Bridging the gap between homopolymer and protein models: A discontinuous molecular dynamics study

Anne Voegler Smith and Carol K. Hall

Department of Chemical Engineering, North Carolina State University, Box 7905, Raleigh, North Carolina 27695-7905

(Received 10 July 2000; accepted 1 September 2000)

A series of seven off-lattice protein models is analyzed that spans a range of chain geometry from a simple, low-resolution homopolymer model to an intermediate-resolution model that accounts for the presence of side chains, the varied character of the individual amino acids, the rigid nature of protein backbone angles, and the length scales that characterize real protein bead sizes and bond lengths. Discontinuous molecular dynamics is used to study the transition temperatures and physical structures resulting from simulations with each protein model. Our results show that each protein model undergoes multiple thermodynamic transitions that roughly correlate with protein transitions during folding to the native state. Other realistic protein behavior, such as burial of hydrophobic side chains and hindered motion due to backbone rigidity, is observed with the more-detailed models. The results suggest that, despite their simplicity when compared with all-atom protein models, the models presented here display a significant amount of protein character and, when coupled with the efficient discontinuous molecular dynamics algorithm, may enable simulation of multiprotein systems over long times. © 2000 American Institute of Physics. [S0021-9606(00)51944-8]

I. INTRODUCTION

We are engaged in a program of research aimed at investigating protein aggregation, a phenomenon that causes serious problems in the biomedical and pharmaceutical industries^{1–5} and has been linked to a number of human conditions including Alzheimer's disease and cataracts.^{6,7} Despite its importance, protein aggregation has received little attention from the theoretical community, and as yet, a general understanding of its physical basis is far from complete. To study aggregation computationally, protein models must be developed that contain enough genuine protein character to mimic real proteins yet are simple enough to allow computer simulation of multiprotein systems over long times. In this article, we describe our efforts to construct a minimalist protein model that is computationally tractable and could eventually serve as the basis for studies of protein aggregation.

The ability of idealized, or minimalist, models to provide insights into the coarse-grained structure and folding mechanisms of proteins has long been recognized.^{8–13} All-atom models are physically more realistic than minimalist models; however, their complexity precludes their use for studies of long-time events. Minimalist models, on the other hand, have played an important role in theoretical and computational studies that describe the shape of the protein folding landscape and the major physical changes experienced by a protein during folding from denatured random conformations to the native state, albeit with a sacrifice in detail.^{14–22} Among the more popular minimalist models are the homo- and heteropolymer chain models in which each amino acid residue is represented by a single sphere with identical (homo) or varied (hetero) interaction parameters.^{23–25} In addition to studies of physical transitions during folding of the

chain,^{20,26,27} homopolymer simulations have been used to probe for molten globule intermediates during the globule-to-coil transition^{28,29} and have successfully produced experimentally-characterized structures.³⁰ Recently, heteropolymer models have gained popularity as they can more accurately represent the character of real protein chains. Both lattice and continuous models have been used to demonstrate that heteropolymers “fold” to a small set of low energy (“native”) structures.^{8,14,15,31–35}

Toward our goal of building a protein model suitable for simulations of protein aggregation, we previously performed single-chain simulations of a freely-jointed, tangent, square-well chain homopolymer model, a very simple protein model in which each amino acid residue is represented by a sphere.^{26,27} Despite the simplicity of the model, the system undergoes multiple equilibrium thermodynamic transitions that roughly correlate with gas-to-liquid, liquid-to-solid, and solid-to-solid transitions. Similar multiple phases are also observed during protein folding wherein proteins transition between denatured, collapsed globule, and native states.³⁶ The gas-to-liquid collapse transition observed in homopolymer simulations mimics the structural collapse in the early stages of protein folding that is often driven by hydrophobic attraction. A homopolymer liquid-to-solid transition qualitatively corresponds to a protein transition from the collapsed, molten-globule state to the native state. The third transition seen for homopolymers at very low temperatures corresponds to complete solidification of the protein and may indicate a cold denaturation process to an inactive state.²⁷ That the simplest of chain models, the homopolymer, can qualitatively describe the stages of protein folding prompted us to consider the behavior of a range of simple protein models based on the homopolymer model studied previously.

In this work, we systematically add selected protein-like features to the square-well homopolymer model studied previously and use computer simulation to evaluate the impact that each new feature has on the phase transitions displayed by the model. Our focus here is on developing a simple representation of protein geometry that accounts for the presence of bulky side chains, the varied character of the individual amino acids, the rigid nature of protein backbone angles, and the length scales that characterize real protein bead sizes and bond lengths. The homopolymer model is modified in four steps by adding the following physical features piece-by-piece: (a) branched structure, (b) heterogeneity, (c) realistic backbone angles, and (d) realistic ratios of bead sizes to bond lengths. By building the model in a step-by-step fashion, we create a hierarchy of distinct protein models that allows us to assess the relative importance of these features as measured by their impact on the phase transitions in the system. We perform discontinuous molecular dynamics simulations on each of the protein models in this hierarchy at several different temperatures, monitoring the radius of gyration, specific heat, internal energy, and overall chain conformation. For each model studied, we report how the particular modification affects the observed thermodynamic phase transitions and the resulting physical structures.

Highlights of our simulation results are the following. We find that introducing branching to the homopolymer model has little effect on either the temperatures or the strengths of the thermodynamic transitions. As with the homopolymer model, branched homopolymers adopt symmetric conformations at moderately low temperatures and tight, spherical conformations at very low temperatures. Heterogeneity, however, causes both a downward temperature shift and a strengthening of the individual transitions which can be attributed to a loss of constraints. Heterogeneous chains assume different types of compact conformations since only a fraction of their segments are subject to hydrophobic collapse. In the systems that incorporate realistic backbone angles and realistic ratios of bead sizes to bond lengths, we find multiple transitions similar to those obtained for the other models studied. The low-temperature structures of the chains, however, are influenced by the increased chain rigidity. In a forthcoming article, we define a potential energy function for use with the most realistic physical model developed here, and we show that this physical model is indeed detailed enough to exhibit protein-like behavior but simple enough to allow for simulations covering very long time scales.³⁷

The remainder of the article is organized as follows. Section II describes the models studied and the discontinuous molecular dynamics simulation technique used. In Sec. III, we present the results associated with the four types of modifications to the homopolymer model considered here: (a) chain branching, (b) chain heterogeneity, (c) chain rigidity, and (d) bead overlap. Section IV provides a brief conclusion.

II. MODELS AND METHODS

We use discontinuous molecular dynamics (DMD) computer simulations to study the effect of different chain geometries on the observed phase transitions and types of equilib-

rium structures. Here we provide a brief review of the DMD algorithm and define the types of interactions present in the simulations. We then describe the seven models studied and discuss the importance of each. At the end of this section, we specify the simulation parameters and the properties measured during the simulations.

The DMD computer simulation technique was developed by Alder and Wainwright to calculate the properties of systems containing hard spheres or square-well spheres.³⁸ DMD is inherently faster than ordinary molecular dynamics, which is based on continuous potentials, because the equations of motion for DMD can be solved analytically. The compromise, however, is that the energy function must be limited to discontinuous functions, such as the hard-sphere and square-well potentials. Pairs of hard-sphere beads interact via the hard-sphere potential,

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

where r is the distance between beads i and j and σ is the bead diameter. A mixed pair, that is a hard-sphere bead and a square-well bead, also interact via the hard-sphere potential. Pairs of square-well beads interact via a square-well potential,

$$u_{ij}(r) = \begin{cases} \infty, & r \leq \sigma, \\ -\epsilon, & \sigma < r \leq \lambda\sigma, \\ 0, & r > \lambda\sigma, \end{cases} \quad (2)$$

where $\lambda\sigma$ is the well diameter and ϵ is the well depth. After the pioneering work of Alder and Wainwright, Rapaport³⁹ and Bellemans, Orban, and Belle⁴⁰ proposed a method of simulating chain systems with discontinuous potentials wherein the bond between neighbors along the chain is allowed to vary freely over a small range around the true bond length. With this method, successive chain segments are essentially decoupled, allowing the chain system to be simulated with the DMD technique.

The speed of the DMD algorithm results from the decoupled nature of the events. Beads influence each other only when the distance between them is equal to a point of discontinuity in their potential, e.g., a collision between hard spheres. At all other distances, they move linearly according to Newton's equation of motion. Therefore, a DMD simulation is broken into a series of events, as opposed to the series of very small time steps used in continuous-potential molecular dynamics simulations. Nonbonded beads move freely in space and experience events at distances of σ (a hard-sphere event) and $\lambda\sigma$ (a square-well event). 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. DMD simulations proceed through time by locating the pair of beads that will be involved in the next event, advancing all beads forward in time to that event, and calculating the event dynamics for the event pair. This process is performed repeatedly. Several efficiency techniques

FIG. 1. Model chains: (a) “homopolymer” model, (b) “homoprotein” model, (c) “heteroprotein” model, (d) “rigid homoprotein” model, (e) “rigid heteroprotein” model, (f) “overlapping rigid homoprotein” model, and (g) “overlapping rigid heteroprotein” model. Filled circles represent square-well beads; open circles represent hard-sphere beads. The extra bonds shown by bold, black lines in (d) through (g) restrain the backbone to a rigid zigzag structure and restrain the side chains to particular locations relative to the backbone.

are used in this work, including neighbor lists, binary trees, and false positioning, and are described in detail by Smith, Hall, and Freeman.⁴¹

Seven types of chain geometry are considered in this study. The first, the “homopolymer” model shown in Fig. 1(a), is a freely-jointed, tangent, square-well chain with 80 beads. Homopolymer models were studied previously in our group for a range of chain lengths^{26,27} and serve as our base structure. The remaining models studied are each built up from this foundation. The second model, the “homoprotein” model shown in Fig. 1(b), is a branched chain composed of a backbone of 60 freely-jointed, tangent, square-well beads and 20 freely-jointed, square-well side chain beads. The ho-

moprotein geometry is reminiscent of protein models in which each amino acid residue is represented by a four-bead monomer unit with three beads along the backbone and one side chain bead.^{42–44} Physically, a four-bead amino acid representation is significantly more realistic than a one-bead representation since it allows for separate backbone and side chain interactions. The third model, the “heteroprotein” model shown in Fig. 1(c) has the same four-bead, branched structure as the homoprotein model but contains hard-sphere backbone beads and square-well side chains. In Fig. 1, hard spheres are depicted by open circles, and square-well spheres are depicted by filled, gray circles. The attractive regions that surround square-well beads mimic attractions within real proteins, such as those due to hydrophobic forces. Heterogeneity is an important feature in proteins. The heteroprotein model, in which the side chains are attractive beads and the backbone beads are hard spheres, is a simple representation of a peptide chain composed of identical hydrophobic residues.

The fourth and fifth models, the “rigid homoprotein” model and the “rigid heteroprotein” model, are shown in Figs. 1(d) and 1(e), respectively. The rigid homoprotein model [Fig. 1(d)] has the same branched, square-well structure as the homoprotein model [Fig. 1(b)]; the rigid heteroprotein model [Fig. 1(e)] has the same branched, mixed hard-sphere and square-well structure as the heteroprotein model [Fig. 1(c)]. In each rigid model, however, extra bonds (bold, black lines in Fig. 1) have been added to implement rigid, angular constraints like those present in real proteins. For the purpose of the simulation, these extra bonds are treated identically to the original, neighbor–neighbor bonds. Extra bonds are added between next-neighbors along the backbone, effectively fixing all bond angles along the backbone. Bonds are also added between the central beads of each four-bead monomer building block. These bonds are used because, in real proteins, the distance between central atoms of neighboring amino acids are relatively fixed due to the fixed backbone bond angles and the *trans* nature of peptide (interresidue) bonds. Finally, the side chain bead is bonded to all three backbone beads in its monomer unit instead of just to the central bead. These extra side chain bonds are necessary because real amino acids have a chiral center atom. Two of the chiral atom’s constituents are the neighboring backbone beads; the other two are a side chain (one of 20 different chemical groups, the side chain mimicked here) and a hydrogen atom. In all models here, we neglect the hydrogen atom side chain. We model the chirality of the central atom by fixing the remaining side chain bead’s position such that the model chain is composed of identical isomers.

The final two models, the “overlapping rigid homoprotein” model and the “overlapping-rigid heteroprotein” model, are shown in Figs. 1(f) and 1(g), respectively. The overlapping rigid homoprotein model [Fig. 1(f)] has the same geometry and set of bonds as the rigid homoprotein model [Fig. 1(d)]; the overlapping rigid heteroprotein model [Fig. 1(g)] has the same geometry and set of bonds as the rigid heteroprotein model [Fig. 1(e)]. However, the overlapping models have enlarged bead diameters. The resulting chains have smoother overall surfaces that more accurately

models (d) and (e):

models (f) and (g):

FIG. 2. Bond lengths, bond angles, and bead sizes for models (d) through (g). Bonds propagate throughout the structure; at least one of each type is shown and labeled.

reflect the atomic-level surface topography of real molecules.

We perform DMD simulations on each of the seven protein models described above. Simulations are performed on DEC Alpha workstations and range in length from 50 million to 1 billion collisions, with longer simulations required to reach equilibrium at lower temperatures. For each simulation, initial chain conformations and velocities are random. In all models, the backbone bond lengths are chosen to be unity and all other lengths are scaled relative to the backbone bond length. The side chain bond length, present in models (b) through (g), is chosen to be larger than the backbone bond length to mimic the geometry of real amino acids in which the center of a side chain group may be extended from the backbone by a distance much larger than that of a backbone bond; the side chain bond length is arbitrarily set to 1.30. The bond lengths and angles for models (d) through (g) are shown schematically in Fig. 2 and include next-neighbor bond lengths of 1.50, chiral center to chiral center bond lengths of 2.35, and side chain to nonchiral center backbone bead bond lengths of 1.92. For models (a) through (e), the bead diameters (σ) are unity; for overlapping models (f) and (g), the bead diameters are 1.50. All square-well beads have well widths ($\lambda\sigma$) of 1.5σ . The bond tolerance, δ , is set to 0.1; therefore, all bonds are held to within 10% of their ideal lengths.

During the simulations, we monitor several thermodynamic and structural properties including radius of gyration, specific heat, and internal energy. The reduced squared radius of gyration, R_g^* , is given by

$$R_g^* = \frac{\langle 1/N \sum_{i=1}^N [(x_i - x_c)^2 + (y_i - y_c)^2 + (z_i - z_c)^2] \rangle}{\sigma^2 N}, \quad (3)$$

where N is the number of beads; x_i , y_i , z_i are the coordi-

nates of bead i , and x_c , y_c , z_c are the center-of-mass coordinates of the chain. The reduced specific heat, C_v^* , is given by

$$C_v^* = \frac{C_v}{k_B} = \frac{[\langle E^2 \rangle - (\langle E \rangle)^2]}{(\epsilon T^*)^2}, \quad (4)$$

where k_B is Boltzmann's constant, E is the internal energy, and T^* is the reduced temperature, defined to be $k_B T / \epsilon$. The reduced internal energy, E^* , is given by

$$E^* = \frac{\langle E \rangle}{\epsilon}. \quad (5)$$

Thermodynamic transitions can be characterized by changes in R_g^* , C_v^* , and E^* with temperature. A sigmoidal shape in an R_g^* vs T^* plot is characteristic of a second-order, gas-to-liquid collapse transition. The number of plateaus and peaks in a C_v^* vs T^* plot corresponds to the number of equilibrium transitions. Discontinuities in an E^* vs T^* plot are evidence of a first-order, liquid-to-solid phase transition.

III. RESULTS AND DISCUSSION

A. Chain branching

To study the effect of branching, we compare the behavior of an unbranched chain [the homopolymer in Fig. 1(a)] with that observed for a chain with separate backbone and side chain beads [the homoprotein in Fig. 1(b)]. Figures 3 and 4 show R_g^* , C_v^* , and E^* vs T^* for the homopolymer and the homoprotein, respectively. In this and in all thermodynamic data shown in this study the symbols represent the average equilibrium value from at least three independent simulations. Error bars are shown on all points and are the standard deviation in the measured values. However, the error bars are only visible when they exceed the size of the symbol. The dashed lines on the R_g^* plots are meant only to guide the eye between the points. The solid lines on the C_v^* and E^* plots are the result of a weighted histogram analysis wherein degeneracy factors are extracted from the results of runs at different temperatures and a partition function is obtained.⁴⁵ The error in the C_v^* measurement can be quite large, especially at low temperatures, since it is a measure of fluctuations about a simulation variable (energy). As a result, the error bars on the C_v^* data can be large and, at times, the histogram data lies outside these error bars. Consequently, we confirm the presence of a phase transition by comparing pairs of plots, either a sigmoidal collapse in the R_g^* vs T^* curve and a peak in C_v^* at the same temperature or a discontinuity in the E^* vs T^* curve and a peak in C_v^* at the same temperature, and by examining the chain structure at the given temperature.

The homopolymer model exhibits multiple phase transitions, as was seen previously for shorter chains.^{26,27} As shown in Fig. 3, at a high temperature (T^* of approximately 2.6 as measured at the midpoint of the transition), the homopolymer system undergoes a collapse, where the R_g^* drops rapidly and the C_v^* curve plateaus. At a lower temperature, T^* of 0.34, we observe a discontinuity in the E^* plot and a corresponding spike in the C_v^* plot, behavior characteristic of a first-order phase transition. As the temperature is

FIG. 3. Reduced-squared radius of gyration (R_g^*), reduced specific heat (C_v^*), and reduced energy (E^*) as a function of reduced temperature (T^*) for the homopolymer.

lowered further, the energy decreases gradually, and a spike in the specific heat at approximately $T^*=0.15$ suggests a third transition. However, considerable variability in C_v data at such low temperatures makes confirming a third transition difficult. The equilibrium structures of chains in this low-temperature regime will be discussed below. The homoprotein system (Fig. 4) displays almost exactly the same transition pattern as the homopolymer system, with a collapse transition at $T^*=2.6$ and a lower-temperature transition at 0.34. Below $T^*=0.34$ the energy continues to decrease, and an increase in C_v at a T^* of approximately 0.15 may indicate a third transition. Transition temperatures for all models are summarized in Table I. Since the only difference between these two models is the way in which the beads are connected (each system contains 80 square-well beads and 79 bonds), it is apparent that varying the connective arrangement of beads has little effect on the transition temperatures or on the strength (steepness) of each transition. These re-

FIG. 4. Reduced-squared radius of gyration (R_g^*), reduced specific heat (C_v^*), and reduced energy (E^*) as a function of reduced temperature (T^*) for the homoprotein.

sults are in agreement with a lattice study of branched polymers which shows that a polymer with short side chains (single beads) adopts a global shape that is similar to that of a linear chain.⁴⁶

Figures 5 and 6 are snapshots of typical conformations of the homopolymer and homoprotein systems at a T^* of 0.30. Backbone beads are dark gray, and side chain beads are light gray. Both figures show symmetric structures with cubic lattice geometry. When viewed from the side, the structures exhibit hexagonal lattice geometry. The mixed-lattice structure results from the square-well diameter of 1.5σ at which cubic and hexagonal lattice geometries are equally stable. At temperatures below approximately 0.15, the chains exhibit less-symmetrical structures, such as the one shown in Fig. 7 for the homoprotein at $T^*=0.125$, in which the simple lattice structure seen at $T^*=0.30$ is sacrificed for a more spherical, lower energy structure. This reorganization at very

TABLE I. Transition temperatures observed for each model studied.

Model	Transition temperatures (reduced units)
Homopolymer [Fig. 1(a)]	2.6,0.34,0.15 ^a
Homoprotein [Fig. 1(b)]	2.6,0.34,0.15 ^a
Heteroprotein [Fig. 1(c)]	0.55,0.19
Rigid homoprotein [Fig. 1(d)]	2.5,0.16,0.12 ^a
Rigid heteroprotein [Fig. 1(e)]	0.50,0.17
Overlapping rigid homoprotein [Fig. 1(f)]	3.2,0.25,0.15 ^a
Overlapping rigid heteroprotein [Fig. 1(g)]	0.65,0.17

^aRough estimates of temperature.

low temperatures, along with the peak in C_v at $T^*=0.15$, suggests a polymorphic solid–solid transition. Similar results were obtained previously with shorter homopolymer models.^{26,27}

The homoprotein model, in which we introduce branching to the homopolymer model, will be useful in future, more realistic protein models because of the increased geometric detail possible due to its discretized backbone and side chains. As shown by the results in this section, the introduction of branching did not disrupt the multiple thermodynamic phase transition behavior seen with homopolymers. This behavior is qualitatively characteristic of phase transitions during protein folding and, consequently, an important behavior in any protein model.

B. Chain heterogeneity

We study the effect of heterogeneity by comparing behavior of the homoprotein, Fig. 1(b), with that of the heteroprotein, Fig. 1(c). Figure 8 shows R_g^* , C_v^* , and E^* vs T^*

FIG. 5. A snapshot of the homopolymer chain at $T^*=0.30$. The square-well beads assume cubic and hexagonal symmetry at this temperature. Chain structures were visualized using the AVS software package from Advanced Visual Systems Inc.

FIG. 6. A snapshot of the homoprotein chain at $T^*=0.30$. Backbone beads are dark gray; side chain beads are pale gray. The square-well beads assume cubic and hexagonal symmetry at this temperature.

for the heteroprotein. While the same phase transitions are present as in the homoprotein system (Fig. 4), introducing heterogeneity causes both a downward temperature shift and a strengthening (steeper curve) of the individual transitions. The collapse occurs at T^* of 0.55, and a first-order transition occurs at approximately 0.19. (Compare with homoprotein results of $T^*=2.6$ and 0.34.) Additional low-temperature transitions for this model, if possible, exist below $T^*=0.1$, the lowest temperature studied here. The shift in transition temperatures from the homoprotein and the heteroprotein models is expected when examined in light of the results for a cluster (no bonds) of square-well beads.²⁷ A comparison of a 64 bead square-well chain with a 64 bead square-well cluster demonstrates that the lack of bonds in the cluster (a lack of physical constraints) leads to sharper, steeper curves, cor-

FIG. 7. A snapshot of the homoprotein chain at $T^*=0.125$. Backbone beads are dark gray; side chain beads are pale gray. The symmetric structure of Fig. 6 is sacrificed in favor of a lower-energy, nonsymmetric structure.

FIG. 8. Reduced-squared radius of gyration (R_g^*), reduced specific heat (C_v^*), and reduced energy (E^*) as a function of reduced temperature (T^*) for the heteroprotein.

responding to stronger physical transitions at lower temperatures. Since the cluster has considerably fewer constraints than the chain, it is necessary to subject it to a lower temperature to force a transition. In this study, the heteroprotein has fewer constraints than the homoprotein in that the beads experiencing the attractions that enable the phase transitions (the square-well side chains) are separated from each other by five bonds. The system can be thought of as 20 square-well beads that are distantly constrained to each other through loops of four hard-sphere beads. The collapse transition occurs via the side chain square-well interactions in spite of the fact that the side chains must drag the hard-sphere backbone along with them.

Figure 9 shows a symmetrical, low-temperature ($T^*=0.17$) conformation for the heteroprotein chain. Since the backbone beads are not square-well, they do not participate in ordering. The size of the backbone beads in Fig. 9 is

FIG. 9. A snapshot of the heteroprotein chain at $T^*=0.17$. Backbone beads are dark gray; side chain beads are pale gray. Backbone bead diameters are reduced to 25% and side chain bead diameters to 50% of their original sizes for ease of viewing the ordered arrangement of the side chains.

significantly reduced to highlight the hexagonal ordering of the side chain beads. As in the homopolymer and homoprotein systems, decreasing the temperature leads to even lower-energy structures (not shown) that do not exhibit lattice symmetry.

Heterogeneity, like branching, is introduced to this hierarchy of models since it will be beneficial in future, more realistic protein models. The addition of heterogeneity allows the study of hydrophobic core formation, an important aspect of protein folding. The low-temperature conformations observed, like the one shown in Fig. 9, readily adopt structures with the hydrophobic side chains clustered at the center and the nonhydrophobic backbone beads buffering them from solvent. However, in contrast to real protein native states, the observed structures are not unique.

C. Chain rigidity

We assess the effect of chain rigidity on phase transitions by comparing the results from the homoprotein model, Fig. 1(b) with those from the rigid homoprotein model, Fig. 1(d). Figure 10 shows R_g^* , C_v^* , and E^* vs T^* for the rigid homoprotein. The rigid homoprotein system displays multiple phase transitions with a collapse transition at a T^* of 2.5 and a first-order transition at a T^* of 0.16. A lower-temperature transition is suggested by the peak in the weighted histogram C_v data at a T^* of approximately 0.12; however there is too much variability in C_v data at such low temperatures to confirm this third transition. These transitions occur at lower temperatures than with the homoprotein model (compare with homoprotein results of $T^*=2.6$, 0.34, and 0.15), demonstrating that lower temperatures are required to force the rigid homoprotein chain into compact conformations. The extra bonds introduced in the rigid model effectively work against compaction by holding next-neighbor and, in some cases, next-next-neighbor beads apart. Although the rigid homoprotein can collapse to a similar overall R_g^* as the homoprotein, it is unable to make the

FIG. 10. Reduced-squared radius of gyration (R_g^*), reduced specific heat (C_v^*), and reduced energy (E^*) as a function of reduced temperature (T^*) for the rigid homoprotein.

subtle structural rearrangements necessary to achieve similar low energies. At the lowest temperature studied ($T^* = 0.1$), the homoprotein has an E^* of -5.25 and the rigid homoprotein has an E^* of -4.25 . This difference can be attributed to decreased conformational freedom due to the increased number of bonds in the chain.

We note that our downward shift in transition temperatures with increased model stiffness is contrary to the upward shift reported elsewhere for lattice⁴⁷ and off-lattice^{48–50} Monte Carlo simulations of stiff square-well and stiff Lennard-Jones homopolymer chains. The difference is most likely due to significant differences between the model chains studied. Other reports have focused on the phase transitions in homopolymer chains with *uniform* stiffness, where stiffness is introduced as an attraction between each bead, i , and its next-neighbor bead, $i+2$. In the models developed here, a nonuniform rigidity results from the three types of

FIG. 11. Nonuniform rigidity in models (d) through (g). Three four-bead monomer building blocks are shown; the central bead of each monomer is striped. Consecutive central beads along the backbone and the two intervening beads are fixed in a plane, as highlighted by the bold parallelogram on the left. All four beads of each monomer are linked in a pyramid shape, as shown with bold lines on the right. The fixed planes and pyramids propagate throughout the structure; one of each type is shown.

bonds used in models (d) through (g) to mimic constraints in proteins [Figs. 1(d)–1(g) and Fig. 2]. The first type of bond, a constraint between each bead, i , and its next-neighbor bead, $i+2$, generates a rigidity similar to that introduced in the other studies and causes the chain to maintain a zigzag shape. Figure 11 shows the geometric constraints resulting from the other two types of bonds. The fixed distance between consecutive central beads of each four-bead monomer building block causes that pair of beads and the two inter-

FIG. 12. Snapshots of the rigid homoprotein chain at T^* of 0.15, (a) and (b), and at T^* of 0.10, (c). Backbone beads are dark gray; side chain beads are pale gray.

FIG. 13. Reduced-squared radius of gyration (R_g^*), reduced specific heat (C_v^*), and reduced energy (E^*) as a function of reduced temperature (T^*) for the rigid heteroprotein.

vening backbone beads to remain fixed in a plane, as shown by the bold parallelogram in Fig. 11. In the final type of constraint, side chain beads are pinned to all three backbone beads of the monomer unit such that each monomer is fixed in a pyramid shape, as shown with bold lines on the right in Fig. 11. The consequence of this nonuniform rigidity is that rotational motion around different bonds is hindered to different degrees, and the hindered motion opposes ordering of the chain.

The structures observed for the rigid homopolymer model reflect the influence of the extra geometric constraints depicted in Fig. 1(d). With the homopolymer model, the collapse transition always results in spherically-shaped structures. However, in rigid homopolymer runs, collapse of the chain results in either spherical or ellipsoidal conformations, such as the ellipsoidal structure shown in Fig. 12(a) (front view) and 12(b) (side view) for a T^* of 0.15. While the packing

FIG. 14. Snapshot of the rigid heteroprotein chain at T^* of 0.16. Backbone beads are dark gray; side chain beads are pale gray. Backbone bead diameters are reduced to 25% and side chain bead diameters to 50% of their original sizes for ease of viewing the ordered arrangement of the side chains.

appears quite dense, lower-energy ordered structures, such as the one shown in Fig. 12(c) for a T^* of 0.10, are possible. The extra bonds present in the rigid homopolymer model prevent the chain from adopting the cubic and hexagonal packed structures like those observed for the homopolymer chain. Ellipsoidal structures similar to those seen here were shown previously to be characteristic of semiflexible polymers.^{50,51}

We also consider the effect of chain rigidity on phase transitions for heterogeneous chains by comparing the results from the heteroprotein model, Fig. 1(c) with those from the rigid heteroprotein model, Fig. 1(e). Figure 13 shows R_g^* , C_v^* , and E^* vs T^* for the rigid heteroprotein. The rigid heteroprotein system displays multiple phase transitions with a collapse transition at a T^* of 0.50 and a first-order transition at a T^* of 0.17. As in the comparison above for homogeneous chains, the transitions for the rigid heteroprotein system occur at lower temperatures than for the (nonrigid) heteroprotein system (compare with heteroprotein results of $T^* = 0.55$ and 0.19). Lower temperatures are required to force rigid chains into compact conformations. In contrast to the homogeneous results above, the rigid heteroprotein is able to achieve equally low energies as the heteroprotein (compare Figs. 8 and 13). This can be explained by the fact that the side chains alone account for the energy of the system. The extra bonds in the rigid heteroprotein model make overall compaction more difficult; however, once the side chains are clustered in the center of the structure, they are able to make the rearrangements necessary to achieve low energy states.

The structures for the rigid heteroprotein model, such as the one shown in Fig. 14 for $T^* = 0.16$, are notably similar to those observed for the heteroprotein (compare with Fig. 9). Again, it is the side chains that dictate structure in these heterogeneous models; the backbone beads are not involved in stabilizing equilibrium structures. Once the side chain

FIG. 15. Reduced-squared radius of gyration (R_g^*), reduced specific heat (C_v^*), and reduced energy (E^*) as a function of reduced temperature (T^*) for the overlapping rigid homoprotein.

beads have aggregated in the center of the conformation, the pseudobonds have little impact and do not prevent the formation of structures with local ordering of the side chains.

In a real amino acid residue, the backbone bond angles are fixed and the side chain is constrained to a particular location relative to the backbone atoms. These constraints severely limit the conformational freedom of the atoms in the amino acid residue and contribute to important features of folded protein structure, such as α -helices and β -sheets. A protein model that utilizes multiple backbone beads to represent a single amino acid residue should incorporate constraints that mimic the constraints in real amino acid residues. The rigid homoprotein and rigid heteroprotein models here include these important constraints in the form of extra bonds between next-neighbor beads and, in some cases, next-next-neighbor beads. The simulation results show that the rigid models [Figs. 1(d)–1(e)], despite their decreased flexibility when compared with the nonrigid models [Figs.

FIG. 16. Snapshots of the overlapping rigid homoprotein chain at T^* of 0.23, (a) and (b), and at T^* of 0.10, (c). Backbone beads are dark gray; side chain beads are pale gray.

1(b)–1(c)], experience multiple thermodynamic phase transitions which qualitatively correspond to the stages of protein folding and successfully bury hydrophobic side chains, two important characteristics of a protein model.

D. Overlapping beads

To study the effect of bead overlap, we perform simulations of the overlapping rigid homo- and heteroprotein models shown in Figs. 1(f) and 1(g), respectively. These models have bead sizes that are 1.5 times larger than in the previous five model systems, as depicted in Figs. 1 and 2. Figure 15 shows R_g^* , C_v^* , and E^* vs T^* for the overlapping rigid homoprotein. This system displays multiple phase transitions with a collapse transition at a T^* of approximately 3.2, a first-order transition at a T^* of 0.25, and a lower-temperature transition as suggested by the C_v data at T^* approximately equal to 0.15. These transitions occur at higher temperatures than for the rigid homoprotein model ($T^* = 2.5, 0.16$, and

FIG. 17. Reduced-squared radius of gyration (R_g^*), reduced specific heat (C_v^*), and reduced energy (E^*) as a function of reduced temperature (T^*) for the overlapping rigid heteroprotein.

0.12). The shift in transition temperatures and the ability to reach lower overall energies result directly from the greater range of attraction for each square-well bead in the overlapping rigid homopolymer model.

The structures observed for the overlapping rigid homopolymer model are similar to those for the rigid homopolymer model and reflect the influence of the next-neighbor and next-next-neighbor bonds depicted in Fig. 1(f). In some overlapping rigid homopolymer runs, collapse of the chain results in rod-shaped conformations, such as the one shown in Fig. 16(a) (side view) and 16(b) (top view) for a T^* of 0.23. Rod-like structures are shown elsewhere to be characteristic of semiflexible polymers.^{50,51} Other low-temperature structures, such as the one shown in Fig. 16(c) for a T^* of 0.10 tend to be ordered and spherical. The extra bonds present in the overlapping rigid homopolymer model hinder the flexibility of the chain and prevent it from adopting struc-

FIG. 18. Snapshots of the overlapping rigid heteroprotein chain at T^* of 0.16. Backbone beads are dark gray; side chain beads are pale gray. (a) Backbone bead diameters are reduced to 25% and side chain bead diameters to 50% of their original sizes for ease of viewing the ordered arrangement of the side chains. (b) Beads are full-size.

tures with cubic and hexagonal packing similar to those observed for the simple homopolymer chain.

Figure 17 shows R_g^* , C_v^* , and E^* vs T^* for the overlapping rigid heteroprotein model shown in Fig. 1(g). The overlapping rigid heteroprotein system displays multiple phase transitions with a collapse transition at a T^* of 0.65 and a first-order transition at a T^* of 0.17. The collapse occurs at a higher temperature than for the rigid heteroprotein system (compare to T^* of 0.50), which can again be attributed to the longer-range attraction of the square-well beads in the overlapping rigid heteroprotein. The low-temperature transition seems to occur at the same temperature as in the rigid heteroprotein system ($T^*=0.17$). This probably results from the small number of hydrophobic beads in these systems (20 side chains) and the heterogeneity of the chain. After the collapse transition, the side chains (pale beads) are clustered in the center of the structure and the backbone (dark beads) snakes around the outside. Subsequent, lower-temperature transitions involve rearrangements of these relatively uncoupled side chain beads. Once the side chains have been segregated and tightly packed in the core of the structure, small changes in the range of the attractive wells are relatively unimportant. In contrast, the comparison above between the rigid homopolymer and the overlapping rigid homopolymer showed a shift in the low-temperature transition. In the homogeneous systems, all beads have attractive wells and small changes in structure can have a dramatic effect on the overall energy depending on the range of the attractive well.

We observe similar side chain packing geometries for all heteroprotein models. As with the heteroprotein and rigid heteroprotein models, the overlapping rigid heteroprotein model displays symmetrical side chain conformations, such as the one shown in Fig. 18 for a T^* of 0.16. Figure 18(a) shows the chain with reduced bead diameters to highlight the symmetrical arrangement of side chain beads. Figure 18(b) shows the same structure from a different angle with full-size beads to demonstrate the way in which the backbone snakes

along the outer surface of the structure, encasing the hydrophobic side chains.

Real protein atoms have significant overlaps. Space-filling pictures of proteins at the atomic level look more like the overlapping models introduced in this section than the five tangent-bead models discussed previously. To properly mimic real protein steric interactions, appropriate bead size to bond length ratios should be chosen when constructing a protein model. The results shown here indicate that overlapping chains are capable of displaying protein-like character. The overlapping rigid models are able to selectively bury hydrophobic side chains and experience multiple thermodynamic phase transitions which qualitatively correspond to the stages of protein folding.

IV. CONCLUSIONS

We have presented results from a series of simulations with different protein models in which we build from a simple, tangent, homopolymer chain to an intermediate resolution model containing several important modifications: branched geometry, heterogeneity, backbone rigidity, and overlapping beads. These four modifications provide a significant amount of physical protein-like character that is lacking in common tangent linear chain models. Although still simple representations of a protein, the models introduced in this work display some degree of protein-like behavior in that the multiple transitions observed qualitatively correspond to structurally significant changes in the protein during folding.

We are currently investigating other modifications to the heteroprotein model, including the effect of a hydrogen bonding term and the competition between backbone-backbone hydrogen bond formation and side chain-side chain hydrophobic interactions. We find that the simple features introduced in this study, in concert with specific hydrogen bonding and hydrophobic potentials, provide a significant amount of protein-like character.³⁷ The resulting model may serve as the basis for studies of protein aggregation.

ACKNOWLEDGMENTS

We would like to thank Dr. Yaoqi Zhou for helpful discussions and for providing the code for the weighted histogram analysis of simulation results. This work was supported by the GAANN Biotechnology Fellowship of the U.S. Department of Education. Partial funding was also provided by the National Science Foundation under Grant No. CTS-9704044 and the National Institutes of Health under Grant No. GM-56766.

¹W. Wei, *Int. J. Pharm.* **185**, 129 (1999).

²J. King, *Chem. Eng. News*, April 10, 52 (1989).

³E. DeBernardis-Clark and G. Georgiou, *ACS Symp. Ser.* **470**, 1 (1991).

⁴H. R. Costantino, R. Langer, and A. M. Klibanov, *J. Pharm. Sci.* **83**, 1662 (1994).

- ⁵S. P. Schwendeman, H. R. Costantino, R. K. Gupta, G. R. Siber, A. M. Klibanov, and R. Langer, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11234 (1995).
- ⁶D. J. Selkoe, *J. NIH Res.* **7**, 57 (1995).
- ⁷M. D. Perng, *J. Biol. Chem.* **274**, 33235 (1999).
- ⁸K. A. Dill, S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan, *Protein Sci.* **4**, 561 (1995).
- ⁹K. A. Dill and H. S. Chan, *Nat. Struct. Biol.* **4**, 10 (1997).
- ¹⁰C. Wilson and S. Doniach, *Proteins: Struct., Funct., Genet.* **6**, 193 (1989).
- ¹¹N. D. Socci and J. N. Onuchic, *J. Chem. Phys.* **103**, 4732 (1995).
- ¹²D. Thirumalai and D. K. Klimov, *Curr. Opin. Struct. Biol.* **9**, 197 (1999).
- ¹³E. I. Shakhnovich, *Folding Des.* **1**, R50 (1996).
- ¹⁴A. R. Dinner and M. Karplus, *J. Phys. Chem. B* **103**, 7976 (1999).
- ¹⁵A. Irbach, C. Peterson, F. Potthast, and O. Sommerli, *J. Chem. Phys.* **107**, 273 (1997).
- ¹⁶A. Sali, E. Shakhnovich, and M. Karplus, *Nature (London)* **369**, 248 (1994).
- ¹⁷S. He and H. A. Scheraga, *J. Chem. Phys.* **108**, 287 (1998).
- ¹⁸J. N. Onuchic, N. D. Socci, Z. Luthey-Schulten, and P. G. Wolynes, *Folding Des.* **1**, 441 (1996).
- ¹⁹T. Veitshans, D. Klimov, and D. Thirumalai, *Folding Des.* **2**, 1 (1997).
- ²⁰S. Doniach, T. Garel, and H. Orland, *J. Chem. Phys.* **105**, 1601 (1996).
- ²¹Y. Zhou and M. Karplus, *Nature (London)* **401**, 400 (1999).
- ²²Y. Zhou and M. Karplus, *J. Mol. Biol.* **293**, 917 (1999).
- ²³K. F. Lau and K. A. Dill, *Macromolecules* **22**, 3986 (1989).
- ²⁴E. I. Shakhnovich and A. M. Gutin, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7195 (1993).
- ²⁵N. D. Socci and J. N. Onuchic, *J. Chem. Phys.* **101**, 1519 (1994).
- ²⁶Y. Zhou, C. K. Hall, and M. Karplus, *Phys. Rev. Lett.* **77**, 2822 (1996).
- ²⁷Y. Zhou, M. Karplus, J. M. Wichert, and C. K. Hall, *J. Chem. Phys.* **107**, 10691 (1997).
- ²⁸W. Hu, *J. Chem. Phys.* **109**, 3686 (1998).
- ²⁹C. Wu and X. Wang, *Phys. Rev. Lett.* **80**, 4092 (1998).
- ³⁰C. Wu and S. Zhou, *Phys. Rev. Lett.* **77**, 3053 (1996).
- ³¹D. Bratko, A. K. Chakraborty, and E. I. Shakhnovich, *J. Chem. Phys.* **106**, 1264 (1997).
- ³²C. J. Camacho and D. Thirumalai, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6369 (1993).
- ³³H. Nymeyer, A. E. Garcia, and J. N. Onuchic, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5921 (1998).
- ³⁴V. P. Zhdanov and B. Kasemo, *Proteins: Struct., Funct., Genet.* **29**, 508 (1997).
- ³⁵G. Iori, E. Marinari, G. Parisi, and M. V. Struglia, *Physica A* **185**, 98 (1992).
- ³⁶O. B. Ptitsyn, *Adv. Protein Chem.* **47**, 83 (1995).
- ³⁷A. V. Smith and C. K. Hall, *Proteins: Struct., Funct., Genet.* (submitted).
- ³⁸B. J. Alder and T. E. Wainwright, *J. Chem. Phys.* **31**, 459 (1959).
- ³⁹D. C. Rapaport, *J. Phys. A* **11**, L213 (1978).
- ⁴⁰A. Bellemans, J. Orban, and D. Van Belle, *Mol. Phys.* **39**, 781 (1980).
- ⁴¹S. W. Smith, C. K. Hall, and B. D. Freeman, *J. Comput. Phys.* **134**, 16 (1997).
- ⁴²S. Sun, P. D. Thomas, and K. A. Dill, *Protein Eng.* **8**, 769 (1995).
- ⁴³S. Sun, *Protein Sci.* **2**, 762 (1993).
- ⁴⁴S. Takada, Z. Luthey-Schulten, and P. G. Wolynes, *J. Chem. Phys.* **110**, 11616 (1999).
- ⁴⁵A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **63**, 1195 (1989).
- ⁴⁶K. Shiokawa, K. Itoh, and N. Nemoto, *J. Chem. Phys.* **111**, 8165 (1999).
- ⁴⁷J. P. K. Doye, R. P. Sear, and D. Frenkel, *J. Chem. Phys.* **108**, 2134 (1998).
- ⁴⁸Yu. A. Kuznetsov and E. G. Timoshenko, *J. Chem. Phys.* **111**, 3744 (1999).
- ⁴⁹Yu. A. Kuznetsov, E. G. Timoshenko, and K. A. Dawson, *J. Chem. Phys.* **105**, 7116 (1996).
- ⁵⁰H. Noguchi and K. Yoshikawa, *J. Chem. Phys.* **109**, 5070 (1998).
- ⁵¹E. G. Timoshenko, Yu. A. Kuznetsov, and K. A. Dawson, *Phys. Rev. E* **57**, 6801 (1998).