Virtual atom representation of hydrogen bonds in minimal off-lattice models of α helices: effect on stability, cooperativity and kinetics

DK Klimov, MR Betancourt and D Thirumalai

Background: The most conspicuous feature of a right-handed α helix is the presence of hydrogen bonds between the backbone carbonyl oxygen and NH groups along the chain. A simple off-lattice model that includes hydrogen bond interactions using virtual atoms is used to examine the stability, cooperativity and kinetics of the helix–coil transition.

Results: We have studied the thermodynamics (using multiple histogram method) and kinetics (by Brownian dynamics simulations) of 16-mer minimal off-lattice models of four-turn α-helix sequences. The carbonyl and NH groups are represented as virtual moieties located between two α-carbon atoms along the polypeptide chain. The characteristics of the native conformations of the model helices, such as the helical pitch and angular correlations, coincide with those found in real proteins. The transition from coil to helix is quite broad, which is typical of these finite-sized systems. The cooperativity, as measured by a dimensionless parameter, Ω_c, that takes into account the width and the slope of the transition curves, is enhanced when hydrogen bonds are taken into account. The value of Ω_c for our model is consistent with that inferred from experiment for an alanine-based helix-forming peptide. The folding time τ_F ranges from 6 to 1000 ns in the temperature range 0.7–1.9 T_F, where T_F is the helix–coil transition temperature. These values are in excellent agreement with the results from recent fast folding experiments. The temperature dependence of τ_F exhibits a nearly Arrhenius behavior. Thermally induced unfolding occurs on a time scale that is less than 40–170 ps depending on the final temperature. Our calculations also predict that, although τ_F can be altered by changes in the sequence, the dynamic range over which such changes take place is not as large as that predicted for β-turn formation.

Conclusions: Hydrogen bonds not only affect the stability of α-helix formation but also have profound influence on the kinetics. The excellent agreement between our calculations and experiments suggests that these models can be used to investigate the effects of sequence, temperature and viscosity on the helix–coil transition.

Address: Institute for Physical Science and Technology and Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA.

Correspondence: DK Klimov
E-mail: klimov@glue.umd.edu

Key words: helix–coil transition thermodynamics and kinetics, hydrogen bonds, Langevin simulations, minimal off-lattice protein models

Received: 17 August 1998
Accepted: 20 October 1998

Published: 25 November 1998
http://biomednet.com/elecref/1359027800300481

Folding & Design 25 November 1998, 3:481–496

Introduction

The use of minimal models of proteins has led to a number of predictions concerning the folding kinetics of proteins [1–5]. One of the most important predictions is that the assembly of proteins, under folding conditions, begins on very short time scales — of the order of tens of nanoseconds. In fact, it is possible that certain small proteins can fold in tens of microseconds. It is likely that on the submicrosecond time scale only secondary structures (helices or turns) form. A key question that arises from these observations is: What are the characteristics of certain sequences that enable them to fold rapidly to the native state? In terms of the energy landscape we can surmise that for these sequences there is a dominant native basin of attraction (NBA) or a folding funnel [1,3,4].

The quest to answer the question posed above has led to significant experimental developments. The answer is beginning to emerge from fast folding experiments pioneered by Eaton and coworkers [6–10] and further provided by others [11,12]. The basic strategy adopted by the experimentalists is to dissect the protein into its secondary structural elements and probe the dependence of the time scales for their formation as a function of sequence and external conditions. Such experiments have been performed for peptides that form α helices and β turns [8,9]. Munoz *et al.* [9] have measured the rate at which β turns can form in a model 16-residue peptide. They found that the β turn can form in about 6 μs. The dynamical range for β-turn formation can span several orders of magnitude depending on the sequence [13].

Williams *et al.* [14] measured the time for forming helical structure in a small 21-residue alanine-based peptide. The folding in this experiment was initiated by laser-induced temperature jump and the kinetics was probed by infrared laser spectroscopy. These experiments suggest that helix formation can occur in about 160 ns.

These local structural elements (α helices and β turns), which form on microsecond or less time scales, have profound influence on the collapse of small proteins [15], so it is crucial to decipher the factors that determine the rates of their formation. In addition, as helices and turns are the most common structural motifs in proteins, a detailed understanding of the factors that contribute to their formation is necessary before a complete picture of the assembly of an entire protein can emerge. In principle, it is necessary to perform fully atomic detailed simulations to understand the folding kinetics of helix- and β-turn-forming peptides. Such simulations of folding of small peptides are beginning to be reported [16–20]. These are extremely important developments and will prove useful in calibrating the success of minimal models of proteins. The room temperature atomic simulations of linear pentapeptides that form a type VI reverse turn show that these can form on the time scale of tens of nanoseconds (D Mohanty, R Elber, D Thirumalai, unpublished data). Such simulations are not yet routine, however. In their absence, the only recourse is to construct a coarse-grained description of polypeptides in the hope that some generic features of the folding kinetics can be gleaned from a detailed study of such models. Much of the focus of minimal models of proteins has been centered on lattice models [2]. Although models based on simple lattices provide some insight into the folding of proteins, they are woefully inadequate for addressing the issues of interest here. It is necessary, therefore, to develop and examine minimal off-lattice models whose ground states represent the elements of the common secondary structure in proteins [13,21–26].

We have recently initiated a series of studies using simple off-lattice models to assess the effects of sequence and external conditions in determining the kinetics of helix and β-turn formation [13,25]. By constructing a number of sequences, all of which have β turn as their folded conformation, we showed that diverse time scales can result depending on the precise sequences [13]. We also showed that under certain optimal circumstances, helices can form on a time scale of a few hundred nanoseconds [25]. These results showed that many of the aspects of folding seen in proteins can be mimicked in relatively small peptides — a conclusion that is in accord with the analysis of Munoz *et al.* [9]. Thus, the study of these sequences will be useful in elucidating some of the factors that control folding of proteins.

The models used in our earlier studies can be thought of as a coarse-grained α-carbon representation of proteins [13]. These models do not include effects due to hydrogen bonding, which are known to be important in stabilizing secondary structures. For these and other reasons, simple bead models of proteins (especially those on a lattice) have been criticized by Honig and Cohen [27]. The purpose of this paper is to include hydrogen bond effects using virtual atoms within the context of simple off-lattice models of α helices. We show, by explicit comparisons to the older models [23,25], that the inclusion of hydrogen bonds not only enhances the stability of the native state but also leads to speed up of the folding itself. More interestingly, the time scales for α-helix formation seen in our simulations are consistent with experimental measurement [10,14]. Thus, these models can be used systematically to predict sequence-dependent mechanisms and rates of formation of secondary structure elements.

Methodology

Model

A simple off-lattice model for protein is obtained by coarse-graining the degrees of freedom of the polypeptide chain in such a way that only the 'essential' features are retained. In addition to self-avoidance, the key degrees of freedom are: dihedral angles that mimic the rotation about the peptide plane; bond angles that describe the local flexibility (responsible for forming helices of appropriate pitch); and long-range attraction between hydrophobic residues. The minimal off-lattice models introduced by Honeycutt and Thirumalai some time ago [21,22], with subsequent development by others [13,23,25,26,28,29], incorporate these characteristics. In particular, they mimic the hydrophobic forces, excluded volume interactions, bond angle and dihedral angle degrees of freedom. In the current study, we incorporate hydrogen-bonding interactions using virtual atoms — an idea that was exploited by Flory in his treatment of polymers and polypeptides [30]. We now present a detailed description of the model.

The polypeptide is modeled as a chain consisting of N connected beads. For the α helix, we assume that a sequence contains two types of residues: hydrophobic (B) and hydrophilic (L). The potential energy of a conformation, which is specified by the set of vectors $\{\vec{r}_i\}$, $i = 1, 2 \ldots N$, is given by:

$$E_p(\{\vec{r}_i\}) = V_{BL} + V_{BA} + V_{DIH} + V_{HB} + V_{NON} \qquad (1)$$

where V_{BL}, V_{BA}, V_{DIH}, V_{HB} and V_{NON} correspond to bond length potential, bond angle potential, dihedral angle potential, hydrogen bond and non-bonded potentials, respectively. A brief summary of these interactions is given below.

Bond length potential. We use a stiff harmonic potential between successive residues, which keeps the bond length approximately fixed, that is:

$$V_{BL} = \sum_{i=1}^{N-1} \frac{k_r}{2}(|\vec{r}_{i+1} - \vec{r}_i| - a)^2 \quad (2)$$

where $k_r = 100\varepsilon_h/a^2$, a is the average bond length between two α-carbon beads, and ε_h, the average strength of the hydrophobic interaction, is the unit of energy in our model.

Bond angle potential. The potential associated with the angle between three successive beads i, i+1, i+2 is taken to be:

$$V_{BA} = \sum_{i=1}^{N-2} \frac{k_\theta}{2}(\theta_i - \theta_0)^2 \quad (3)$$

where $k_\theta = 20\varepsilon_h/(rad)^2$ and $\theta_0 = 1.8326\ rad$ or 105°. This value of the equilibrium bond angle is consistent with analysis based on the protein databank [31].

Dihedral angle potential. This potential describes the ease of rotation around the angle formed between four successive beads. This degree of freedom, together with hydrogen bonding, is largely responsible for determining secondary structures. Following our previous studies [23,25], we use two function forms for dihedral angle potential. The first was introduced in [23] and is given by:

$$V^A_{DIH} = \sum_{i=1}^{N-3} [A_i(1 - cos\phi) + B_i(1 + cos3\phi) + C_i(1 + cos(\phi + \pi/4))] \quad (4)$$

where $A_i = B_i = C_i = 1\varepsilon_h$ for all i, and ϕ is the dihedral angle. With this form of the dihedral potential the folding kinetics and thermodynamics of the *de novo* designed four-helix bundle were investigated [23]. It was shown that the resulting four-helix bundle was stable below the folding transition temperature. The isolated helix is unstable, however, which implies that the stability of the native state topology of the four-helix bundle arises solely due to tertiary interactions. The reason for the instability of the helix is that there is an improper balance of the forces on short length scale (responsible for helix formation) and the interaction between hydrophobic residues which imparts globularity. The parameters of the hydrophobic interactions were chosen so that the four-helix bundle is stable. But for these parameters the ordered one-dimensional structure, namely the isolated helix, is not the preferred native state.

For these reasons (in a recent article addressing the viscosity dependence of the folding rates in proteins [25]), we altered the dihedral angle potential so that a stable and relatively fast folding helix is obtained. The resulting dihedral potential is:

$$V^B_{DIH} = \sum_{i=1}^{N-3} [A_i(1 - cos\phi) + B_i(1 + cos3\phi) + C_i(1 - sin\phi)] \quad (5)$$

where $A_i = 1\varepsilon_h$, $B_i = 1.6\varepsilon_h$, $C_i = 2\varepsilon_h$ for all i. A comparison of V^A_{DIH} and V^B_{DIH} indicates that the latter has a deeper *gauche*$^+$ minimum and the barriers separating g^+, g^- and t states are smaller. This allows a more facile rotation about the putative peptide bonds. Both forms of dihedral angle potential favor g^+ (≈ 60°) conformation. Note that the potential does not depend on the type of residues involved.

Hydrogen bond potential. It is known that backbone hydrogen bonds (HBs) in α helix are formed between the carbonyl oxygen of residue i and the amide hydrogen of the i+4 residue [32]. We model HBs using a virtual atom representation. This is best illustrated in Figure 1. In the minimal off-lattice model the beads approximately represent the α-carbon atoms. We imagine that in between the α-carbons there are two virtual groups — CO and NH — representing the carbonyl and NH groups, respectively. With this representation, shown in Figure 1, the key characteristic features of α helices, namely the presence of HBs, can be easily introduced. The HBs are modeled as bonds between the virtual groups CO_i and NH_{i+4} located on the lines connecting α-carbons (i, i+1) and (i+3, i+4). Thus, as is found in right-hand α helices, there is a hydrogen bond between the carbonyl of each residue and the NH of the fourth residue along the chain. Specifically, we place a CO_i group at the distance $\frac{1}{3}|\vec{r}_{i,i+1}|$ from the residue i, where $|\vec{r}_{i,i+1}| = |\vec{r}_{i+1} - \vec{r}_i|$ is the separation between residues i and i+1. The NH_{i+4} group is put at the distance $\frac{2}{3}|\vec{r}_{i+3,i+4}|$ from the residue i+3. Following the HB geometry found in real proteins, we further assume that the CO_i group of residue i may interact only with the group NH_{i+4} of residue i+4. Our model does not permit formation of HBs between any other residues or CO and NH groups.

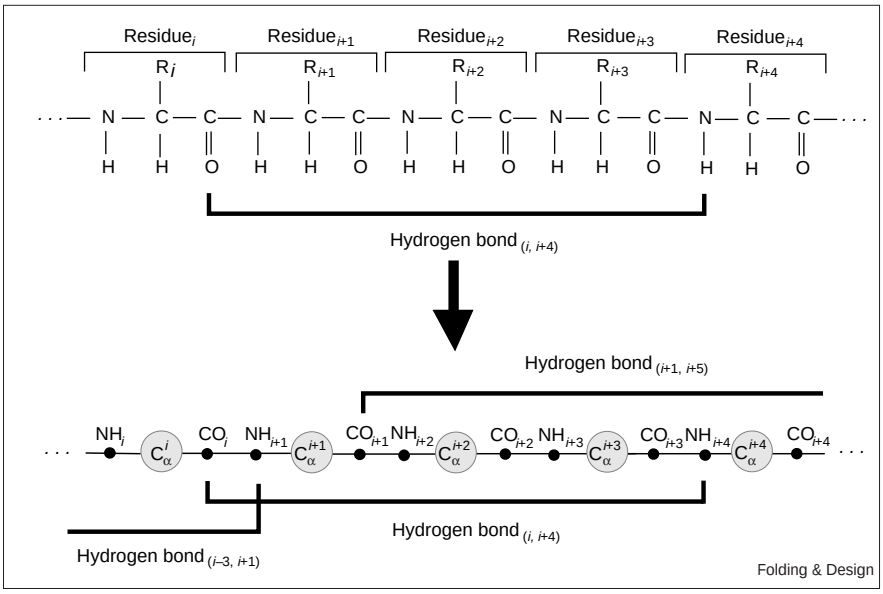
The potential associated with the hydrogen bond between CO_i and NH_{i+4} is given by:

$$V_{HB} = \sum_{i=1}^{N-4} -\varepsilon_{hb} e^{-\alpha_{hb}(cos^2\phi_i + cos^2\psi_i)} \quad (6)$$

where ε_{hb} determines the strength of hydrogen bonding and α_{hb} controls the angular directionality of a hydrogen bond (represented by a vector $\vec{r}_{OH}$ connecting the groups CO_i and NH_{i+4}) with respect to the vectors $\vec{r}_{i,i+1}$ and $\vec{r}_{i+3,i+4}$. The cosines in Equation 6 are given by:

$$\cos\phi_i = \frac{(\vec{r}_{OH} \cdot \vec{r}_{i,i+1})}{|\vec{r}_{OH}||\vec{r}_{i,i+1}|}; \cos\Psi_i = \frac{(\vec{r}_{OH} \cdot \vec{r}_{i+3,i+4})}{|\vec{r}_{OH}||\vec{r}_{i+3,i+4}|} \quad (7)$$

Figure 1



Schematic representation of the coarse-graining employed in the reduction of the polypeptide chain into a minimal off-lattice model. Sidechains are merged into Cα carbon atoms. The NH and CO groups are represented as virtual moieties located between two successive Cα atoms. For α helices, hydrogen bonds are formed between the carbonyl of the *i*th residue and the NH of the (*i*+4)th residue.

The potential function in Equation 6 favors the HB geometry in which the vector $\vec{r}_{OH}$ tends to be perpendicular to both the vectors $\vec{r}_{i,i+1}$ and $\vec{r}_{i+3,i+4}$. Note that Equation 6 does not mandate that the vectors $\vec{r}_{i,i+1}$ and $\vec{r}_{i+3,i+4}$ be parallel as it would result in highly unrealistic helical pitch. We set ε_{hb} to be $\frac{1}{3}\varepsilon_h$. The orientation of HBs is enforced by setting the parameter α_{hb} to 2. The parameter α_{hb}, which penalizes the HB conformations that do not satisfy the required orientation, should be as large as possible. The compromise value of $\alpha_{hb} = 2$ was chosen so that the algorithm used in integrating the equations of motion is stable. There are no quantitative changes in the results when α_{hb} is varied with certain limits.

We have also tried other forms of the hydrogen bond potential that include spatial dependence. In particular, we assumed that the hydrogen bond interaction arises due to dipolar interaction between the CO and NH groups. The resulting helix is also stable (D Klimov, D Thirumalai, unpublished data). But for simplicity, in this study, we present results using Equation 6.

Non-bonded potential. The non-bonded potentials, V_{NON}, arise between pairs of residues that are not covalently bonded [13]. Although interaction between non-bonded residues does not contribute significantly to α-helix stability, the attraction between the hydrophobic residues cannot be so large as to render the helix unstable. Improper balance of forces between non-bonded residues and dihedral angle potential may also give rise to kinetic traps, which invariably slow down helix folding. These qualitative statements are reflected in extensive experimental studies [33–35] which have inferred helix properties of various amino acid residues. The total non-bonded potential is written as [13]:

$$V_{NON} = \sum_{i=1}^{N-3} \sum_{j=i+3}^{N} V_{ij}(r) \tag{8}$$

where $r = |\vec{r}_i - \vec{r}_j|$. The potential between two L beads or between a (L, B) pair is taken to be:

$$V_{L\alpha}(r) = 4\varepsilon_h \left[\left(\frac{a}{r} \right)^{12} + \left(\frac{a}{r} \right)^{6} \right] (\alpha = L \text{ or } B) \tag{9}$$

This potential is purely repulsive and the presence of the r^{-6} term provides longer-range repulsion than the usual r^{-12} term. We find that this larger-range repulsion interaction is effective in destabilizing the kinetic traps in folding. We have also used a short-range version of $V_{L\alpha}$ in which only the r^{-12} term is retained.

$$V_{L\alpha}(r) = 4\varepsilon_h \left(\frac{a}{r} \right)^{12} (\alpha = L \text{ or } B) \tag{10}$$

If both the residues are hydrophobic (B) the potential of interaction is taken to be:

$$V_{BB}(r) = 4\varepsilon_h \left[\left(\frac{a}{r} \right)^{12} - \left(\frac{a}{r} \right)^{6} \right] \tag{11}$$

where ε_h determines the strength of the hydrophobic interaction.

Simulation methods: low friction noisy molecular dynamics and Brownian dynamics algorithms

We assume that the dynamics of the α helix is governed by the Langevin equation. Following our earlier studies of off-lattice models [13,23,25], we include in the equation of motion for a protein sequence a damping term with a friction coefficient ζ and a Gaussian random force Γ, which balances the energy dissipation caused by friction. The equation of motion written for the generalized coordinate x is given by:

$$m\ddot{x} = -\zeta\dot{x} + F_c + \Gamma \tag{12}$$

where $F_c = -\partial E_p/\partial x$ is the conformation force, which is a negative gradient of potential energy with respect to the coordinate x, Γ is the random force which has a white noise spectrum, and m is the mass of a bead.

The numerical integration of Equation 12 depends on the friction coefficient ζ. In the underdamped limit (low friction) we use a velocity form of Verlet algorithm, which is applicable when $h\zeta << 1$, where h is an integration step (we set the mass of a bead to unity) [13]. Generally, the Verlet algorithm can be used at small values of friction coefficient ζ, and it becomes increasingly inefficient at larger ζ because the condition $h\zeta << 1$ requires that progressively smaller values of h be chosen. In the overdamped limit, when the inertial term is negligible with respect to the damping term (large friction), we use a Brownian dynamics algorithm proposed by Ermak and McCammon [36,37]. The position of a bead at the time $t + h$ is given with respect to the first order term of h by:

$$x(t+h) = x(t) + \frac{h}{\zeta}\left(F_c(t) + \Gamma(t)\right) \tag{13}$$

where Γ(t) is a random force that has a white noise spectrum. The autocorrelation function for Γ(t) in the discretized form is [13]:

$$\langle \Gamma(t)\Gamma(t+nh)\rangle = \frac{2\zeta k_B T}{h}\delta_{0,n} \tag{14}$$

where $\delta_{0,n}$ is the Kronecker delta and $n = 0, 1, 2\ldots$. Note that in writing Equation 13 we assume that hydrodynamic interactions are negligible, that is, diffusion tensor D_{ij} does not depend on sequence conformation and only its diagonal elements are non-zero (so there is no coupling between various degrees of freedom). Equation 13 is valid when the time step h satisfies the condition $h\zeta >> 1$. It is clear that Equation 13 may be used with relatively large values of friction coefficient as compared with the Verlet algorithm. The algorithm given by Equation 13 describes Brownian type dynamics, for which:

$$\langle (\Delta x)^2 \rangle = 2Dh \tag{15}$$

where $\Delta x(h) = x(t + h) - x(t)$ and D is a diffusion coefficient equal to T/ζ. Note that the formula for $\langle(\Delta x)^2\rangle$ is obtained in the first order approximation with respect to h. Larger step sizes can be chosen by performing simulations in the dihedral angle space [38].

It is clear that thermodynamics does not depend on the precise value of friction coefficient, and therefore we may turn to a low friction (underdamped) limit to study equilibrium folding transitions. Thus, equilibrium simulations have been performed using a velocity form of the Verlet algorithm at the value of friction coefficient $\zeta = 0.05\tau_L^{-1}$ with the time integration step $h = 0.005\tau_L$ [13], where τ_L is given by Equation 16.

In our previous studies we have shown that the value of friction coefficient $\zeta = 50\tau_L^{-1}$ approximately corresponds to water and the onset of the overdamped limit [13]. For this reason, we have performed kinetic studies of helix formation at this value of ζ using the Brownian dynamics algorithm with the time integration step $h = 0.02\tau_L$. The value of h gives control of the temperature to within 1 or 2%.

Mapping simulation time scales to real time

We measure temperature in units of ε_h/k_B and length in terms of an average distance between two Cα atoms a. In the underdamped limit the natural choice of time unit is:

$$\tau_L = \left[\frac{ma^2}{\varepsilon_h}\right]^{\frac{1}{2}} \tag{16}$$

where m is the mass associated with a residue. For a typical choice of $m \approx (3–5) \times 10^{-22}$ g, $a \approx (5–6) \times 10^{-8}$ cm, $\varepsilon \approx (1–2)$ kcal/mol $\tau_L \approx (2.5–4)$ ps. For purposes of converting the simulation times to real times we will take $\tau_L = 3$ ps. In the overdamped limit the time unit is:

$$\tau_H = \frac{\zeta\varepsilon_h}{T}\tau_L \tag{17}$$

As τ_H is dependent on temperature, we use $\tau_L = 3$ ps as the unit of time in this study.

Calculation of thermodynamic functions

The native conformation of a sequence is determined from multiple slow cooling trajectories [13]. The conformation with the lowest energy at $T = 0.0$ is assumed to be the native state. The thermodynamic quantities of sequences are computed using the multiple histogram technique [39]. From slow cooling simulations we have an estimate of the folding transition temperature T_F. This allows us to select the temperature range over which histograms are to be collected. We generated $M = 50$ trajectories with different

initial conditions. Each trajectory starts at high temperature $T_h = 1.5 > T_F$ and ends at the low temperature $T_l = 0.02 < T_F$. In the course of a trajectory, temperature is decreased by the decrement $\Delta T = 0.02$ every $2{,}500\tau_L$. The histograms were collected separately at the temperatures T_r, $r = 0, \ldots, (T_h - T_l)/\Delta T$ using all M trajectories. In all we have $R = (T_h - T_l)/\Delta T + 1$ histograms. A portion of the trajectory immediately following a temperature change must be excluded in order to allow sequence to equilibrate at a new T_r. This interval of equilibration, which does not exceed $500\tau_L$ at the lowest temperature, was kept constant at all temperatures.

The following variables were used to collect histograms: the overlap function χ [13,26], the potential energy E_p (see Equation 1), and the helical content H. The overlap function is:

$$\chi = 1 - \frac{2}{N^2 - 5N + 6} \sum_{i=1}^{N-3} \sum_{j=i+3}^{N} \Theta(\varepsilon - |r_{ij} - r_{ij}^N|) \qquad (18)$$

where r_{ij} is the distance between the beads i and j for a given conformation, r_{ij}^N is the corresponding distance in the native conformation, and $\Theta(x)$ is the Heavyside function. If $|r_{ij} - r_{ij}^N| \le \varepsilon$, then the beads i and j are assumed to be at the native distance. We take $\varepsilon = 0.2a$. The helical content H of a conformation is defined as:

$$H = \frac{\sum_{i=1}^{N-3} \Theta(\delta\phi_i - |\phi_i - \phi_i^N|)}{N - 3} \qquad (19)$$

where ϕ_i and ϕ_i^N are the dihedral angles in a given conformation and in the native helix, respectively. The dihedral angle tolerance $\delta\phi$ indicates the maximum deviation of dihedral angles from their native position, which still corresponds to a native state. We set $\delta\phi$ to 10°. According to Equation 19, helical content reflects the fraction of the sequence in a helical conformation. Since only a relatively small tolerance is allowed this is a rather stringent criterion.

We have also studied the thermodynamics of HB formation. To this end, we assumed that a HB i is formed if $|r_{\mathrm{OH},i} - r_{OH,i}^N| \le \varepsilon_{hb}$, where $|r_{\mathrm{OH},i}|$ and $|r_{OH,i}^N|$ are the distances between CO_i and NH_{i+4} groups in a given conformation and in the native helix, respectively. The HB tolerance ε_{hb} is set to 0.15. This definition allows us to construct histograms for all 12 HBs with respect to two states — a hydrogen bond is 'on' and 'off'. In practice, this is implemented in the form of a unified histogram for the variables E_p and B_m, $m = 1,2,\ldots$, where B_m, a HB indicator, is either 0 or 1. The thermodynamic probability of the formation of HB i is calculated as:

$$P_{hb}^i = \frac{\sum_k \delta(B_i = 1) e^{-\frac{E_k}{k_B T}}}{\sum_k e^{-\frac{E_k}{k_B T}}} \qquad (20)$$

where the sum is taken over all conformation space. Note that this is an equilibrium definition for HB probability P_{hb}^i. Below we will introduce a kinetic analog of P_{hb}^i, which probes the dynamics of HB formation.

The histograms were collected using a 0.1 grid interval for the energy E_p, 92 exact discrete values for the overlap χ, and 14 exact values for the helical content H. Using multiple histogram technique, we calculated the following equilibrium quantities as a function of temperature: the overlap function $\langle\chi\rangle$ and its fluctuation $\Delta\chi$, energy $\langle E_p\rangle$, specific heat C_v, the helical content $\langle H\rangle$, the probabilities of HB formation P_{hb}^i, and the probability of occupancy of the NBA [26,40]:

$$P_{NBA} = \frac{\sum_k \delta(\chi_k \le \chi_{NBA}) e^{-\frac{E_k}{k_B T}}}{\sum_k e^{-\frac{E_k}{k_B T}}} \qquad (21)$$

where $\chi_{NBA} = \langle\chi(T_F)\rangle$ and the T_F is a folding transition temperature.

Sequence selection

For the present study we selected two sequences. We assume that sequences consist of residues of only two types — hydrophobic (B) and hydrophilic (L). The composition of the sequences shown in the caption to Figure 2 is identical and corresponds to a 16-mer helical fragment taken from [23].

The difference between the sequences is due to the different versions of potential functions, namely the dihedral angle and the non-bonded potentials used. Sequence A has the dihedral angle potential given by Equation 4 and the short-range repulsion potential involving L residues (Equation 10). Sequence B has the dihedral angle potential given in Equation 5 and the long-range repulsion potential involving L residues given by Equation 9. In all other aspects sequences A and B are identical.

Results and discussion

Characteristics of the native state

The native state for sequence B is shown in Figure 2. It can be clearly seen that the native state of sequence B has a helical native topology. The average helical pitch for sequence B is 3.83, while the average dihedral angle is 59.8°. These values reflect the adoption of g^+ conformations by all the dihedral angles. To evaluate the number of native contacts we assume that residues i and j ($|i - j| \ge 3$) form a native contact if the space separation between them is less than 2. Using this definition, we found that the native conformation of B has 13 native contacts. The native conformation of sequence A is very similar to that of sequence B (data not shown); the average pitch for sequence A is 3.86, and the average dihedral angle is

58.9°. We have computed the average values of the dihedral angles in helical proteins using PDB coordinates for bovine acyl-coenzyme A binding protein (ACBP) and equine cytochrome *c* proteins. The dihedral angle is associated with the one found between the planes defined by four successive Cα atoms. The average values for bovine ACBP and equine cytochrome *c* are 53.4° and 56.8°. These values are remarkably close to those found in our models, which suggests that structurally the helix shown in Figure 2 indeed corresponds to that found in proteins.

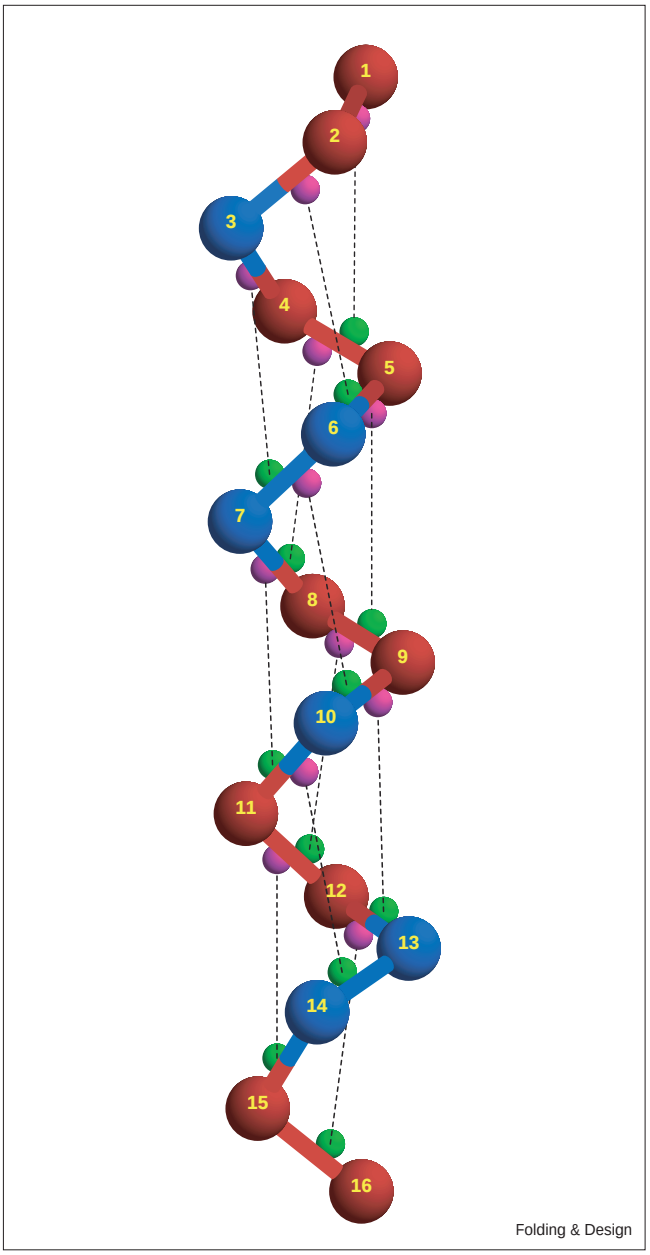
The native structure for sequence A also has 13 native contacts. The structures are not exactly identical, however. For example, the overlap between the native states of the two sequences measured in terms of χ is about 0.16. The difference arises because the native conformation of sequence A is somewhat stretched compared to that of sequence B.

Thermodynamics of helix formation

The temperature dependence of various quantities that probe the thermodynamics of helix formation is shown in Figure 3a. We find that at low temperatures the polypeptide chain adopts a helical conformation. The folding transition for sequence B associated with the location of the maximum in $\Delta\chi$ occurs at T_F = 0.31 (see Figure 3a). As seen in experiments on small helix-forming peptides, the transition is quite broad [8,14]. We assume that the conformations with an overlap smaller than $\chi_{NBA} = \langle\chi(T_F)\rangle$ comprise the NBA (see Equation 21) [26,40]. In Figure 3a, we display the probability P_{NBA} of being in the NBA as a function of temperature. We can alternatively compute T_F using $P_{NBA}(T_F) = 0.5$. This condition gives $T_F = 0.31$, which coincides with the folding temperature obtained from the peak of $\Delta\chi$. In Figure 3a, we also display the helical content $\langle H\rangle$ as a function of temperature. This measure shows that the fraction of the chain in the native helical topology gradually increases from about 0.3 at $T = 1.5$ to almost 0.8 at $T = T_F$. These data indicate that at high temperatures (well above T_F), the sequence still contains a considerable amount of helical structure. Interestingly, the derivative of $\langle H\rangle$ with respect to temperature reaches a maximum at a much lower temperature of $T = 0.16$ (with respect to T_F). This result suggests that even below T_F, that is, when the sequence is confined to the NBA, it experiences substantial fluctuations, which is consistent with a broad transition (see Figure 3a). The sequence becomes effectively 'frozen' at $T \lesssim 0.16$, when almost 95% of the sequence as measured by $\langle H\rangle$ adopts helical conformation (see Figure 3a).

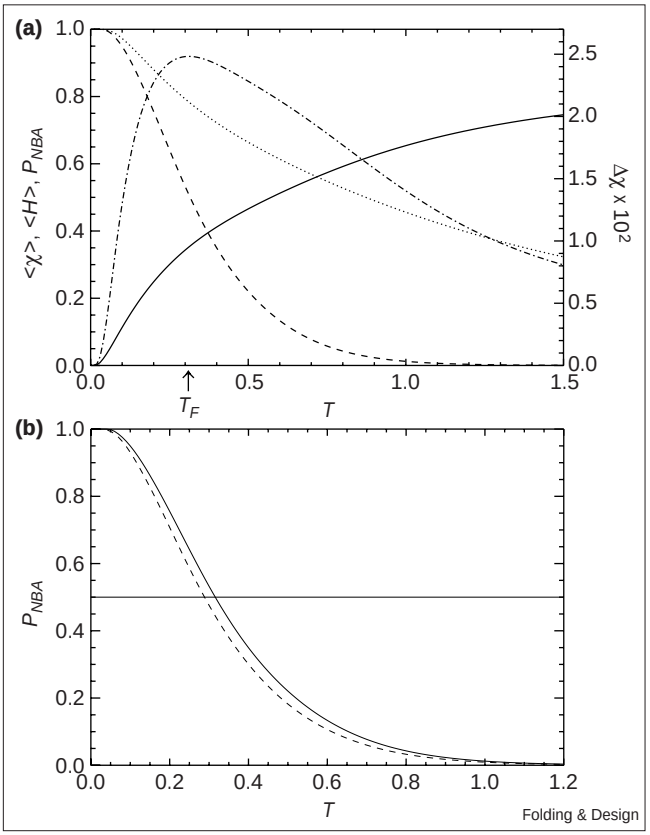
The specific heat C_v (data not shown) reveals a very poorly defined peak at 1.13. Due to the one-dimensional ordering in the case of helix, the peak in C_v (or changes in $\langle R_g\rangle$) cannot be associated with the collapse transition [16]. In fact, the radius of gyration $\langle R_g\rangle$ slightly increases upon the formation of helix (data not shown). At high temperatures well above T_F, $\langle R_g\rangle$ is approximately 2.4, while at T close to zero it approaches the value of 2.6. This result is expected

Figure 2



The native conformation of the α helix determined by a slow cooling method. The sequence for this α helix is *LLBLLBBLLBLLBBLL*, where *L* is a hydrophilic residue (shown in red) and *B* is a hydrophobic residue (shown in blue). The groups CO and NH participating in HBs are given by small magenta and green beads, respectively. All 12 native HBs formed between CO_i and NH_{i+4} groups are indicated by black dashed lines. The structural characteristics, such as the pitch of the helix and the dihedral angles for four successive C_α atoms, are in accord with those found in real proteins. The figure has been created using RasMol v2.6 (R Sayle, 1995).

Figure 3



Thermodynamic characteristics of sequence B. (a) Temperature dependence of the helical content $\langle H \rangle$ (dotted line), probability of being in the native basin of attraction, P_{NBA} (dashed line), the overlap function $\langle \chi \rangle$ (solid line), and fluctuation in the overlap function, $\Delta\chi$ (dash-dot line). The scale for $\Delta\chi$ is given on the right. The folding temperature T_F (= 0.31) indicated by an arrow is determined from the peak in $\Delta\chi$. This value coincides with the condition $P_{NBA}(T_F) = 0.5$. (b) Comparison of $P_{NBA}(T)$ for sequence B with HBs (solid line) and without HBs (dashed line). The inclusion of HBs increases the cooperativity of the helix–coil transition (as measured by Ω_c (Equation 23)).

because a helix is essentially a one-dimensional stretched system as opposed to three-dimensional random coil. The potential energy as a function of temperature shows near-linear dependence, which suggests that helix formation spans a broad temperature region and that the helix–coil transition is weakly cooperative. This conclusion is further substantiated by the data discussed below.

It is interesting to examine the formation of HBs as a function of temperature. In Figure 4a, we plot 12 curves representing the probabilities of HB formation $P^i_{hb}(T)$. It is very clear that HBs fall into two groups — HBs 1,4,5,8,11,12 and HBs 2,3,6,7,9,10 — which differ with respect to HB stabilities. It is also evident that at high temperatures $T \gtrsim 1.0$, the probabilities $P^i_{hb}(T)$ do not exceed 0.4, whereas at $T \simeq T_F$, the probabilities $P^i_{hb}(T)$ are around 0.7. This figure reveals that a considerable amount of helical structure is preserved at elevated temperatures just as indicated by the temperature dependence of $\langle H \rangle$. More interestingly, this plot points to clear inhomogeneity among HBs. Figure 4b provides further evidence for this and suggests a plausible explanation for differences in $P^i_{hb}(T)$. In this figure we plot $P^i_{hb}(T)$ taken at $T = 0.30$, which periodically alters between approximately 0.67 and 0.72. These variations closely match those in the number of hydrophobic beads in the vicinity of HBs N^i_B. (Specifically, N^i_B is the number of B residues in the sequence fragment $(i, i+4)$.) Furthermore, as the peaks in N^i_B correspond to minima in $P^i_{hb}(T)$, it appears that hydrophobic beads tend to weaken the proximal HBs. Thus, it seems likely that sequence heterogeneity is responsible for the variations in $P^i_{hb}(T)$ along the sequence.

We may further contrast this result by studying the thermodynamics of a purely hydrophilic sequence, referred to as sequence C. In the inset to Figure 4a we display $P^i_{hb}(T)$ calculated for such a sequence. It follows from this figure that the probabilities $P^i_{hb}(T)$ show almost no variation along the sequence as T changes, hence lending additional support to the conclusion that sequence heterogeneity causes variations in $P^i_{hb}(T)$. These observations are consistent with detailed experiments showing that helix propensities depend in rather subtle ways on residue type and the environment.

The thermodynamics for sequence A is similar to that of sequence B (data not shown). The folding transition temperature T_F is found to be 0.30, which is very close to the temperature at which P_{NBA} is 0.5.

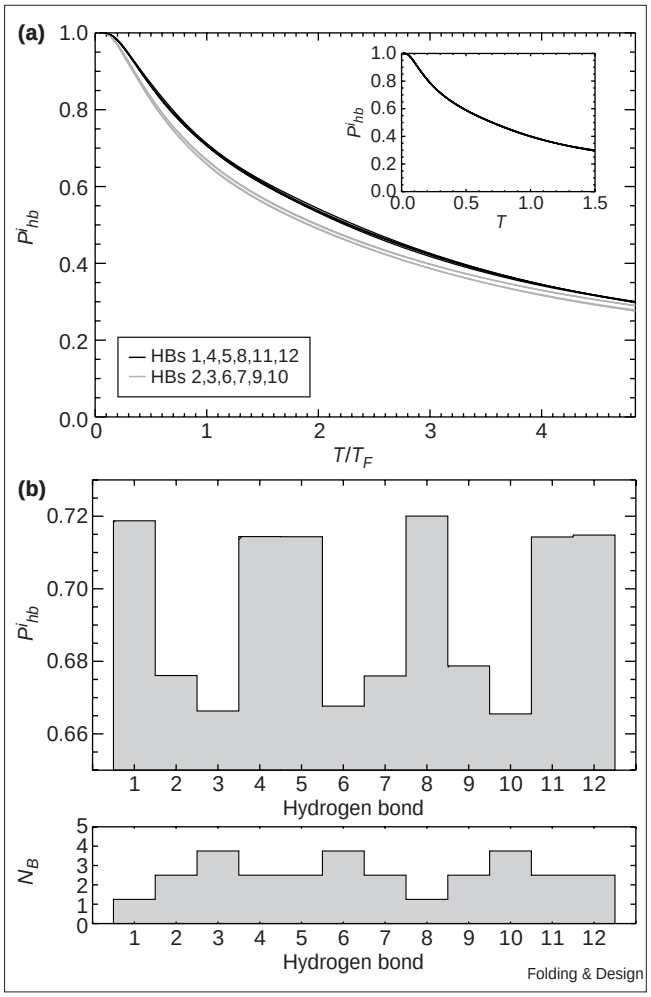
Effects of hydrogen bonds on stability

The effect of HBs on the stability of the α-helical structure is most vivid when comparing sequence A with and without HBs. Apart from the inclusion of HBs, sequence A is identical to the helix from the previously studied four-helix bundle. In the absence of HBs, the α-helix conformation for this sequence is unstable, because one can encounter collapsed structures with lower energy. Thus, the thermodynamic stability of the native helix for sequence A is entirely due to the inclusion of hydrogen bonding. In contrast, the helical structure of sequence B is stable even without HBs, but hydrogen bonds do provide enhanced stability to the helical structure. This can be shown by comparing the free energy of stability, ΔF, at T_s, where $\Delta F = -k_B T_s / nK(T_s)$ with:

$$K(T_s) = \frac{P_{NBA}(T_s)}{1 - P_{NBA}(T_s)} \qquad (22)$$

Here we have assumed that the folding is two state. The value of $\Delta F(T_s)/k_B T_s$ for sequence B without HBs is –0.88

Figure 4



(a) The temperature dependence of the probabilities of forming HBs for sequence B. For this sequence the various probabilities cluster into two groups. At all temperatures the first group (HBs 1,4,5,8,11,12, shown by black lines) has a small but discernible gain in P^i_{hb} as compared with the second group (HBs 2,3,6,7,9,10, shown by grey lines). The inset shows the probabilities P^i_{hb} for L_{16} helix (sequence C). Here all the probabilities show exactly the same dependence.
(b) Hydrogen bond formation probabilities P^i_{hb} as a function of the HB number for sequence B at $T \approx T_F$. The figure shows that the probability of forming HBs decreases marginally in the vicinity of hydrophobic residues. The lower panel of (b) displays the number of *B* residues in the fragment (i, i+4). It shows that as the number of *B* beads increases in the (i, i+4) fragment the probability of P^i_{hb} accordingly decreases in the upper panel.

at $T_s = 0.20$, whereas $\Delta F(T_s)/k_B T_s$ for sequence B with HBs is –1.13 at the same temperature. This observation is also illustrated in Figure 3b, where we plot P_{NBA} for sequence B with and without HBs. It follows from this that at any temperature P_{NBA} for the sequence with HBs exceeds that for the sequence without HBs. More importantly, we show below that the inclusion of HBs has a dramatic effect on the folding kinetics.

Cooperativity of helix–coil transition

Our study suggests that helix formation in these small systems is only weakly cooperative, which is consistent with the broad transition region observed experimentally [14]. This observation is in accord with the time-honored theories of the helix–coil transition [41–44] which have shown that only in the limit of infinite length is the transition sharp. We can quantify this by measuring the degree of cooperativity in terms of the dimensionless parameter [40]:

$$\Omega_c = \frac{T_F^2 \, max\left[\frac{dP_{NBA}}{dT}\right]}{\Delta T} \qquad (23)$$

where $max[dP_{NBA}/dT]$ is the maximum value of the derivative of P_{NBA} (see Equation 21) with respect to T, and ΔT is the full width at half the maximum of dP_{NBA}/dT. The values of Ω_c for sequences A and B are 0.38 and 0.30, respectively. Alternatively, instead of P_{NBA}, we may use helical content $\langle H \rangle$ in Equation 23. With this measure, we estimate $\Omega_c = 0.63$. It is clear that both quantities, P_{NBA} and $\langle H \rangle$, give consistent results, revealing weak cooperativity of helix formation. We believe that helical content $\langle H \rangle$ is perhaps more appropriate for the calculation of Ω_c because P_{NBA} measures the acquisition of the native state with respect to entire structure (as distances between all residues i and j are taken into account), while $\langle H \rangle$ relies only on the formation of 13 native dihedral angles.

The very small values of Ω_c are indicative of a very broad transition. This is in complete agreement with experimental studies on helix thermodynamics [14]. In Figure 5 we show the probability of being in the native state f_N (assuming two-state transition) using the van't Hoff parameters for the 21-residue alanine-based peptide investigated using T-jump experiments by Williams *et al.* [14]. For comparison we also show the temperature dependence of P_{NBA} for sequence B scaled to the same range of f_N. From the experimental data, we compute $\Omega_c = 0.70$, which is consistent with the models studied here. In contrast, Ω_c for proteins is in general greater than about 5, which implies that thermodynamic transitions for these small helical peptides are only weakly cooperative.

Folding kinetics

In this section we present our results on the kinetics of helix formation starting from a random coil. We studied the kinetics of folding over a range of temperatures using several probes, such as the overlap function $\langle \chi(t) \rangle$, helix content $\langle H(t) \rangle$, and the fraction of unfolded molecules $P_u(t)$ [13]:

$$P_u(t) = 1 - \int_0^t P_{fp}(s)ds \qquad (24)$$

where $P_{fp}(s)$ is the distribution of first passage times. The first passage time τ_{1i} corresponds to the first occurrence of $\chi(t = \tau_{1i}) = 0$ for a folding trajectory i. These quantities have been obtained by averaging over a number of independent initial conditions. In Figure 6 we display the function $P_u(t)$ for sequence B calculated at $T_s = 0.275 < T_F$. This function can be well fit with the sum of three exponentials of the form:

$$P_u^{fit}(t) = \Phi exp\left(-\frac{t}{\tau_{FAST}}\right) + a_1 exp\left(-\frac{t}{\tau_{SLOW,1}}\right) + a_2 exp\left(-\frac{t}{\tau_{SLOW,2}}\right) \quad (25)$$

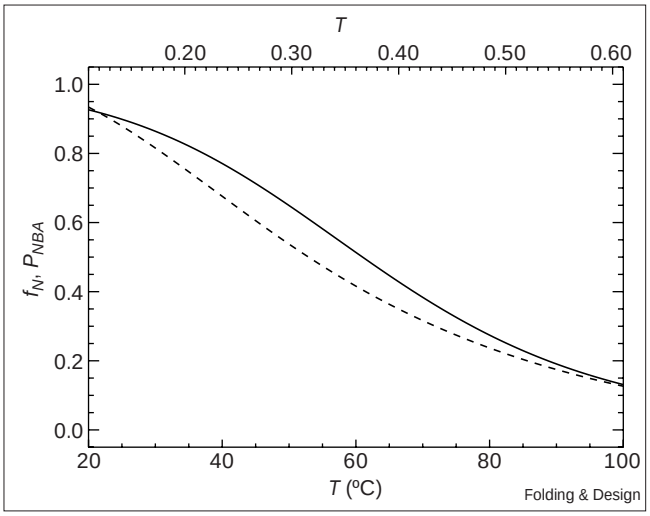
where Φ is the amplitude of the fast (nucleation) folding trajectories with the time scale τ_{FAST}, and $\tau_{SLOW,1}$ and $\tau_{SLOW,2}$ are the time scales of the slow (off-pathway) processes, which are typically associated with 'trapping' in local energy minima [13]. The corresponding fit is shown by a solid line in Figure 6. Note that $\Phi + a_1 + a_2 = 1$. At most temperatures for which $\Phi < 1$, however, $P_u(t)$ can be well fit by sum of two exponentials.

We have shown in our earlier studies on the kinetics of β-turn formation [13] that when the function $P_u(t)$ is given by a sum of exponentials (Equation 25) the folding proceeds by a kinetic partitioning mechanism [4,13] — a characteristic of moderate folding proteins. The three exponential fit (given by a solid line in Figure 6) suggests that the slow phases with the amplitude $(1 - \Phi)$ may correspond to transient trapping in helical conformation with defects. We find that at $T_s = 0.275$, $\Phi = 0.79$, which implies that 79% of folding trajectories reach the native helix on a fast time scale without being trapped in any intermediate state. Presumably folding in these trajectories occurs via nucleation followed by propagation of the helical structures as envisaged in standard helix–coil theories [41]. The time scale $\tau_{FAST} = 16$ ns, whereas $\tau_{SLOW,1} = 106$ ns and $\tau_{SLOW,2} = 774$ ns. The folding time τ_F can be readily calculated from $P_u(t)$ as $\tau_F = \int_0^\infty P_u(t)dt$. This yields $\tau_F = 66$ ns.

Inclusion of hydrogen bonds decreases the probability of transient trapping

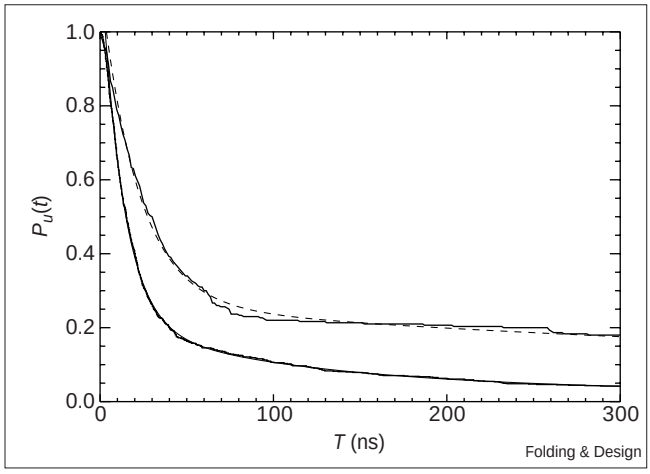
It is interesting to contrast the folding kinetics of sequence B with that of the same sequence without HBs (sequence B′) studied in [25]. To make such a comparison meaningful, similar folding conditions have to be chosen for both sequences. At the simulation temperature $T_s = 0.275$, the equilibrium value of $\langle\chi\rangle$ for sequence B is 0.323. Sequence B′ attains the same value of $\langle\chi\rangle$ at $T_s = 0.24$. The fraction of unfolded molecules $P_u(t)$ is plotted for that sequence in Figure 6. Although the majority of folding trajectories find the native state rather rapidly, a small fraction (about 26%) become trapped in intermediates. The escape time from these far exceeds the simulation time. For this reason, it is impossible to

Figure 5



Comparison of the temperature dependence of $P_{NBA}(T)$ for sequence B (dashed line) with the fraction of native state $f_N(T)$ (solid line) as a function of temperature for a 21-mer helix reported in [14]. Assuming a two-state transition we computed $f_N(T)$ using the thermodynamic parameters given in [14]. The temperature scales for the 21-mer helix and sequence B are given on the lower and upper horizontal axes, respectively. The values of Ω_c (see Equation 23) for both sequence B (0.30) and the 21-mer helix (0.70) are small, which implies that the transition is not cooperative.

Figure 6



The time dependence of the fraction of molecules $P_u(t)$ that have not found the native helix. The solid line represents a three exponential fit (see Equation 25) for sequence B at $T_s = 0.275$ (= 0.89 T_F). The dashed line corresponds to the fit to the function $P_u(t)$ for sequence B without HBs. The inclusion of HBs accelerates escape from transient traps by at least one order of magnitude in folding times.

provide a reliable estimate for τ_{SLOW}. However, we can easily calculate τ_{FAST} and obtain a lower bound for τ_F by substituting τ_{1i} with τ_s for the trajectories in which the native state is not found within the simulation time $\tau_s = 2.4$ μs. It turns out that the fast folding time scale is 23 ns, which is larger by only a factor of 1.4 than τ_{FAST} for sequence B with HBs. In contrast, the estimate of τ_F for sequence B′ is > 556 ns, which is almost an order of magnitude larger than for sequence B.

These results underscore a dramatic difference in the folding kinetics of both sequences. Incorporation of HBs speeds up the formation of helix by at least one order of magnitude. It is interesting that the inclusion of HBs enables easy escape from transient kinetic traps. The fast folding trajectories, on the other hand, reach the native state on roughly equal time scales. We believe that the kinetic consequences of incorporation of HBs are associated with significant destabilization of partially folded states that reduces the chances of trapping in local minima en route to the native state. These observations suggest that in natural helical fragments even moderate enhancement of stability of native interactions over non-native contacts may be sufficient to (nearly) destabilize any transient traps.

Temperature dependence of folding times of helix formation

The temperature dependence of the folding times may be readily obtained from the time dependence of the fraction of unfolded molecules $P_u(t)$ (see Equation 24) at different temperatures. We calculated τ_F over the temperature range 0.225–0.575. At all temperature values except 0.225, the statistical error in calculating τ_F is less than 10%. At 0.225 the error is about 10% because of the sharp increase in folding time scales. To achieve this level of accuracy, the number of trajectories used to compute $P_u(t)$ varies between 500 and 900. In Figure 7a we plot the dependence of τ_F (in ns) on temperature. It is seen that the folding times reach a minimum at a temperature of about 0.45 and increase at lower and higher temperatures. There is a dramatic increase in the folding times at temperatures below T_F. When the temperature changes by a factor of about 1.3 (from 0.30 to 0.225), for example, the folding time increases by a factor of about 24. Eaton and coworkers [10] have measured the temperature dependence of helix formation using tryptophan fluorescence and found that when the temperature changes by a factor of 1.2 (from 278K to 333K) the rate of helix formation increases by a factor of nearly 10. This is in complete accord with our findings. A qualitatively similar temperature dependence of τ_F to that shown in Figure 7a has been observed for other protein models [45], and it appears to be a generic feature of polypeptide chains. In all such cases it was found that folding rates reach maximum values at $T \gtrsim T_F$. A roughly linear (on semi-log plot) increase in τ_F with temperature signifies the Arrhenius-like behavior at $T < T_F$.

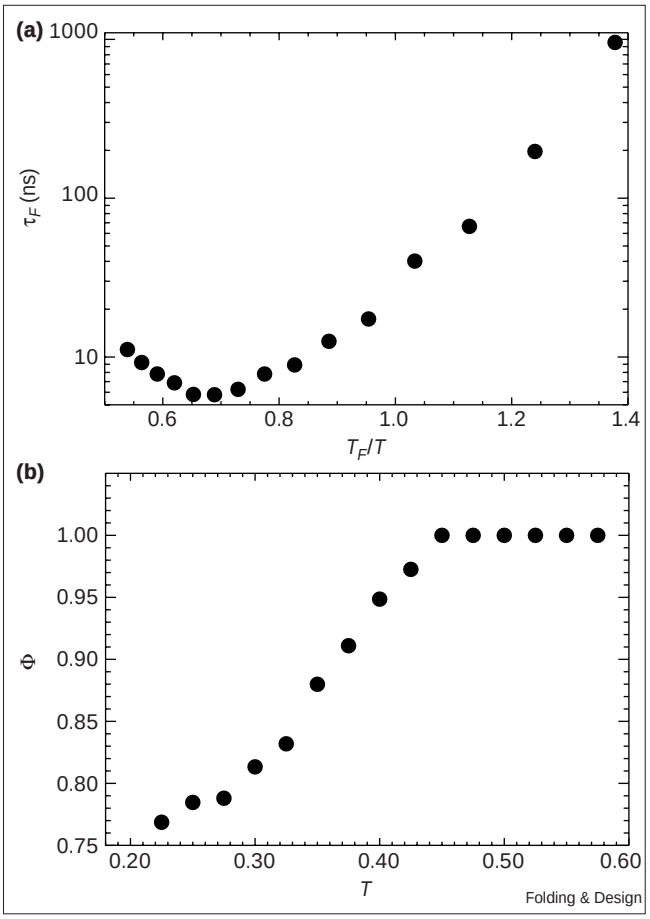
The typical time for helix formation at temperatures less than T_F is in the range 40–1000 ns. In our previous study, in which we did not include HBs [25], we showed that helices can form on time scales faster than about 500 ns. The time scales for helix formation reported here are in reasonable agreement with the experiments on a model 21-mer polypeptide [14]. These experiments suggest that helix–coil transition can occur on the time scale of the order of 150–200 ns. The good agreement between experiments and the Langevin simulations suggests that the simple models could be used to calibrate sequence-dependent properties of helix formation.

Progressive trapping in a local minima at low temperatures is also reflected in the plot displayed in Figure 7b, where we present the amplitude of the fast phase (see Equation 25) as a function of T. It is clear that at $T \gtrsim 0.45$, $\Phi = 1.0$, which implies that all folding trajectories reach the native state synchronously on the same scale τ_{FAST}. At $T \lesssim 0.45$, however, the fraction of fast folding trajectories is less than unity and monotonically decreases with temperature. At $T \lesssim 0.275$, for example, Φ becomes less than 0.8. It is interesting that even in the formation of these simple structures one encounters alternative routes. The steady decline in Φ at low temperatures indicates the onset of kinetic partitioning, according to which $(1 - \Phi)$ fraction of trajectories get trapped in intermediates and proceed to the native state via activation transition. It is interesting that Φ becomes less than unity approximately at temperatures at which the folding times τ_F reach a minimum.

Dynamics of hydrogen bond formation

Now let us focus our attention on the formation of individual HBs. We monitor the formation of HBs in the following way. First, we calculate the distances $r_{OH,i}$ (i= 1, .., 12) between virtual atomic groups CO and NH in conformations sampled along a folding trajectory. Then we assume that a hydrogen bond i is formed in a given conformation if the distance $r_{OH,i}$ satisfies the condition $|r_{OH,i} - r^N_{OH,i}| < \varepsilon_{hb}$, where $\varepsilon_{hb} = 0.15$ and $r^N_{OH,i}$ is the $r_{OH,i}$ distance in the native helix. We varied ε_{hb} within reasonable limits and found qualitatively similar results. We look for the presence of HBs at every step of the Brownian dynamics algorithm and then determine the kinetic probability $P^i_{hb}(t)$ that a hydrogen bond i is formed at time t by calculating the fraction of time, in which a HB i is on, over a small interval $(t - \Delta\tau/2, t + \Delta\tau/2)$ with $\Delta\tau = 60$ ps. We performed kinetic simulations at $T_s = 0.30 < T_F$ for 200 trajectories each of length 600 ns. In Figure 8 we display the probabilities for one of the folding trajectories — values of $P^i_{hb}(t)$ are encoded by colors according to the scale shown on the right — (red represents the

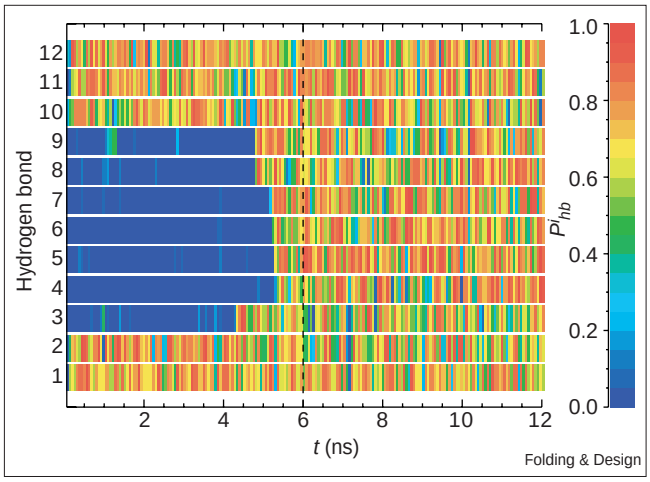
Figure 7



(a) The dependence of the folding time τ_F on temperature for sequence B. τ_F displays a nearly Arrhenius temperature dependence for $T \leq T_F$. The coil–helix transition times for $T \leq T_F$ range between 40 and 1000 ns depending on the temperature. (b) The fraction of molecules that are not transiently trapped as a function of temperature. The probability of getting trapped in some misfolded helix states is enhanced at low temperatures.

highest $P^i_{hb}(t)$ values, and blue indicates low $P^i_{hb}(t)$ values). It followed from Figure 8 and other trajectories (not shown) that formation of HBs is generally initiated near terminal residues (bonds 1 or 12), that is, the propagation of the helix begins at the termini. Interestingly, when the terminal bonds are already in place, but the central ones are still not formed, the sequence adopts conformations in which helical segments are already present near sequence ends (dihedral angles are in g^+ positions), but a few dihedral angles in the middle of a sequence are non-helical. This is illustrated by conformation snapshots displayed in Figure 9 for the trajectory shown in Figure 8. In Figure 9 we plot conformations every 1.2 ns until the first passage time at $\tau_{1i} = 6$ ns. It is evident that HBs start to form near sequence ends and gradually propagate towards its center.

Figure 8



Dynamics of the formation of HBs for a fast folding trajectory as a function of time in nanoseconds obtained at $T_s = 0.30$. The color codes for the kinetic probabilities of forming the HBs P^i_{hb} are given on the right and the HB labels are shown on the left. This and other trajectories show that in general (but not always), the HBs are initiated near the termini.

The initiation of HBs near terminal residues does not occur in all trajectories. In some instances HBs 9–12 near one sequence end and HBs 5,6 in the center stabilize rapidly, while HBs 1–4 near the other sequence end are established much later. It is clear from Figure 8 and similar analysis of other trajectories (not shown) that all HBs become stable at a time approximately equal to the first passage time τ_{1i} (marked by the vertical dashed line at $\tau_{1i} = 6$ ns in Figure 8). Furthermore, Figure 8 shows that after the first passage time is reached, native HBs may still undergo significant fluctuations and even occasionally break for a short time. We expect such fluctuations because the simulation temperature $T_s = 0.30$ is close to folding transition temperature T_F. All these conclusions are further illustrated in Figure 10, in which we plot the probabilities $\langle P^i_{hb}(t) \rangle$ averaged over 200 independent initial conditions. This figure provides direct evidence that the formation of HBs starts near the sequence ends, whereas the central HBs become stable much later.

Origins of the slow phase

Kinetic results show that at temperatures $T \lesssim 0.45$, the partition factor Φ is less than unity. This implies that a fraction of molecules reach the native conformation on a larger time scale. It is interesting to probe the origin of slow phase in folding trajectories under these conditions. Analysis of several slow folding trajectories reveals that although most of the HBs are formed very rapidly, some of them in the interior are not readily established. Close examination of such trajectories shows that folding is blocked by the formation of extremely stable non-native

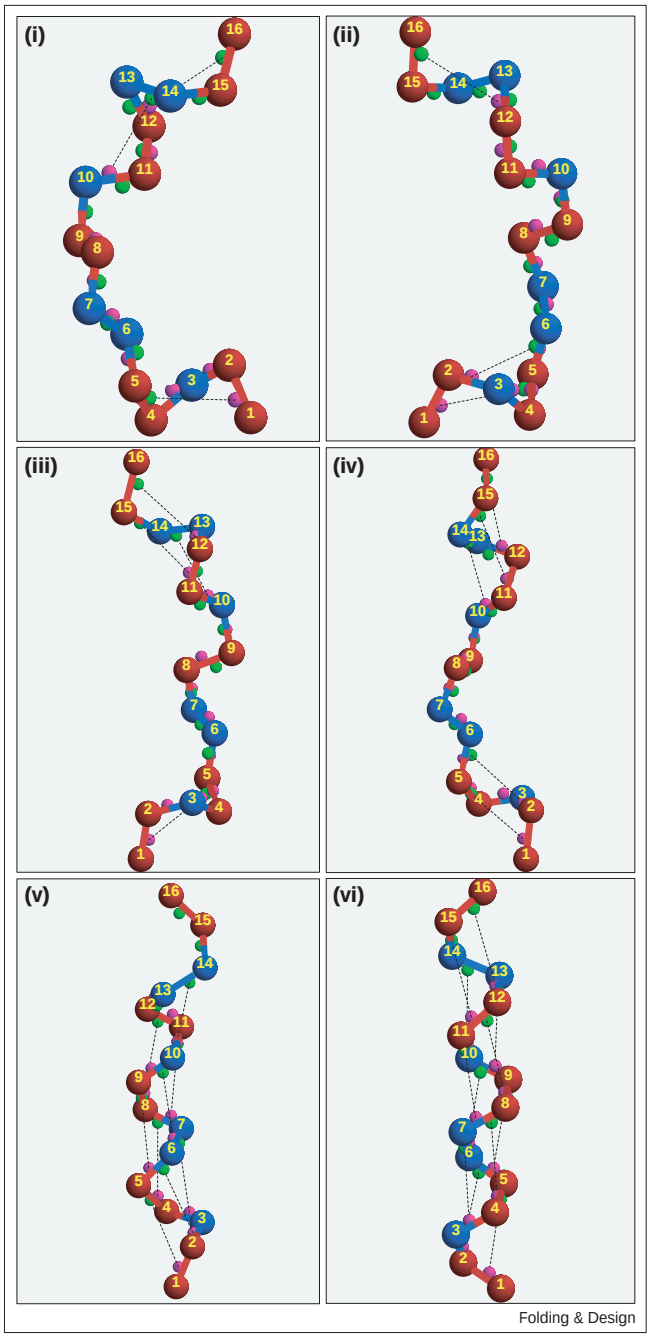


Figure 9

Time development of the transition from a nearly random coil to the α-helical conformation for the trajectory shown in Figure 8. The figure shows the formation of helix in 1.2 ns intervals after a temperature jump. The conformational changes (in terms of the number of HBs and overlap function) are as follows: (i) t = 0.6 ns: 3 HBs, χ = 0.74; (ii) t = 1.8 ns: 3 HBs, χ = 0.63; (iii) t = 3.0 ns: 4 HBs, χ = 0.59; (iv) t = 4.2 ns: 5 HBs, χ = 0.53; (v) t = 5.4 ns: 9 HBs, χ = 0.26; (vi) t = 6.0 ns: 12 HBs, χ = 0.0. Even at the earliest times certain HBs (shown for clarity by black lines) are formed. As time progresses other helical fragments emerge. At t = 4.2 ns, however, only 5 HBs are still formed. This structure is dramatically different from the native helix conformation (χ = 0.53). In the next 1.8 ns, the rest of the structure is assembled and the polypeptide chain reaches the native conformation, when all HBs are established and the value of χ becomes zero. The snapshot pictures have been created using RasMol v2.6 (R Sayle, 1995).

hydrophobic contacts (e.g. contact between residues 6 and 10), which survive with small fluctuations over the interval exceeding 60 ns. This can be seen readily from the time dependence of the fraction of native dihedral angles (D Klimov, D Thirumalai, unpublished data). Conformations containing strong interactions between hydrophobic beads serve as kinetic traps, the escape from which requires breaking non-native contacts. Generally, we observe that in slow folding trajectories, conformations acting as kinetic traps are more compact than the native helix.

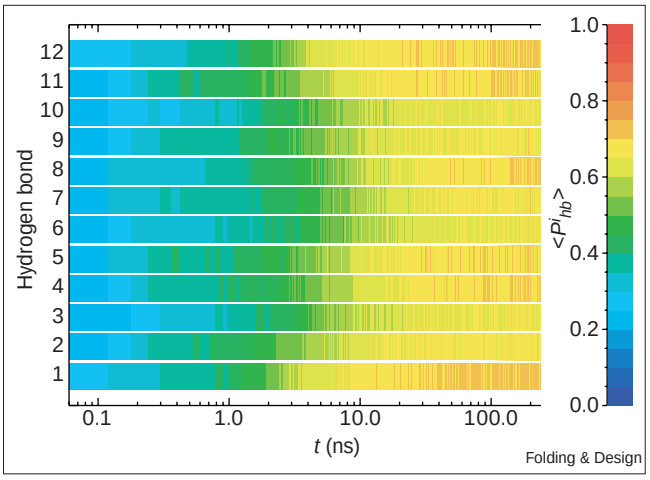
The observations described above suggest that the origin of slow phase is primarily associated with sequence heterogeneity. Therefore, if there is a strong interaction between sidechains, either due to proximity or due to intrinsic attraction, then this would stabilize the kinetic trap and lead to slower helix formation. Conversely, it follows that if the helix propensities of two residues are comparable, the homopolymer analogue would fold faster provided topological frustration does not inhibit helix formation. We have computed the rate of helix formation for sequence C — consisting only of L beads. This sequence folds almost twice as fast as sequence B and, more importantly, the slow phase amplitude and time scale are drastically reduced.

Temperature-induced helix denaturation

Further insights into the process of helix formation may also be obtained by performing unfolding simulations. These simulations have been done in the following way. First, we performed simulations in which we took the zero-temperature native structure as the initial conformation and raised the temperature to $T_w = 0.1$. The equilibration time at T_w was set to 6 ns, which was sufficient for adequate sampling of conformations that are proximal to the native state. The final conformations of this 'warming' phase are thermally distributed in the NBA and have an average overlap of about 0.137, which is very close to the equilibrium value of $\langle\chi\rangle$ at T_w. When the warming phase is completed, the temperature is raised to $T_u > T_F$. Thus, in terms of experimental technique, we conducted a sudden temperature jump in the system.

We have tested two values for T_u: $T_u = 0.45$ ($\approx 1.5T_F$) and $T_u = 0.60$ ($\approx 2T_F$). Qualitatively we found very similar results. In what follows we will describe in detail the unfolding at $T_u = 0.45$. The kinetic probabilities $P^i_{hb}(t)$ were calculated in the same way as in refolding simulations. The only difference is in the choice of the small interval over which $P^i_{hb}(t)$ probabilities are to be averaged. As the unfolding process takes place on a shorter time scale, we

Figure 10

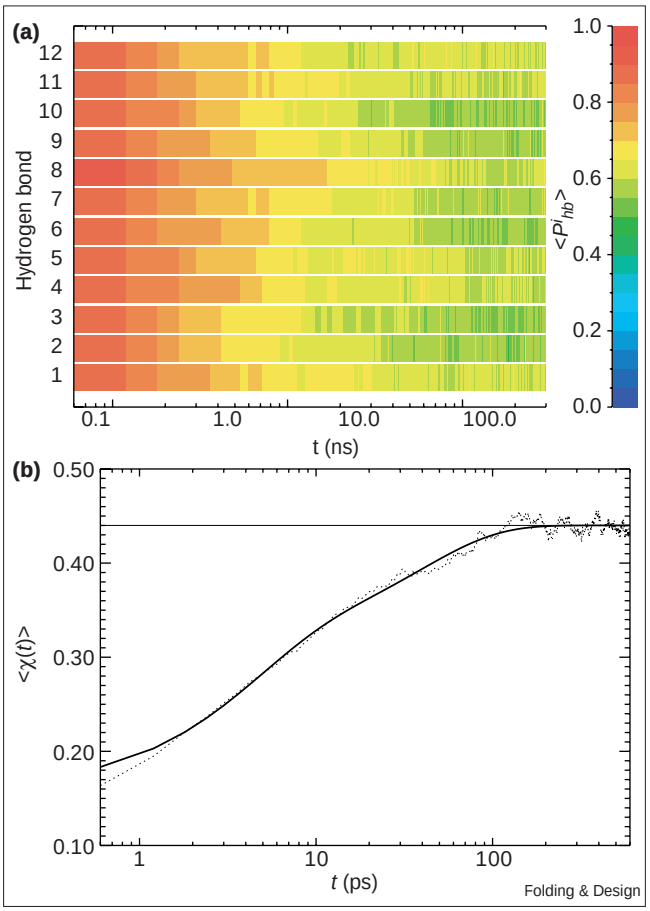


This figure shows the time dependence of probabilities of HB formation $\langle P^i_{hb} \rangle$ averaged over 200 trajectories at $T_s = 0.30$. The color code is the same as in Figure 8. HB formation – average – is initiated at helix termini and gradually propagates towards the center. This is revealed by a delay in the onset of yellow/red bars for central HBs.

chose $\Delta\tau$ to be 0.6 ps. In Figure 11a we display $\langle P^i_{hb}(t) \rangle$ obtained at $T_u = 0.45$ by averaging over 500 trajectories. Two conclusions can be drawn from this plot. First, the unfolding process is very rapid. In fact, the very first events in helix melting occur on a time scale of about 6 ps, when local rearrangements in dihedral angles and HBs presumably take place. This time scale is consistent with that of dihedral angle transitions, which were observed in full atomic unfolding simulations of isolated helix [16]. The complete melting of helix occurs over larger time scales which are discussed below. Second, unlike helix formation in refolding simulations, which is clearly non-cooperative, the helix melting appears to be dynamically cooperative. This is illustrated by the almost synchronous onset of yellow regions in Figure 11a, that is, the hydrogen bonds are almost simultaneously broken.

A complementary description of the unfolding process can be gained from monitoring the overlap function $\langle \chi(t) \rangle$ as a function of time (Figure 11b). The overlap $\langle \chi(t) \rangle$ approaches the equilibrium value $\langle \chi(T_u = 0.45) \rangle = 0.44$ in two steps as suggested by an excellent biexponential fit to $\langle \chi(t) \rangle$. The time scales for the fast and slow phases are $\tau_1 = 4$ ps and $\tau_2 = 40$ ps. Apparently, the first time scale corresponds to the local events in helix melting discussed above. The second time scale of 40 ps may be taken as the time scale of helix unfolding, because at $t \gtrsim \tau_2$ the helix reaches equilibrium. These results suggest that helix melting is considerably faster than helix formation under folding conditions. The typical folding time τ_F is of the order of 100 ns at $T < T_F$. In sharp contrast, unfolding requires less than 100 ps.

Figure 11



(a) The dynamics of break up of HBs as a function of time upon thermal denaturation of the helix averaged over 500 trajectories. The HB labels are given in the left vertical axis and the associated color codes for the probabilities $\langle P^i_{hb} \rangle$ are shown on the right-hand side. The unfolding is initially local. Around $t \geq 10$ ps a dynamic global unfolding begins due to fraying of the ends of the helix. On this time scale the probabilities $\langle P^i_{hb} \rangle$ dramatically decrease. (b) The time dependence of $\langle \chi(t) \rangle$ averaged over 500 trajectories. The solid line represents a biexponential fit to $\langle \chi(t) \rangle$. The first time scale $\tau_1 = 4$ ps corresponds to initial local events in helix meeting (see upper panel). On the second time scale of $\tau_2 = 40$ ps complete (global) unfolding takes place.

It is useful to compare the results of our unfolding kinetics to the detailed atomic simulations of thermally induced helix denaturation reported by Daggett and Levitt some time ago [16]. These authors used molecular dynamics simulations to monitor the melting of a 13-mer which forms a three-turn α helix. They estimated that unfolding occurs on a time scale of 25–200 ps depending on the final temperature. This is in excellent agreement with our results on a reduced description model. Microscopically, we find that at short times ($\lesssim 10$ ps) the melting involves only local loss of hydrogen bonds. The global unfolding occurs at about 40 ps, when the hydrogen bonds at the ends of the helix are broken. After that, "fraying of ends" [16] occurs and the

helix melting proceeds in a dynamically cooperative manner. These microscopic events were also seen in the simulations of Daggett and Levitt [16].

Conclusions

It is interesting to put the results presented here in the context of developments in the understanding of protein folding that have taken place over the past several years. On the one hand, theoretical advances have provided a conceptual framework for thinking about global aspects of protein folding [1–5]. Experiments on very fast time scales are beginning to dissect the assembly of proteins into the formation of very simple fragments such as α helices and β turns [8,9,14]. Detailed atomic simulations of peptides and unfolding simulations of proteins are providing a picture of protein folding that is quite consistent with both experiments and theories based on minimal models [16,18,20,46,47]. The time scales on which helices form and melt are accessible both experimentally and in molecular dynamics simulations. Thus, the validity of the minimal models can be directly assessed by making detailed comparisons with experiments and atomic simulations. In this paper, we have shown that the models studied here provide very accurate descriptions of both the thermodynamics and kinetics of helix–coil transition. In particular, the refolding times obtained here and their temperature dependence are nearly in quantitative agreement with experiments on similar-sized helices [14]. Comparison of unfolding simulations with full atomic molecular dynamics simulations on small helical fragments also shows consistent mechanisms and unfolding time scales [16]. In addition, the structure of the helix is nearly the same as the classical α helix found in proteins. These comparisons serve to calibrate the model and suggest that the minimal models, on which very detailed computations are possible, can be used to predict folding kinetics in helix and β-turn motifs.

Acknowledgements

We are very grateful to Victor Munoz for useful discussion. This work was supported in part by a grant from the National Science Foundation through grant number CHE96-29845. M.R.B. is a National Science Foundation Minority Postdoctoral Fellow.

References

1. Wolynes, P.G., Onuchic, J.N. & Thirumalai, D. (1995). Navigating the folding routes. *Science* **267**, 1619-1620.
2. Dill, K.A., Bromberg, S., Yue, K., Fiebeg, K.M., Yee, D.P., Thomas, P.D. & Chan, H.S. (1995). Principles of protein folding – a perspective from simple exact models. *Protein Sci.* **4**, 561-602.
3. Dill, K.A. & Chan, H.S. (1997). From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10-19.
4. Thirumalai, D., Klimov, D.K. & Woodson, S.A. (1997). Kinetic partitioning mechanism as unifying theme in the folding of biomolecules. *Theor. Chem. Acct.* **1**, 23-30.
5. Bryngelson, J.D., Onuchic, J.N., Socci, N.D. & Wolynes, P.G. (1995). Funnels, pathways and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167-195.
6. Jones, C.M., Henry, E.R., Hu, Y., Chan, C-K., Luck, S.D., Bhuyan, A.K., Roder, H., Hofrichter, J. & Eaton, W.A. (1993). Fast events in protein folding initiated by nanosecond laser photolysis. *Proc. Natl. Acad. Sci. USA* **90**, 11860-11864.
7. Eaton, W.A., Munoz, V., Thompson, P.A., Chan, C-K. & Hofrichter, J. (1997). Sub-millisecond kinetics of protein folding. *Curr. Opin. Struct. Biol.* **7**, 10-14.
8. Thompson, P.A., Eaton, W.A. & Hofrichter, J. (1997). Laser temperature jump study of the helix-coil kinetics of an alanine peptide interpreted with a "kinetic-zipper" model. *Biochemistry* **36**, 9200-9210.
9. Munoz, V., Thompson, P.A., Hofrichter, J. & Eaton, W.A. (1997). Folding dynamics and mechanisms of β-hairpin formation. *Nature* **390**, 196-199.
10. Eaton, W.A., Munoz, V., Thompson, P.A., Henry, E.R. & Hofrichter, J. (1998). Kinetics and dynamics of loops, α-helices, β-hairpins, and fast folding proteins. *Acc. Chem. Res.* in press.
11. Bopp, M.A., Jia, Y, Li, L., Cogdell, R.J. & Hochstrasser, R.M. (1997). Fluorescence and photobleaching dynamics of single light-harvesting complexes. *Proc. Natl. Acad. Sci. USA* **94**, 10630-10635.
12. Huang, G.S. & Oas, T.G. (1995). Submillisecond folding of monomeric λ repressor. *Proc. Natl. Acad. Sci. USA* **92**, 6878-6882.
13. Veitshans, T., Klimov, D.K., & Thirumalai, D. (1996). Protein folding kinetics: time scales, pathways, and energy landscapes in terms of sequence dependent properties. *Fold. Des.* **2**, 1-22.
14. Williams, S., Causgrove, T.P., Gillmanshin, R., Fang, K.S., Callender, R.H., Woodruff, W.H. & Dyer, R.B. (1996). Fast events in protein folding: helix melting and formation in a small peptide. *Biochemistry* **35**, 691-697.
15. Ladurner, A.G. & Fersht, A.R. (1998). Upper limit of the timescale for diffusion and chain collapse in chymotrypsin inhibitor 2. *Nat. Struct. Biol.* in press.
16. Daggett, V. & Levitt, M. (1992). Molecular dynamics simulations of helix denaturation. *J. Mol. Biol.* **223**, 1121-1138.
17. Boczko, E.M. & Brooks III, C.L. (1995). First principle calculation of the folding free energy of a three helix bundle protein. *Science* **21**, 393-396.
18. Guo, Z., Brooks III, C.L. & Boczko, E.M. (1997). Exploring the folding free energy surface of a three-helix bundle protein. *Proc. Natl. Acad. Sci. USA* **94**, 10161-10166.
19. Demchuk, E., Bashford, D. & Case, D.A. (1997). Dynamics of a type VI reverse turn in a linear peptide in aqueous solution. *Fold. Des.* **2**, 35-46.
20. Mohanty, D., Elber R, Thirumalai, D., Beglov, D. & Roux, B. (1997). Kinetics of peptide folding: computer simulations of SYPFDV and peptide variants in water. *J. Mol. Biol.* **272**, 423-442.
21. Honeycutt, J.D. & Thirumalai, D. (1990). Metastability of the folded states of globular proteins. *Proc. Natl. Acad. Sci. USA* **87**, 3526-3529.
22. Honeycutt, J.D. & Thirumalai, D. (1992). The nature of folded states of globular proteins. *Biopolymers* **32**, 695-709.
23. Guo, Z. & Thirumalai, D. (1996). Kinetics and thermodynamics of folding of a *de novo* designed four-helix bundle protein. *J. Mol. Biol.* **263**, 323-343.
24. Irback, A., Peterson, C., Potthast, F., Sommelius, O. (1997). Local interactions and protein folding: a 3D off-lattice approach. *J. Chem. Phys.* **107**, 273-290.
25. Klimov, D.K. & Thirumalai, D. (1997). Viscosity dependence of the folding rates of proteins. *Phys. Rev. Lett.* **79**, 317-320.
26. Guo, Z. & Brooks III, C.L. (1997). Thermodynamics of protein folding: a statistical mechanical study of a small all β protein. *Biopolymers* **42**, 745-757.
27. Honig, B. & Cohen, F.E. (1996). Adding backbone to protein folding: why proteins are polypeptides. *Fold. Des.* **1**, R17-R20.
28. Shea, J.E., Nochomovitz, Y.D., Guo, Z. & Brooks III, C.L. (1998). Exploring the space of protein folding hamiltonians: the balance of forces in a minimalist β-barrel model. *J. Chem. Phys.* **109**, 2895-2903.
29. Nymeyer, H., Garcia, A.E. & Onuchic, J.N. (1998). Folding funnels and frustration in off-lattice minimalist protein landscapes. *Proc. Natl. Acad. Sci. USA* **95**, 5921-5928.
30. Flory, P.J. (1988). *Statistical Mechanics of Chain Molecules.* Hanser Publishers, New York.
31. Oldfield, T.J. & Hubbard, R.E. (1994). Analysis of *C*α geometry in protein structures. *Proteins* **18**, 324-337.
32. Creighton, T.E. (1993). *Proteins: Structures and Molecular Principles.* W.H. Freeman & Co., New York.
33. Munoz, V. & Serrano, L. (1995). Elucidating the folding problem of helical peptides using empirical parameters. III. Temperatures and pH dependence. *J. Mol Biol* **245**, 247-308.
34. Munoz, V. & Serano, L. (1995). Analysis of *i*, *i*+5 and *i*, *i*+8 hydrophobic interactions in a helical model peptide bearing the hydrophobic staple motif. *J. Mol. Biol.* **245**, 275-296.

35. Kumar, S. & Bansal, M. (1998). Dissecting α-helices: position-specific analysis of α-helices in globular proteins. *Proteins*, 31, 460-476.
36. Ermack, D.L. & McCammon, J.A. (1978). Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.* 69:1352-1369.
37. Rey, A. & Skolnick J. (1991). Comparison of lattice Monte Carlo dynamics and Brownian dynamics folding pathways of α-helical hairpins. *Chem. Phys.* 158, 199-219.
38. He, S. & Scheraga, H.A. (1998). Brownian dynamics simulations of protein folding. *J. Chem. Phys.* 108, 287-300.
39. Ferrenberg, A.M. & Swendsen, R.H. (1989). Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* 63, 1195-1198.
40. Klimov D.K. & Thirumalai, D. (1998). Cooperativity in protein folding: from lattice models with side chains to real proteins. *Fold. Des.* 3, 127-139.
41. Zimm, B.H. & Bragg, J.K. (1959). Theory of the phase transition between helix and random coil in polypeptide chains. *J. Chem. Phys.* 31, 526-535.
42. Flory, P.J. & Miller, W.G. (1966). A general treatment of helix-coil equilibria in macromolecular systems. *J. Mol. Biol.* 15, 284-297.
43. Poland, D. & Scheraga, H.A. (1970). In *Theory of Helix-Coil Transitions in Biopolymers*. Academic Press, New York.
44. Qian, H. & Schellman, J.A. (1992). Helix-coil theories. A comparative study for finite length peptides. *J. Phys. Chem.* 96, 3987-3994.
45. Socci, N.D. & Onuchic, J.N. (1995). Kinetic and thermodynamic analysis of protein-like heteropolymers: Monte Carlo histogram technique. *J. Chem. Phys.* 103, 4732-4744.
46. Daggett, V., Li, A., Itzhaki, L.S., Otzen, D.E. & Fersht, A.R. (1996). Structure of the transition state for folding of a protein derived from experiment and simulation. *J. Mol. Biol.* 257, 430-440.
47. Li, A. & Daggett, V. (1996). Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 using molecular dynamics simulations. *J. Mol. Biol.* 257, 412-429.

Because *Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/fad – for further information, see the explanation on the contents pages.