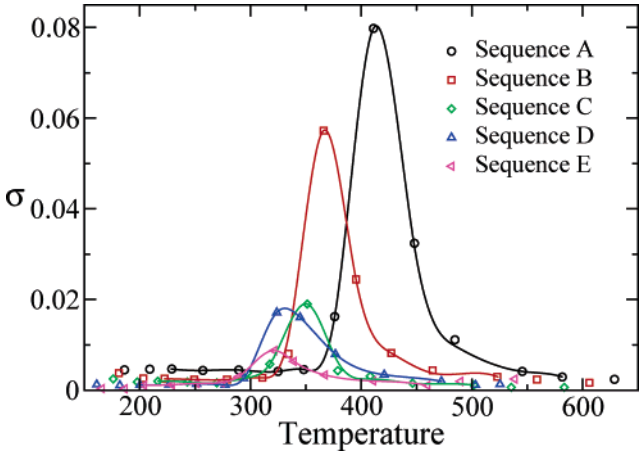

Figure 8. The Zimm−Bragg parameter σ shows a strong peak at the folding temperature. The magnitude of the peak increases as the hydrophobic content increases. The symbols and colors are the same as in Figure 3.

at $\sigma = 0.08$, while the most hydrophilic sequence has a peak of only $\sigma = 0.009$. The behavior for σ of sequence A is similar to that seen by Ohkubo and Brooks.[36] However, the value of σ at high temperatures determined here is considerably smaller than seen in that work.

4.2. Kinetics. To determine the sequence dependence of the kinetics of the coil-to-helix transition, a procedure similar to that developed by Veitshans et al.[30] was used. For each sequence, $M = 400$ independent initial configurations, generated from a high-temperature simulation run, were quenched to the folding temperature and were allowed to propagate forward via eqs 25 and 26 until the overlap function χ was equal to 1.0, whereupon the simulation was stopped. The fraction of unfolded peptides as a function of time, $P_u(t)$, was then used to characterize the folding kinetics. $P_u(t)$ is defined as

$$P_u(t) = 1 - \int_0^t P_{fp}(s)ds \tag{37}$$

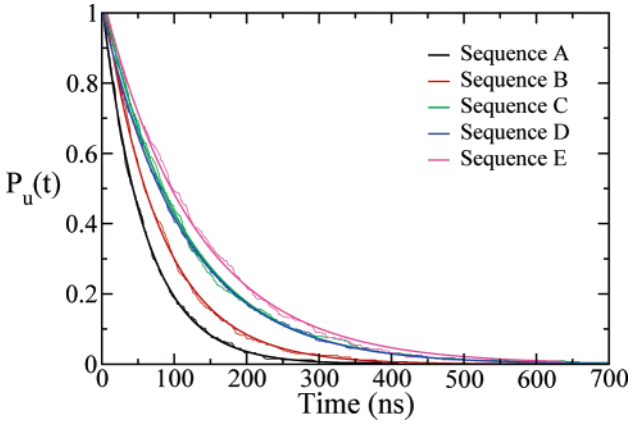

Figure 9. A plot of the fraction of unfolded peptides $P_u(t)$ as a function of simulation time (in nanoseconds). Each sequence displays single-exponential kinetics. Black is sequence A, red is sequence B, green is sequence C, blue is sequence D, and magenta is sequence E.

where $P_{fp}(t)$ is the distribution of first passage times

$$P_{fp}(s) = \frac{1}{M}\sum_{i=1}^{M}\delta(s - \tau_i) \tag{38}$$

and where τ_i is the first passage time of trajectory i. $P_u(t)$ for each sequence was then fit to a single exponential

$$P(t) = A_0\exp(-t/\tau_0) \tag{39}$$

where we shall refer to τ_0 as the mean folding time. Unlike Veitshans et al.,[30] $P_u(t)$ was best fit by a *single* exponential. Single-exponential folding kinetics were also seen by Bredenbeck et al.[40] in the folding of a 16-residue helix-forming peptide. Both $P_u(t)$ and the best-fit exponential for each sequence are shown in Figure 9. That the kinetics is best described by a single exponential is in line with what would be predicted from the value of σ_f, a measure of the "foldability" of the peptide. The parameter σ_f introduced by Thirumalai is defined by

$$\sigma_f = \frac{T_\theta - T_f}{T_\theta} \tag{40}$$

where T_θ and T_f are the collapse and folding temperatures determined in the previous section. For these model peptides, $T_\theta = T_f$, and $\sigma_f = 0$. Peptides with $\sigma_f \approx 1$ encounter misfolded structures, some of which can be very stable and which serve as kinetic traps.[30] Peptides with $\sigma_f \approx 0$ have folding kinetics that show two-state behavior,[30] which is best fit by a single exponential.

As expected, the most hydrophobic sequence has the fastest folding time, while the most hydrophilic sequence has the slowest. The folding times are given in Table 1 and vary from 59 ns for the fastest to 132 ns for the slowest. Sequences B, C, and D behave as expected, with the folding time increasing with increasing hydrophilic content. The data for sequences C and D lie on top of each other, though sequence C does fold slightly faster.

Two main folding pathways were observed for the coil-to-helix transition. The first, by which approximately 80% of the trajectories folded, began with a single helix nucleation site, followed by the growth of the helix to encompass the entire peptide. This mechanism is illustrated in Figure 10. In this figure, the unfolded peptide is shown in part A, a structure with a single helical hydrogen bond near the center of the peptide is shown in part B. In part C, the helix has grown in both directions and includes one of the end termini, and finally in part D, the full helix. The second pathway is the more interesting of the two, as helix formation begins with *two* helix nucleation sites. This pathway is shown in Figure 11. In part A of that figure, the unfolded peptide is shown. In part B, the first helix nucleation site is formed at the C-terminus. Before this helix can grow to include the entire peptide, a second helix is nucleated at the N-terminus, shown in part C. Given the small size of the peptide, multiple helix nucleation can only occur with a nucleation site at each terminus. Finally, in part D, the two helices meet to form a full helix. Multiple folding pathways have been seen in larger proteins,[41] and peptides with multiple helical structures were seen by Nymeyer and Garcia at low temperatures in simulations of the folding of a 21-residue helical peptide.[42]

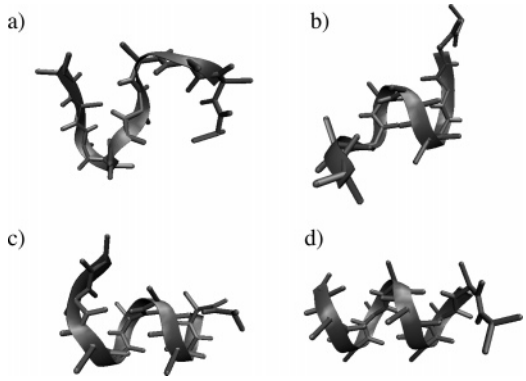

Figure 10. An illustration of a folding pathway that proceeds from a single nucleation site. Part (a) is the unfolded peptide, part (b) shows a single helical turn, in part (c) the peptide has grown in each direction, and part (d) shows the fully folded peptide.

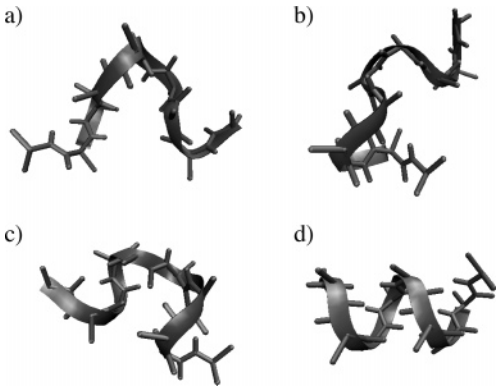

Figure 11. An illustration of a folding pathway that proceeds via multiple helix nucleation sites. Part (a) is the unfolded peptide, part (b) shows a single nucleated helix at the C-terminus, in part (c) the peptide has *two* helical segments, one at each terminus, and part (d) shows the fully folded peptide.

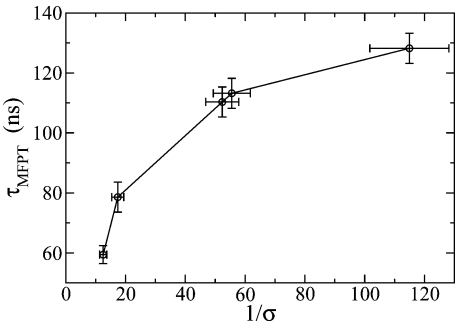

Figure 12. A plot of the mean folding time (in nanoseconds) versus the inverse of the Zimm−Bragg parameter σ. In agreement with the work of Buchete and Staub,[26] we find that τ is not simply proportional to $1/\sigma$.

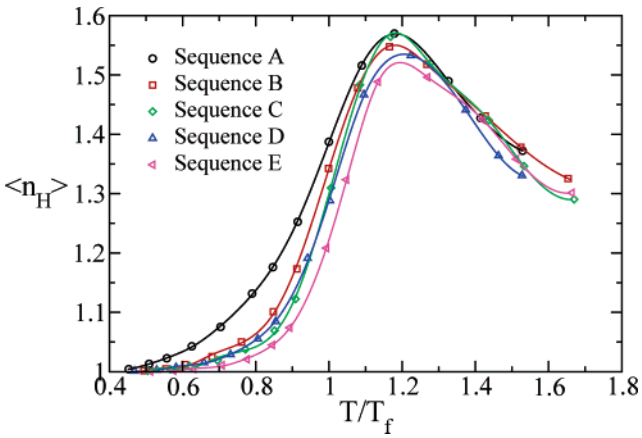

Figure 13. A plot of the average number of helical segments versus the reduced temperature T/T_f for each sequence. As the hydrophobic content of the sequence increases, so too does the average number of helices. The symbols and colors are the same as in Figure 3.

Finally, a plot of the mean folding time versus the Zimm−Bragg parameter $1/\sigma$ for all five sequences at their folding temperature is shown in Figure 12. In agreement with the results of Buchete and Straub,[26] we find that τ is not linear in $1/\sigma$. Sequences C and D lie very close to each other in Figure 12. They have very similar folding temperatures, very similar values of σ at the folding temperature, and very similar mean folding times. The origin of this similarity lies in the details of their sequences and in the relatively short chain length of the peptides studied. Sequences C and D differ only by 2 residues, residues 3 and 7, which are hydrophobic in sequence C and hydrophilic in sequence D. Changing the identity of two internal residues (which is 20% of the sequence) does not have much of an effect. In contrast, the sequence B differs from sequence C by 8 residues, and sequence A from sequence B by 3 residues, one of which is the N-terminus. There are marked differences between the behavior of these sequence pairs. Using longer chain lengths would give a clearer indication of the importance of amino acid sequence versus hydrophobic/hydrophilic content. It is certain, however, that increasing hydrophobic content increases the helix stability and decreases the mean folding time.

In their work, Buchete and Straub[26] identify the inclusion of multiple nucleation sites as the origin of the nonlinear behavior. To test whether this is a possible mechanism in these simulations, we can look at the average number of helical segments, $\langle n_H \rangle$ (defined as three or more consecutive residues in a helical state), at the folding temperature for each sequence. This is plotted versus the reduced temperature T/T_f in Figure 13. The more hydrophobic sequences, which have higher values of σ, do indeed have a higher average number of helical segments at the folding temperature than the more hydrophilic sequences. For example, sequence A has $\langle n_H \rangle = 1.39$, while sequence E has $\langle n_H \rangle = 1.21$. That all sequences have $\langle n_H \rangle > 1.0$ at the folding temperature indicates that multiple helix nucleation sites are important even for short chains such as those studied here.

5. Conclusion

A coarse-grained model peptide was introduced and used in a series of molecular dynamics simulations. For five different peptide sequences, each 10 residues in length and of varying hydrophobic/hydrophilic content, the thermodynamics of the coil-to-helix transition was characterized and the folding temperature was determined. The folding temperature, and hence the helix stability, increased with increasing hydrophobic content of the peptide. For each sequence, 400 independent configurations were simulated at their folding temperature, and the time was measured for each conformation to go from a random coil to an α-helix. From this distribution of folding times, the kinetics of the coil-to-helix transition was characterized. Folding times varied from 59 to 132 ns, which is slightly faster than the 200 ns established by experiment. However, given the small size (10 residues) of the peptides in this work, and the longer (16−21 residues) peptides used in the experimental work, this faster folding time is not unreasonable. While the coarse-grained interaction potentials used in this work are clearly approximate, especially the directionality of the hydrogen bond, the thermodynamic and kinetic results presented here indicate that it offers a realistic description of the coil-to-helix transition. For all five peptide sequences studied, single-exponential kinetics were observed, indicating a "two-state" folding process. Finally, the mean folding time was plotted versus the inverse of the Zimm−Bragg parameter σ, and a nonlinear dependence was found. The origin of the nonlinearity was ascribed to multiple helix nucleation sites, and pathways proceeding from single and from multiple helix nucleation sites were discussed. It was further shown that increasing values of σ correlated with an increasing average number of helical segments, indicating that even for small peptides such as those studied here, multiple helix nucleation sites play an important role in the folding kinetics.

Acknowledgment. This work was funded by the NIH (R01 NS41356) and the Center for Computational Science at Boston University.

References

(1) Schellman, J. A. The factors affecting the stability of hydrogen-bonded polypeptide structures in solution. *J. Phys. Chem.* **1958**, *62*, 1485−1494.

(2) Zimm, B. H.; Bragg, J. K. Theory of the phase transition between helix and random coil in polypeptide chains. *J. Chem. Phys.* **1959**, *31*, 526−535.

(3) Lifson S.; Roig, A. On the theory of helix-coil transition in polypeptides. *J. Chem. Phys.* **1960**, *34*, 1963−1974.

(4) Doig, A. J. Recent advances in helix-coil theory. *Biophys. Chem.* **2002**, *101*−*102*, 281−293.

(5) Takano, M.; Nagayama, K.; Suyama, A. Investigating a link between all-atom model simulation and the Ising-based theory on the helix-coil transition: Equilibrium statistical mechanics. *J. Chem. Phys.* **2002**, *116*, 2219−2228.

(6) Zana, R. On the rate-determining step for helix propagation in the helix-coil transition of polypeptides in solution. *Biopolymers* **1975**, *14*, 2425−2428.

(7) Williams, S.; Causgrove, T. P.; Gilmanshin, R.; Fang, K. S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B. Fast events in protein folding: helix melting and formation in a small peptide. *Biochemistry* **1996**, *35*, 691−697.

(8) Thompson, P. A.; Eaton, W. A.; Hofrichter, J. Laser temperature jump study of the helix-coil kinetics of an Alanine peptide interpreted with a 'kinetic zipper' model. *Biochemistry* **1997**, *36*, 9200−9210.

(9) Lednev, I. K.; Karnoup, A. S.; Sparrow, M. C.; Asher, S. A. α-helix peptide folding and unfolding activation barriers: a nanosecond UV resonance Raman study. *J. Am. Chem. Soc.* **1999**, *121*, 8074−8086.

(10) Clarke, D. T.; Doig, A. J.; Stapley, B. J.; Jones, G. R. The α-helix folds on the millisecond time scale. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 722−7237.

(11) Gilmanshin, R.; Williams, S.; Callender, R.; Woodruff, W.; Dyer, R. Fast events in protein folding: Relaxation dynamics of secondary and tertiary structure in native apomyoglobin. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 3709−3713.

(12) Dyer, R.; Gai, F.; Woodruff, W.; Gilmanshin, R.; Callender, R. Infrared Studies of Fast Events in Protein Folding. *Acc. Chem. Res.* **1998**, *31*, 709−716.

(13) Callender, R.; Dyer, R.; Gilmanshin, R.; Woodruff, W. FAST EVENTS IN PROTEIN FOLDING: The Time Evolution of Primary Processes. *Annu. Rev. Phys. Chem.* **1998**, *49*, 173−202.

(14) Huang, C.-Y.; Klemke, J. W.; Getahun, Z.; DeGrado, W. F.; Gai F. Temperature-dependent helix-coil transition of an alanine based peptide. *J. Am. Chem. Soc.* **2001**, *123*, 9235−9238.

(15) Huang, C.-Y.; Getahun, Z.; Wang, T.; DeGrado, W. F.; Gai, F. Time-resolved infrared study of the helix-coil transition using ^{13}C-labeled helical peptides. *J. Am. Chem. Soc.* **2001**, *123*, 12111−12112.

(16) Huang, C.-Y.; Getahun, Z.; Zhu, Y.; Klemke, J. W.; DeGrado, W. F.; Gai, F. Helix formation via conformation diffusion search. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2788−2793.

(17) Wang, T.; Du, D.; Gai F. Helix-coil kinetics of two 14-residue peptides. *Chem. Phys. Lett.* **2003**, *370*, 842−848.

(18) Wang, T.; Zhu, Y.; Getahun, Z.; Du, D.; Huang, C.-Y.; DeGrado, W. F.; Gai, F. Length dependent helix-coil transition kinetics of nine alanine-based peptides. *J. Phys. Chem. B* **2004**, *108*, 15301−15310.

(19) Daggett, V.; Levitt, M. Molecular dynamics simulations of helix denaturation. *J. Mol. Biol.* **1992**, *223*, 1121−1138.

(20) Klimov, D.; Betancourt, M.; Thirumalai, D. Virtual atom representation of hydrogen bonds in minimal off-lattice models of alpha-helices: Effects on stability, cooperativity and kinetics. *Fold. Des.* **1998**, *3*, 481−496.

(21) Takada, S.; Luthey-Schulten, Z.; Wolynes, P. G. Folding dynamics with nonadditive forces: a simulation study of a designed helical protein and a random heteropolymer. *J. Chem. Phys.* **1999**, *110*, 11616−11629.

(22) Margulis, C. J.; Stern, H. A.; Berne, B. J. Helix unfolding and intramolecular hydrogen bond dynamics in small α-helices in explicit solvent. *J. Phys. Chem. B* **2002**, *106*, 10748−10752.

(23) Dohsi, U. R.; Muñoz, V. The principles of α-helix formation: explaining complex kinetics with nucleation-elongation theory. *J. Phys. Chem. B* **2004**, *108*, 8497−8506.

(24) Schwartz, G. Kinetics of the helix-coil transition. *J. Mol. Biol.* **1965**, *11*, 64−77.

(25) Brooks, C. Helix-Coil Kinetics: Folding Time Scales for Helical Peptides from a Sequential Kinetic Model. *J. Phys. Chem.* **1996**, *100*, 2546−2549.

(26) Buchete, N.-V.; Straub, J. E. Mean first-passage time calculations for the coil-to-helix transition: the active helix Ising model. *J. Phys. Chem.* **2001**, *105*, 6684−6697.

(27) Voegler Smith, A.; Hall, C. K. Protein refolding versus aggregation: computer simulations on an intermediate-resolution protein model. *J. Mol. Biol.* **2001**, *312*, 187−202.

(28) Honig, B.; Cohen, F. E. Adding backbone to protein folding: why proteins are polypeptides. *Fold. Des.* **1996**, *1*, R17−R20.

(29) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry; Part I*; Freedman: New York, 1980.

(30) Veitshans, T.; Klimov, D.; Thirumalai, D. Protein folding kinetics: time scales, pathways and energy landscapes in terms of sequence-dependent properties. *Fold. Des.* **1996**, *2*, 1−22.

(31) Swenson, R. H.; Wang, J.-S. Replica Monte Carlo simulation of spin-glasses. *Phys. Rev. Lett.* **1986**, *57*, 2607−2609.

(32) Hukushima, K.; Nemoto, K. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604−1608.

(33) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.

(34) Sugita, Y.; Okamoto, Y. Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscapes. *Chem. Phys. Lett.* **2000**, *329*, 261−270.

(35) Press: W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed.; Cambridge University Press: Cambridge, 2002.

(36) Ohkubo, Y. Z.; Brooks, C. L., III Exploring Flory's isolated-pair hypothesis: Statistical mechanics of helix-coil transitions in polyalanine and the C-peptide from RNase A. *Proc. Natl. Acad. Sci.* **2003**, *100*, 13916−13921.

(37) Hansmann, U. H. E.; Okamoto, Y. Finite-size scaling of helix-coil transitions in poly-alanine studied by multicanonical simulations. *J. Chem. Phys.* **1999**, *110*, 1267−1276.

(38) Mitsutake, A.; Okamoto, Y. Helix-coil transitions of amino-acid homo-oligomers in aqueous solution studied by multicanonical simulations. *J. Chem. Phys.* **1999**, *122*, 10638−10647.

(39) van Giessen, A. E.; J. E. Straub, J. E. Monte Carlo simulations of polyalanine using a reduced model and statistics-based interaction potentials. *J. Chem. Phys.* **2005**, *122*, 024904.

(40) Bredenbeck, J.; Helbing, J.; Kumita, J. R.; Woolley, G. A.; Hamm, P. α-Helix formation in a photoswitchable peptide tracked from picoseconds to microseconds by time-resolved IR spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2379−2384.

(41) Goldbeck, R. A.; Thomas, Y. G.; Chen, E.; Esquerra, R. M.; D. S. Kliger, D. S. Multiple pathways on a protein-folding energy landscape: kinetic evidence. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 2782−2787.

(42) Nymeyer, H.; Garcia, A. E. Simulation of the folding equilibrium of α-helical peptides: a comparison of the generalized Born approximation with explicit solvent. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13934−13939.

CT0503318