PROTEINS: Structure, Function, and Genetics 44:344–360 (2001)

α-Helix Formation: Discontinuous Molecular Dynamics on an Intermediate-Resolution Protein Model

Anne Voegler Smith and Carol K. Hall
Department of Chemical Engineering, North Carolina State University, Raleigh, North Carolina

ABSTRACT **An intermediate-resolution model of small, homogeneous peptides is introduced, and discontinuous molecular dynamics simulation is applied to study secondary structure formation. Physically, each model residue consists of a detailed three-bead backbone and a simplified single-bead side-chain. Excluded volume and hydrogen bond interactions are constructed with discontinuous (i.e., hard-sphere and square-well) potentials. Simulation results show that the backbone motion of the model is limited to realistic regions of Φ–Ψ conformational space. Model polyalanine chains undergo a locally cooperative transition to form α-helices that are stabilized by backbone hydrogen bonding, while model polyglycine chains tend to adopt nonhelical structures. When side-chain size is increased beyond a critical diameter, steric interactions prevent formation of long α-helices. These trends in helicity as a function of residue type have been well documented by experimental, theoretical, and simulation studies and demonstrate the ability of the intermediate-resolution model developed in this work to accurately mimic realistic peptide behavior. The efficient algorithm used permits observation of the complete helix–coil transition within 15 min on a single-processor workstation, suggesting that simulations of very long times are possible with this model. Proteins 2001;44:344–360.**

Key words: helix–coil transition; computer simulation; four-bead protein model

INTRODUCTION

In general, commonly used protein models for simulations of protein folding and dynamics can be categorized as one of two types: low-resolution models, or high-resolution models. The approach in developing a low-resolution model is to strip the protein down to its minimal form. These idealized models have been valuable in providing insights into the coarse-grained structure and folding mechanisms of proteins. However, simple models lack the detail necessary to study the interplay between different forces during protein folding. High-resolution models, which aim to include as much detail as is known about the physical geometry and interactions in real proteins, have been used successfully in studies of very short time dynamics.[2] Unfortunately, the detail that makes high-resolution models so realistic also makes them extremely computationally intensive. All-atom simulations are so complex that the protein folding community was amazed when Duan and Kollman[3] reported in 1998 on the simulation of the folding of a solvated small protein for a full microsecond. A single trajectory of this length required 2 months on a massively parallel supercomputer. Within the context of all-atom protein folding, their work is a remarkable accomplishment. However, simpler models are required for simulations over biologically relevant time scales and for simulations of complex multiprotein systems.

Intermediate-resolution protein models provide a balance between low-resolution models (few details, computationally fast) and high-resolution models (many details, computationally slow) for simulating proteins. Many of the existing intermediate-resolution models are "two-bead" models in which an amino acid residue is composed of a one-bead backbone and a one-bead side-chain.[4–10] Wallqvist and Ullner[11] extended the two-bead approach with the design of a three-bead model, one bead for the residue backbone and one or two beads for the side-chain, that improves the detail of the residue side-chains without resorting to an all-atom model. In efforts to improve the detail of the backbone structure, four-bead models,[12,13] in which each residue is represented by three backbone beads and one side-chain bead, and six-bead models,[14,15] in which each residue is represented by five backbone beads and one side-chain bead have been developed. The success with four- and six-bead protein models is promising and suggests that complete folding of small proteins from denatured conformations may be possible with intermediate-resolution models.

The goal of our work is to develop an off-lattice, intermediate-resolution protein model suitable for use in computer simulations of complex, multiprotein systems. We aim to create a model that contains enough genuine protein character to mimic real protein dynamics yet is simple enough to be computationally tractable for future studies of protein aggregation. In a previous article, we described a physical protein representation similar to the one used in the present study and demonstrated its ability to undergo

Grant sponsor: GAANN Biotechnology Fellowship program/U.S. Department of Education; Grant sponsor: National Science Foundation; Grant number: CTS-9704044; Grant sponsor: National Institutes of Health; Grant number: GM-56766.

*Correspondence to: Carol K. Hall, Department of Chemical Engineering, North Carolina State University, Raleigh, NC 27695-7905. E-mail: hall@turbo.che.ncsu.edu

Received 23 September 2000; Accepted 13 April 2001

multiple thermodynamic transitions that qualitatively correspond to the structural transitions a protein undergoes during folding.[16] We now describe the addition of a new hydrogen-bonding potential to this four-bead protein model and demonstrate its impact on the formation of secondary structural elements, particularly the helix–coil transition. The formation of α-helices, a common structural motif in proteins, has been studied extensively in an effort to understand the larger issue of protein folding; and a wealth of simulation,[17–30] experimental,[31–38] and theoretical[34,39–42] results has been published describing helix stability and the helix–coil transition. In this study, we focus on helix stability and formation to evaluate our model's ability to mimic realistic protein behavior.

Our model is designed for use with the discontinuous molecular dynamics (DMD) algorithm,[43–45] a fast alternative to the continuous-potential molecular dynamics algorithms commonly used for protein folding simulations. DMD simulations have been performed previously on low-resolution protein models using a biased potential that varies the strength of native (correct) contacts relative to non-native (incorrect) contacts.[46–50] To our knowledge, the results we present in this work represent the first application of DMD to an intermediate-resolution protein model and the first unbiased use of DMD for model peptides. The intermediate-resolution protein model reported in this work serves as a powerful compromise between the low-resolution protein models and the complex all-atom models currently favored by the protein folding community.

In this article, we present an off-lattice intermediate-resolution protein model with backbone hydrogen-bonding capability. In our model, amino acid residues are represented by four beads. The backbone of the model is relatively detailed, with three beads representing six atoms, allowing a realistic representation of backbone bond angles and dihedral angles. Side-chains are modeled using a more idealized, single-bead approach; however, the side-chain bead size may be varied to represent small residues, such as alanine, as well as larger, bulkier residues. Beads are subject to three types of forces: repulsion due to excluded volume effects, permanent attraction between covalently bonded beads, and attraction between pairs of backbone beads during hydrogen bond formation. Hydrophobicity, another dominant force in protein folding, is not included in this model but could easily be incorporated as an attraction between pairs of side-chain beads. The effects of hydrophobicity are explored in detail in forthcoming publications.[51,52] In this work, we perform simulations on model chains with a range of side-chain sizes. Polyalanine was chosen as the main model peptide because of its experimentally observed intrinsic tendency to form α-helices stabilized by hydrogen bonds. Unlike polyalanine peptides, polyglycine peptides do not tend toward α-helical structures; therefore, model polyglycine chains are also simulated to establish the unbiased nature of our hydrogen bonding potential. The model presented in this article is limited to steric interactions and backbone–backbone hydrogen bonding. As a result, the model is currently not suited to simulations of β-sheet and β-hairpin formation, as these structures are stabilized to a large extent by interactions between side-chains. Although β-sheet and β-hairpin structures are possible and are frequently observed in our simulations, they do not represent the lowest energy structures for the model chains simulated in this work. Each of the forces in the model is described by a discontinuous potential, e.g., a hard core or a square well, which enables the use of the DMD algorithm.[43–45] Although only a handful of protein simulations have been performed using DMD, it is a powerful computational algorithm that has been used extensively by our group[53–56] and others[57,58] to study polymer systems.

Highlights of our simulation results follow. Owing to the physical parameters chosen for the model, realistic Ramachandran plots are produced for both glycine (which, in the model, does not have a side-chain bead) and non-glycine residues. Much of the work presented focuses on polyalanine, which is known to have a strong tendency toward α-helical structures. Isolated polyalanine chains are shown to undergo cooperative transitions to form α-helices; and the efficiency of our model allows observation of the helix–coil transition within a 15-min simulation on a 500-MHz workstation. Helix nucleation occurs at any point along the chain, with multiple nucleation sites possible. The length of the polyalanine chain is shown to have a small effect on the optimal conditions for helix formation, probably because of the relative effect of end fluctuations on the overall stability of the helix. We observe that, as side-chain size is increased, helix formation becomes more difficult and, ultimately, impossible because of excluded volume constraints. Finally, contrary to our models with side-chains, our model polyglycine chains sample α-helical space but do not form stable α-helices. The protein model developed in this work successfully mimics real protein behavior in its selective formation of secondary structure elements.

The following section describes the model developed for this work, including the physical representation and the potential energy function, and the DMD simulation technique. The results obtained for (1) Φ–Ψ conformational freedom, (2) polyalanine helix formation, (3) helix formation as a function of side-chain size, and (4) polyglycine dynamics are presented. The results are then discussed, followed by a summary and a brief description of our future work using the model developed in this study.

MODELS AND METHODS

Physical Chain Representation

The protein model used in this study has a fairly realistic backbone structure and a very simplified side-chain structure. Each amino acid residue is modeled with four beads, as depicted by the four dashed circles in Figure 1. An N united atom represents the amide nitrogen and hydrogen of an amino acid, a C_α united atom represents the α-carbon and its hydrogen, and a C united atom represents the carbonyl carbon and oxygen. The fourth bead in the model, R, represents the side-chain group. This physical structure, a three-bead backbone and one-bead

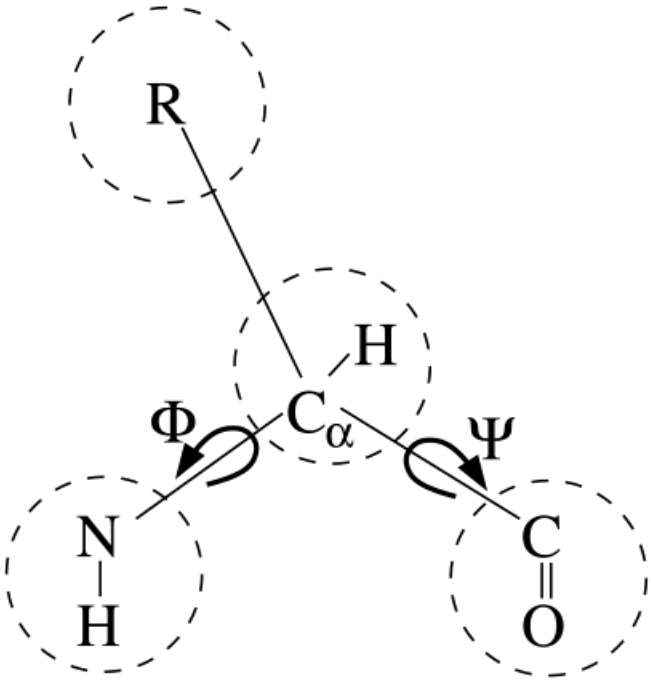

Fig. 1. Amino acid residue. The side-chain group is denoted R and represents one of 20 different chemical groups. Dashed circles depict atom groups, each of which is represented by a sphere in the model. Φ is rotation around the bond between nitrogen and α-carbon atoms. Ψ is rotation around the bond between α-carbon and carbonyl carbon atoms.

side-chain, has been used successfully elsewhere[12,13,59] in concert with different search algorithms and potential energy functions. In our model, the interpeptide bond is assumed to be in the *trans* configuration, all backbone bond lengths and bond angles are fixed at their ideal values, and the distance between consecutive C_α united atoms is fixed in accordance with empirical observations. The side-chains in the model vary in size and distance from C_α, depending on the particular amino acid residues being modeled, and are held in positions relative to the backbone so that all residues are L-isomers. The values of the bond lengths and angles and the method used to maintain these values and chirality are given below. Solvent molecules are not explicitly included in the model. The effect of solvent is factored into the energy function as a potential of mean force.

Forces and Interactions

Beads in the protein model are subject to three different types of forces: repulsion due to excluded volume effects, attraction between bonded beads and pseudobonded beads (as defined below), and attraction between pairs of backbone beads during hydrogen bond formation. Each of these forces is represented by a discontinuous potential force, either a hard-sphere potential

$$u_{ij}(r) = \begin{cases} \infty, & r \le \sigma \\ 0, & r > \sigma \end{cases} \qquad (1)$$

where r is the distance between beads i and j, and σ is the bead diameter, or a square-well potential

TABLE I. Simulation Parameters

Bead diameters, σ	(Å)
N	3.300
C_α	3.700
C	4.000
R_{alanine}	4.408
Well diameters, λσ	(Å)
N	4.200
C	4.200
Bond lengths, l	(Å)
$N_i - C_{\alpha,i}$	1.460
$C_{\alpha,i} - C_i$	1.510
$C_i - N_{i+1}$	1.330
$C_{\alpha,i} - R_{\text{alanine},i}$	1.531
Pseudobond lengths, l	(Å)
$N_i - C_i$	2.45
$C_{\alpha,i} - N_{i+1}$	2.41
$C_i - C_{\alpha,i+1}$	2.45
$C_{\alpha,i} - C_{\alpha,i+1}$	3.80
$N_i - R_{\text{alanine},i}$	2.44
$C_i - R_{\text{alanine},i}$	2.49
Bond angles	(°)
$\angle N_i - C_{\alpha,i} - C_i$	111.0
$\angle C_{\alpha,i} - C_i - N_{i+1}$	116.0
$\angle C_i - N_{i+1} - C_{\alpha,i+1}$	122.0
$\angle R_{\text{alanine},i} - C_{\alpha,i} - N_i$	109.6
$\angle R_{\text{alanine},i} - C_{\alpha,i} - C_i$	110.1

$$u_{ij}(r) = \begin{cases} \infty, & r \le \sigma \\ -\varepsilon, & \sigma < r \le \lambda\sigma \\ 0, & r > \lambda\sigma \end{cases} \qquad (2)$$

where λσ is the well diameter and ε is the well depth. Conceptually, a "hard sphere" refers to an impenetrable, solid sphere, and a "square well" refers to an attractive region of thickness λ that envelops that sphere. Deeper wells correspond to stronger attractive interactions between square-well beads, and shallower wells correspond to weaker attractive interactions. The well depth parameter is coupled to the temperature, so that a single-parameter reduced well depth (ϵ^*) equal to $\epsilon/\kappa_B T$ characterizes the protein environment. For example, strong attractive interactions (deep wells) can be considered to characterize a low-temperature or poor solvent environment.

Excluded volume

Pairs of beads collide and repel when the distance between them becomes so small that their surfaces touch (when $r_{ij} = \sigma$). Diameters for each of the three types of backbone beads are chosen to be reasonable estimates for the sizes of the atoms they represent; they are given in Table I and are the same as those used by Takada et al.[13] However, as explained in the work by Takada et al.[13] interactions involving a pair of beads that are separated along the chain by three or fewer bonds are more faithfully represented by an interaction between the atoms them-

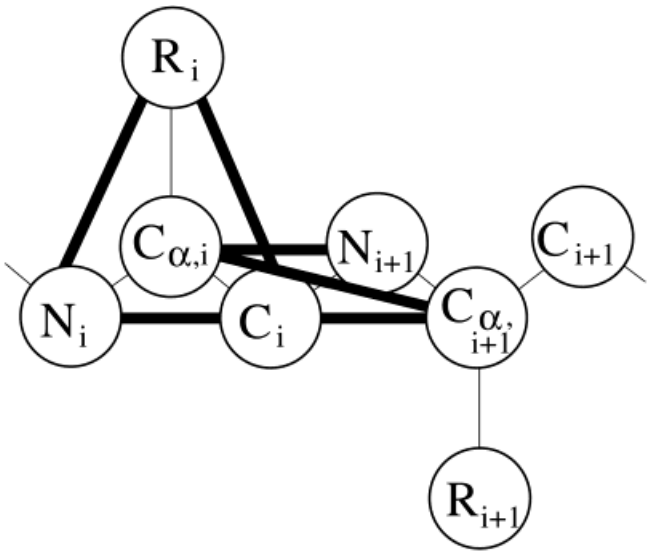

Fig. 2. Covalent bonds are shown with narrow lines connecting beads. At least one of each type of pseudobond is shown with a bold line. Pseudobonds are used to maintain backbone bond angles ($N_iC_{\alpha,i}C_i$, $C_{\alpha,i}C_iN_{i+1}$, and $C_iN_{i+1}C_{\alpha,i+1}$), consecutive C_α distances ($C_{\alpha,i}$ to $C_{\alpha,i+1}$), and residue L-isomerization (through a concerted effort of pseudobonds N_i—R_i, C_i—R_i, and N_i—C_i and covalent bonds N_i—$C_{\alpha,i}$ and $C_{\alpha,i}$—C_i).

selves, not the united atoms. In other words, neighboring carbonyl carbon atoms experience interactions based on the diameter of a carbon atom, not on the diameter of a united carbon–oxygen bead. Consequently, for interactions between pairs of beads separated by three or fewer bonds, we allow the beads to overlap by up to 25% of their bead diameters. Our amount of overlap is slightly different from that used by Takada and colleagues but is chosen with the same goal in mind: to limit motion around the N—C_α and C_α—C bonds in the protein models to a physically reasonable extent. The limited motion in our model is demonstrated by the Φ–Ψ conformational freedom results in a later section.

Bonds

Covalent bonds are maintained between neighboring beads along the chain backbone and between the C_α and R united atoms. Bonded beads move freely over a small range between $(1 - \delta)l$ and $(1 + \delta)l$, where δ is the bond tolerance and l is the ideal bond length. The choice of δ defines the acceptable range of fluctuation in the bond length. In this work, δ is chosen to be 0.02. In effect, bonds in the simulation fluctuate within 2% of their assigned lengths by experiencing a hard-sphere repulsion at $(1 - \delta)l$ and an infinitely strong square-well attraction at $(1 + \delta)l$. All covalent bond lengths in the model are assigned ideal values[60] and are given in Table I. Ideal backbone bond angles, C_α—C_α distances, and residue L-isomerization are achieved by using the set of pseudobonds shown in Figure 2 by the wide bold lines. Like the covalent bonds described above, the pseudobonds fluctuate within 2% of their assigned lengths by experiencing a hard-sphere repulsion at $(1 - \delta)l$ and an infinitely strong square-well attraction at $(1 + \delta)l$. Pseudobonds are used between next-neighbor beads along the backbone of the chain to hold backbone angles fixed. For example, a pseudobond is placed between N_i and C_i in Figure 2 to force the N_i—$C_{\alpha,i}$—C_i angle to be near its ideal value. Pseudobonds are also included between neighboring pairs of C_α beads, such as shown in Figure 2 between $C_{\alpha,i}$ and $C_{\alpha,i+1}$, to maintain their distances near the experimentally determined constant value. The combination of the $C_{\alpha,i}$—N_{i+1}, C_i—$C_{\alpha,i+1}$, and $C_{\alpha,i}$—$C_{\alpha,i+1}$ pseudobonds and the $C_{\alpha,i}$—C_i, C_i—N_{i+1}, and N_{i+1}—$C_{\alpha,i+1}$ covalent bonds holds the interpeptide group ($C_{\alpha,i}$—C_i—N_{i+1}—$C_{\alpha,i+1}$) in the *trans* configuration. Finally, pseudobonds are placed between side-chains and backbone N and C united atoms, such as shown in Figure 2 between N_i and R_i and between R_i and C_i, to hold the side-chain beads fixed relative to the backbone so that all model residues are L-isomers. The assigned bond angles and corresponding pseudobond lengths are shown in Table I.

Hydrogen bonds

We have developed a hydrogen-bonding method for use with the physical protein model described above. Our method treats all hydrogen bonds equally. Although this article focuses on α-helical structures, no bias is given toward α-helical hydrogen bonds over all other hydrogen bonds. Backbone hydrogen bonds form in real proteins between amide hydrogen atoms and carbonyl oxygen atoms. Because hydrogen and oxygen atoms are not explicitly represented in our model, hydrogen bonds must occur as a square-well attraction between the N united atom and the C united atom. However, hydrogen bonds in proteins are strongly angle-dependent, with the strongest association resulting from a coaxial geometry (N—H· · ·O—C, where the nitrogen–hydrogen–oxygen and hydrogen–oxygen–carbon angles are each 180°) with a hydrogen-to-oxygen distance of ~2 Å[60] Although a coaxial geometry is ideal, deviations are common, and nitrogen–hydrogen–oxygen and hydrogen–oxygen–carbon angles as small as 110° are observed in proteins.[61]

Backbone hydrogen bonding is permitted in our model as shown in Figure 3, which illustrates a hydrogen bond between N_i and C_j. Since the distance between consecutive C_α beads is fixed with a pseudobond and the peptide bond is held in the *trans* configuration, the positions of C_{i-1}, N_i, and $C_{\alpha,i}$ are sufficient for determining precisely where the ith amide hydrogen should be.

The projection that determines the location of the ith hydrogen atom is depicted in Figure 3 as an arrow extending through N_i. Similarly, the positions of $C_{\alpha,j}$, C_j, and N_{j+1} can be used to locate the jth carbonyl oxygen atom. The use of these projections during the simulation makes it possible to represent a hydrogen bond as a square-well attraction between N_i and C_j subject to reasonable angles between projections. In effect, the hydrogen bond occurs between virtual H_i and O_j atoms, atoms that are not explicitly in the simulation but instead lie within the N_i and C_j united atom spheres.

In our simulations, a hydrogen bond forms between N and C when four criteria are met: N and C are separated by

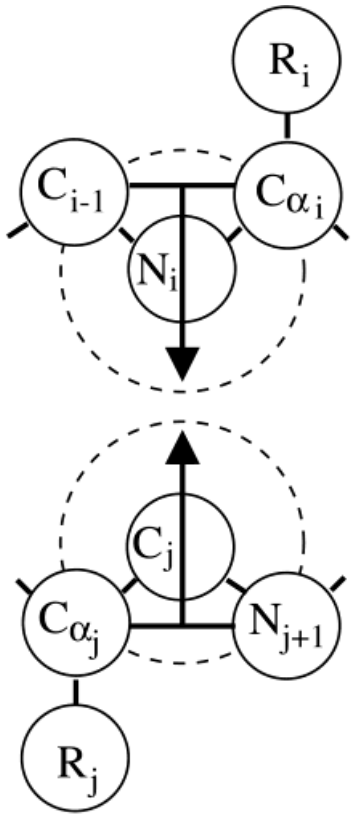

Fig. 3. Backbone hydrogen bonding is achieved through a square-well interaction between N and C. The dashed circles symbolize the square-well attractive shells around N_i and C_j. The location of H_i can be determined by examining the geometry of C_{i-1}, N_i, and $C_{\alpha,i}$ and then projecting down through N_i as shown by the arrow through N_i. The location of O_j can be determined by examining the geometry of $C_{\alpha,j}$, C_j, and N_{j+1} and then projecting up through C_j as shown by the arrow through C_j.

an appropriate distance, the nitrogen-to-hydrogen and carbon-to-oxygen vectors are reasonably close to coaxial, neither N nor C is already involved in a hydrogen bond with a different partner, and N and C are separated by at least three intervening residues along the protein chain. The first requirement, that N and C are separated by an appropriate distance, is met by assigning suitable square-well diameters for N and C. The well diameters for N and C are chosen to be 4.2 Å for this study, as shown in Table I. This value is based on the nitrogen-to-carbon distance in a coaxial hydrogen bond (N—H· · ·O—C) where the nitrogen-to-hydrogen bond is 1 Å,[60] the hydrogen-to-oxygen distance is 2 Å,[61] and the oxygen-to-carbon bond is 1.2 Å.[60] During the simulation, if the attractive wells around N_i and C_j touch (if the N to C distance is 4.2 Å), an evaluation is made to determine whether the other three hydrogen-bond criteria are met. The second requirement, that the nitrogen-to-hydrogen and carbon-to-oxygen vectors are aligned so as to be characteristic of a hydrogen bond, is evaluated by determining the locations of the oxygen and hydrogen atoms with the vector projections as described above. To form a hydrogen bond in our model, we require that both the NHO and the HOC angles be between 120° and 180°. The third and fourth requirements, that neither N nor C is already involved in a hydrogen bond and that N and C are separated by at least three intervening residues, are implemented to limit hydrogen bonding to physically realistic situations. If all four criteria are met, a hydrogen bond forms between N_i and C_j via an attractive event between the N_i and C_j square-well beads.

In a method similar to that used by Takada et al.,[13] when a hydrogen bond occurs between N and C, we add stability to the hydrogen bond by increasing the range of the repulsive forces between the beads that neighbor N and C. Referring again to Figure 3, upon the formation of a bond between N_i and C_j, the diameters of the C_{i-1}, $C_{\alpha,i}$, $C_{\alpha,j}$, and N_{j+1} beads temporarily increase. The increase is felt only when N_i interacts with $C_{\alpha,i}$ or N_{j+1} and when C_j interacts with C_{i-1} or $C_{\alpha,i}$; it does not affect interactions with the other beads on the peptide. Increasing the range of the repulsive forces between the beads surrounding the hydrogen bonded pair helps the nitrogen-to-hydrogen and carbon-to-oxygen vectors stay aligned in a stable hydrogen bond. The diameters of C_{i-1}, $C_{\alpha,i}$, $C_{\alpha,j}$, and N_{j+1} return to their original sizes upon breaking of the hydrogen bond.

The criteria above lead to hydrogen bonds with realistic OH distances and NHO angles in our simulations, as illustrated in Figure 4. Figure 4(a) shows the fluctuation in the NHO angle over the course of a simulation. Time, t^*, is in reduced units.[62] The NHO angle varies between 115 and 180 degrees, with an average of 150 degrees. The HOC angle also varies between 115 and 180 degrees, with an average of 150 degrees (data not shown). These results are similar to the hydrogen bond properties reported by Baker and Hubbard based on an extensive analysis of α-helical and β-sheet structures in a high-resolution protein database.[61] For NHO, they report a range of 110–180 degrees, with a mean of approximately 159 degrees; for HOC, they report a range of 110–180 degrees, with a mean of approximately 150 degrees. Fluctuations in our hydrogen bond length (hydrogen-to-oxygen distance) over the course of a simulation are shown in Figure 4(b) to vary between 1.8 and 3.1 Å, with an average of 2.4 Å. This distribution is somewhat different from the range (1.6–2.5 Å) and mean (2 Å) reported by Baker and Hubbard. The hydrogen-to-oxygen distances in our simulations are a direct result of our choice of N and C well diameter (4.2 Å), and smaller well diameters would lower the average hydrogen-to-oxygen distance. We do not arbitrarily reduce the N and C well diameters because, as mentioned above, the nitrogen-to-carbon distance in a coaxial (ideal) hydrogen bond is 4.2 Å. Therefore, 4.2 Å is the minimum N and C well diameter that, when the hydrogen-to-oxygen distance is 2 Å, will recognize an alignment that corresponds to a coaxial hydrogen bond and will appropriately permit formation of the hydrogen bond.

Discontinuous Molecular Dynamics

Simulations are performed using a discontinuous molecular dynamics (DMD) simulation algorithm.[43–45] Molecular dynamics is a popular technique for simulating complex systems. By monitoring the progression of the system over

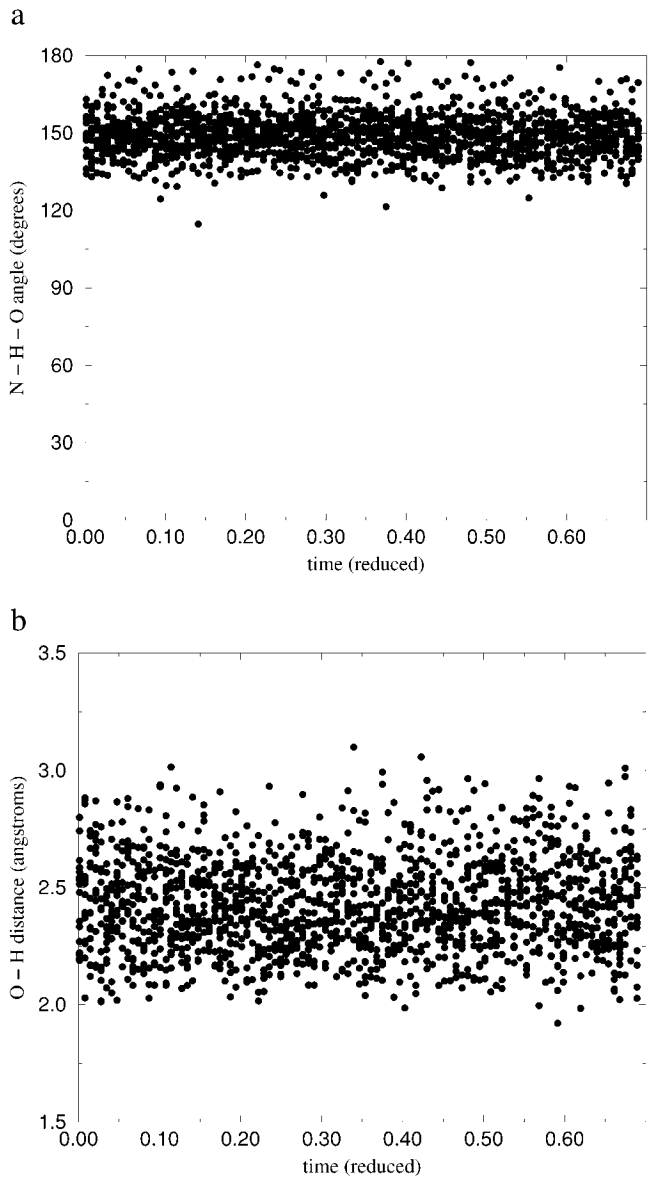

Fig. 4. Fluctuations in hydrogen bond (**a**) NHO angle, and (**b**) O—H distance over time.

time, macroscopic properties of the system may be computed and the evolution of important conformations may be observed. DMD simulations are conducted as follows. Each bead of the model protein chain is assigned an initial position and an initial velocity. The initial positions are random, although they may not violate any of the assigned bond lengths and angles listed in Table I. The initial velocities are chosen at random from a Maxwell–Boltzmann distribution at the simulation temperature. When the simulation begins, each bead moves according to its individual velocity. The simulation proceeds according to the following schedule: identify the first event, move forward in time until that event occurs, calculate new velocities for the pair of beads involved in the event, and calculate any changes in system energy or hydrogen bonding status, find the second event, and so on. The event-to-event nature of DMD offers significant computational advantages over standard, continuous-potential molecular dynamics techniques that must proceed through time by taking very small steps.[54]

Types of events include excluded volume events, bond events, and square-well events. An excluded volume event occurs when the surfaces of two hard-sphere beads collide and repel each other. Bond (or pseudobond) events include a hard-sphere repulsion event which occurs when the bond length is $(1 - \delta)l$ and an infinite square-well attraction event that occurs with the bond length is $(1 + \delta)l$. Square-well events include capture, bounce, and dissociation events, which occur when the square wells of N and C beads touch. An attractive square-well event, such as is experienced between an N bead and a C bead when separated by the average of their well widths, is often referred to as a square-well capture event. In the simulation, the attraction results in an increase in kinetic energy (beads N and C move faster toward each other) and a decrease in potential energy (in accordance with the depth of the N and C square wells). In essence, the capture event causes the beads to become partners. After a partnership has formed, when the simulation reaches a time when the distance between the beads is again equal to the average of their well diameters, an evaluation must be made to determine whether the pair has enough kinetic energy to dissolve the partnership (through a square-well dissociation event) or will remain partnered (through a square-well bounce event). A dissociation event is the reverse of a capture event; N and C move away from each other and lose velocity (lowering the kinetic energy of the system), while the system gains potential energy. If the beads experience a bounce event instead of a dissociation event, they move toward each other, maintaining their partnership, and the system's potential energy remains constant. Both energy and momentum are conserved during all types of events. For details on DMD simulations with square-well potentials, see the articles by Alder and Wainwright[43] and Smith et al.[54]

Square-well events are used in our simulations to mimic hydrogen bonding interactions. However, hydrogen bond partnerships are slightly more complicated than the standard square-well partnership described above. Consider, for example, a situation where the N—C pair does not have enough kinetic energy to break the partnership but the hydrogen bond properties (except the O to H distance) have become so distorted that the bond should no longer exist. In these cases, we force the pair to dissolve the partnership by temporarily making the square wells just shallow enough to allow a dissociation event. System velocities are then rescaled to conserve energy. Forced dissociation events are rare, accounting for less than one hundredth of 1% of all hydrogen-bond square-well events. As in the events described above, both energy and momentum of the system are conserved during forced dissociation events.

Simulations are performed in the canonical ensemble, which means that the number of particles, volume, and temperature are held constant. The number of particles

and volume are automatically fixed, since simulations are performed on a single chain of fixed length. There are a number of ways to preserve temperature during a molecular dynamics simulation, including periodic scaling techniques,[63] extended system methods,[64] and hybrid Monte Carlo methods.[65–67] We chose to implement the Andersen thermostat method,[68] as has been used in DMD simulations of homopolymers.[69] With this procedure, all beads in the simulation are subject to random, infrequent collisions with ghost particles. The post-event velocity of a bead colliding with a ghost particle is chosen randomly from a Maxwell–Boltzmann distribution at the simulation temperature. We have implemented several optimization techniques in this work, including neighbor lists and false positioning, which have been described elsewhere.[54]

Model Peptides

We perform DMD simulations on model peptides of different lengths. Simulations are performed on 10-, 20-, and 30-mer peptides with small side-chains chosen to represent the methyl group of alanine. We also perform simulations on peptides with bigger side-chains and on a peptide with no side-chains, representing polyglycine. In each case, the chain is homogeneous; that is, each peptide is composed of residues with identical side-chains.

Simulations are performed on α workstations and range in length from 250 million to one billion events, the length chosen in each case based on the progress of the system. For example, we often observed relatively fast folding wherein the chain adopts a stable, low-energy structure without significant barriers and remains within small fluctuations of that state. However, we occasionally observed instances of slower folding, with the peptide exploring many moderately low energy conformations or stalling in a non-native conformation before settling into the lowest energy state. Data are averaged over the final 50% of each run. On average, each simulation consists of approximately 7% hard-sphere repulsion events, 89% bond and pseudobond events, 3% hydrogen bonding events, and 1% ghost collision events.

RESULTS

In this section, we present our results for simulations of isolated, model protein chains. As the major components of the model are steric repulsion and backbone hydrogen bonding, we evaluate the realism of the model by studying the Φ–Ψ freedom exhibited by the chain and the amount of hydrogen-bond stabilized secondary structure produced for different types of model peptides.

Φ–Ψ Conformational Freedom

As a test of our values for bead diameters and for the degree of overlap permitted between pairs of beads separated by three or fewer bonds, we generate Ramachandran plots of Φ versus Ψ, where Φ and Ψ refer to the rotation around the backbone bonds as shown in Figure 1. Figure 5 plots Φ versus Ψ for residues (a) with side-chains (model non-glycine residues) and (b) without side-chains (model glycine residues) which show that our model chain backbones are confined to particular regions of Φ–Ψ space. The regions in our Ramachandran plots agree well with standard, published data that describe the regions of sterically allowed Φ–Ψ space in real proteins.[60] This agreement indicates that the bead diameters and the degree of overlap between near-neighbors used in our model are reasonable values. The data in Figure 5(a) were obtained from a simulation on a peptide with small, alanine-sized side-chains. The patterns within Figure 5(a) are a direct result of the intrinsic α-helical propensity of alanine and are germane to a discussion of hydrogen bonding results which is presented in the next section.

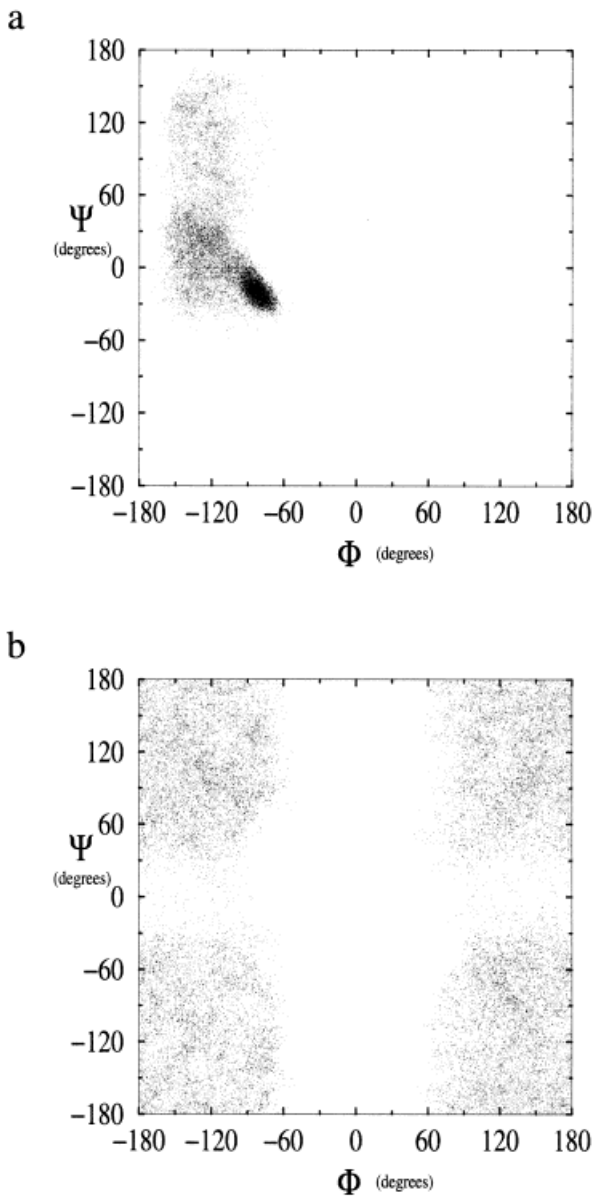

Fig. 5. Ramachandran plots for simulations of (**a**) non-glycine residues, and (**b**) glycine residues.

Polyalanine Helix Formation

The formation of α-helices is studied over the course of our simulations. Initial simulations are performed with a

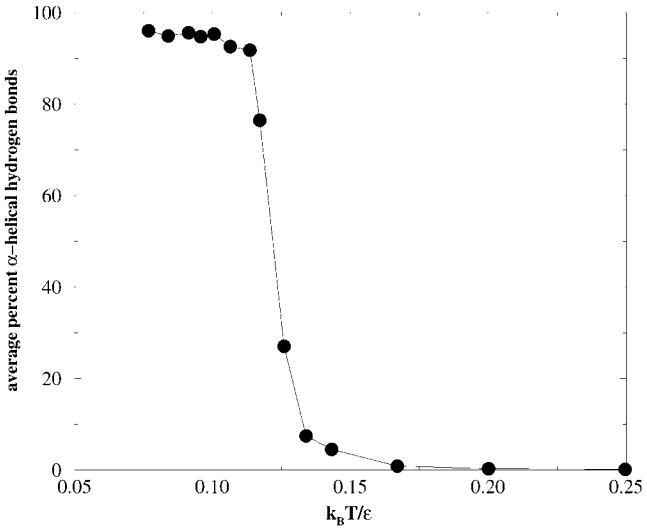

Fig. 6. Percentage of α-helical hydrogen bonds formed for the 20-residue polyalanine chain as a function of reduced temperature.

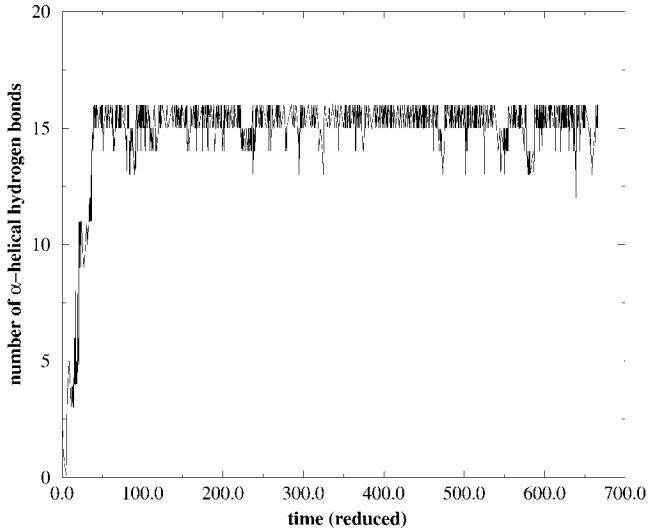

Fig. 7. Number of α-helical hydrogen bonds formed versus reduced time for a 20-residue polyalanine chain at a relatively low temperature ($T^* = 0.10$). The chain folds into a stable, full-length helix.

20-residue chain with side-chain diameters of 4.408 Å, a value representative of the methyl group of alanine. Polyalanine is known to have a strong tendency toward α-helical structures; with our model, the lowest energy structure for a model polyalanine chain is an α-helix. Although alanine is hydrophobic, we do not include side-chain–side-chain interactions in this article. Our main concern is to develop a model that forms hydrogen bonds and generates stable α-helical secondary structure.

Polyalanine helix formation as a function of temperature

Figure 6 shows the observed percentage of α-helical hydrogen bonds formed for a 20-residue polyalanine (polyA) chain as a function of reduced temperature. Hydrogen bonds are defined as α-helical if N_{i+4} is bound to C_i; with this definition, the maximum number of α-helical bonds is the chain length minus four, or 16, and the percentage plotted is relative to this maximum. Reduced temperature, T^*, is defined as $k_B T/\epsilon$, where k_B is Boltzmann's constant and T is temperature. Low temperatures are shown in Figure 6 to allow formation of stable α-helices. Native α-helix formation was often achieved within approximately 15 min on a single-processor workstation.

Figure 7 presents the number of α-helical hydrogen bonds formed over the course of a single simulation starting from a random initial conformation at a relatively low reduced temperature, T^*, of 0.1. The type of folding curve shown in Figure 7 is extremely reproducible and is independent of the initial conformation. In Figure 8, we show snapshots of the polyA chain as it evolves toward the native state, i.e., at times corresponding to the early stages of the simulation shown in Figure 7. The backbone beads are light gray, the side-chains are dark gray, and the N-terminal N beads are black and shown at the top of each structure. The structure at a reduced time of 8.02 has three α-helical hydrogen bonds in the middle of the chain. By $t^* = 21.4$, a second helical segment at the C-terminus of the chain has formed; at $t^* = 26.2$, the chain has two distinct separate helical segments, a 7 bond α-helix involving residues 2–12 and a 4-bond α-helix at the C-terminus. At $t^* = 41.5$, the chain has adopted a full-length, 16-bond α-helix. The α-helix persists through the remaining 630 time units. In this low-temperature environment, the formation of stable α-helices appears effortless.

Figure 9 plots the number of α-helical hydrogen bonds formed versus reduced time for polyA at a higher temperature ($T^* = 0.125$). Under these conditions, there is broad sampling of conformational space. The peptide chain repeatedly forms and dissolves long α-helices. Figure 10 highlights the range of motion experienced during the high temperature run shown in Figure 9. At a reduced time of 61.3, the chain is nearly entirely α-helical with just the C-terminal end not formed. At $t^* = 146$, the chain has lost all its α-helical hydrogen bonds and is extended; at $t^* = 362$, it assumes a full-length α-helix; at $t^* = 607$, the chain adopts a β-turn stabilized by five hydrogen bonds. Although the chain explores β-sheet regions, a full-length α-helix has more hydrogen bonds, and therefore a lower energy in our model, than any type of β-sheet. Consequently, with a small, alanine-sized side-chain, the peptide is consistently able to form an α-helix. However, the high-temperature environment prevents stable, long-lasting α-helices.

Fast and slow folding of polyalanine peptides

Many of the simulations at low temperatures exhibit a very rapid transition from the random coil to the full-length α-helix, with misfolded structures or intermediates absent or fleeting; however, less productive folding processes are also observed. Figure 11 shows the average percentage of α-helical hydrogen bonds formed during two simulations of 20-residue polyA at $T^* = 0.077$, one of a "fast folder" and one of a "slow folder." In the fast-folding

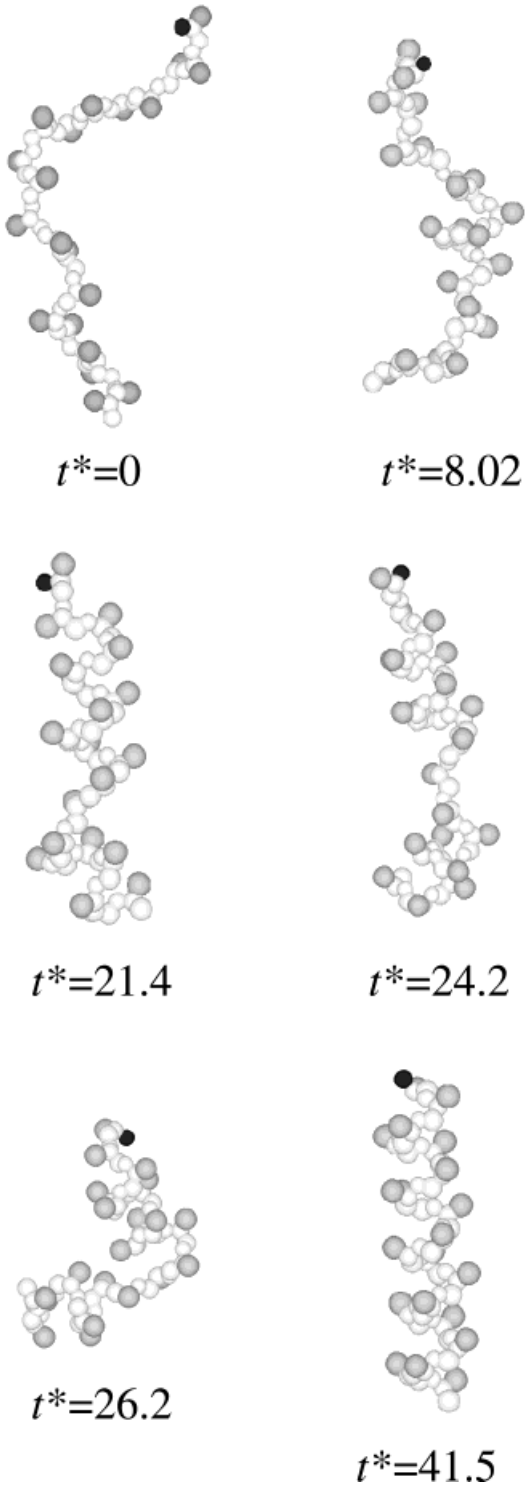

Fig. 8. Snapshots of the polyalanine chain during the low-temperature run shown in Fig. 7. Beads are reduced by half for ease of viewing. The backbone is light gray, the side-chains are dark gray, and the N-terminal end bead is black and shown at the top of each structure.

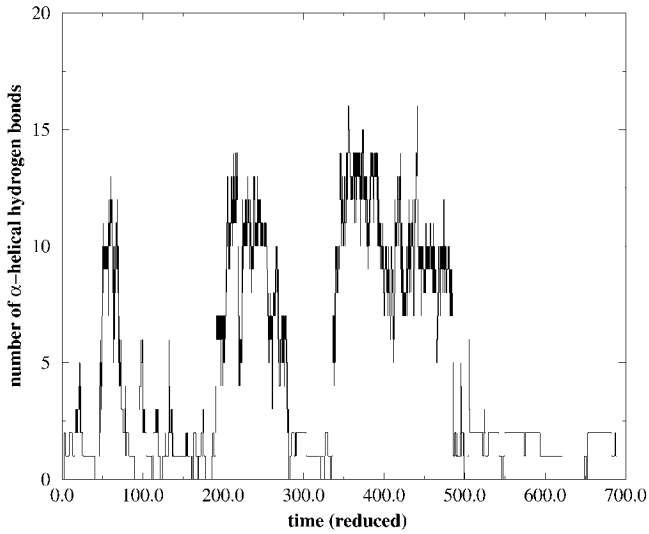

Fig. 9. Number of α-helical hydrogen bonds formed versus reduced time for a 20-residue polyalanine chain at a relatively high temperature ($T^* = 0.125$). The helix is repeatedly formed and dissolved.

run, shown by open circles, the chain quickly adopts a long, stable α-helix. In Figure 12, the time course of this simulation is divided into blocks, and the time each potential α-helical bond spends as an intact hydrogen bond during each block is shown in the separate panels of the figure. Bond 1 refers to the N-terminal hydrogen bond, N_5—C_1. The top five panels in Figure 12 span the simulation from $t^* = 0$ to $t^* = 38.1$. The first full-length α-helix is observed at $t^* = 38.1$. The bottom panel in Figure 12 shows the time immediately following the first observation of a full-length α-helix. Figure 12 shows that, for this particular simulation, the α-helix initiates at the center of the chain (bonds 7 and 8, which correspond to N_{11}—C_7 and N_{12}—C_8, respectively) and progresses faster toward the C-terminus than toward the N-terminus. After the helix forms, there are fluctuations in the bonds, particularly at the N-terminus. These fluctuations can be seen in the bottom of Figure 12, where bonds 1 and 2 spend 20% and 70% of their time in the bonded state, respectively.

We also observe slow-folding trajectories, very different from the one shown above, such as the one shown by filled squares in Figure 11 and in greater detail in Figure 13. In Figure 13, the α-helix is shown to initiate in two places, one near each end. By $t^* = 79.9$, already twice the length of the "fast folding" simulation, each end has formed a very stable α-helical segment. However, for the next 320 time units, the intermittent presence of an i to $i + 5$ hydrogen bond between N_{12} to C_7 prevents propagation of the α-helices to the center of the chain. At $t^* = 380$, N_{12}—C_7 the bond breaks for the last time; and within the next 20 time units, the final four α-helical hydrogen bonds form, connecting the α-helical end segments.

Although all the polyalanine chains studied in this report eventually produce full-length α-helices, we observe very different folding trajectories to the same final structure. In many cases, multiple helix nucleation events along the chain are followed by fusion of the independent helical segments; in other cases, the helix results from a single nucleation event that propagates throughout the chain. In fact, we observe helix nucleation events at each of

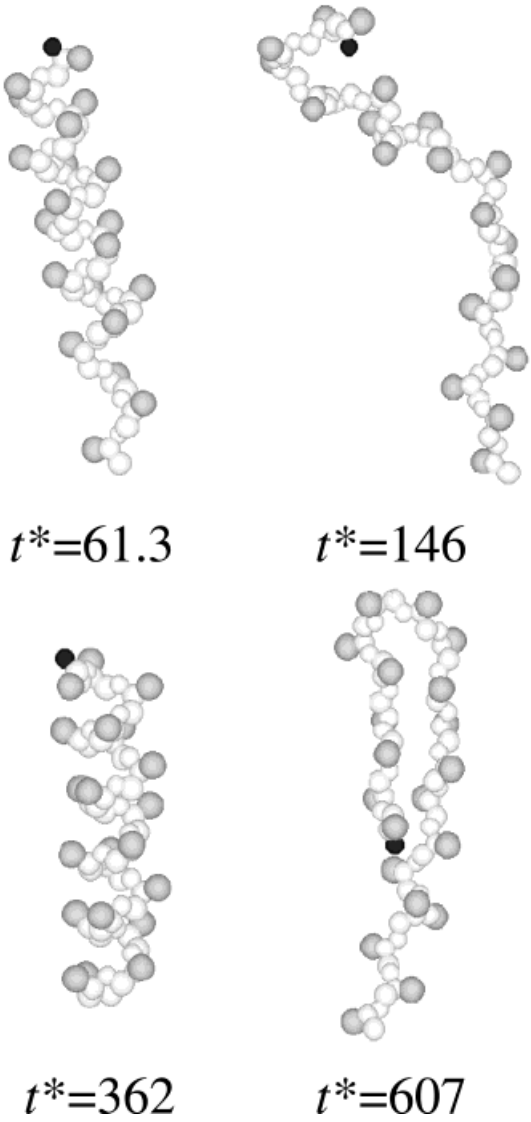

Fig. 10. Snapshots of the polyalanine chain during the high-temperature run shown in Fig. 9. Bead size and color as in Fig. 8.

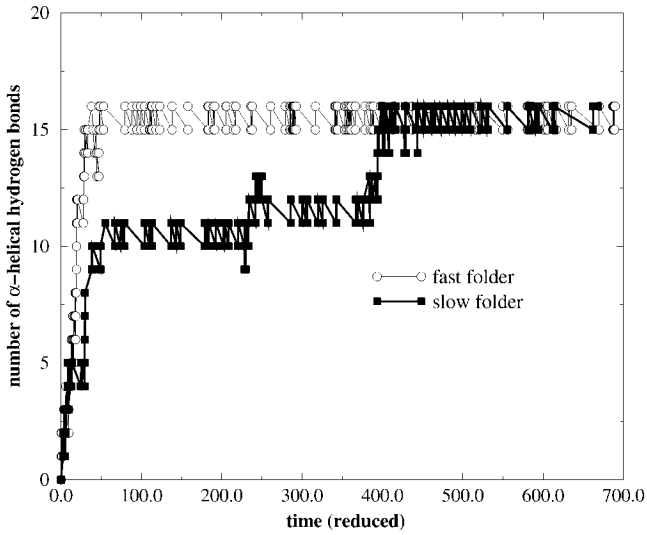

Fig. 11. Average number of α-helical hydrogen bonds formed as a function of reduced time for a 20-bead polyalanine chain at $T^* = 0.077$. Two separate simulations are shown, one an example of fast-folding to the native state and the other an example of slow-folding.

the 16 α-helical hydrogen bonds. There is no obvious trend toward preferred nucleation regions.

Conformations with β-turns are occasionally observed and tend to delay α-helix formation. An abundance of different folding pathways has been seen in simulations elsewhere[17,23,30,70] and is consistent with recent experimental observations[71,72] and with the landscape theories of protein folding in which a protein chain experiences a folding landscape that is dotted with hills and bumps.[73–76] To fold successfully, the chain must maneuver around or over the hills during its descent to its native state. The landscapes of the simple, homogeneous peptides studied in this report are much less rugged than for real proteins. Consequently, we observe far fewer non-native traps and long-lived intermediate structures than would be present for longer, heterogeneous proteins.

Local cooperativity in polyalanine helix formation

Local cooperativity along the peptide is believed to facilitate the helix–coil transition and has been observed previously in simulations of polyalanine.[17,24] In our simulations, the presence of an α-helical hydrogen bond significantly influences the formation of neighboring α-helical hydrogen bonds. Figure 14 plots the time that elapses between the formation of an α-helical bond at position zero on the x-axis and the formation of neighboring α-helical bonds. Negative 1 on the x-axis refers to the nearest potential α-helical hydrogen bond in the N-terminal direction, and positive 1 on the x-axis refers to the nearest potential α-helical hydrogen bond in the C-terminal direction. (If the bond at $x = 0$ is denoted N_{i+4}—C_i, a value of negative 1 on the x-axis refers to bond N_{i+3}—C_{i-1} and a value of positive 1 on the x-axis refers to bond N_{i+5}—C_{i+1}.) Data are averaged over the 5,083 α-helical hydrogen bonds formed (via capture events) in 30 independent polyA simulations at $T^* = 0.1$, each of which involves a fast folding trajectory to a stable α-helix. The U shape of the data in Figure 14 indicates that the presence of an α-helical hydrogen bond at position zero on the x-axis causes nearest neighbor bonds to form sooner than distant bond formation, corresponding to a local cooperative effect in the transition to an α-helix. Given that multiple nucleation sites along the chain are possible, the times shown for distant bonds are a combined result from both the influence of an α-helical hydrogen bond at position zero on the x-axis and the possible influence of separate α-helical nucleation sites. However, the effect of multiple nucleation sites would tend to underestimate the times reported for distant bonds and, therefore, does not affect the general observation of a U-shaped curve and the conclusion of local cooperative behavior.

Figure 14 also suggests that propagation is faster toward the C-terminus than toward the N-terminus. This phenomenon has been observed elsewhere and has been attributed to the asymmetrical physical structure of chiral

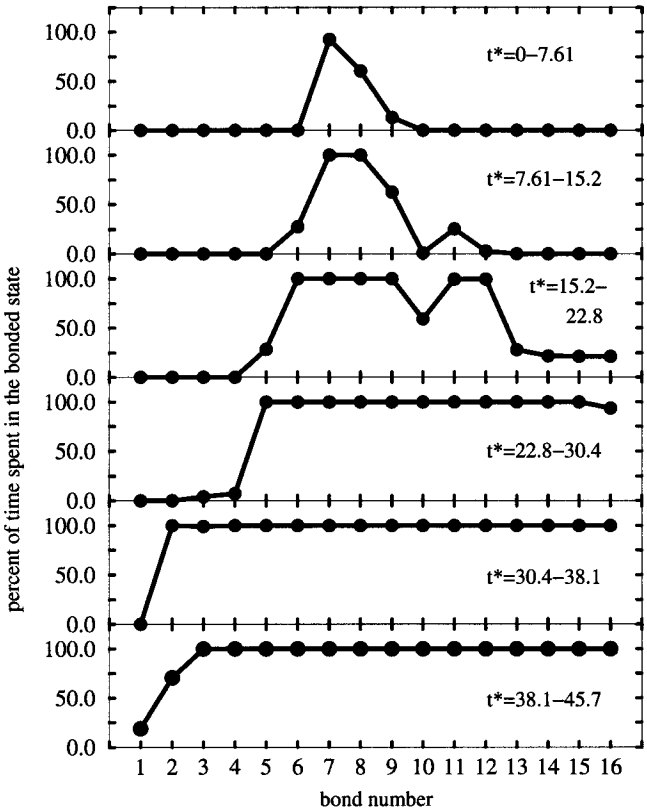

Fig. 12. The "fast-folding" simulation of Fig. 11 divided into 6 time blocks, 5 leading up to the time at which the first full-length α-helix is observed (t^* = 38.1) and 1 time block afterward. Each time block is shown in a panel in this figure. The percentage of time each potential α-helical hydrogen bond spends as an α-helical hydrogen bond is plotted. Bond 1 on the *x*-axis refers to the N-terminal α-helical hydrogen bond, N_5—C_1.

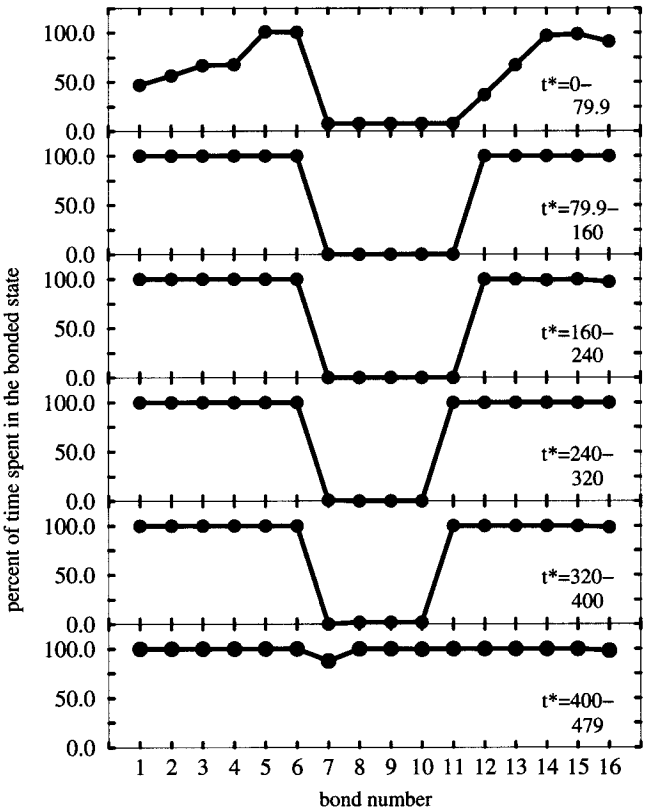

Fig. 13. As in Fig. 12, but for the "slow-folding" simulation of Fig. 11 in which the first full-length α-helix is observed at t^* = 400.

amino acid residues and to the induced dipole of a helical structure.[25,27] As we have not included charges in our model, the results presented here confirm that the asymmetry of residues and the local backbone orientation imposed by α-helical hydrogen bonds preferentially facilitate helix propagation toward the C-terminus.

Intrinsic helical propensity of alanine

Model alanine residues have a strong tendency to form α-helical (Φ,Ψ) pairs as shown by the density of data at approximately (−75°,−37°) in Figure 5(a). These data were generated during a simulation on a polyA chain at a temperature favoring α-helix formation. Therefore, Figure 5(a) shows that although our model alanine residues are able to sample the entire kidney-shaped, sterically allowed region, as was discussed earlier, they have a specific preference for the α-helical region of Φ–Ψ space.

The relative ability of individual amino acids to stabilize an α-helix has been studied using a variety of theoretical,[77,78] simulation,[23,26,29,79–84] and experimental[85–93] approaches. The parameters in the Zimm–Bragg model[40] provide a convenient means of comparing α-helical propensities as measured by different methods. According to this model, the average number of helical residues ($\langle n \rangle$) and the average length of a helical segment ($\langle l_{ZB} \rangle$) for a chain of length N are related to the nucleation parameter (σ_{ZB}), a measure of the propensity for nucleating a helical segment, and the propagation parameter (s), a measure of helix propensity, according to equations (3) and (4)[26,40,97]:

$$\frac{\langle n \rangle}{N} = \frac{1}{2} - \frac{1 - s}{2\sqrt{(1 - s)^2 + 4s\sigma_{ZB}}} \tag{3}$$

$$\langle l_{ZB} \rangle = 1 + \frac{2s}{1 - s + \sqrt{(1 - s)^2 + 4s\sigma_{ZB}}} \tag{4}$$

We calculate the Zimm–Bragg parameters, s and σ_{ZB}, from simulations on our model polyalanine chain. In accord with the work of others,[21,27,30] we consider a residue to be helical if it is part of a segment of at least three residues whose Φ and Ψ angles lie within 30° of (−75°,−37°), the average (Φ, Ψ) angles observed in our simulations for stable α-helices. The N- and C-terminal residues are neglected, since Φ and Ψ angles cannot both be defined for each. Table II shows the average number of helical residues, $\langle n \rangle$, average length of a helical segment, $\langle l_{ZB} \rangle$, and average number of helical segments, $\langle n \rangle / \langle l_{ZB} \rangle$, observed in our simulations at different reduced temperatures, T^*. Table II also lists the associated values of the Zimm–Bragg parameters, which are calculated using equations (3) and (4) for each run, and the temperature of the simulations in degrees Kelvin. The propagation parameter, s, decreases

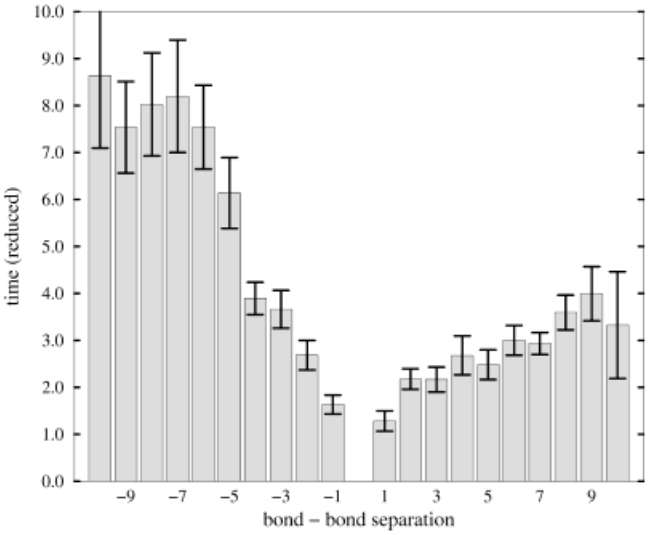

Fig. 14. The time that elapses before α-helical hydrogen bonds form, measured from the time at which the α-helical hydrogen bond at position zero on the *x*-axis forms. If the bond at position zero on the *x*-axis is denoted N_{i+4}—C_i, a value of positive 1 on the *x*-axis refers to bond N_{i+5}—C_{i+1} and a value of negative 1 on the *x*-axis refers to bond N_{i+3}—C_{i-1}. Therefore, positive *x*-values represent bonds toward the C-terminus and negative *x*-values represent bonds toward the N-terminus. Data are averaged over all α-helical hydrogen bonds formed in 30 independent polyA simulations at $T^* = 0.1$ (a total of 5083 observations); error bars represent the standard deviation in the observations.

TABLE II. Helix Properties Measured During Simulations*

T^*	$\langle n \rangle$	$\langle l_{ZB} \rangle$	$\frac{\langle n \rangle}{\langle l_{ZB} \rangle}$	s	σ_{ZB}	$T(K)$
0.060	16.1	14.7	1.09	2.15	0.0949	200
0.100	15.5	11.9	1.30	1.87	0.0952	330
0.115	14.7	9.61	1.53	1.65	0.0983	380
0.125	14.3	8.66	1.65	1.59	0.104	415
0.135	13.7	7.69	1.78	1.49	0.106	450
0.175	3.52	1.31	2.68	0.291	0.731	585

*Average number of α-Helical residues, $\langle n \rangle$, average length of an α-helical segment, $\langle l_{ZB} \rangle$, and average number of helical segments, $\langle n \rangle / \langle l_{ZB} \rangle$) and calculated Zimm–Bragg parameters (propagation parameter, s, and nucleation parameter, σ_{ZB}) for different values of temperature (reduced, T^*, and in Kelvin[a]).

[a]Very rough assignment of temperature (Kelvin) based on the results of Okamoto and Hansmann,[26] as described in text.

with increasing temperature, since increased temperature causes increased fluctuations within the chain that are detrimental to the process of elongating an established helical segment. The nucleation parameter, σ_{ZB}, increases with increasing temperature, indicating that nucleation is easier at higher temperatures. The average number of helical segments, $\langle n \rangle / \langle l_{ZB} \rangle$, also increases with temperature. At high temperatures, the ensemble of structures is populated by conformations with multiple short helical segments. These trends in σ_{ZB} and σ with temperature are consistent with theoretical calculations from Yang and Honig[78] and with results from Okamoto and Hansmann for simulations of 20-residue polyalanine chains.[26]

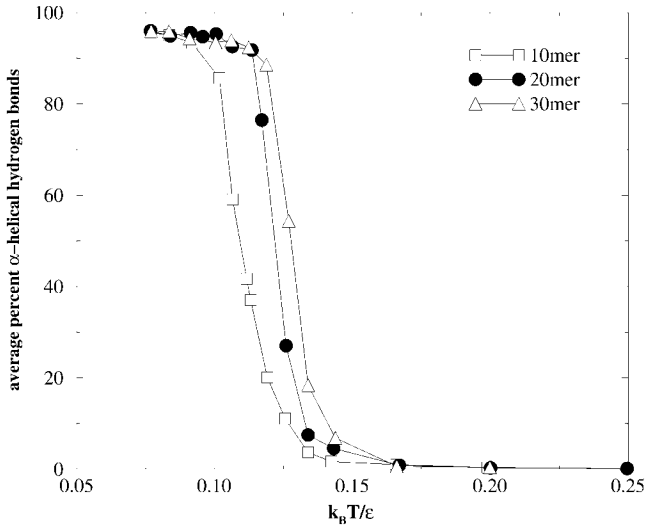

Fig. 15. Average percentage α-helical hydrogen bonds formed as a function of reduced temperature for polyalanine chains with 10-, 20-, and 30 4-bead residues.

The temperatures in Kelvin presented in Table II are assigned on the basis of a comparison between our results and the work of Okamoto and Hansmann.[26] This comparison, in which we determine an effective value of ϵ (approximately 6 kcal/mol) and convert between our reduced temperature (T^*) and temperature in Kelvin ($T(K) = T^*\epsilon/k_B$), allows us to estimate that $s \approx 1.97$ and $\sigma_{ZB} \approx 0.095$ at 270K. The s value from our study is in good agreement with experimental results that yield estimates of s for alanine of 1.33–2.19.[90,94] Our values of s at different temperatures are also similar to published simulation results from Okamoto and Hansmann.[26] Our value of σ_{ZB} at 270K (≈0.095) is much larger than the experimental estimate[90] (≈0.003) but is less than twice that found by Okamoto and Hansmann[26,29] ($\sigma_{ZB} \approx 0.051$) and by Sung and Wu[28] ($\sigma_{ZB} \approx 0.056$) and is smaller than estimates from other simulations.[17,21] The disagreement probably results from simplifications in our model that affect the chain environment, such as our lack of solvent. Our estimate of σ_{ZB} is probably also affected by the definition of helicity, where we have arbitrarily chosen to define a residue as helical if its Φ and Ψ angles are within 30° of ideal angle values.

Polyalanine helix formation as a function of chain length

To ensure that our model polyalanine chain consistently forms a helix over a range of peptide lengths, we also report on helix formation for 10mer and 30mer polyA chains. Figure 15 plots the average percentage of α-helical hydrogen bonds formed over a range of reduced temperatures for polyA of lengths 10, 20, and 30. (In each case, the maximum number of α-helical bonds is the chain length minus four, and the percentage plotted is relative to this maximum.) With such small side-chains, α-helices form relatively easily. As discussed above, β-sheet conformations are occasionally observed, but the full-length α-helix is the lowest possible energy conformation.

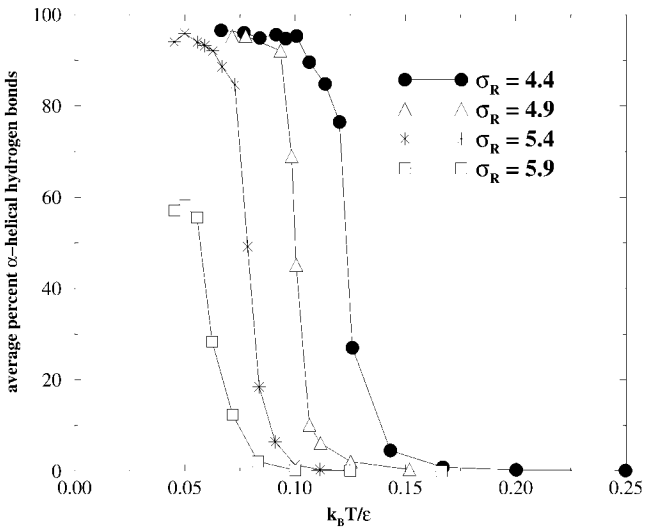

Fig. 16. Average percentage α-helical hydrogen bonds formed as a function of reduced temperature for chains with polyalanine-sized (σ_R = 4.4) and larger side-chains.

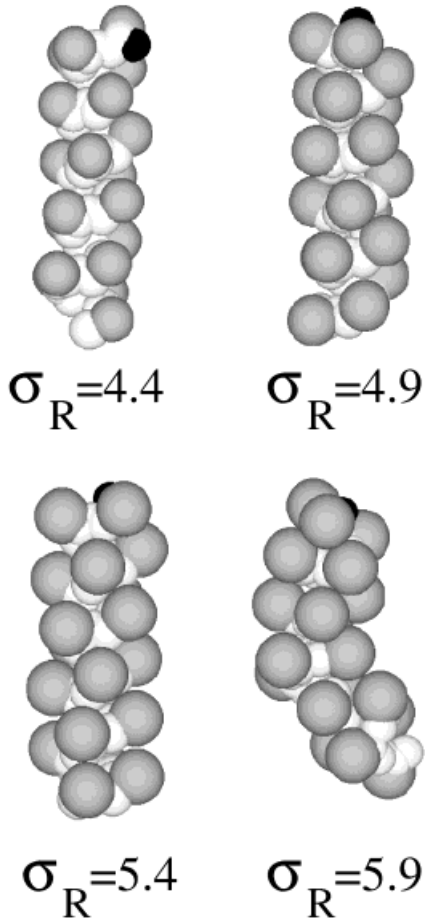

Fig. 17. Snapshots of the lowest-energy states of homogeneous peptide chains with different size side-chains, as in Fig. 16. Beads are shown full-size; bead color as in Fig. 8.

For each chain length, there is a sharp transition from the random coil state at high temperatures to the folded, α-helical state at low temperatures. Shorter chains require lower temperatures to force the transition, but the length of the chain has only a slight effect on the transition temperature. This is most likely due to chain end effects. The α-helical hydrogen bonds at the end of a chain are less stable than those in the chain center since the four N-terminal amide hydrogens and the four C-terminal carbonyl oxygens remain unbound in a full-length α-helix. In contrast, both the amide hydrogen and the carbonyl oxygen atoms of each central residue are involved in α-helical hydrogen bonds, creating more stability in the vicinity of these central residues. Repeated fraying and reformation of the ends of otherwise stable helices has been reported elsewhere.[24,25,30,86] Since the relatively large flexibility in the chain ends can weaken neighboring hydrogen bonds, the shorter the chain, the more likely that unraveling of the ends will cause the whole helix to dissolve. The same trend in helicity as a function of chain length was observed by Okamoto and Hansmann for polyalanine simulations of 10, 15, 20, and 30 residues.[26,29]

Helix Formation as a Function of Side-Chain Size

As expected, the ability to form an α-helix is affected by the size of the side-chains. Figure 16 plots the average percentage of α-helical hydrogen bonds formed as a function of temperature for 20-residue peptides with side-chain diameters, s_R, of 4.4 (polyalanine), 4.9, 5.4, and 5.9. For the peptides with smallest side-chains, α-helix formation is seemingly effortless. As the side-chain diameter is increased, lower temperatures are required to force the transition to the α-helical state. At the largest side-chain, σ_R = 5.9, the formation of more than two consecutive α-helical hydrogen bonds is sterically impossible due to side-chain–side-chain overlaps; therefore, the lowest energy conformation for this chain is one with approximately two-thirds of its α-helical bonds formed. The minimum-energy structure for each peptide is depicted in Figure 17. The first three chains form full-length α-helices. The σ_R = 5.9 chain forms pairs of α-helical hydrogen bonds separated by a missing bond, resulting in a kinked structure. These results are in agreement with experimental[94,96] studies demonstrating that steric hindrance causes residues with larger, bulkier side-chains to have lower tendencies toward helical structures.

Polyglycine Dynamics

Because model glycine residues lack side-chains, polyglycine (polyG) chains have significantly more conformational freedom than do non-glycine chains and do not tend to form α-helical structures. This freedom is evident in Figure 5(b), where model glycine residues sample most of Φ–Ψ space. Figure 18 shows the number of α-helical hydrogen bonds (black line) and the total number of hydrogen bonds (gray line) over the course of a single polyG simulation at T^* = 0.1. Despite the high degree of hydrogen bonding, polyG does not form extensive or long-lasting α-helical structures. Snapshots of the polyG chain during this simulation are shown in Figure 19. The first

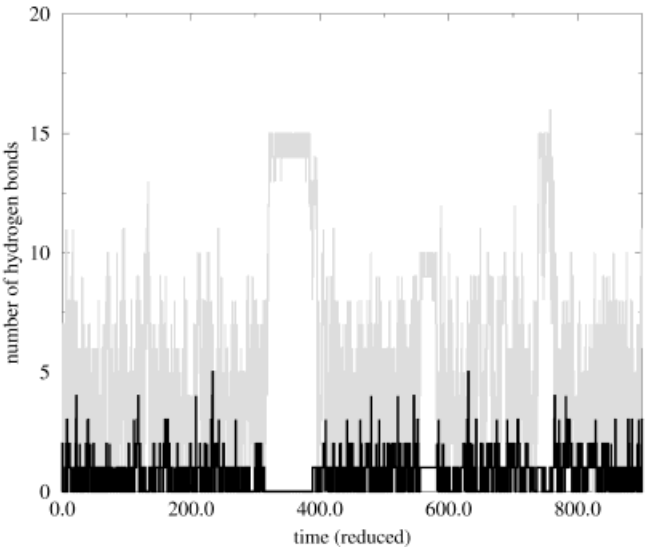

Fig. 18. Total number of hydrogen bonds formed (gray line) and number of α-helical hydrogen bonds formed (black line) for the polyglycine homopolymer over reduced time.

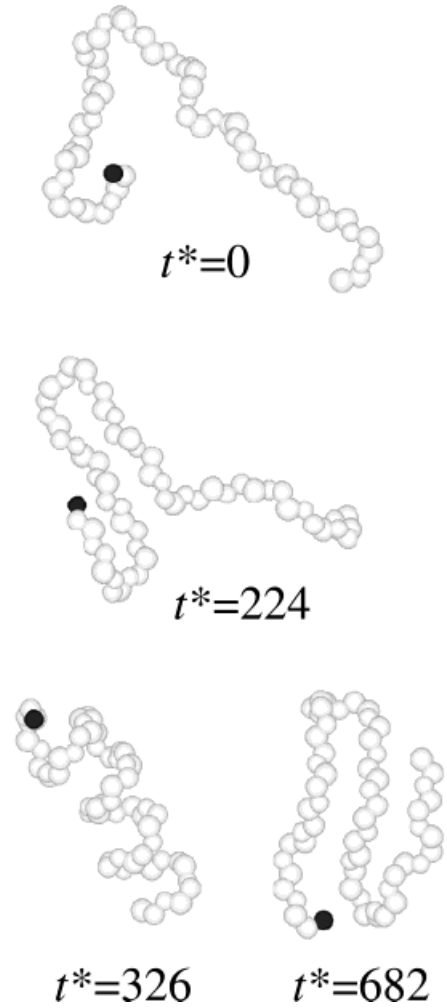

Fig. 19. Snapshots of the polyglycine homopolymer during the run shown in Fig. 18. Beads are reduced to one-half their true size for ease of viewing. The N-terminal end bead is black.

snapshot is the random initial conformation. At $t^* = 224$, there is a small amount of β-sheet structure. In the third snapshot ($t^* = 326$), the chain adopts a structure stabilized by 15 hydrogen bonds that is partially helical (although not α-helical) with a β-turn near the N-terminus. The last snapshot ($t^* = 682$) shows a three-stranded β-sheet stabilized by 9 hydrogen bonds. Overall, polyG displays much more conformational freedom and a much smaller preference for α-helical structure than does polyA, as has been shown elsewhere.[23,25,26,78,83,94] Model polyglycine chains do not favor α-helices, confirming that the hydrogen bond potential developed for this work does not have a bias toward α-helical structures.

DISCUSSION

The intermediate-resolution protein model presented in this work serves as a valuable compromise between computationally intensive models with all-atom detail and computationally fast models with minimalist physical detail. The efficiency of the proposed model results from the speed inherent to the DMD algorithm. Using our model, polyalanine chains assume α-helical conformations within 15 min on a 500-MHz workstation. Since α-helices are estimated experimentally to form within 100 ns,[31–34] a 100-ns simulation can be completed in less than 15 min. The time calculation is imprecise; however, this approximation indicates an ability to simulate nearly ten microseconds per day with our simplified model, a marked improvement over high-resolution, continuous-potential models. Simulations spanning such long times are only possible with idealized models, such as the one developed in this work.

Our ultimate goal is to study multiprotein systems, focussing in particular on aggregation phenomena in these complex systems. We are encouraged by the realistic backbone–backbone hydrogen bonding and secondary structure formation displayed by our model. An abnormally high level of β-sheet structure, possibly resulting from an undesired α-helix to β-sheet transition, has been implicated in several medically important instances of aberrant aggregation,[98–101] the assembly of β-amyloid proteins into amyloid plaques, the hallmark of Alzheimer's disease,[99,102,103] and the conversion of soluble prion proteins into fibrils, a precursor to fibrillar aggregation in the prion diseases.[104] It is possible that backbone–backbone hydrogen bonds, which are known to stabilize α-helices and β-sheets, play a role in these instances of aggregation.

Although the backbone of our model is detailed enough to display genuine protein character, the model side-chains are very simplified. This lack of side-chain detail is a potential shortcoming of the model. In a modification to the model that will be presented elsewhere, we add a hydrophobic force by allowing side-chain beads to be either hydrophobic or polar. This simple addition to the model enables simulation of residue-specific hydrophobic collapse (in terms of a three-letter alphabet: H, hydrophobic; P, polar; and G, glycine).

Ultimately, we are interested in being able to distinguish between native and aggregated states, where "native" can merely mean stably folded to a structure that is close enough to the true native conformation so as not to aggregate. For this purpose, we believe a united-atom side-chain (either hydrophobic or not hydrophobic) will suffice. However, other aspects of side-chain character are ignored, including charges and hydrogen bonding capability (between pairs of side-chains and between the backbone and a side-chain). Consequently, the model presented here is unlikely to adequately simulate intricate side-chain–side-chain interactions or ligand–receptor interactions and is not currently suited for studies that involve differentiating between largely similar folded structures.

Another potential shortcoming of the model is the lack of solvent. We incorporate solvent as a potential of mean force, having an equal effect on all chain beads regardless of the chain conformation. Consequently, beads in the interior of a compact structure are just as affected by solvent as if they were fully exposed in a random coil structure. A more accurate solvation effect would allow forces, such as the hydrogen bonding force, in the core of a collapsed chain to be different from those at the surface. It is unclear, however, that this shortcoming is critical in the work presented here, since our peptides are so small that the concept of a "core" is hard to define. Nonetheless, the effect will become more significant with longer chains or multiprotein systems.

CONCLUSIONS

In this work, we present an intermediate-resolution protein model for use with discontinuous molecular dynamics simulations. We have concentrated on the steric repulsion and hydrogen bonding terms for this study, particularly as related to α-helix formation in model protein chains. With physical parameters chosen to best represent polyalanine, we see a characteristic α-helical structure stabilized by backbone hydrogen bonding. We observe a slight dependence of chain length on α-helix formation for polyalanine chains of length 10, 20, and 30; and we find that very large side–chains prevent steric α-helix formation. Polyglycine chains explore broad regions of conformational space and do not tend to form α-helices.

Ongoing research is aimed at using this simplified protein model to study the effects of side-chain heterogeneity and hydrophobicity on secondary and tertiary structure formation and to study the competition between protein folding and aggregation in multiprotein environments. Given the current power of computers and supercomputers, we believe that intermediate-resolution protein models are valuable tools for researchers. By sacrificing some of the physical detail present in all-atom protein models, intermediate-resolution models will allow simulations of significantly larger and more complicated protein systems. The key, of course, is that newly proposed intermediate-resolution models are rigorously tested to establish that they possess enough realistic character to enable protein-like dynamics.

REFERENCES

1. Thirumalai D, Klimov DK. Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models. Curr Opin Struct Biol 1999;9:197–207.
2. Doniach S, Eastman P. Protein dynamics simulations from nanoseconds to microseconds. Curr Opin Struct Biol 1999;9:157–163.
3. Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science 1998;282:740–744.
4. Kolinski A, Skolnick J. Discretized model of proteins. I. Monte Carlo study of cooperativity in homopolypeptides. J Chem Phys 1992;97:9412–9426.
5. Kolinski A, Skolnick J. Monte Carlo simulations of protein folding. I. Lattice model and interaction scheme. Proteins 1994;18: 338–352.
6. Kolinski A, Skolnick J. Monte Carlo simulations of protein folding. II. Application to protein A, ROP, and crambin. Proteins 1994;18:353–366.
7. Sikorski A, Kolinski A, Skolnick J. Computer simulations of de novo designed helical proteins. Biophys J 1998;75:92–105.
8. Sikorski A, Kolinski A, Skolnick J. Computer simulations of the properties of the α_2, α_2C, and α_2D de novo designed helical proteins. Proteins 2000;38:17–28.
9. Li L, Mirny LA, Shakhnovich EI. Kinetics, thermodynamics and evolution of non-native interactions in a protein folding nucleus. Nature Struct Biol 2000;7:336–342.
10. Klimov DK, Thirumalai D. Cooperativity in protein folding: from lattice models with sidechains to real proteins. Fold Design 1998;3:127–139.
11. Wallqvist A, Ullner M. A simplified amino acid potential for use in structure predictions of proteins. Proteins 1994;18:267–280.
12. Sun S. Reduced representation model of protein structure prediction: statistical potential and genetic algorithms. Protein Sci 1993;2:762–785.
13. Takada S, Luthey-Schulten Z, Wolynes PG. Folding dynamics with nonadditive forces: a simulation study of a designed helical protein and a random heteropolymer. J Chem Phys 1999;110: 11616–11629.
14. Gunn JR, Friesner RA. Parallel implementation of a protein structure refinement algorithm. J Comput Chem 1996;17:1217–1228.
15. Irback A, Sjunnesson F, Wallin S. Three-helix-bundle protein in a Ramachandran model. Proc Natl Acad Sci USA 2000;97:13614–13618.
16. Smith AV, Hall CK. Bridging the gap between homopolymer and protein models: a discontinuous molecular dynamics study. J Chem Phys 2000;113:9331–9342.
17. Daggett V, Kollman PA, Kuntz ID. A molecular dynamics simulation of polyalanine: an analysis of equilibrium motions and helix–coil transitions. Biopolymers 1991;31:1115–1134.
18. DiCapua FM, Swaminathan S, Beveridge DL. Theoretical evidence for water insertion in α-helix bending: molecular dynamics of Gly_{30} and Ala_{30} in vacuo and in solution. J Am Chem Soc 1991;113:6145–6155.
19. Soman KV, Karimi A, Case DA. Unfolding of an α-helix in water. Biopolymers 1991;31:1351–1361.
20. Tirado-Rives J, Jorgensen WL. Molecular dynamics simulations of the unfolding of an α-helical analogue of ribonuclease A S-peptide in water. Biochemistry 1991;30:3864–3871.
21. Daggett V, Levitt M. Molecular dynamics simulations of helix denaturation. J Mol Biol 1992;223:1121–1138.
22. Hirst JD, Brooks CL III. Molecular dynamics simulations of isolated helices of myoglobin. Biochemistry 1995;34:7614–7621.
23. Bertsch RA, Vaidehi N, Chan SI, Goddard WA III. Kinetic steps for α-helix formation. Proteins 1998;33:343–357.
24. Takano M, Yamato T, Higo J, Suyama A, Nagayama K. Molecular dynamics of a 15-residue poly(L-alanine) in water: helix formation and energetics. J Am Chem Soc 1999;121:605–612.
25. Sung S-S. Helix folding simulations with various initial conformations. Biophys J 1994;66:1796–1803.
26. Okamoto Y, Hansmann UHE. Thermodynamics of helix–coil transitions studied by multicanonical algorithms. J Phys Chem 1995;99:11276–11287.
27. Sung S-S. Folding simulations of alanine-based peptides with lysine residues. Biophys J 1995;68:826–834.

28. Sung S-S, Wu X-W. Molecular dynamics simulations of synthetic peptide folding. Proteins 1996;25:202–214.
29. Hansmann UHE, Okamoto Y. Finite-size scaling of helix–coil transitions in poly-alanine studied by multicanonical simulations. J Chem Phys 1999;110:1267–1339.
30. Ferrara P, Apostolakis J, Caflisch A. Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations. J Phys Chem B 2000;104:5000–5010.
31. Williams S, Causgrove TP, Gilmanshin R, Fang KS, Callender RH, Woodruff WH, Dyer RB. Biochemistry 1996;35:691–697.
32. Eaton WA, Munoz V, Thompson PA, Chan C-K, Hofrichter J. Submillisecond kinetics of protein folding. Curr Opin Struct Biol 1997;7:10–14.
33. Callender RH, Dyer RB, Gilmanshin R, Woodruff WH. Fast events in protein folding: the time evolution of primary processes. Annu Rev Phys Chem 1998;49:173–202.
34. Thompson PA, Munoz V, Jas GS, Henry ER, Eaton WA, Hofrichter J. The helix–coil kinetics of a heteropeptide. J Phys Chem B 2000;104:378–389.
35. Ballew RM, Sabelko J, Gruebele M. Direct observation of fast protein folding: the initial collapse of apomyoglobin. Proc Natl Acad Sci USA 1996;93:5759–5764.
36. Chan C-K, Hu Y, Takahashi S, Rousseau DL, Eaton WA, Hofrichter J. Submillisecond protein folding kinetics studied by ultrarapid mixing. Proc Natl Acad Sci USA 1997;94:1779–1784.
37. Shastry MCR, Roder H. Evidence for barrier-limited protein folding kinetics on the microsecond time scale. Nature Struct Biol 1998;5:385–392.
38. Pascher T, Chesick JP, Winkler JR, Gray HB. Protein folding triggered by electron transfer. Science 1996;271:1558–1560.
39. Schellman JA. The factors affecting the stability of hydrogen-bonded polypeptide structures in solution. J Phys Chem 1958;62: 1485–1494.
40. Zimm BH, Bragg JK. Theory of the phase transition between helix and random coil in polypeptide chains. J Chem Phys 1959;31:526–535.
41. Lifson S, Roig A. On the theory of helix–coil transition in polypeptides. J Chem Phys 1961;34:1963–1974.
42. Munoz V, Serrano L. Elucidating the folding problem of helical peptides using empirical parameters. Nature Struct Biol 1994;1: 399–409.
43. Alder BJ, Wainwright TE. Studies in molecular dynamics. I. General method. J Chem Phys 1959;31:459–466.
44. Rapaport DC, Molecular dynamics simulation of polymer chains with excluded volume. J Phys A Math Gen 1978;11:L213–L217.
45. Bellemans A, Orban J, Van Belle D. Molecular dynamics of rigid and non-rigid necklaces of hard discs. Mol Phys 1980;39:781–782.
46. Zhou Y, Karplus M. Folding thermodynamics of a model three-helix-bundle protein. Proc Natl Acad Sci USA 1997;94:14429–14432.
47. Zhou Y, Karplus M. Interpreting the folding kinetics of helical proteins. Nature 1999;401:400–403.
48. Zhou Y, Karplus M. Folding of a model three-helix bundle protein: a thermodynamic and kinetic analysis. J Mol Biol 1999;293:917–951.
49. Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI. Discrete molecular dynamics studies of the folding of a protein-like model. Fold Design 1998;3:577–587.
50. Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI. Identifying the protein folding nucleus using molecular dynamics. J Mol Biol 2000;296:1183–1188.
51. Smith AV, Hall CK. Assembly of a tetrameric α-helical bundle: computer simulations on an intermediate-resolution protein model. Proteins 2001;42:376–391.
52. Smith AV, Hall CK. Protein refolding versus aggregation: computer simulations on an intermediate-resolution protein model. J Mol Biol 2001 (in press).
53. Smith SW, Hall CK, Freeman BD. Molecular dynamics study of entangled hard-chain fluids. J Chem Phys 1996;104:5616–5637.
54. Smith SW, Hall CK, Freeman BD. Molecular dynamics for polymeric fluids using discontinuous potentials. J Comp Phys 1997;134:16–30.
55. Kenkare NR, Smith SW, Hall CK, Khan SA. Discontinuous molecular dynamics studies of end-linked polymer networks. Macromolecules 1998;31:5861–5879.
56. Kenkare NR, Smith SW, Hall CK. Theory and simulation of the swelling of polymer gels. J Chem Phys 2000;113:404–418.
57. Hu L, Rangwalla H, Cui J, Elliott JR Jr. Vapor–liquid equilibria of vibrating square well chains. J Chem Phys 1999;111:1293–1301.
58. Pawloski AR, Torres JA, Nealey PF, de Pablo JJ. Applications of molecular modeling in nanolithography. J Vac Sci Technol B 1999;17:3371–3378.
59. Sun S, Thomas PD, Dill KA. A simple protein folding algorithm using a binary code and secondary structure constraints. Protein Eng 1995;8:769–778.
60. Voet D, Voet JG. Biochemistry. New York: John Wiley & Sons; 1990.
61. Baker EN, Hubbard RE. Hydrogen bonding in globular proteins. Prog Biophys Mol Biol 1984;44:97–179.
62. Allen MP, Tildesley DJ. Computer simulation of liquids. New York: Oxford University Press; 1987.
63. Woodcock LV. Isothermal molecular dynamics calculations for liquid salts. Chem Phys Lett 1971;10:257–261.
64. Nose S. A molecular dynamics method for simulations in the canonical ensemble. Mol Phys 1984;52:255–268.
65. Krajci M. Markov chain algorithms for canonical ensemble simulation. Comput Phys Commun 1986;42:29–35.
66. Forrest BM, Suter UW. Hybrid Monte Carlo simulations of dense polymer systems. J Chem Phys 1994;101:2616–2629.
67. Gromov DG, de Pablo JJ. Structure of binary polymer blends: multiple time step hybrid Monte Carlo simulations and self-consistent integral-equation theory. J Chem Phys 1995;103:8247–8256.
68. Andersen HC. Molecular dynamics simulations at constant temperature and/or pressure. J Chem Phys 1980;72:2384–2393.
69. Zhou Y, Karplus M, Wichert JM, Hall CK. Equilibrium thermodynamics of homopolymers and clusters: molecular dynamics and Monte Carlo simulations of systems with square-well interactions. J Chem Phys 1997;107:10691–10708.
70. Huo S, Straub JE. Direct computation of long time processes in peptides and proteins: reaction path study of the coil-to-helix transition in polyalanine. Proteins 1999;36:249–161.
71. Goldbeck RA, Thomas YG, Chen E, Esquerra RM, Kliger DS. Multiple pathways on a protein-folding energy landscape: kinetic evidence. Proc Natl Acad Sci USA 1999;96:2782–2787.
72. Sabelko J, Ervin J, Gruebele M. Observation of strange kinetics in protein folding. Proc Natl Acad Sci USA 1999;96:6031–6036.
73. Dill KA, Chan HS. From Levinthal to pathways to funnels. Nature Struct Biol 1997;4:10–19.
74. Baldwin RL. Matching speed and stability. Nature 1994;369:183–184.
75. Baldwin RL. The nature of protein folding pathways: the classical versus the new view. J Biomol NMR 1995;5:103–109.
76. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways and the energy landscape of protein folding: a synthesis. Proteins 1995;21:167–195.
77. Ooi T, Oobatake M. Prediction of the thermodynamics of protein unfolding: the helix–coil transition of poly(L-alanine). Proc Natl Acad Sci USA 1991;88:2859–2863.
78. Yang A-S, Honig B. Free energy determinants of secondary structure formation. I. α-helices. J Mol Biol 1995;252:351–365.
79. Hermans J, Anderson AG, Yun RH. Differential helix propensity of small apolar side chains studied by molecular dynamics simulations. Biochemistry 1992;31:5646–5653.
80. Creamer TP, Rose GD. Side-chain entropy opposes α-helix formation but rationalizes experimentally determined helix-forming propensities. Proc Natl Acad Sci USA 1992;89:5937–5941.
81. Creamer TP, Rose GD. α-helix-forming propensities in peptides and proteins. Proteins 1994;19:85–97.
82. Creamer TP, Rose GD. Interactions between hydrophobic side chains within α-helices. Protein Sci 1995;4:1305–1314.
83. Okamoto Y. Helix-forming tendencies of nonpolar amino acids predicted by Monte Carlo simulated annealing. Proteins 1994;19: 14–23.
84. Sung S-S, Wu X-W. Molecular dynamics simulations of helix folding: the effects of amino acid substitution. Biopolymers 1997;42:633–644.
85. Marqusee S, Baldwin RL. Helix stabilization by Glu^{-}. . .Lys^{+} salt bridges in short peptides of de novo design. Proc Natl Acad Sci USA 1987;84:8898–8902.
86. Marqusee S, Robbins VH, Baldwin RL. Unusually stable helix

formation in short alanine-based peptides. Proc Natl Acad Sci USA 1989;86:5286–5290.
87. Lyu PC, Liff MI, Marky LA, Kallenbach NR. Side chain contributions to the stability of alpha-helical structure in peptides. Science 1990;250:669–673.
88. O'Neil KT, DeGrado WF. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. Science 1990;250:646–651.
89. Wojcik J, Altmann K-H, Scheraga HA. Helix–coil stability constants for the naturally occurring amino acids in water. XXIV. Half-cystine parameters from random poly(hydroxybutylglutamine-co-S-methylthio-L-cysteine). Biopolymers 1990;30:121–134.
90. 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;31:1463–1470.
91. Park SH, Shalongo W, Stellwagen E. Residue helix parameters obtained from dichroic analysis of peptides of defined sequence. Biochemistry 1993;32:7048–7053.
92. Park SH, Shalongo W, Stellwagen E. Modulation of the helical stability of a model peptide by ionic residues. Biochemistry 1993;32:12901–12905.
93. 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;3:843–852.
94. Chakrabartty A, Baldwin RL. Stability of α-helices. Adv Protein Chem 1995;46:141–176.
95. 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;92:10924–10928.
96. Luo P, Baldwin RL. Interaction between water and polar groups of the helix backbone: an important determinant of helix propensities. Proc Natl Acad Sci USA 1999;96:4930–4935.
97. Poland D, Scheraga HA. Theory of helix–coil transitions in biopolymers. New York: Academic Press; 1970.
98. Pan K-M, Baldwin M, Nguyen J, Gasset M, Serban A, Groth D, Mehlhorn I, Huang Z, Fletterick RJ, Cohen FE, Prusiner SB. Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. Proc Natl Acad Sci USA 1993;90:10962–10966.
99. Fink AL. Protein aggregation: folding aggregates, inclusion bodies and amyloid. Fold Design 1998;3:R9–R23.
100. Jacchieri SG. Study of α-helix to β-strand to β-sheet transitions in amyloid: the role of segregated hydrophobic β-strands. Biophys Chem 1998;74:23–34.
101. Radford SE, Dobson CM. From computer simulations to human disease: emerging themes in protein folding. Cell 1999;97:291–298.
102. Koo EH, Lansbury PT Jr, Kelly JW. Amyloid diseases: abnormal protein aggregation in neurodegeneration. Proc Natl Acad Sci USA 1999;96:9989–9990.
103. Sinha S, Lieberburg I. Cellular mechanisms of β-amyloid production and secretion. Proc Natl Acad Sci USA 1999;96:11049–11053.
104. Harrison PM, Chan HS, Prusiner SB, Cohen FE. Thermodynamics of model prions and its implications for the problem of prion protein folding. J Mol Biol 1999;286:593–606. = 0.10). The chain folds into a stable, full-length helix.