Modeling self-assembling of proteins: Assembled structures, relaxation dynamics, and phase coexistence

Benjamin Vekhter and R. Stephen Berry[a)]
Department of Chemistry, The University of Chicago, Chicago, Illinois 60637

(Received 24 July 1998; accepted 22 October 1998)

A string-of-beads model used previously to describe folding of a polypeptide into a β-barrel is transformed into four-strand model of a self-assembling system, which also produces a β-barrel. In molecular dynamics (MD) simulations, both isothermal and variable-temperature (annealing), the system behaves much like a typical small cluster, insofar as it exhibits the dynamic coexistence of several phase-like forms over ranges of temperature. A three-state analytic model, then a four-state model, supplemented by degeneracies inferred from the MD simulations, yield partition functions and phase diagrams that reproduce the simulations rather well. © *1999 American Institute of Physics.* [S0021-9606(99)70104-2]

INTRODUCTION

Intense effort has gone and continues to go toward understanding protein folding, i.e., on finding out how the parts of a long polypeptide chain, often located rather far apart in the sequence along the chain, find each other and form a biologically active, three-dimensional structure predetermined by that sequence. At the same time, producing and understanding the self-assembly of free, complex molecules into organized structures such as monolayers of parallel chains has also become a vigorous field. Here, with a model system as the vehicle, we investigate the propensity of chains to self-assemble when those chains are idealized fragments composed of amino acid sequences. The model is indeed highly idealized: it is based on a chain-of-beads representation of β-barrel formation, first introduced as a lattice model,[1,2] then as a continuum model[3,4] and, in the latter form, as a device to investigate how topography induces folding.[5] The questions we address are these: do free strands self-assemble in the same way that they fit themselves together if those same strands are linked into a long chain? If the answer is ''yes,'' what are the dynamics and thermodynamics of the assembling process, and to what degree do the assembled structures resemble those of the fully bonded chain?

To answer these questions we have performed molecular dynamics simulations using the three-color 46-bead model just cited, with a modification. We have cut the four strands of the model apart between pairs of ''neutral'' beads, i.e., we have broken the links in the very flexible sections to make four rather stiff, isolated strands. Thus, apart from the ''neutral'' beads at the ends, two strands, 1–11 and 23–34, contain only hydrophobic beads and two others, 12–22 and 35–46, alternating hydrophobic and hydrophilic beads, which are abbreviated ''*B*'' and ''*L*,'' respectively. Apart from detaching the four segments and confining them within a spheric volume, the model is the same as that used by Ball *et al.*, and almost the same as that of Guo and Thirumalai. (The difference here is in the stiff bead–bead stretching force constant, which does not affect either folding or self-assembly dynamics.)

The goals of the simulations were: (a) determining what conditions, if any, would lead to self-assembly of the strands into a β-barrel structure or any other structure; (b) finding what kinds of intermediate structure, if any, the system would take on in the process of self-assembly, (c) determining from time histories what kind of phase equilibrium behavior the system would exhibit at low, intermediate and high temperatures; and (d) how the system would relax under a temperature-controlled annealing program. The original 46-bead model is somewhat artificial, first in being a bead model at all and second, in the set of force constants which are chosen because they induce formation of the β-barrel structure, rather than because they mirror known interatomic forces. This investigation thus addresses the general question of whether and how a set of strands, known to form a specific structure when linked together, form a similar structure under conditions of free self-assembly. Anticipating the results, we see that the self-assembling fragments indeed reach the same kind of structure as the linked segments, and even find minima more stable than those accessible to the linked system.

THE SIMULATIONS

The investigation is based entirely on molecular dynamics (MD) simulations. The algorithm is the ''velocity Verlet'' method for solving the classical Newtonian equations of motion for the chains of beads.[6,7] The beads interact only through pairwise, two-body forces. The forces between bonded pairs of beads are harmonic; the forces between nonbonded pairs are those adopted by Guo *et al.*,[3,4] and used by Ball *et al.*[5] Time scale has been scaled: if the masses of all beads are set to $m=40$ a.u., one time step corresponds to 10^{-14} s; the scale increases as $1/\sqrt{m}$. Typical runs took from 2 to 10×10^{6} time steps. Some of these were done isothermally, with the stochastic (Kast) algorithm.[8,9] Others simulated annealing processes; these were isothermal (for rather

a)Electronic mail: berry@rainbow.uchicago.edu

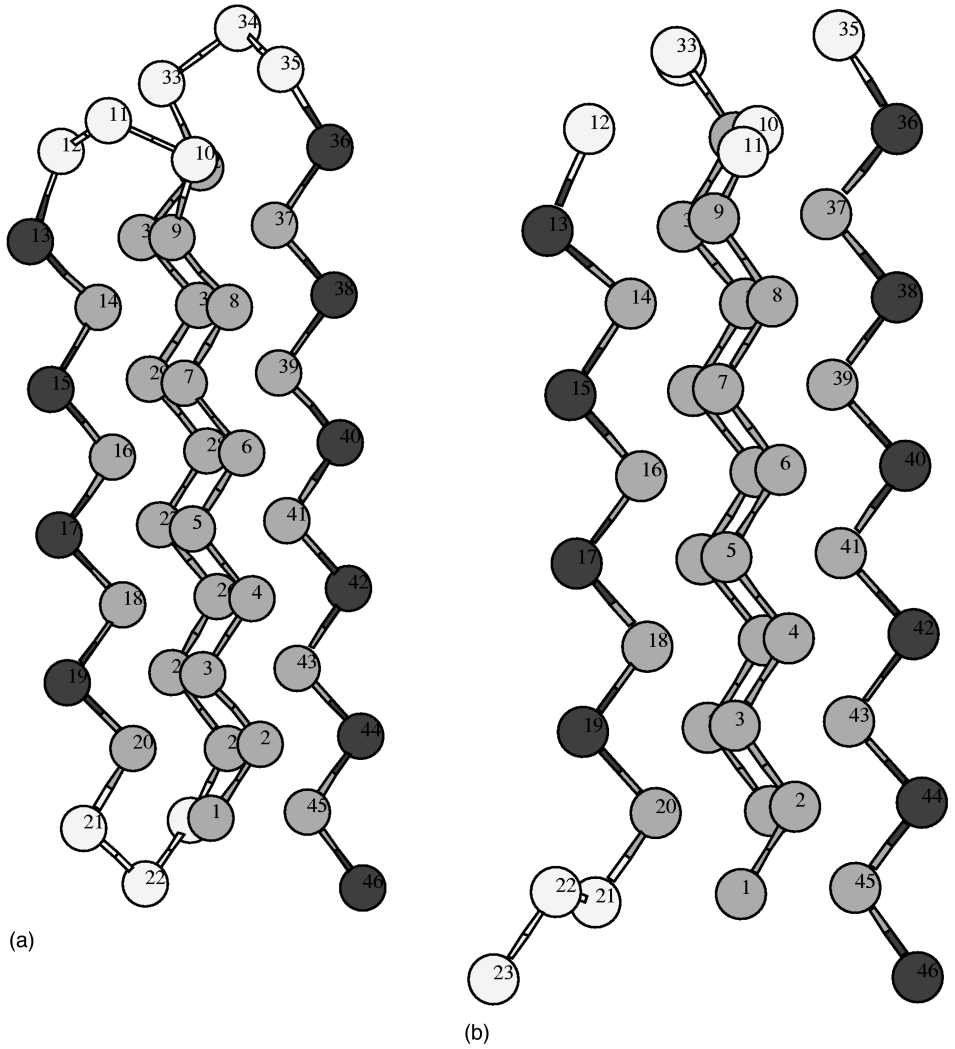

FIG. 1. Structure of low-energy configurations of the four-strand, 46-bead model: (a) the global minimum configuration of the fully linked chain, energy $E=-0.536$ eV; (b) the structure of the configuration of the free-strand model with energy $E=-0.580$; (c) the free-strand configuration with $E=-0.551$, illustrating by comparison with (b), an $n/2$ rotation of one B strand around another B strand; (d) the free-strand configuration with $E=-0.531$ illustrating, by comparison with (b), the sliding of hydrophobic (B) and alternating hydrophobic–hydrophobic (L) strands.

high temperatures) for some initial time intervals, then the temperature was monotonically dropped and again held constant for another lower-temperature isothermal interval for equilibration.

We started the MD isothermal and annealing runs from high-energy configurations with four separated strands. At low temperatures, assembling is usually very slow because the attractions between distant strands are rather weak, making accelerations slow; since fusion happens only when strands are near one another, it takes long times for strands to come close enough to join. On the other hand, at high temperatures, because the forces between nonbonded beads are weak, the strands remain close only briefly, separating from each other without fusing into composite structures. This can be interpreted as a consequence of the high entropy contribution to the free energy at high temperatures, which makes the separated strands free to explore all the available volume. The free-strand form, at high temperatures, has a free energy much lower than that of any compound configuration. To minimize this free-volume entropy effect, we have performed simulations within a confined spherical volume of radius R, $40<R<100$ (here we use the length units in which the bead–bead distance along the chain is equal to 2), as well as other simulations with the volume unrestricted. These have been carried out both for isothermal and variable-temperature, annealing runs. No shapes other than spherical were used, but the volume was shown to be large enough that the only influence of the radius was a simple shift in the temperature scale of aggregation.

Most of the runs, both isothermal (for temperatures T lower than 45 K) and annealing (from $T\sim100$ K to $T<45$ K) have terminated with structures whose energies are very close to that of the global minimum of the 46-bead chain polymer (Fig. 1). The energies of the deepest four-strand structures (~-0.58 eV) are lower than the global minimum (-0.53 eV) of the intact 46-bead chain. The main reason is that, in contrast to the fully connected chain, the separated strands need not ''spend'' energy to twist and fold in the bending sections. The full, 46-bead model requires some (small amount of) energy to bring the dihedral angles

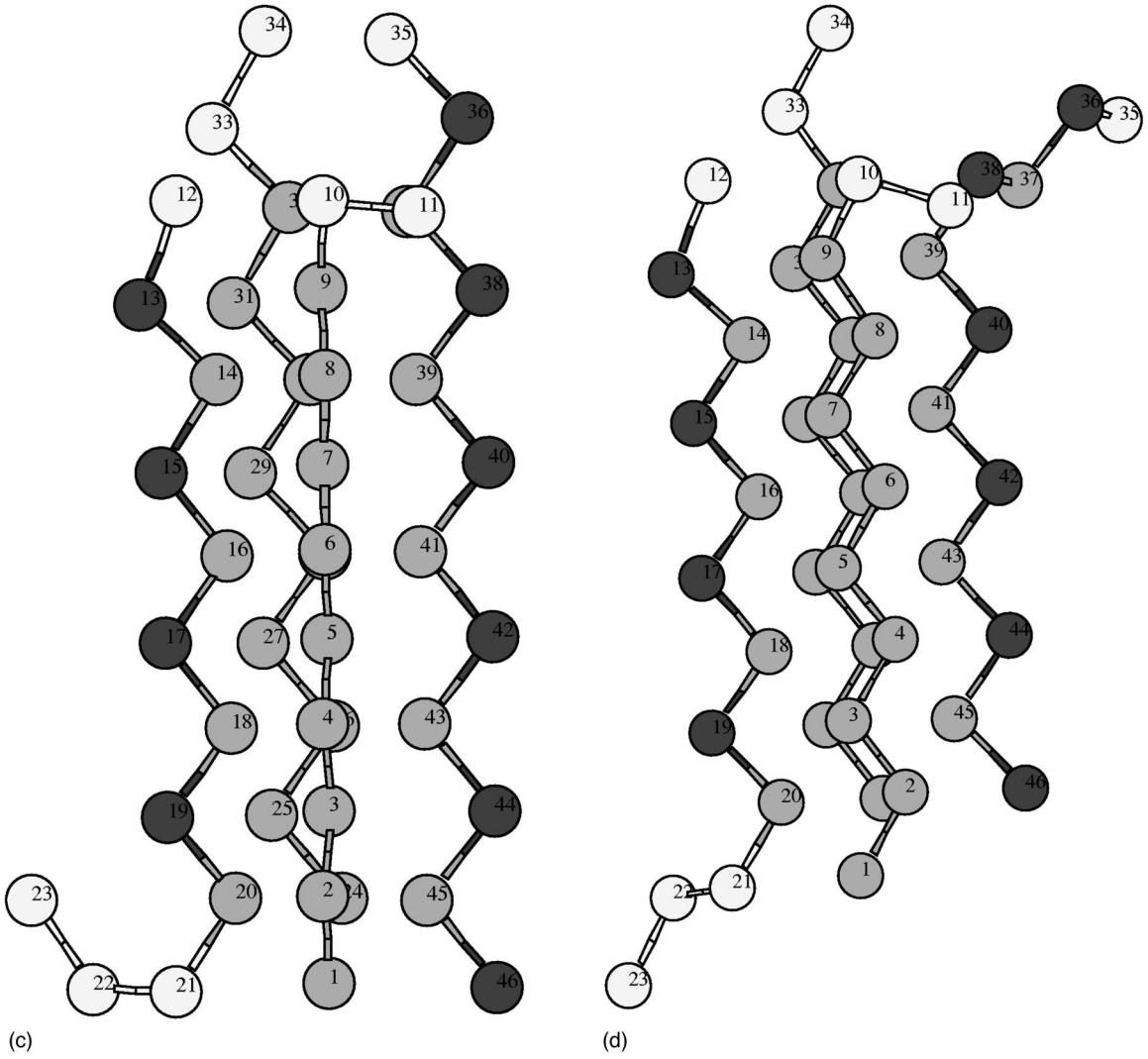

FIG. 1. (*Continued.*)

of the ''neutral'' sections into the hairpin conformations requisite for making the β-barrel.

The four-strand configurations follow the same general kinds of pathways as the fully linked chain does, in passing from one minimum to another: the favored motions to accomplish transitions between different low-energy structures are rotations and slidings of strands. The lower the temperature (in an isothermal relaxation) or the faster the annealing is (in a variable-T simulation) the higher in energy and farther from the global minimum structure the final relaxed configurations are (Fig. 2). This is much like what was found for relaxation of the 46-bead chain. As for the dynamics of assembling strands, it takes a much longer time for separated strands to find each other than for an unfolded chain to collapse to a compact three-dimensional structure. The reason is that the chain (even unfolded) necessarily keeps its strands close. Usually the two *B* strands—entirely hydrophobic strands—find each another first, after which the mixed *BL* (hydrophobe and hydrophil) strands join them one by one.

MD runs show that at low and high temperatures the system exists in a single form (state) only, in the fully condensed form with all four strands together for $T<T_1=50\,\mathrm{K}$, or with all four strands separated for $T>T_h=130\,\mathrm{K}$. In the intermediate temperature range, $T_1<T<T_h$, there is a nonzero probability for finding the system in any of two or even three different forms. This phase-coexistence behavior for finite systems has been well known for clusters of atoms and simple molecules of virtually every kind, among them clusters bound by Lennard-Jones, Morse and Born–Mayer (ionic) interactions among atoms, and sums of atom–atom interactions for polyhedral molecules. These are all, by comparison with the three-color polymer bead model, rather simple in the sense that the building elements are spheres or simple polyhedra. Results presented in Fig. 3 illustrate that the much more complicated system of four strands, each containing more than ten different beads, possesses the same phase-coexistence behavior as the clusters, namely a finite band of temperature (at a fixed pressure or volume) within which phase-like forms are in dynamic equilibrium.

Figure 4 shows the temperature dependence of the coexistence of three phase-like forms of the four strands: the completely packed (4), the $(3+1)$, and the $(2+1+1)$

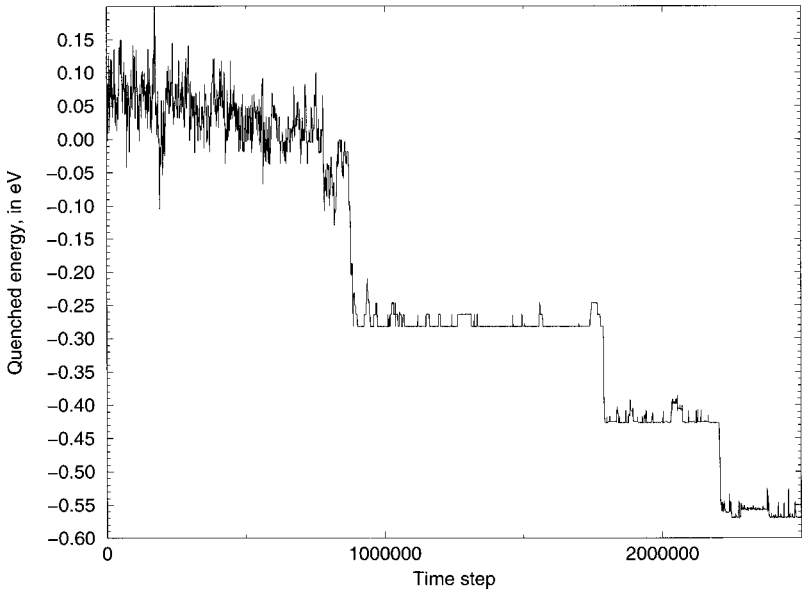

FIG. 2. Examples of time-resolved relaxation dynamics with annealing from $T=100$ to $T=40$ on the time interval 0–1 200 000 time steps: states with energies below -0.43, between -0.43 and -0.27, and between -0.27 and -0.12 eV, corresponding, respectively, to the fully fused (4), the $(3+1)$, and the $(2+1+1)$ structures.

phases, obtained from constant-temperature simulations. To describe this behavior analytically we use the same three-level model we applied previously to the Ar_{55} cluster.[10] This model supposes that the system has only three energy levels, one corresponding to each of the forms. Their degeneracies are $g(3)$, $g(2)$, and $g(1)$, respectively. Then the occupancy of each level, i.e., of each form at equilibrium is

$$n(i)=g(i)e^{-E(i)/k_BT}/Z, \quad Z=\sum_i g(i)e^{-E(i)/k_BT}. \tag{1}$$

Equation (1) can be rewritten as

$$\begin{aligned} n_1&=1/Z,\\ n_2&=xe^{-\Delta E_{21}/k_BT}/Z,\\ n_3&=xye^{-\Delta E_{32}/k_BT}/Z, \end{aligned} \tag{2}$$

where

$$x=g(2)/g(1), \quad y=g(3)/g(2), \quad \Delta E_{ij}=E_i-E_j,$$

$$Z=1+xe^{-\Delta E_{21}/k_BT}+xye^{-\Delta E_{32}/k_BT}.$$

If we relate $E(i)$ to the sharp boundaries found in MD runs for intervals of stable existence of each form (see Fig. 3), we find $E(1)=-0.58\,\text{eV}$, $E(2)=-0.43\,\text{eV}$, and $E(3)=-0.28\,\text{eV}$. Long enough MD runs provide us with the ratio $n(1):n(2):n(3)$ also, which enables us to estimate the ratios of the degeneracies $g(i)$,

$$\begin{aligned} g(2)/g(1)&=[n(2)/n(1)]e^{-\Delta E_{21}/k_BT},\\ g(3)/g(2)&=[n(3)/n(2)]e^{-\Delta E_{32}/k_BT}. \end{aligned} \tag{3}$$

From MD runs for $T=58\,\text{K}$ and $R=60$ we found $n(1):n(2):n(3)=0.21:0.57:0.22$, which gives $g(2)/g(1)=e^{31}$ and $g(3)/g(2)=e^{29}$. Figure 4 shows the analytical dependence of $n(i;T)$ given by Eqs. (2) and the ''experimental'' behavior from MD runs; one can see that they agree

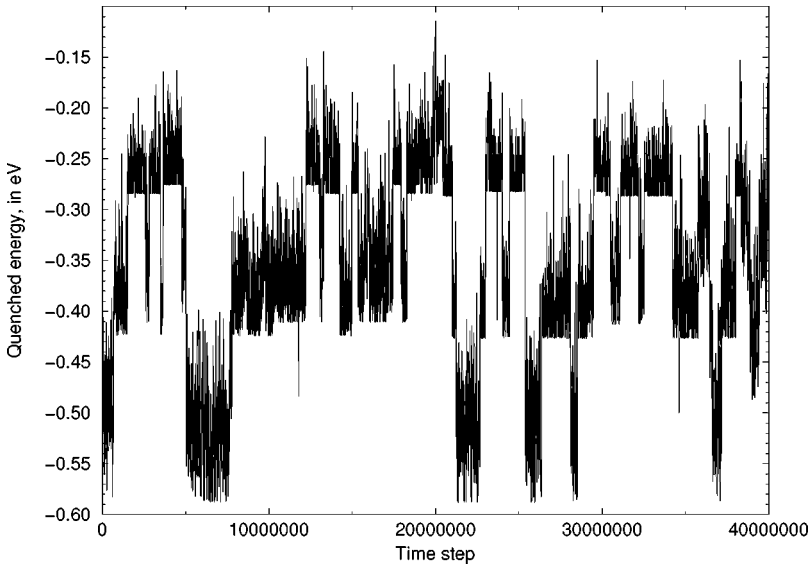

FIG. 3. A typical time history of quenched energies in molecular dynamics runs, illustrating phase coexistence behavior at $T=58$ for $R=75$.

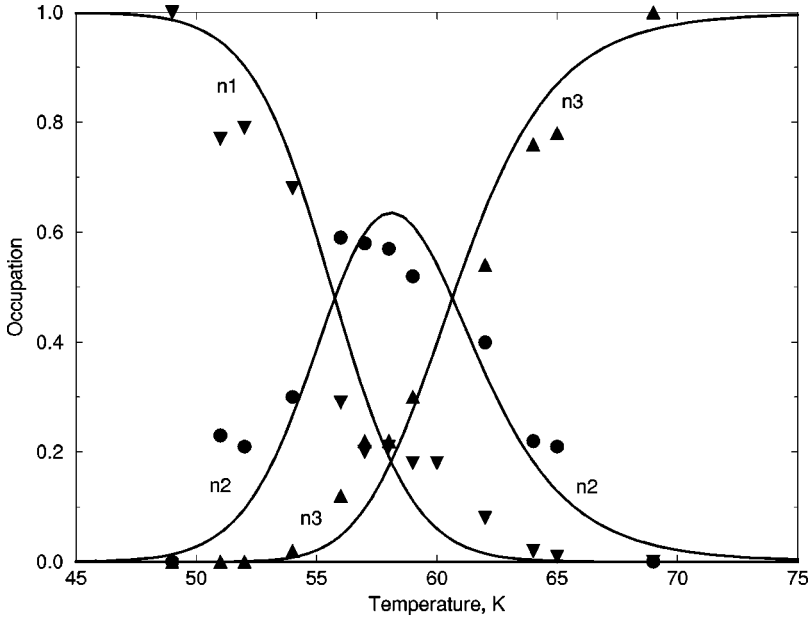

FIG. 4. A phase diagram for a confining volume with radius $R=60$, illustrating three-form coexistence: results of MD-''experiments'': ▼, ●, and ▲ are for $n(1)$, $n(2)$, and $n(3)$ respectively and the three-level model (solid lines) for $\ln[g(2)/g(1)]=31$, $\ln[g(4)/g(3)]=28.5$.

rather well, and the ratio of degeneracies $g(i)$ that fits the simulation results is very close to that inferred from the three-state model.

The values of $g(i)$ reflect the number of allowed states (or volume of available phase space) for each form. In a more elaborate model, $g(i)$ would be replaced by a partition function for all the states of the translations and the internal rotational and vibrational degrees of freedom which, in the simple three-state model, are lumped together as if all these states associated with each level had the same energy. In the $4\leftrightarrow(3+1)$ and $(3+1)\leftrightarrow(2+1+1)$ transitions, the numbers of all three vibrational, rotational and translational degrees of freedom change. Among these, it is reasonable to assume that the states related to translational degrees of freedom are the most drastically changed. Each time a strand is torn off the compound structure, three more translational degrees of freedom appear, generating a free center of mass which can explore the whole allowed volume. As the largest contribution to the partition function under most ordinary conditions, the translational degrees of freedom are the most important contribution to the ratio of degeneracies (number of allowed states). The change from $g(i)$ to $g(i+1)$ reflects the addition of three translational degrees of freedom. Hence the ratios of $g(2)/g(1)$ and $g(3)/g(2)$ should be similar, and the results of the simulations show that they indeed are. From that we may conclude that the ratio of degeneracies of $(1+1+1+1)$ and $(2+1+1)$ forms, $g(4)/g(3)$, should have a similar value of order e^{29} also. This assumption gives us the opportunity to predict the temperature range of coexistence of phases 3 and 4, and thus to construct the total phase diagram of the self-assembling four-strand species (Fig. 5) which is in good agreement with the results of high-temperature MD runs. The assumption that $g(2)$ is proportional to the allowed volume V that $g(3)\sim V^2$ and $g(4)$

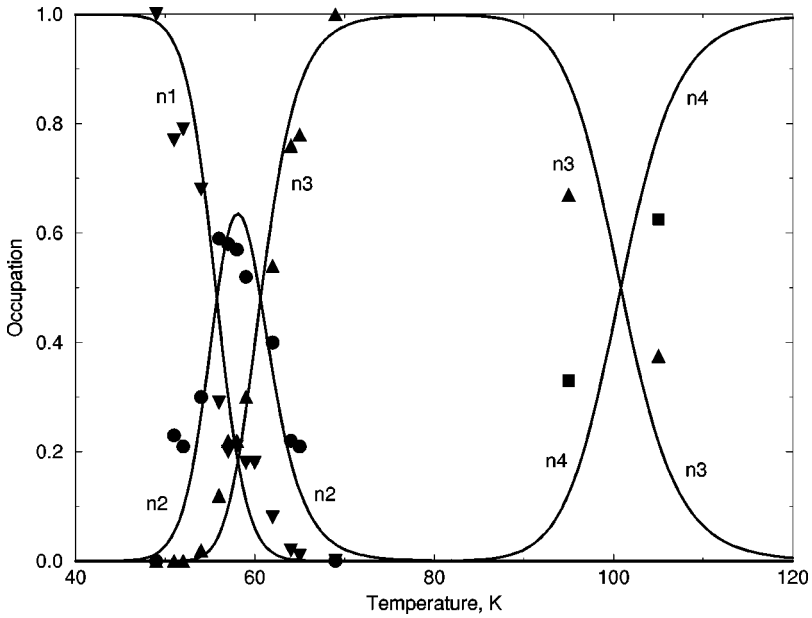

FIG. 5. A phase diagram for $R=60$ that includes the high-temperature region of coexistence of phase 3 and 4, and the results of a four-level model; the notations as in Fig. 4 with ■ for MD $n(4)$; the model uses the same degeneracy ratios as for Fig. 4 with the additional assumption that $\ln[g(4)/g(3)]=31$.

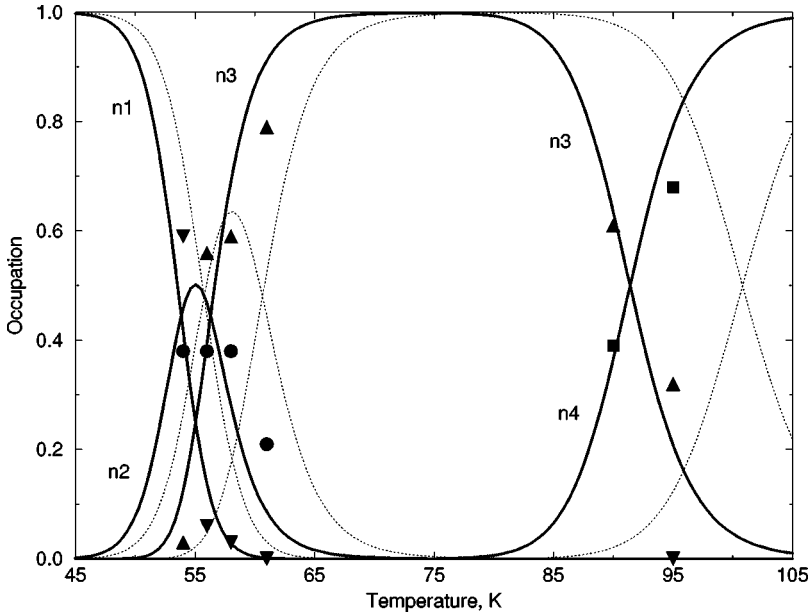

FIG. 6. A phase diagram for $R=90$; the dotted lines are the model's results for $R=60$ illustrating the shift of the phase diagram to higher temperatures when the free volume decreases and, as a result, the free energy of the disassembled structures increases.

$\sim V^3$ (see below) allows us to predict the phase diagram for various values of the radius of the confining volume; for example we have computed this diagram for $R=90$ (Fig. 6). (Two of the moments of inertia of the strands are large, so that the next most important contributions to the partition functions, after translation, are the rotational degrees set free by dissociation of a strand.)

The same three-state model can be used to describe the kinetics of self-assembly by solving corresponding kinetic equations

$$dn(i)/dt=\sum_j\,[-k_{ij}n(i)+k_{ji}n(j)], \tag{4}$$

where k_{ij} is the probability of an $i\rightarrow j$ transition. In the region $45<T<70\,\mathrm{K}$ we have observed coexistence of only the compound forms $(3+1)$ and $(2+1+1)$. It means that in Eq. (4) $i,j\leqslant 3$. We have never observed any direct $1\leftrightarrow 3$ transitions. Taking this into account also, i.e., supposing that $k_{13}=k_{31}=0$, we can rewrite Eq. (4) as

$$\begin{aligned} dn(1)/dt&=-k_{12}n(1)+k_{21}n(2),\\ dn(2)/dt&=-n(2)(k_{21}+k_{23})+k_{12}n(1)+k_{32}n(3),\\ dn(3)/dt&=-k_{32}n(3)+k_{23}n(2). \end{aligned} \tag{5}$$

The sum of $n(1)+n(2)+n(3)$ must be constant and can be normalized to unity. [Note that the numbers 1, 2 and 3 are only indices of the three states, and that $n(j)$ is the number of monomer strands in state n.] The transition probabilities k_{ij} determine the kinetics of self-assembly and the equilibrium values of the $n(i)$. At equilibrium $dn(i)/dt=0$, so that equilibrium implies that

$$n(1)=1/\zeta,\quad n(2)=\xi/\zeta,\quad n(3)=\xi\eta/\zeta, \tag{6}$$

where

$$\xi=k_{12}/k_{21},\quad \eta=k_{23}/k_{32},\quad \zeta=1+\xi+\xi\eta. \tag{7}$$

[Note that Eqs. (6) look very similar to Eqs. (2).] Hence if we find values of $n(i)$ from an isothermal run at temperature T, we can calculate values of k_{ij} at that T.

The dynamics (kinetics) of self-assembly depends of course on R, the value of the radius of the containing volume, so we performed MD simulations for various values of R. As one expects, the larger the allowed volume, the lower the temperature at which each of the separated forms appears, in full correspondence with the aforementioned volume dependence of the entropy contribution to the free energy. Because the algorithm we used for MD runs allows us to evaluate vibrational frequencies we can, in principle, calculate the free energies of the completely assembled structure with all four strands together and of the remaining three, partly and totally dissociated structures $(3+1)$, $(2+1+1)$ and $(1+1+1+1)$. However all we would actually like to know are the relative changes of free energy with changes of volume, and not the actual values of the free energy. Hence we may start with a simpler and easier approach. We make the natural assumption that the probabilities of the fusion transitions: $(2+1+1)\rightarrow(3+1)$ and $(3+1)\rightarrow(4)$ are inversely proportional to the allowed volume, while the probability of a decay process is independent of V, at least to some low order. This implies that for any two volumes V_1 and V_2:

$$\begin{aligned} k_{12}(V_1)/k_{12}(V_2)&=k_{23}(V_1)/k_{23}(V_2)=1,\\ k_{21}(V_1)/k_{21}(V_2)&=k_{32}(V_1)/k_{32}(V_2)=V_2/V_1, \end{aligned} \tag{8}$$

and that

$$\xi(V_2)=\xi(V_1)Q,\quad \eta(V_2)=\eta(V_1)Q, \tag{9}$$

where $Q=V_2/V_1$. From Eqs. (7)–(9) one finds

$$\begin{aligned} n(1;V_2)&=1/\zeta(V_2),\\ n(2;V_2)&=Q\xi(V_1)/\zeta(V_2),\\ n(3,V_2)&=Q^2\xi(V_1)\,\eta(V_1)/\zeta(V_2),\\ \zeta(V_2)&=1+Q\xi(V_1)+Q^2\xi(V_1)\,\eta(V_1). \end{aligned} \tag{10}$$

Equations (10) allow us to predict, without any adjustable parameter, values of $n(i,V_2)$ at volume V_2 from a knowledge of $n(i,V_1)$, i.e., they describe the volume dependence of occupations $n(i)$ at constant temperature.

From MD run at $T=58$ K we have estimated $n(i)$ for $R=75$ as 0.09:0.47:0.44. Using these values we have found from Eq. (10) that the ''theoretical'' ratios of $n(i)$ for $R=60$ and 90 are 0.20:0.54:0.26 and 0.04:0.37:0.59. These values are close to the ''experimental'' values from MD runs: 0.21:0.57:0.22 and 0.03:0.38:0.59.

There can be several reasons for the remaining differences. First of all it can be that MD runs were not long enough to provide accurate equilibrium values of $n(i)$. However because our runs have been rather long (on the order of 400 ns) we believe that this is not likely to be the source of the remaining differences, at least for a system of this small size. There is some uncertainty in the definition of free volume. We have not used absolutely hard spheres or a hard outer wall in this model so the effective volume of each bead is slightly more ambiguous than it would be for a hard wall. Specifically, if a bead reaches a point at a distance greater than R from the center of the confining volume, then it is subject to a steep exponential repulsion returning it into the sphere. Furthermore the excluded volume of the strands should strictly be subtracted from the sphere volume in determining the real available volume. However, as these two effects contribute opposite changes to the computed available volume, it is likely that at the large values of R that we used, their net effect must be rather small. It can be that the very initial assumption, namely that $k_{ij}\sim V^{-1}$ if $i>j$, is not accurate enough. It is based on the assumption that fusion may happen at any point inside the sphere with the same probability. However it may be that fusion happens more frequently near the sphere boundary when the velocities of colliding strands are small. The probability of two-strand collisions at the surface varies as $\sim R^{-4}$. It is possible also to explore a more general dependence of k_{ij} on the volume, e.g., as

$$k_{32}(V_2)/k_{32}(V_1)=\alpha(R_1/R_2)^m$$

with two free parameters: α and m. These parameters can be found from an MD run at a second volume, say with $R=60$, and used to predict $n(i)$ for $R=90$.

SUMMARY

A four-strand bead model for a self-assembling system forming a β-barrel shows the dynamic coexistence of phase-like forms. Each form is characterized by the number of strands fused together by nonbonded interactions. No distinction is made in the model among different configurations with the same number of fused strands. At highest temperatures, only four separated strands appear $(1+1+1+1)$; as the temperature drops, first this form comes into dynamic equilibrium with a form in which the two all-hydrophobic strands fuse, i.e., $(1+1+1+1)$ and $(2+1+1)$ coexist. As the temperature drops further, the form with three strands fused appears; i.e., $(1+1+1+1)$ and $(2+1+1)$ coexist with $(3+1)$. At still lower temperatures, the forms with 2, 1 and no free strands coexist, i.e., $(2+1+1)$, $(3+1)$ and (4) coexist, and at still lower temperatures, only the last two and eventually only the fully fused forms are found. No conditions were seen in which all four forms, or any other pairings, appeared. This is consistent with the possible situations of coexisting multiple phases delineated previously.[11,12] A three-state model with degeneracies inferred from the simulations yielded partition functions that reproduce the results of the simulations well, and also provide rather accurate predictions of the volume dependence of the equilibrium compositions.

ACKNOWLEDGMENT

This research was supported by the National Science Foundation.

[1] J. Skolnick, A. Kolinski, and R. Yaris, Proc. Natl. Acad. Sci. USA **85**, 5057 (1989).
[2] J. Skolnick and A. Kolinski, Science **250**, 1121 (1990).
[3] Z. Guo, D. Thirumalai, and J. D. Honeycutt, J. Chem. Phys. **97**, 525 (1992).
[4] Z. Guo and D. Thirumalai. ''Energy landscape and folding mechanisms in proteins,'' in *Protein Folds*, edited by H. Bohr and S. Brunak (Chemical Rubber Corporation, Cleveland, OH, 1995), p. 233.
[5] K. D. Ball, R. S. Berry, A. Proykova, R. E. Kunz, and D. J. Wales, Science **271**, 963 (1996).
[6] M. P. Tildesley and D. J. Allen, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1989).
[7] D. W. Heermann, *Computer Simulation Methods in Theoretical Physics* (Springer, Berlin, 1990).
[8] S. M. Kast, L. Nicklas, H.-J. Bär, and J. Brickmann, J. Chem. Phys. **100**, 566 (1994).
[9] S. M. Kast and J. Brickmann, J. Chem. Phys. **104**, 3732 (1996).
[10] B. Vekhter and R. S. Berry, J. Chem. Phys. **106**, 6456 (1997).
[11] R. E. Kunz and R. S. Berry, Phys. Rev. Lett. **71**, 3987 (1993).
[12] R. E. Kunz and R. S. Berry, Phys. Rev. E **49**, 1895 (1994).