PROTEINS: Structure, Function, and Genetics 36:249–261 (1999)

# Direct Computation of Long Time Processes in Peptides and Proteins: Reaction Path Study of the Coil-to-Helix Transition in Polyalanine

**Shuanghong Huo**[1] **and John E. Straub**[1,2]*
[1]*Department of Chemistry, Boston University, Boston, Massachusetts*
[2]*Institute for Advanced Studies, The Hebrew University of Jerusalem, Givat Ram, Israel*

***ABSTRACT*** **The MaxFlux reaction path algorithm was used to isolate optimal transition pathways for the coil-to-helix transition in polyalanine. Eighteen transition pathways, each connecting one random coil configuration with an ideal α-helical configuration, were computed and analyzed. The transition pathway energetics and mechanism were analyzed in terms of the progression of the peptide nonbonded contact formation, helicity, end-to-end distance and energetics. It was found that (1) localized turns characterized by i, i + 3 hydrogen bonds form in the early stages of the coil-to-helix transition, (2) the peptide first collapses and then becomes somewhat more extended in the final stage of helix formation, and (3) $3_{10}$-helix formation does not appear to be a necessary step in the transition from coil to helix. These conclusions are in agreement with the results of more computationally intensive direct molecular dynamics simulations. Proteins 1999;36:249–261.** © 1999 Wiley-Liss, Inc.

**Key words: helix-coil transition; peptide dynamics; diffusion; polyalanine; reaction paths; continuum solvation**

## BACKGROUND

The helix-coil transition is the most extensively studied biomolecular conformational transition. Simple analytical models, such as the Zimm-Bragg model[1] and its generalizations,[2] have provided a valuable way of thinking about the thermodynamics of the helix-coil equilibrium. The peptide is typically represented as a chain of residues or turns that can take on either a helical or coil configuration. A nucleation parameter provides a measure of the difficulty in first forming a helical segment. A coupling parameter measures how each residue or helical turn is stabilized by being adjacent to a helical residue or turn. Such models predict the cooperative formation of helix as has been seen in experiments.[3–5] Both experiment and computer simulation have been used to define the parameters of such theories.[6–10]

Computer simulations have now been used to investigate the dynamics of atomistic models of a peptide in solution for extended times.[11,12] The simulations have been used to parameterize simple models for the kinetics of helix formation. However, at the moment, microsecond dynamics of helix formation is still at the outer limits of computation.[12]

It is worth asking what can be said about the dynamics and thermodynamics of the helix-coil transition using methods that stand between simple analytical models and full dynamical simulations employing atomistic models. One possibility is to use a coarse-grained model where each or every other residue is represented by a single interaction site. This type of model has been criticized for its lack of "backbone" or explicit amide and carbonyl groups for hydrogen bonding.[13,14] However, a proper choice of torsional potential[15–17] or the use of a virtual bond model[14] can partially incorporate the important hydrogen bonding stabilization resulting from the formation of α-helical peptide segments in a realistic fashion. By removing many of the higher frequency motions, time steps used in the simulation of the dynamics of such minimal models of proteins can be on the order of picoseconds making it possible to directly stimulate coil-to-helix dynamical transitions.

A shortcoming of such simple models is that many questions of interest can only be clearly answered by atomistic simulations. What role does the solvent dynamics play in the transition? What role do the side chains play in the thermodynamics and dynamics of the transition? How do the relative stability of the helix and rate of helix formation depend on the peptide sequence?

In this work we present a computational method that stands between the dynamical simulation of atomistic and minimal models. We employ an atomistic model of the peptide that allows us to explore the sequence dependence of helix formation. We replace the solvent by a solvation potential proposed by Wesson and Eisenberg.[18] Most importantly, we replace the actual dynamical simulation with a variational estimate of the mean first passage time to move between coil and helical configurations. Starting from known reactant (coil) and product (helix) conformations, we find an optimal pathway connecting the two states that minimizes an estimate of the mean first passage time. This MaxFlux algorithm[19] assumes a diffusive dynamics (ignoring inertial motion) of the peptide

Grant sponsor: the National Science Foundation; Grant number: CHE-9632236.
*Correspondence to: John Straub, Department of Chemistry, Boston University, Boston, MA 02215
Received 19 January 1998; Accepted 15 January 1999

which should be well justified for the coil-to-helix transition in solvent for sufficiently long peptides.

One motivation for this coil-helix transition study is role played by secondary structure formation in protein folding. Both the diffusion-collision mechanism[20,21] and framework model[22] assume that the secondary structural elements form first followed by coalescence of secondary structural elements to form the tertiary structure. Recent microsecond molecular dynamics simulation of the folding of the fully solvated villin headpiece subdomain[12] did suggest the dominance of the secondary structure formation during the early stage of folding. Since the α-helix is a common secondary structural element, the study of helix formation is necessary if we are to understand the details of protein folding. A second motivation is that the existence of long time molecular dynamics simulations allows us to test the accuracy of our results against true dynamical trajectories.

Previous molecular dynamics simulation of polyalanine[10,23] helical denaturation and equilibrium fluctuation demonstrated that the global helix unraveling started either at the termini of the helix or within the helix where a kink or turn formed by creation of a non-i,i+4 hydrogen bond. The refolding was often found to begin from the turn regions. Is the dominant mechanism in helical formation the nucleation of helical structure by an internal turn? What is the exact role of the $3_{10}$-helical motif in the random coil to α-helix transition? A number of theoretical studies[24–26] showed that α-helix is energetically more stable than the $3_{10}$-helix. The result was contrary to the ESR spectral data obtained by the Millhauser group[27,28] which implied the coexistence of $3_{10}$-helix and α-helix. These are the questions that we have addressed in this study.

In the following sections we describe the MaxFlux algorithm and our computational model of the peptide. We then describe results of the calculation of a number of reaction pathways for the coil-to-helix transition in polyalanine. The resulting "bundle" of transition pathways is then analyzed using a number of plausible reaction coordinates to clarify the thermodynamics, dynamics, and mechanism of helix formation.

## ACETYL-$(Ala)_{10}$-N-METHYL PEPTIDE MODEL

The total system energy consisted of the intramolecular potential energy for the peptide (CHARMM potential[29]) and the solvation energy term. An implicit solvent model developed by Wesson and Eisenberg[18] was employed. The solvation model assumes that the solvation free energy of a given atom in the peptide is proportional to the atom's solvent-accessible surface area. The atomic solvation energies per unit exposed atomic area are C ($12 \pm 3$), O/N ($-116 \pm 13$), S ($-18 \pm 21$), O− ($-175 \pm 36$) and N+ ($-186 \pm 22$) in units of cal $mol^{-1}$ $Å^{-2}$ as given in Table 3 of the paper by Wesson and Eisenberg.[18] There are two important features of the solvation parameters. First, the atomic solvation parameter for C is positive (hydrophobic) while all others are negative (hydrophilic). Second, the magnitude of the atomic solvation parameters of C and S are roughly one order of magnitude smaller than those of the polar atoms N and O and the ions O− and N+.

The model was applied to molecular dynamics simulations of synthetic peptide folding[30] and reasonable results were obtained. Abagyan and coworkers[31] also employed this model in their biased probability Monte Carlo conformational search method. The implicit solvent model not only reduced the computational burden of an explicit water model but also avoided the difficulties in energy minimization caused by disordered water molecule orientations.

### Advantages of a Continuum Solvation Model

Computing the optimized transition pathways involves energy minimization of a set of intermediate structures along the transition pathway. When an explicit water model is used, there will be a variety of disordered water structures of low energy that are compatible with any given peptide configuration. Typically, no single solvent configuration will provide an accurate estimate of the solvation energy which is an average over the distribution of water configurations consistent with the peptide configuration. Using an explicit solvent model, two similar peptide configurations may have different surrounding solvent configurations with significantly different energies. Such large variations are a result of a lack of averaging over the distribution of solvent configurations consistent with a given peptide structure and can lead to wild variations in the solvation energy along the transition pathway.

The use of a solvation potential such as that of Wesson and Eisenberg has the advantage of implicitly including the average over all solvent configurations consistent with a given peptide configuration at room temperature. Two similar peptide configurations will have similar solvation energies. As a result, the solvation energy along the transition pathway will be a consistent and smoothly varying function of the peptide geometry.

### Intial Guess at the Transition Pathways

The ideal extended state of the α-helix of decaalanine was built using Quanta97.[32] Subsequently, the acetyl and N-methyl groups were patched at the N-terminus and C-terminus, respectively, using the CHARMM simulation program.[29] Energy minimization was then used to relieve strain in the idealized model structure. To generate a set of random coil configurations, 18 trajectories of 200 ps in length were computed using molecular dynamics simulation at an elevated temperature of 3,000K. Each trajectory originated from an ideal, extended configuration with a random set of initial atomic velocities. Strain in the end point structures was relieved using energy minimization where the peptide bond ω dihedral angle was constrained to be 180°. A distance-dependent dielectric constant was used to mimic the solvent screening. The resulting 18 random coil structures were used as "reactant" configurations in the reaction path calculations. The mass weighted

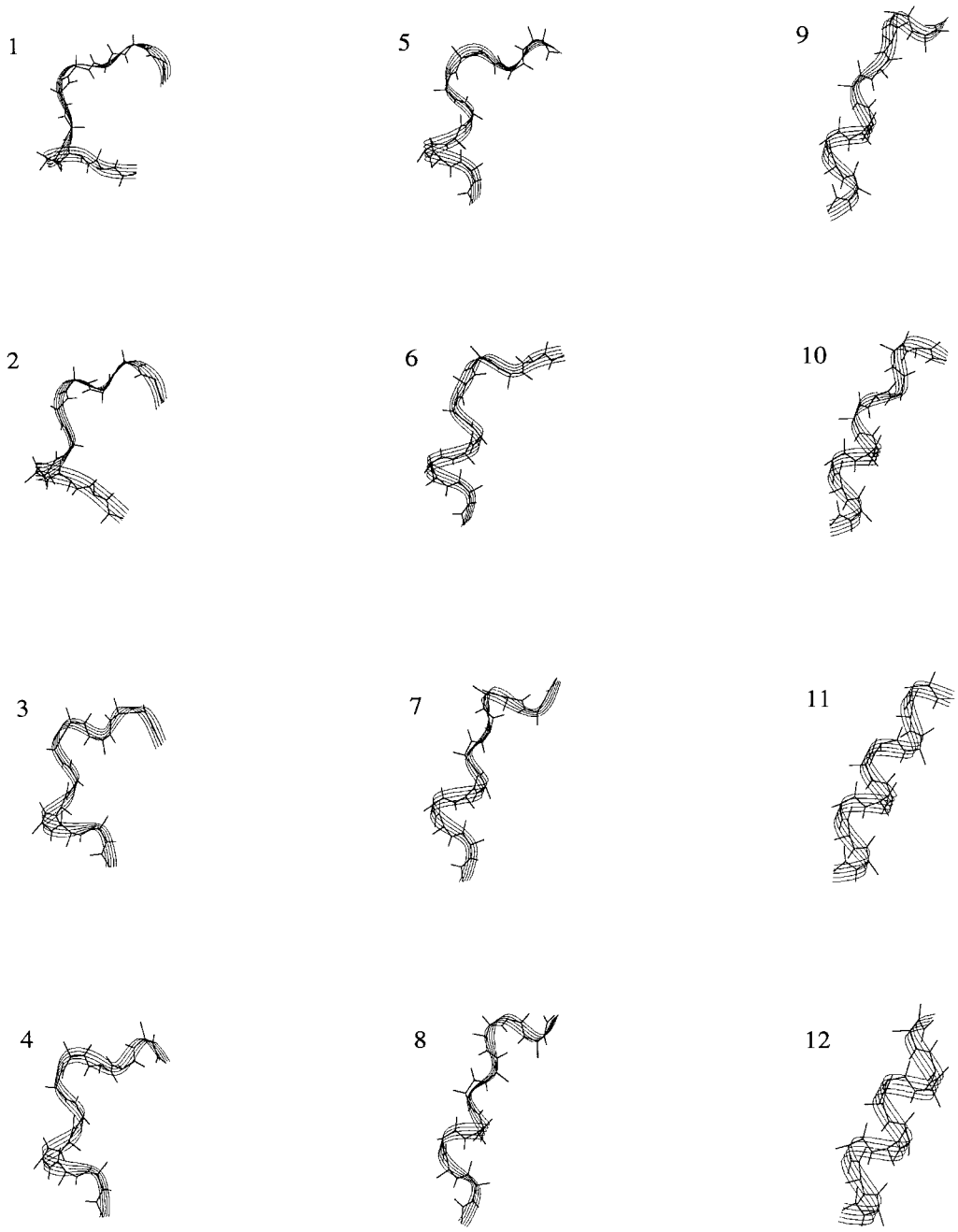


Fig. 1. Structures along the optimized transition pathway connecting a random coil structure (#1) and the ideal α-helical structure (#12). This transition pathway shows no initial collapse transition. The helix propagates from the N-terminus upwards to completion.

root-mean-square deviation between each pair of random coil configurations was greater than 3.4 Å. In each case, the "product" configuration was taken to be an ideal α-helix.

A linear interpolation in ψϕ space was performed to generate the initial guess for the coil-to-helix transition path. Ten intermediate structures were generated along the path connecting a random coil reactant to the α-helical product. To relieve strain in the initial pathway, restrained energy minimization was performed with the ϕ and ψ dihedral angles fixed. Conjugate gradient energy minimization was used to relieve strain in the initial, guessed path. The parameter set for the path restraints was taken to be $\beta = 0.1$ and $\lambda = 2.0$. The parameters κ and ρ were set by trials. The range of values which were used for the eighteen trajectories are $\kappa \in [800, 5000]$ and $\rho \in [8000, 30000]$ in standard units. These paths correspond to high temperature ($1/\beta$) reaction paths.

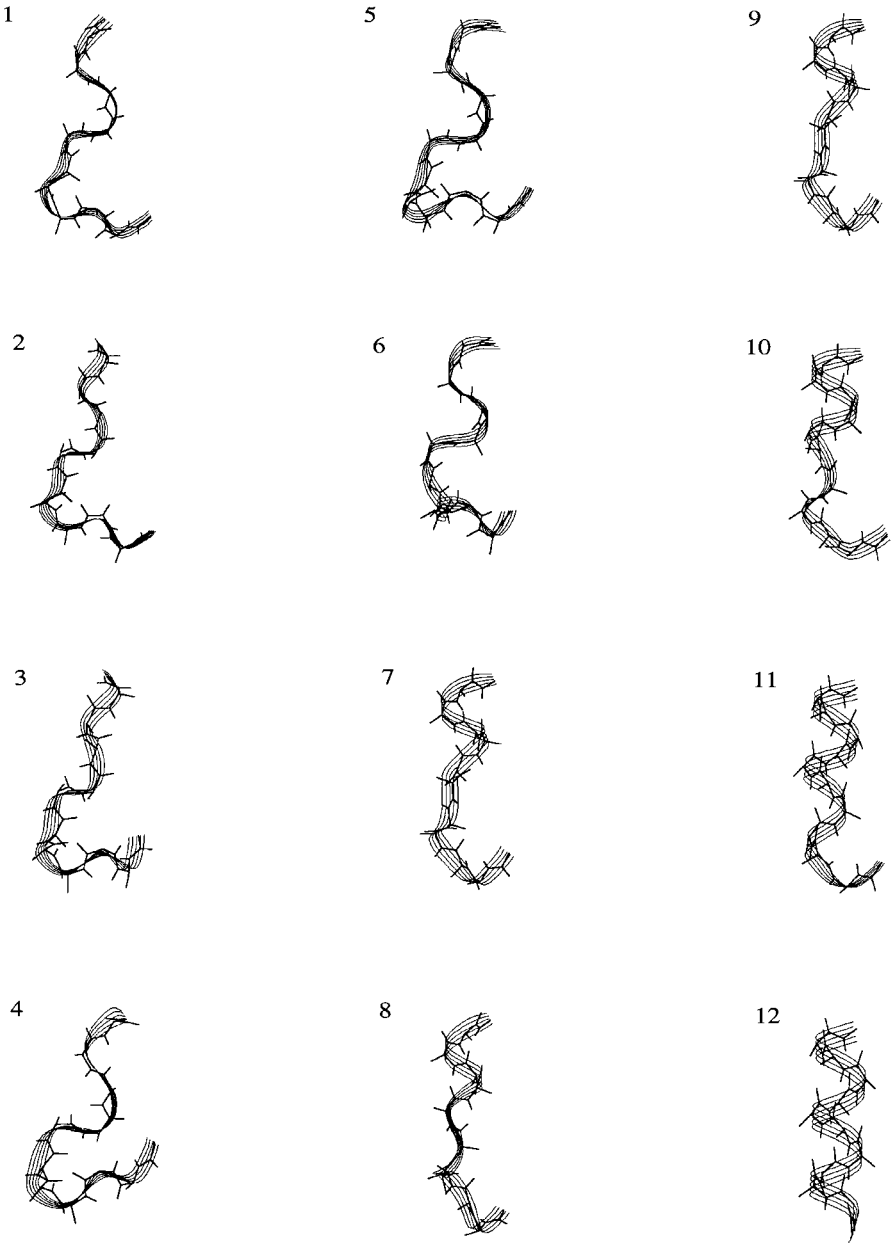


Fig. 2. Structures along the optimized transition pathway connecting a random coil structure (#1) and the ideal α-helical structure (#12). This transition pathway shows an initial collapse transition followed by formation of nascent, $3_{10}$-helix and then final and complete propagation of α-helix.

## COMPUTATIONAL MAXFLUX METHOD FOR ISOLATING IMPORTANT TRANSITION PATHWAYS

For large conformational transitions in biomolecules it is reasonable to approximate the dynamics of the system as overdamped and diffusive as described by the Smoluchowski equation.[33] The system moves on a multidimensional potential of mean force U(r) damped by a friction γ(r) at a temperature T. When the friction is isotropic in space the diffusion tensor $D(r) = (k_BT/m\gamma(r))I(r)$. We can define the pathway l(r) in terms of a series of peptide configurations leading from the reactant configuration to product configuration. Berkowitz et al.[34] defined the optimal reaction path as the path l(r) of "minimum resistance" where the resistance is proportional to

$$\mathscr{R} = \int_{r_R}^{r_P} e^{\beta \mathscr{W}(r)}\, dl(r). \quad (1)$$

The effective potential $\mathscr{W}(r) = U(r) + \kappa_B T \ln(\gamma(r)/\gamma(r_R))$ where $\gamma(r_R)$ is the friction at an arbitrarily chosen reference point. This definition is consistent with the fact that there is a strong bias for the system to remain in configurations of low energy. If there are two configurations differing in energy by ΔU, the configuration of lower energy will be more probable by a factor of exp (βΔU). This principle provides a definition of the optimal transition pathway l(r)

between $r_R$ and $r_P$ as one of minimum resistance or maximum reactive flux.

If one further assumes that the friction is spatially invariant so that $\gamma(r)$ = constant, the effective potential $\mathscr{W}(r) = U(r)$ and the optimal transition pathway corresponds to the minimum of the line integral

$$\mathscr{P} = \int_{r_R}^{r_P} e^{\beta U(r)}\, dl(r). \tag{2}$$

This underlying assumption will be exact in a one dimensional system. It will be accurate for a many dimensional system when the group of transition pathways (both the path of minimum resistance, and the adjacent paths of greater resistance) are parallel and non-intersecting.[34]

We have proposed a computational algorithm which can be used to isolate solutions to this variational problem. It is referred to as the MaxFlux method. Following the "path" protocol of Elber and coworkers,[35] we minimize a discretized form of the line integral Eq. (3) with added restraints. The discretized integral

$$\mathscr{P}(R) = \sum_{k=0}^{M-1} e^{\beta U(r_k)} |r_{k+1} - r_k| \,. \tag{3}$$

can be thought of as a chain of M + 1 "snap shot" structures at a series of positions $R = (r_R, r_1 \ldots r_{M-1}, r_P)$ along a transition pathway connecting reactant and product. The objective is to minimize Eq. (2) by adjusting the positions of the intermediates in the chain, all the while encouraging the distances between successive monomers in the chain (the dl(r) increments) to be equal in magnitude. This is accomplished through the use of a number of restraints.[35] (1) One restraint acts as a bond between nearest neighbor intermediate structures to encourage the mean-square distances between adjacent structures to be approximately constant

$$\mathscr{C}_A(R) = \kappa \sum_{k=1}^{M} [r_k - r_{k-1})^2 - d_{ave}^2]^2 \tag{4}$$

where $d_{ave}^2 = \Sigma_{k=1}^{M}(r_k - r_{k-1})^2/M$. (2) A repulsive interaction between intermediates along the path

$$\mathscr{C}_R(R) = \frac{\rho}{\lambda} \sum_{j>k+1} \exp\,[-\lambda(r_j - r_k)^2/\langle d\rangle^2] \tag{5}$$

prevents two intermediates from coming too close to one another where $\langle d\rangle = \Sigma_{k=1}^{M}(r_k - r_{k-1})/M$. This makes the path a self-avoiding walk. (3) For molecular systems there are constraints that eliminate rigid body translations and rotations[35]

$$\sum_{\mu=1}^{N} m_\mu (r_\mu - r_\mu^{fix}) = 0 \tag{6}$$

$$\sum_{\mu=1}^{N} m_\mu r_\mu \times r_\mu^{fix} = 0 \tag{7}$$

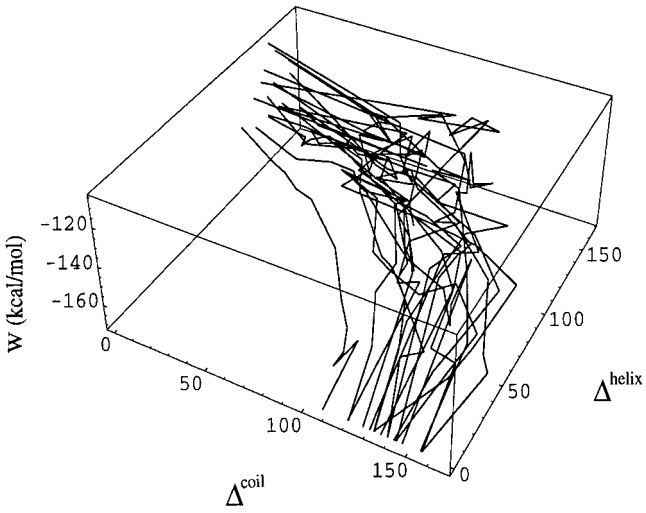


Fig. 3. The bundle of transition pathways for the helix coil transition plotted as a function of the mean distance from the ideal helical geometry in (ψ, ϕ) space from reactant (X) and product (Y) against the total energy (Z) which is a sum of internal, nonbonded, and solvation energies. The results seem to support an underlying assumption of the MaxFlux method that the trajectories are roughly parallel as they move through space.

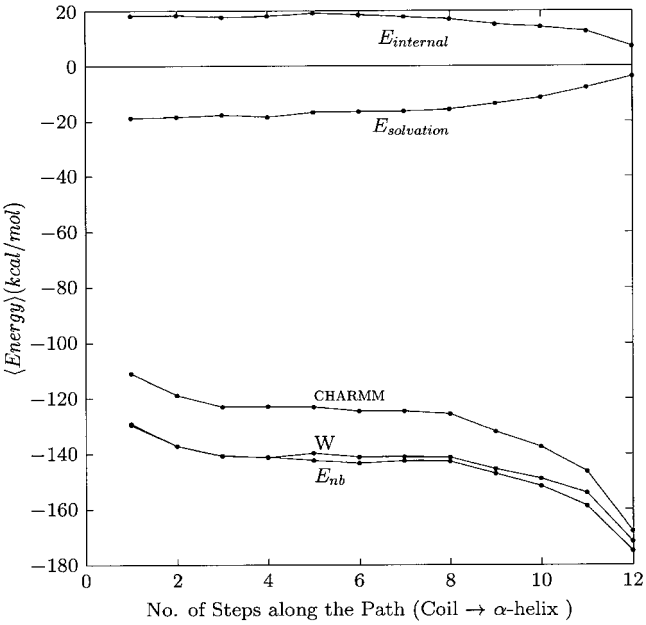


Fig. 4. The energetics of the peptide along the optimized transition pathways is decomposed into contributions from the nonbonded interaction potential energy, the solvation energy and the intramolecular potential energy.

where N is the number of atoms in the system, $m_\mu$ is the atomic mass and $r_\mu$ the Cartesian coordinates for the μth atom. $\{r_\mu^{fix}\}_{\mu=1,N}$ is the arithmetic average of the coordinate of the ith atom in the reactant and product configurations.

Now the objective function is defined as $\mathscr{O}(R) = \mathscr{P}(R)$ + constraints. Our job is to find the global minimum value of the objective function in the space of all possible reaction paths. This is a computationally demanding task in a large

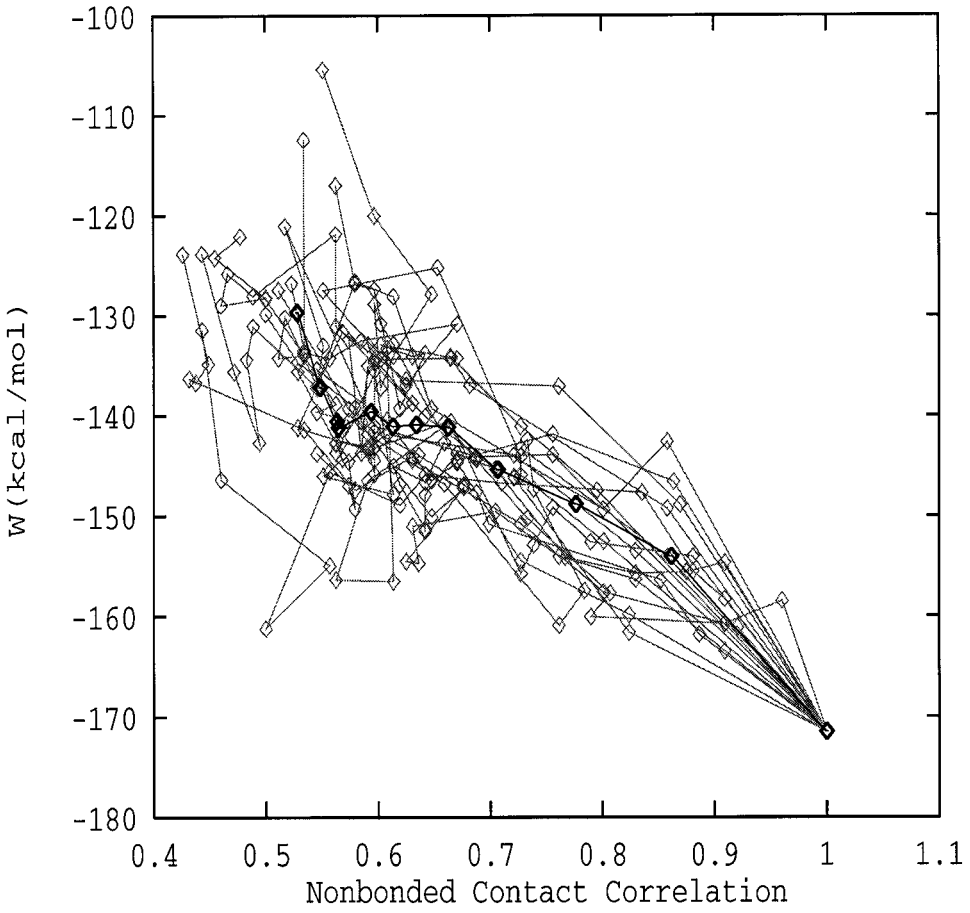


Fig. 5. Plot of the energy as a function of the nonbond contact correlator, $C_{nbc}(k)$, for each transition pathway from the random coil structure to the final $\alpha$-helical structure. The average over all pathways is shown as the dark line.

biomolecule. It is best to use a global minimization method which effectively explores all possible reaction pathways. This approach will be employed in the future. For reasons of computational efficiency, in this work we optimize the cost function using a conjugate gradient minimization for the chain $R = (r_R, r_1 \ldots r_{M-1}, r_P)$ of intermediate configurations $r_\kappa = (r_{1k}, \ldots, r_{\mu k}, \ldots, r_{Nk})$ each with N particles where the reactant ($r_R$) and product ($r_P$) configurations are held fixed.

## OVERALL ANALYSIS OF COIL-TO-HELIX TRANSITION AND ENERGETICS

The resulting optimized transition pathway consists of a series of intermediate structures or "snapshots" along the reaction path. "Movies" showing the progression from unique random coil (#1) to $\alpha$ helix (#12) configurations are displayed in Figures 1 and 2. An important feature common to both pathways is that the helix formation occurs by nucleation of initial helical turns followed by propagation along the peptide. An alternative mechanism would be a more uniform progression, at each residue, from the initial coil configuration to the helical structure. Such a uniform progression, which is in fact our initial guessed path, is not seen in the final trajectories. This indicates that the variational optimization of the pathway isolates a path of minimum resistance leading to significant changes in the transition pathway from the initial guessed path.

A general feature that emerges from our analysis is that a majority of pathways show an initial collapse and subsequent helix propagation while others show direct propagation from an extended state. In Figure 1 we see an example of a rare trajectory which proceeds directly to helix formation without a collapse transition (path #6). The first helical segment forms near the N-terminus in the 8th intermediate configuration out of a turn region that is present in the initial reactant structure. In the 8th intermediate the helix begins to propagate upward toward the C-terminus until there is complete formation of $\alpha$-helix.

A typical transition pathway which shows the initial collapse and subsequent helix formation is shown in Figure 2 (path #7). The initial random coil state becomes slightly compacted over the first three steps in the transition. On reaching the 7th configuration the first helical turn is formed near the C-terminus. The helix is then seen to propagate downward toward the N-terminus until the helix is fully formed. At intermediate stages the peptide is rich in i,i + 3 hydrogen bonds and forms $3_{10}$ helical segments starting at step #8.

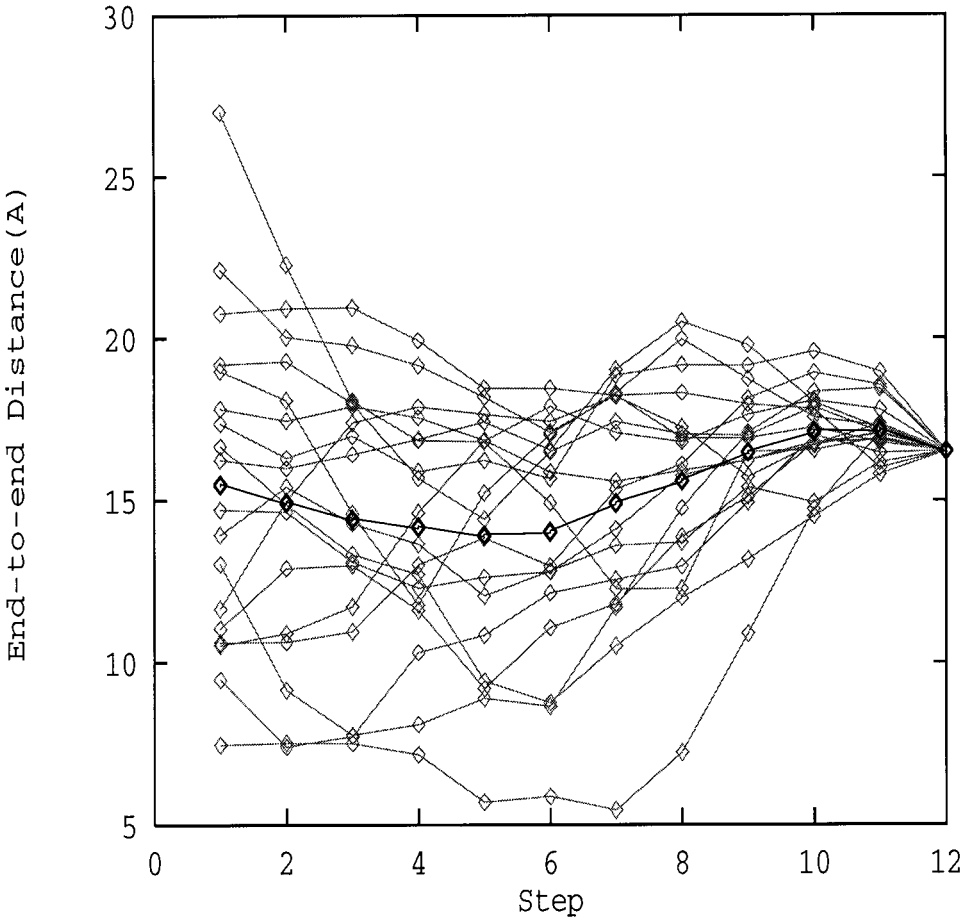


Fig. 6. Plot of the end-to-end distance (in Å) of the peptide as a function of the intermediate position along the transition pathway from the random coil structure to the final α-helical structure. The average over all pathways is shown as the dark line.

**Reaction Diagram of Trajectory Bundle**

To monitor the overall progress of the transition from coil to helix of the bundle of eighteen transition pathways, we employ the residue average root-mean-square deviation in ψφ space. The deviation between the kth intermediate structure and the helical product structure is defined

$$\Delta^{\text{helix}}(k) = \frac{1}{N}\sum_{i}^{N}[(\phi_i(k) - \phi_i^{\text{helix}}(k))^2 + (\psi_i(k) - \psi_i^{\text{helix}}(k))^2]^{1/2} \quad (8)$$

A similar definition exists for $\Delta^{\text{coil}}(k)$ for each pathway. This function provides one measure of the distance between intermediate structures along the coil-to-helix transition connecting the initial random coil (reactant) and α-helix (product) configurations. In Figure 3, we present the effective energy of the solvated peptide as a function of the distance from the coil reactant and helical product configurations along the transition pathway. The effective energy of initial random coils is in the range of −105 kcal/mol to −145 kcal/mol. Considering the rms deviation in Cartesian coordinates, where the pairwise root-mean-square distance between pairs of random coil structures is at least 3.4 Å, it is clear that the initial random coil structures are widely distributed. Furthermore, Figure 3 shows that the paths are almost parallel after passing through the relatively irregular range at the beginning of the coil-helix transition. This supports the idea that the underlying assumption of a collection of parallel and effectively one-dimensional transition paths is justified.

**Energetics of the Coil-to-Helix Transition**

Figure 4 shows the energy as a function of the extent of transition averaged over all computed pathways. The energy shows an initial decrease associated with the collapse transition. This decrease is followed by a plateau region with relatively little change in the energy as the peptides reorganize within a set of compact states. The final precipitous decrease in energy coincides with the propagation of α helix which proceeds in a downhill fashion to the final helical structure. Over the transition pathways the solvation energy is on average a monotonically increasing function. The rise in solvation energy shows the enthalpic cost associated with the burial of polar backbone groups in the formation of hydrogen bonds and helix. The mostly monotonic decrease in internal energy of the peptide demonstrates that the helix is relatively unstrained compared with the random coil configurations. The largest contribution to the total energy is found to be

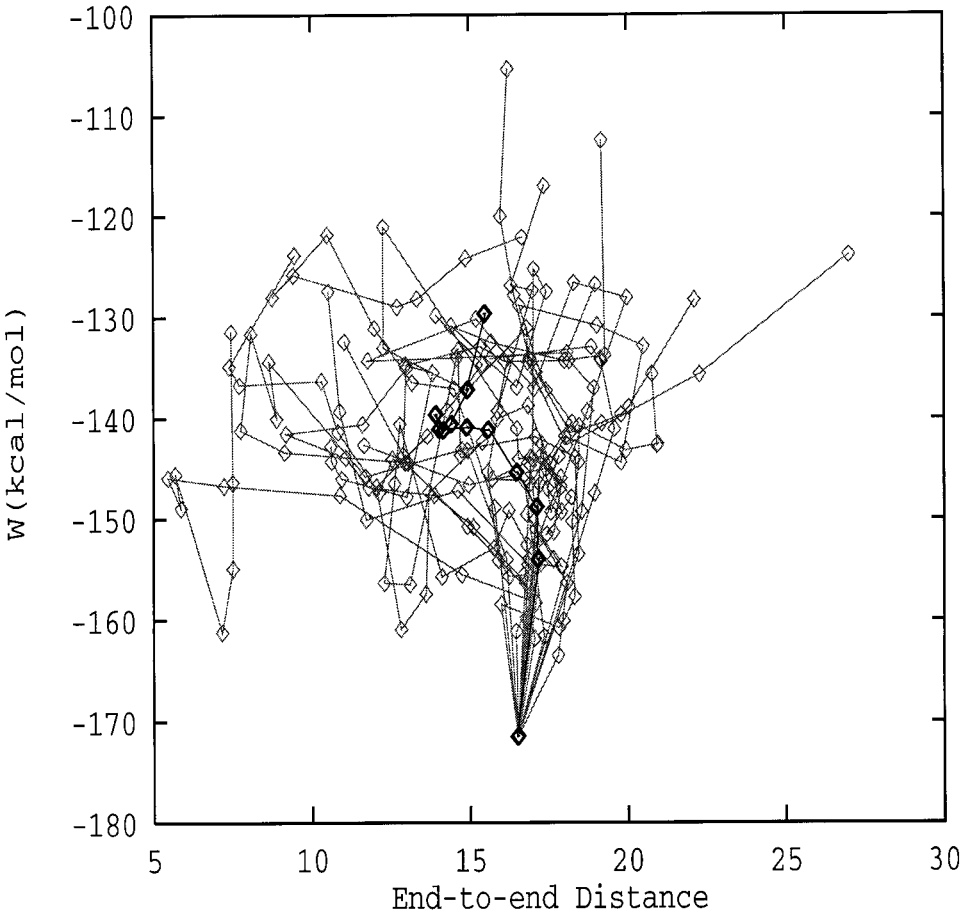


Fig. 7. Plot of the total energy of the peptide versus the end-to-end distance (in Å) of the peptide along the transition pathway from the random coil structure to the final α-helical structure. The average over all pathways is shown as the dark line.

the nonbonded energy. The initial drop in nonbonded energy is associated with the formation of compact structures and $3_{10}$ and α helical hydrogen bonds. The final and precipitous drop in the nonbonded energy comes with the formation of significant helical structure with backbone geometries ideal for hydrogen bonding.

## ANALYSIS OF TRANSITION PATHWAYS

A useful measure of the progression of the peptide transition is the nonbonded contact correlation function. It was shown to be a good measure of the compactness of an off-lattice model protein.[36] In this context the measure is defined as

$$C_{nbc}(k) = \frac{1}{N_{nbc}} \sum_{pairs} \Theta\,[\rho - r_{ij}(k)]\,\Theta\,[\rho - r_{ij}(0)]$$

$$\Theta(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad (9)$$

where $r_{ij}(k)$ is the distance between the ith and jth atom in the kth structure. The "target" structure corresponds to $k = 0$ and

$$N_{nbc} = \sum_{pairs} \Theta\,[\rho - r_{ij}(0)] \qquad (10)$$

**TABLE I. Comparison of the Point of Appearance of the Shortest Peptide Intermediate (in Terms of the End-to-End Distance) With the First Appearance of Partial Helical Structure (an i, i + 4 Hydrogen Bond)**

| Pathway number | Shortest intermediate | First helical intermediate |
|---|---|---|
| 1 | 6 | 9 |
| 2 | 12 | 8 |
| 3 | 5 | 11 |
| 4 | 4 | 7 |
| 5 | 1 | 11 |
| 6 | 1 | 8 |
| 7 | 4 | 8 |
| 8 | 5 | 10 |
| 9 | 6 | 7 |
| 10 | 2 | 9 |
| 11 | 6 | 9 |
| 12 | 8 | 11 |
| 13 | 5 | 6 |
| 14 | 7 | 8 |
| 15 | 3 | 11 |
| 16 | 7 | 8 |
| 17 | 1 | 9 |
| 18 | 2 | 8 |

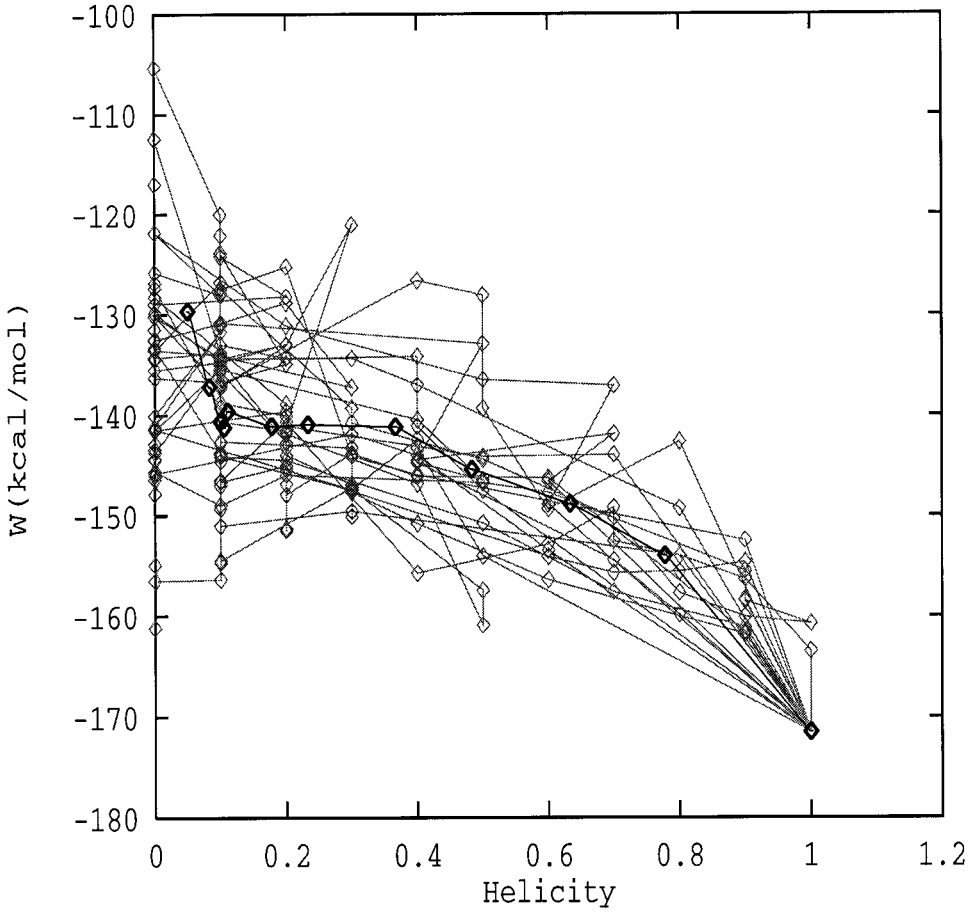


Fig. 8. Plot of the total energy of the peptide versus the helicity of the peptide along the transition pathway from the random coil structure to the final α-helical structure. The average over all pathways is shown as the dark line.

is the number of nonbonded contact atom pairs in the target structure. ρ is set to 5.4 Å corresponding to one turn of α helix. Only the backbone atoms are included in the measurement.

Figure 5 shows the effective energy as a function of nonbonded contact correlation for the backbone atoms. Viewed as a "bundle," the paths form a funnel-like reaction channel. The head of the channel is relatively broad. The channel becomes narrow when the nonbonded contact correlation reaches 0.66. A structure with $C_{nbc}$ greater than 0.8 can be considered to be in a native-like compact state.

In the kinetic partitioning mechanism of protein folding proposed by Thirumalai[37,38] and coworkers, a fraction of trajectories undergo rapid transition to the folded state while the remaining fraction misfold into a low energy intermediate structure. Those misfolded structures must overcome energetic barriers to reach the correctly folded state. Analysis of individual transition pathways shows that roughly $\Phi = \frac{2}{3}$ of the paths move energetically downhill in the helix formation from the native-like compact state. Roughly $1 - \Phi = \frac{1}{3}$ of the paths must overcome an effective energy barrier on forming the ideal α-helix from the native-like compactness. Direct dynamical simulations of coarse-grained models for helix formation[14] have found that a fraction $\Phi = 0.79$ of the folding trajectories reach the native state on a fast time scale while the remainder are trapped in an intermediate state. Our findings are in reasonable agreement with this result.

### Peptide Collapse Preceeds Helix Formation

The end-to-end distance is a commonly used monitor of the overall configuration of the peptide. Figure 6 presents the end-to-end distance along each of the computed transition pathways. The distance between the C-terminal main chain O atom and the N-terminal N atom is defined as the end-to-end distance. The length of the ideal α-helical configuration is 16.5 Å while the length of the fully extended state is 34.9 Å. Figure 6 shows that typically the end-to-end distance first becomes short before subsequently increasing to reach the ideal α-helix value. Daggett and Levitt[10] found the same pattern of motion, in reverse, in their molecular dynamics simulation of the denaturation of a 13-residue polyalanine helix in aqueous solution. During the first 50 ps of dynamics at 373 K, the end-to-end distance fluctuated and then began to decrease. It reached a minimum value after 120 ps of dynamics. At that point, the helicity was less than 18%. Subsequently, the peptide's end-to-end distance and overall helicity began to increase. Our computed transition pathways capture the same features of the transition.

**TABLE II. Hydrogen Bond Propagation Along the Coil-Helix Transition Pathways†**

| Pathway number | H-bond type | Intermediates along coil-helix transition pathway | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 1 | $3_{10}$ | | | | 7 | 6, 7 | 5, 6 | 5 |
| | α | | | | 7 | 7 | 6–8 | 3, 5–8 |
| 2 | $3_{10}$ | 5 | 5, 8 | 2, 3, 5, 8 | 2–5, 8 | 1–5, 8, 9 | 1, 2, 5, 8, 9 | 1, 8 |
| | α | | | 5 | 5 | 5 | 2, 5, 6 | 1–6 |
| 3 | $3_{10}$ | | | | | | 1, 3, 5, 7, 8 | 1, 3, 5, 7 |
| | α | | | | | | 8 | 1, 7, 8 |
| 4 | $3_{10}$ | | 2, 3, 6 | 2, 6, 8 | 2, 7, 8 | 1, 2, 7, 9 | 1, 2, 7–9 | 1, 2, 7, 8 |
| | α | 2 | 2 | 2 | 2 | 2 | 2, 3, 8 | |
| 5 | $3_{10}$ | | | | | | 1, 2, 7, 8 | 1–7 |
| | α | | | | | 7 | 7, 8 | |
| 6 | $3_{10}$ | | | 2, 8 | 2, 4, 8 | 1, 2, 4, 5, 8, 9 | 1, 2, 5, 6, 8, 9 | 1, 7–9 |
| | α | | | 2 | 2 | 2 | 2, 3 | 2–5 |
| 7 | $3_{10}$ | | | 8 | 7–9 | 6–8 | 5, 6 | 3–5 |
| | α | | | | | 8 | 6–8 | 5–8 |
| 8 | $3_{10}$ | | | | | 1, 2, 4, 7 | 1–4, 6, 7 | 1, 2, 5, 6 |
| | α | | | | | 7 | 3, 7, 8 | 2, 3, 6–8 |
| 9 | $3_{10}$ | | — | 8, 9 | 2, 3, 8, 9 | 2, 8, 9 | 1, 2, 5, 8, 9 | 1, 4, 7–9 |
| | α | | | | | | 2, 5 | 1, 2, 4, 5 |
| 10 | $3_{10}$ | | | | — | 3–5, 9 | 2–6, 9 | 2, 9 |
| | α | | | | | | 6 | 2–7 |
| 11 | $3_{10}$ | | | | 1, 5, 6 | 1, 5, 6 | 1, 5, 7 | 1, 3, 5 |
| | α | | | | 7 | 6, 7 | 5–7 | 5, 7, 8 |
| 12 | $3_{10}$ | | | | | | 2, 4, 9 | 1, 8, 9 |
| | α | | | | | | 4 | 1, 6 |
| 13 | $3_{10}$ | 3, 7 | | 3 | 1, 4, 7 | 1, 4 | 1, 4 | |
| | α | 1 | 1, 3, 4 | 1, 3–5 | 1, 4, 5 | 1, 4, 5 | 1, 4, 5, 7 | 3–5, 7, 8 |
| 14 | $3_{10}$ | | | 6, 7 | 5–7 | 3–5 | 3 | 3 |
| | α | | | | | 5–8 | 3–8 | 3–8 |
| 15 | $3_{10}$ | | | | | | 5, 6, 9 | 1, 3, 5 |
| | α | | | | | | 6, 7 | 5–8 |
| 16 | $3_{10}$ | | | 3 | 3, 7, 8 | 7, 8 | 7 | 3 |
| | α | | | 1, 4 | 1, 4 | 1, 3, 4 | 1, 3, 4, 7, 8 | 1, 3–8 |
| 17 | $3_{10}$ | | | | | 1, 4 | 2, 4, 5, 7 | 2, 7, 9 |
| | α | | | | 8 | 4 | 4 | 2–4 |
| 18 | $3_{10}$ | | | 8 | 3, 7, 8 | 3, 5–8 | 1, 3, 5–8 | 1, 5, 8 |
| | α | | | | | | | 2, 3, 5, 6 |

†Each number indicates the residue at which the helical turn originates. An entry of i in column $3_{10}$ (α) indicates the presence of an i, i + 3 (i, i + 4) hydrogen bond. Consecutive occurance of three such hydrogen bonds indicates the presence of a $3_{10}$ or α helical segment.

Figure 7 shows the total energy of the peptide as a function of the end-to-end distance for all computed transition pathways. The plot demonstrates the extent of the energy landscape explored by the bundle of trajectories. The plot shows a deep energy minimum positioned at the α-helical end-to-end distance. The more extended peptide conformations are all of higher energy. However, there are distinct minima on the surface at shorter end-to-end distances associated with collapsed intermediate states with low van der Waals energy by higher electrostatic energy due to the incomplete formation of hydrogen bonds relative to the helical configuration. These collapsed states also appear in Figure 5 as being states of incomplete nonbond contact formation.

In Table I we compare the point of appearance of the shortest end-to-end distance with that of the first helical segment along the coil-helix transition pathway. A residue is defined as having a configuration compatible with the formation of a helical turn when its backbone torsional angles are in the range $-100° < \phi < -30°$ and $-80° < \psi < -5°$.[10] A helical segment is defined as three continuous residues having a helical conformational region. Comparing the end-to-end distance along each pathway with the degree of helix formation, one can see that the random coils collapse and become more compact before any helical segments form. This is true for seventeen of the eighteen trajectories.

In Figure 8 we show the measure of helicity versus the total energy along the computed transition pathways. The correlation between the helicity and energy is quite strong; in general, a greater degree of helicity implies a lower energy. This is not the case for either the nonbonded contact correlator (Fig. 5) or the end-to-end distance (Fig. 7) which show distinct energy minima along the transition pathway.

### Local Turns Form With a Mixture of $3_{10}$ and α Helical Segments

Table II shows how the $O_i$, $H_{i+3}$ and $O_i$, $H_{i+4}$ hydrogen bonds developed along the transition pathways. The backbone hydrogen bonds were identified using Quanta97[32] employing the definition that the maximum distance between the amide H atom and acceptor carbonyl O atom is 2.5 Å and the minimum angle between the CO and NH bond vectors is 90.0. The first step appearing in the table represents the step before the first formation of a helical segment. If no i, i+3 or i,i+4 hydrogen bonds formed before the first helical segment appeared, a "−" appears in the corresponding column. The $3_{10}$-helical segment was defined as one containing three or more consecutive (i, i + 3) hydrogen bonds. We found that four out of eighteen trajectories contained pure short $3_{10}$-helical segments.

The general mechanism of α helix formation suggested by the result in Table II is the following. 1) In the first state, defined as a transition from random coil ↔ nascent helix, the localized i,i + 3 hydrogen bonds are easily formed and more highly populated than i,i + 4 hydrogen bonds. 2) In the second stage, nascent helix ↔ $3_{10}$-helix, for eight (1,3,11,12,13,15,16,17) out of eighteen trajectories, we did not observe three or more sequential i,i + 3 hydrogen bonds. This implies that the $3_{10}$-helical segments did not form. For six (2,4,5,6,8,9) out of the complementary ten trajectories, we found $3_{10}$-helix mixed with either localized i,i + 4 hydrogen bonds or α-helical segments.

In Figure 9 we plot the probability of seeing a certain measure of helicity at a fixed intermediate step and location in the peptide chain computed for the bundle of transition pathways. The plot demonstrates that the helix formation begins internal to the C-terminal and N-terminal positions. This same pattern is observed in the probability of degree of nonbonded contact formation shown in Figure 10 where the internal residues to the N-terminal and C-terminal sides of the peptide center show a greater degree of compactness than the peptide center or termini.

### $3_{10}$ Helix Formation is Not a Necessary Intermediate in α Helix Formation

Millhauser proposed a mechanism for helix formation which includes the $3_{10}$-helix as an intermediate along the transition pathway.[27] His mechanism can be written as a progression of types of hydrogen bonds (HBs) that predominate at points along the transition pathway. The mechanism is random coil (no HB) ↔ nascent helix (transient HB i → i + 3) ↔ $3_{10}$-helix (HB i → i + 3) ↔ α-helix (HB i → i + 4). From Table II, it appears that five trajectories (2,7,10,14,18) fit a loosely defined version of Millhauser's mechanism. An example of such a trajectory is shown in the "movie" of Figure 2 (path #7).

However, most of the trajectories show a mixture of characteristic i,i + 3 and i,i + 4 hydrogen bonds. The so-called nascent helix, which corresponds to the intermediate conformer, advanced to a configuration containing helical segments with localized turns and a mixture of i,i + 3 and i,i + 4 hydrogen bonds. One can argue that this observation is dependent on the exact definition of a hydrogen bond used in the calculation. However, it is clear that during the propagation of hydrogen bond formation, we observed intermediate configurations containing a mixture of i,i + 3 and i,i + 4 hydrogen bonds rather than a pure $3_{10}$-helix segment.

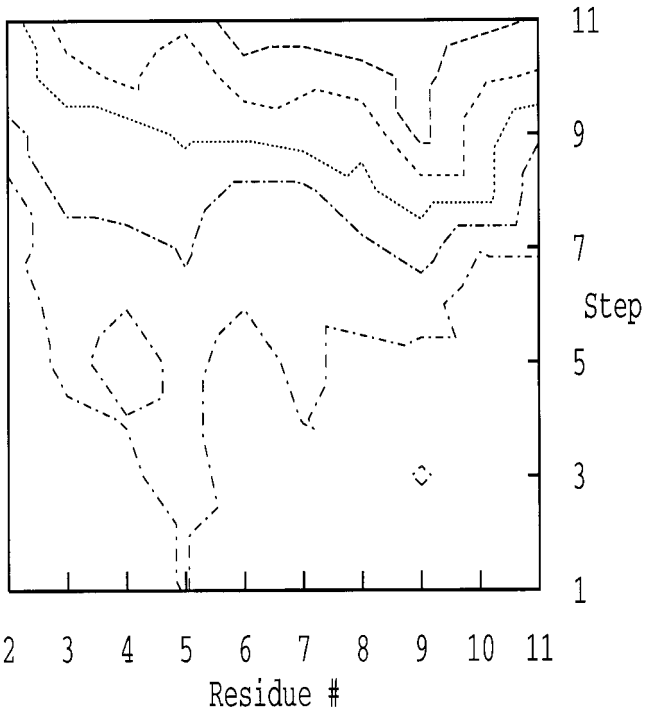


Fig. 9. The average percentage of helix in the peptide averaged over all computed transition pathways (Z-axis contours) as a function of the number of the participating residue (X-axis) and the step along the transition pathway (Y-axis).

It should be noted that previous molecular dynamics simulations are consistent with these results. Sung and coworkers[30] observed that i,i + 3 hydrogen bonds formed frequently during the helix formation and denaturation when they studied the synthetic alanine-based peptide folding. They rarely found a complete $3_{10}$-helix. However, they often observed i,i + 3 hydrogen bonds amidst i,i + 4 hydrogen bonds in a helical segment. Brooks and coworkers[39] noted the breaking of i,i + 4 hydrogen bonds accompanied by the formation of i,i + 3 hydrogen bonds in their MD simulation of alanine-based peptides. These observations lead to the conclusion that $3_{10}$-helix may act as an intermediate in a transition pathway connecting random coil and α-helical configurations. However, it is unlikely that it can be identified as a unique intermediate. This is consistent with the fact that in the study of N- and C-terminal helix growth in polyalanine, Brooks and coworkers[9] did not find that the $3_{10}$-helix corresponded to a local minimum on their free energy surface for the terminal ψ dihedral angle transition but they did observe a small population of $3_{10}$-helix. They suggested that $3_{10}$-helix was a kinetic intermediate.

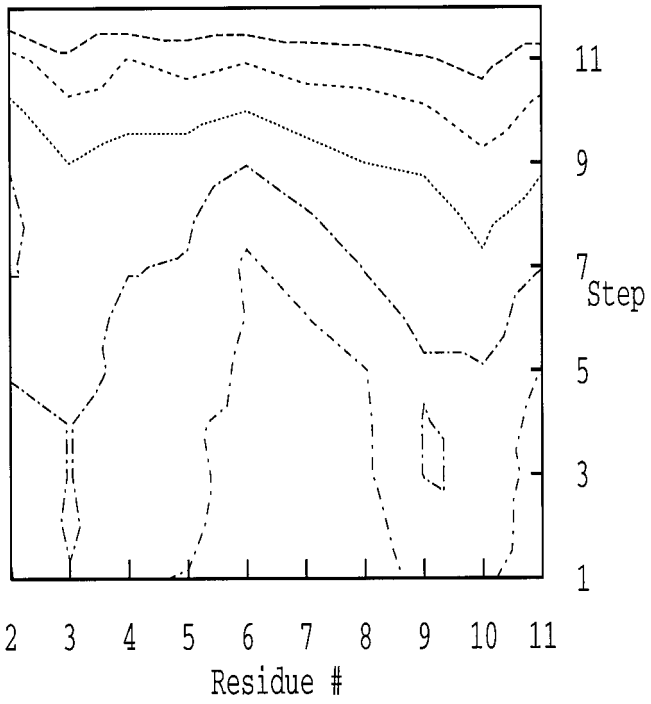


Fig. 10. The average percentage of the nonbond contact correlator in the peptide averaged over all computed transition pathways (Z-axis contours) as a function of the number of the participating residue (X-axis) and the step along the transition pathway (Y-axis).

## CONCLUSIONS

The MaxFlux algorithm provides an alternative to molecular dynamics and Monte Carlo for the simulation of large scale configurational transitions in macromolecules. We have explored the transition from random coil to α-helix in the acetyl-$(Ala)_{10}$-N-methyl peptide. Each optimized transition pathway connecting random coil and α-helical configurations was found to be unique when examined using various measures of the extent of transition including the number of helical turns, the extent of formation of nonbonded contacts, the number of torsional angle transitions, and the peptide end-to-end distance. These measures were useful in demonstrating that the unique transition pathways shared certain common features. (1) Highly populated, localized turns, characterized by i,i + 3 hydrogen bonds, were found to form in the early stages of the coil-to-helix transition. (2) The peptide was found to collapse and become more compact before extending slightly in the final states of helix formation. (3) $3_{10}$-helix formation during the coil-helix transition does not appear to be a necessary step in the transition from coil to helix.

## ACKNOWLEDGMENTS

We thank D. Eisenberg and L. Wesson for providing the implicit solvation energy calculation source code. S. Huo is grateful for helpful discussion with L. Wesson. JES gratefully acknowledges the National Science Foundation for support (CHE-9632236) and the Center for Scientific Computing and Visualization at Boston University for computational resources.

## REFERENCES

1. Zimm BH, Bragg JK. Theory of phase transition between helix and random coil in polypeptide chains. J Chem Phys 1959;31:526.
2. Qian H, Schellman J. Helix-coil theories: a comparative study for finit length polypeptides. J Phys Chem 1992;96:3987.
3. Poland D, Scheraga HA. Theory of helix-coil transitions. New York. Academic Press, 1970.
4. Kemp DS, Boyd JG, Muendel CC. The helical s constant for alanine in water derived from template-nucleated helices. Nature 1991;352:451.
5. Chakrabartty A, Kortemme T, Baldwin RL. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. Protein Sci 1994;3:843.
6. Scholtz JM, Qian H, York EJ, Stewart JM, Baldwin RL. Parameters of helix-coil transition theory for alanine-based peptides of varying chain lengths in water. Biopolymers 1991;31:1463.
7. Wójcik J, Altmann KH, Scheraga HA. Helix-coil stability constants for the naturally occuring amino acids in water. xxiv. half-cystine parameters from random poly (hydroxybutlglutamine-co-s-methylthio-l-cysteine). Biopolymers 1990;30:121.
8. Wang L, O'Connell T, Tropsha A, Hermans J. Thermodynamic parameters for the helix-coil transition of oligopeptides: molecular dynamics simulation with the peptide growth method. Proc Natl Acad Sci USA 1995;92:10924.
9. Young WS, Brooks CL III. A microscopic view of helix propagation: N and c-terminal helix growth in alanine helices. J Mol Biol 1996;259:260.
10. Daggett V, Levitt M. Molecular dynamics simulations of helix denaturation. J Mol Biol 1992;223:1121.
11. Mohanty D, Elber R, Thirumalai D, Beglov D, Roux B. Kinetics of peptide folding: computer simulations of SYPFDV and peptide variants in water. J Mol Biol 1997;272:423.
12. Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Proc Natl Acad Sci USA 1998;282:740.
13. Honig B, Cohen FE. Adding backbone to protein folding: why proteins are polypeptides. Fold Des 1996;1:R17–R20.
14. Klimov DK, Betancourt MR, Virtual atom representation of hydrogen bonds in minimal off-lattice models of α-helices: effect on stability, cooperativity and kinetics. Thirumalai D. Fold Des 1998;3:481–496.
15. Rey J, Skolnick J. Comparison of lattice Monte Carlo dynamics and Brownian dynamics folding pathways of α-helical hairpins. Chem Phys 1991;158:199.
16. Honeycutt JD, Thirumalai D. The nature of folded states in globular proteins. Biopolymers 1992;32:695.
17. Guo Z, Thirumalai D. Kinetics and thermodynamics of folding of a *de novo* designed four-helix bundle protein. J Mol Biol 1996;263: 323.
18. Wesson L, Eisenberg D. Atomic solvation parameters applied to molecular dynamics of proteins in solution. Protein Sci 1992;1: 227.
19. Huo S, Straub JE. The maxflux algorithm for calculating variationally optimized reaction paths for conformational transitions in many body system at finite temperature. J Chem Phys 1997;107: 5000.
20. Karplus M, Weaver DL. Protein-folding dynamics. Nature 1976; 260:404.
21. Karplus M, Weaver DL. Diffusion-collision model for protein folding. Biopolymers 1979;18:1421.
22. Baldwin RL. How does protein folding get started? Trends Biochem Sci 1989;14:291.
23. Daggett V, Kollman PA, Kuntz ID. A molecular dynamics simulations of polyalanine: an analysis of equilibrium motions and helix-coil transitions. Biopolymers 1991;31:1115.
24. Tirado-Rives J, Maxwell DS, Jorgensen WL. Molecular dynamics and Monte Carlo simulations favor the *alpha*-helical form for alanine-based peptides in water. J Am Chem Soc 1993;115:11590.
25. Smythe ML, Huston SE, Marshall GR. Free energy profile of a $3_{10}$- to α-helical transition of an oligopeptide in various solvents. J Am Chem Soc 1993;115:11594.

26. Zhang L, Hermans J. $3_{10}$ helix versus α-helix: a molecular dynamics study of conformational preferences of aib and alanine. J Am Chem Soc 1994;116:11915.
27. Millhauser GL. Views of helical peptides: a proposal for the position of $3_{10}$-helix along the thermodynamic folding pathway. Biochemistry 1995;34:3873.
28. Miick MS, Martinez GV, Fiori WR, Todd AP, Millhauser GL. Short alanine-based peptides may form $3_{10}$-helices and not α-helices in aqueous-solution. Nature 1992;359:653.
29. Brooks BR, Bruccoleri R, Olafson B, States D, Swaninathan S, Karplus M. CHARMM: a program for macromolecular energy minimization and dynamics calculations. J Comp Chem 1983;4: 187.
30. Sung S, Wu X. Molecular dynamics simulations of synthetic peptide folding. Proteins 1996;25:202.
31. Abagyan R, Totrov M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. J Mol Biol 1994;235:983.
32. Molecular Simulations, Inc. (1997).
33. Gardiner CW. Handbook of stochatic methods. New York: Springer Verlag; 1983.
34. Berkowitz M, Morgan JD, McCammon JA, Northrup SH. Diffusion-controlled reactions: a variational formula for the optimum reaction coordinate. J Chem Phys 1983;79:5563.
35. Czerminski R, Elber R. Self-avoiding walk between two fixed points as a tool to calculate reaction paths in large molecular systems. Int J Quant Chem 1990;24:167.
36. Guo Z, Thirumalai D, Honeycutt JD. Folding kinetics of proteins: a model study. J Chem Phys 1992;97:525.
37. Guo Z, Thirumalai D. Nucleation mechanism for protein-folding and theoretical predictions for hydrogen-exchange labeling experiments. Biopolymers 1994;35:137.
38. Thirumalai D, Klimov DK, Woodson SA. Kinetics partitioning mechanism as a unifying theme in the folding of biomolecules. Theor Chem Acc 1997;1:23.
39. Shirley WA, Brooks CL III. Curious structure in canonical alanine-based peptides. Proteins 1997;28:59.