Effect of chain connectivity on the structure of Lennard-Jones liquid and its implication on statistical potentials for protein folding

W. C. Lu,[1,2] C. Z. Wang,[1,*] and K. M. Ho[1]
[1]*Ames Laboratory and Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011, USA*
[2]*State Key Laboratory of Theoretical and Computational Chemistry, Jilin University, Changchun 130023, People's Republic of China*
(Received 29 May 2002; revised manuscript received 21 November 2003; published 23 June 2004)

Statistical contact potentials and bead-spring models have been widely used for computational studies of protein folding. However, there has been speculation that systematic error may arise in the contact energy calculations when the statistical potentials are deduced under the assumption that the chain connectivity in proteins can be ignored. To address this issue, we have performed molecular-dynamics simulations to study the structure and dynamics of a simple liquid system in which the beads are either connected or unconnected with springs. Results from the present study provide useful information for assessing the accuracy of the statistical potentials for protein structure simulations.

DOI: 10.1103/PhysRevE.69.061920 PACS number(s): 87.15.Aa, 87.14.Ee, 61.20.Ja, 02.70.Ns

I. INTRODUCTION

Empirical energy functions are widely employed to study the protein folding problem. Because of the complexity, a detailed description of interactions in proteins at the atomistic level will require a large amount of computational work load. Therefore, statistical potentials [1,2] based on the concept of contact energy [3] have been the most commonly used "knowledge-based" energy functions to provide a simple coarse-grained description at the residue level in many studies of protein structure recognition and prediction [4–6] and protein folding simulations [7–10]. The statistical potentials also have been used to study protein docking [11,12], and to study designability of protein structures [13–15].

In the "statistical potential" approach, the effective contact energies between protein residues are estimated directly from the numbers of residue-residue contacts observed in the known protein structures by regarding them as statistical averages in quasichemical approximation. This idea was first proposed by Tanaka and Scheraga in 1976 [16]. Miyazawa and Jernigan made an important step forward to include solvent effects in statistical potentials [1]. Meanwhile, statistical potentials for protein folding were developed in several other aspects, e.g., incorporating distance-dependent forces and multibody interactions [17–20], and adding terms to describe dihedral angles and secondary structures [21–23].

There are two essential steps in deriving the statistical potentials for proteins from the residue pair distributions. First, the numbers of residue-residue contacts derived from protein crystal structures are compared with those expected in a random mixture state. Next, a quasichemical approximation was employed to connect these normalized contact numbers with effective inter-residue contact energies E_{ij} via the relation [24]

$$E_{ij} = -T \ln n_{ij}, \tag{1}$$

where i and j denote amino acid types, T is temperature in the unit of energy, and n_{ij} is the normalized contact number. It is clear that the effective contact energy in Eq. (1) is dependent only on residue-residue contact numbers or coordination numbers for each residue.

The procedure of deriving statistical potentials relies on two basic approximations [1], which have been noticed by Thomas and Dill [25,26]. The first one is the assumption that chain connectivity imposed by the protein sequence can be neglected so that the statistics of contacts in a connected chain ensemble is the same as in a liquid of disconnected amino acids. It was argued that for a large sample of proteins, the effects of specific sequences would be averaged out and characteristics of residue-residue contacts would reflect intrinsic differences of interactions among residues [1]. However, even if this is correct, there still remains the question of whether the pair distribution of a connected chain is the same as a liquid of disconnected monomers. The second approximation in statistical potentials is the quasichemical approximation or the Boltzmann distribution assumption, in which the residue-residue contact numbers meet the Boltzmann distribution law [Eq. (1)]. Some evidence supports the use of a Boltzmann distribution, e.g., some protein substructures have about the same frequencies as they would have in thermodynamic equilibrium.

In order to see how accurate are the statistical potentials for describing the energies of the protein folding problem, Thomas and Dill [25] devised a rigorous test using two-dimensional (2D) lattice models consisting of chains of two monomer types H (hydrophobic) and P (polar). They set up a database of the native structures, from which statistical potentials were extracted. Comparing the known true energies with the statistical energies, they found that statistical potentials often correctly rank the orders of the relative strengths of inter-residue interactions, but they do not reflect the true underlying energies because of systematic errors arising from the neglect of excluded volume effects. There is also error in neglecting indirect correlation which changes

*Corresponding author. Email address: wangcz@ameslab.gov

the pair distribution of two given residues because of their interactions with a common third residue. Such effects are negligible in the low-density limit (e.g., gas phase), but can be significant in the liquid phase, where the density is not low. Since the study of Thomas and Dill was done with a 2D lattice model, it is not clear how serious these errors will be in 3D structures more appropriate to real proteins. Although successful and unsuccessful applications of the "knowledge-based" statistical potentials to 3D and off-lattice protein structure simulations have also been discussed extensively in the literature [5,27–32], the possible error in the statistical potentials for protein folding due to neglecting chain connectivity has not been well addressed.

In this work, we aim at understanding how good is the approximation of neglecting the chain connectivity in the statistical potentials for 3D protein structure modeling. It has been well known from polymer simulations that chain connectivity does play an important role in determining the structures and dynamics of polymers [33–35]. Since most of the statistical potentials for protein folding are derived using the residue pair correlation functions of proteins, our present studies will be focused on understanding the effect of chain connectivity on the pair correlation function $g(r)$ and coordination numbers of the residues in proteins through simulation studies of a liquid system where the proteins are represented by a bead-spring model [36,37] and the residues are interacting with the Lennard-Jones potential. We will compare the pair correlation functions $g(r)$ for the residue beads with and without chain connections. Although such a bead-spring model is a very simplified representation of proteins, we believe that the effects of chain connectivity obtained from such model studies should produce useful information for assessing the accuracy of statistical potentials for protein folding. Our paper is organized as follows. In Sec. II, we will describe the models used in the simulations. More simulation details including the choice of density and the correction to $g(r)$ due to excluded volume will be given in Sec. III, followed by the simulation results in Sec. IV. Finally, conclusions are given in Sec. V.

II. MODELS

A. Reference system

Our reference system consists of 512 identical particles interacting with the Lennard-Jones (LJ) potential,

$$U_{LJ} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]. \tag{2}$$

The potential parameters ε and σ are chosen to be 240 K and 5.0 Å, respectively. The choice of $\sigma = 5.0$ Å will give the first peak of $g(r)$ around 5.45 Å, which is close to the average residue contact distance in proteins. The cutoff distance of the LJ interaction is chosen to be 12.0 Å, which is found to be large enough to give a smooth pair correlation function $g(r)$ in the liquid state. Although this cutoff distance (R_{cut}) is much larger than 6.5 Å, which is commonly used as contact distance to determine the statistical potential for proteins, we found that the pair correlation function is not very sensitive to the cutoff distance. Therefore, the choice of R_{cut} is not crucial for our present purpose of study.

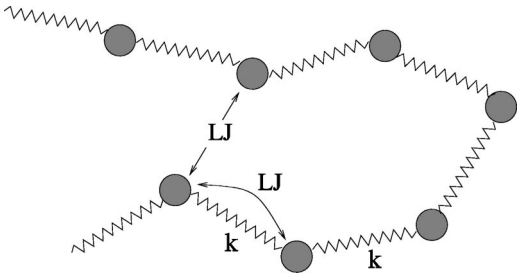

FIG. 1. Cartoon view of the single-bead model.

B. Single-bead model

The first model we used to investigate the effects of chain connectivity is a single-bead model as illustrated in Fig. 1. The 512 LJ particles as described in the reference system are connected by nearest-neighbor harmonic springs: $U = \frac{1}{2}k(r - r_{\text{equil}})^2$, where r is the distance between particles and r_{equil}(5.45 Å) represents its equilibrium distance. Such a model is similar to the single-bead model for protein simulations widely used in the literature except that our beads are identical in character. The force constant is chosen to be $k = 0.5$, 1.0, and 1.5, respectively, in order to investigate the effects of the strength of the chain connectivity on the structures of the system.

C. Double-bead model

Chain connectivity in proteins is mainly imposed by the strong peptide bonds on backbone atoms. The side-chain residues interact mostly through hydrophobic interactions and hydrogen bonds. The protein structure is therefore better modeled with a double-bead model as illustrated in Fig. 2. In this double-bead model, the backbone C_α atoms are represented by the set of A-type particles which are connected by harmonic springs (main chain) $U_{AA} = \frac{1}{2}k(r - r_\alpha)^2$, where $r_\alpha = 3.84$ Å is the distance between C_α atoms in proteins. The force constant k is set to be 0.5, 1.0, and 1.5, respectively, for the same reason as discussed in Sec. II B. For better representing the backbone structure in protein, interactions among

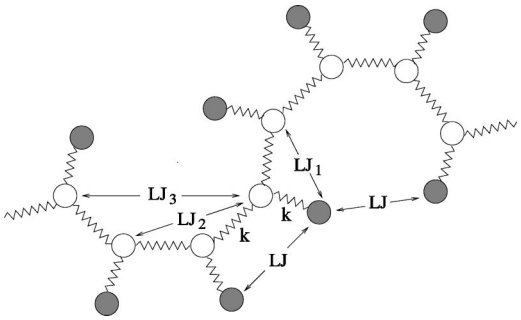

FIG. 2. Cartoon view of the double-bead model. The white balls are the A-type beads and the gray balls are the side-chain B-type beads.

the second and third neighbors of the main-chain *A*-type particles are also included. These interactions are also modeled by LJ potentials. The second-neighbor interactions LJ_2 (with $\varepsilon=960$ K and $\sigma=4.91$ Å) keep the dihedral angles between the C_α atoms close to that in real protein. The third-neighbor interactions LJ_3 (with $\varepsilon=960$ K and $\sigma=4.4$ Å) mimic the hydrogen bonds in the α-helix backbone environment. The residues of the protein are modeled by the set of LJ particles as described in the reference system. These side-chain particles (*B*-type particles) are also connected by a spring to the corresponding *A*-type particles in the main chain, respectively, i.e., $U_{AB}=\frac{1}{2}k(r-r_\beta)^2$ ($r_\beta=3.0$ Å). Moreover, a hard-core repulsion, $U_{\text{rep}}=4\varepsilon(\sigma/r)^{12}$ ($\varepsilon=960$ K, $\sigma=4.95$ Å), is used for the *A*-*A* pair beyond the third neighbors along the chain, and a weak LJ potential $LJ_1=4\varepsilon[(\sigma/r)^{12}-(\sigma/r)^6]$($\varepsilon=48$ K, $\sigma=5.0$ Å) is used for the *A*-*B* pairs that are not connected by the spring. These interactions are illustrated in Fig. 2. In this study, 512 *A*-type and 512 *B*-type particles are used so that the results can be compared with those of the reference system.

III. SIMULATION DETAILS

A. Choice of density

In order to perform the simulations at the density regime that is relevant to protein, we need the information about the residue density in proteins. We have estimated such a density profile using the protein structures from the Protein Data Bank (PDB). We approximate the shape of a protein by an ellipsoid, and the residue masses are assumed to be uniformly distributed within the ellipsoid. By calculating the moment of inertia I_x, I_y, and I_z of a protein along the three principal axes and using the relationship between the I_x, I_y, and I_z and axis a, b, and c of the ellipsoid,

$$I_x = \tfrac{1}{5}nm(b^2 + c^2),$$

$$I_y = \tfrac{1}{5}nm(c^2 + a^2),$$

$$I_z = \tfrac{1}{5}nm(a^2 + b^2) \tag{3}$$

the volume and thus the residue density of the protein can be estimated. In Eq. (3), nm is the total mass of the protein. Our calculations were performed on 853 representative protein structures selected from the PDB database. The densities of these proteins obtained from our analysis are plotted in Fig. 3. We see from the plot that the densities of most of the proteins are between 4.0 and 6.0 residues/nm^3. The average density of the proteins calculated from the plot is about 4.8/nm^3. We note that [38] the protein volume obtained from our method is very close to the envelope volume (which is defined as van der Waals volume divided by packing density) when we use the van der Waals volume from the VOLUME package calculation [39] and take the packing density to be 0.75 [40]. In our present study, we will perform simulations with densities of 4.0, 4.8, and 6.0/nm^3, respectively.

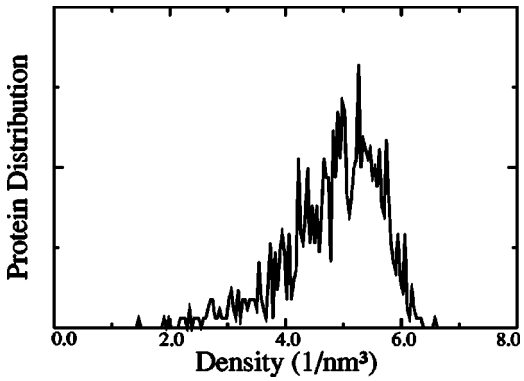

FIG. 3. Density profile of side-chain residues in proteins obtained from our estimations as described in the text.

B. Correction to the pair correlation function

The constraint of spring connectivity between particles in our models is expected to have much more severe restrictions on nearest-neighbor beads in comparison with the others. We note that in the single-bead model, although the second-neighbor beads along the chain are not connected directly by covalent bonds, they are connected by springs to a common atom between them. Therefore, the covalent bonding contributions from the second neighbor along the chain may also not be negligible. Such covalent bonding contributions are usually excluded when contact potentials are constructed. Our simulation results as will be discussed in the following indicate that exclusion of second neighbors along the chain is necessary for the single-bead model in order to minimize the covalent bonding contributions. On the other hand, in the double-bead model, the residues (type-*B* beads) are not directly connected by springs and the covalent bonding effects on the pair correlation function of the residue beads are much weaker. Thus only first neighbors due to the chain connections are omitted in our analysis for the double-bead model. Note that in protein statistical potential modeling, the contacts due to nearest neighbors along the sequence are also explicitly omitted in the estimation of effective contact energies.

Because the excluded beads do occupy a certain volume, such volume should also be deducted when these neighbors are excluded from the calculation of the pair distribution function, otherwise the density of the system will be underestimated. In other words, the pair correlation function $g_0(r)$ obtained from the molecular-dynamics (MD) simulations with neighbors along the chain excluded has to be renormalized according to the excluded volume. As shown in Fig. 4, suppose the neighboring particles that are to be excluded (shaded balls) have a hard-core radius of r_0 and are a distance r' away from the center particle located at O. When the $g(r)$ is calculated at a distance r from the center particle, the area of the intersection between the sphere of radius r and the bodies of the excluded neighbors has to be deducted from the overall sphere surface area of $4\pi r^2$. The ratio of the effective volume to volume of the whole system can be approximated as

$$\xi(r,r') = [4\pi - 2\Delta\Omega(r,r')]/4\pi \tag{4}$$

with

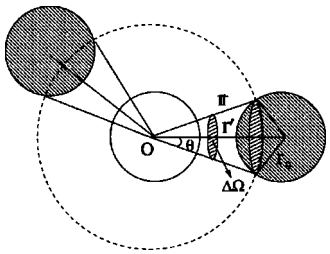

FIG. 4. Illustration of the scheme for renormalizing the pair correlation function $g(r)$ due to the exclusion of nearest (or next nearest) neighbors along the chain.

$$\Delta\Omega(r,r') = 2\pi[1 - \cos\ \theta(r,r')] = 2\pi\left(1 - \frac{r^2 + r'^2 - r_0^2}{2rr'}\right). \tag{5}$$

We see from Eqs. (4) and (5) that ξ is not only dependent on r, but is also a function of r' the distance between the center particle and the excluded neighbors. Therefore, the corrected radial distribution function $g(r)$ can be calculated using the reduced distribution function $g_0(r)$ and the distribution function of the neighboring particles $f(r')$ obtained from the MD simulation,

$$g(r) = g_0(r)\int_{r'} \frac{1}{\xi(r,r')} f(r')dr', \tag{6}$$

where

$$f(r') = g_{\mathrm{excl}}(r')r'^2 \Big/ \int_r g_{\mathrm{excl}}(r)r^2 dr. \tag{7}$$

$g_{\mathrm{excl}}(r)$ is the radial distribution function of the excluded neighboring beads which can also be obtained from the same MD simulation. Note that such a renormalization scheme can be applied to either first-neighbor correction in which the nearest neighbors along the chain are excluded or two-neighbor correction where both first and second neighbors along the chain are omitted.

We further verify the reliability of this correction scheme by comparing the $g(r)$ of the free-particle system with those of the chain-connected models in which the force constant of the springs is set to be very small ($k=0.001$). In the double-bead model, the interaction strength of LJ_1 between B-type and A-type particles is also weakened to 0.005 ε. Figure 5 shows that the corrected $g(r)$ for the chain-connected systems is very close to that of the free-particle system, as they should be, because the springs are very weak. In comparison, the uncorrected pair correlation function $g_0(r)$ is clearly below that of the free-particle system. This result indicates that the correction scheme we performed (with the optimized r_0 of 4.2 Å for single-bead two-neighbor correction and 3.5 Å for double-bead one-neighbor correction) is adequate.

As shown in Fig. 6(a), the raw data of pair correlation functions (without exclusion and correction) of the single-bead model are very sensitive to the spring constant. However, after the exclusion and correction procedures as discussed above, the corrected $g(r)$ becomes almost

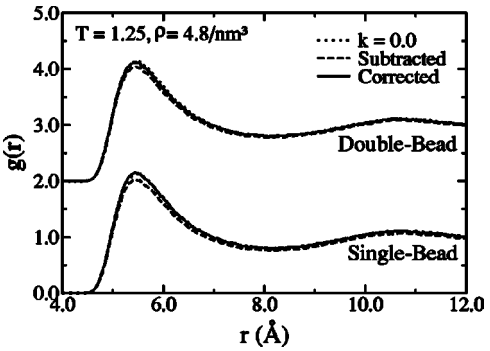

FIG. 5. The corrected $g(r)$ (solid lines) for the single-bead model and double-bead model with $k=0.001$ are compared with that of the unconnected LJ system (dotted lines). The reduced $g_0(r)$ (dashed lines) are also plotted. Note that $g_0(r)$ is clearly below the corrected $g(r)$.

k-independent as shown in Fig. 6(b). These results indicate that the scheme we used is effective for removing the covalent bond contributions. On the other hand, the effects of spring constant k (and thus the effects of the covalent bond) are much less pronounced in the double-bead model, as one can see from Figs. 6(c) and 6(d). Therefore, only the results from the simulations with $k=1.0$ will be shown in Sec. IV, although we have performed the simulations with several spring constants ($k=0.5$, 1.0, and 1.5).

C. Simulation procedure

MD methods have been extensively used in the past to study a variety of physical systems. The MD technique has been fully documented elsewhere (see, for example, Refs. [41,42]). In the present simulation, most properties (otherwise specified) are described in reduced units, in which σ, ε, and k_B/ε are used as the units of length, energy, and temperature, respectively [41]. The equations of motions of the particles were solved with a time step of 0.02 reduced units, using the fifth-order predictor-corrector algorithm [43]. The particles were confined to a cubic cell and subjected to periodic boundary conditions. The velocity-scaling method was used to control the temperature of the system. For each model described in Sec. II, we performed the MD simulations with three different densities $\rho=4.0$, 4.8, and $6.0(1/\mathrm{nm}^3)$ and with three different temperatures $T=0.83(200\ \mathrm{K})$, 1.25 (300 K), and 2.5 (600 K), respectively. For each case, we run the simulation for over 20 000 steps to allow the system to approach the thermal equilibrium state, followed by 20 000 steps for statistical average of the properties of interest. The structure of the system is monitored by the pair correlation function $g(r)$ of the particles representing the residues of proteins. The dynamical properties of the system are also studied by calculating the time dependence of the mean-square displacement of the residue particles,

$$\langle R^2(t)\rangle = \left\langle \frac{1}{N}\sum_{i=1}^{N} |\mathbf{r}_i(t+\tau) - \mathbf{r}_i(\tau)|^2 \right\rangle_{\tau}. \tag{8}$$

Our simulation temperatures are much higher than the estimated triple point of $T=0.68$ for the LJ potential system

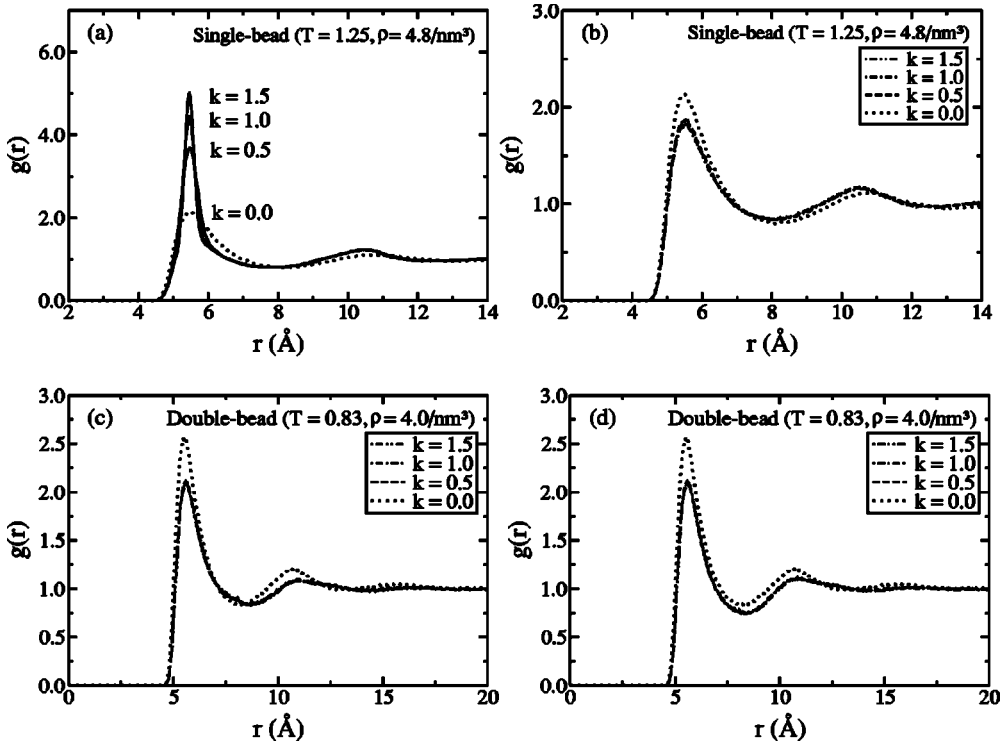

FIG. 6. Pair correlation functions as a function of spring constant k. (a) The raw data (without exclusion and correction) from the single-bead model; (b) corrected $g(r)$ of the single-bead model; (c) the raw data from the double-bead model; (d) corrected $g(r)$ of the double-bead model.

[44,45], therefore the systems in our simulations are in the liquid state.

IV. RESULTS AND DISCUSSION

A. Single-bead model

We have investigated the effects of chain connectivity at different densities. In addition to the average density of $4.8/\mathrm{nm}^3$, we select both low and high densities from the protein density distribution profile (Fig. 3): $4.0/\mathrm{nm}^3$ for the low-density regime and $6.0/\mathrm{nm}^3$ for the high-density regime, respectively.

First, we discuss the chain effect at the average density. One of the noticeable differences in the results of the single-dead model as compared to those of the unconnected system is that the chain connections in the single-bead model tend to cause segregation of the particles in the system and result in an inhomogeneous distribution of the particles over the simulation unit cell, particularly at lower temperatures. The simulation indicates that the particles at the density $(4.8/\mathrm{nm}^3)$ and low temperature (0.83) are not uniformly distributed, but segregate into a more compact structure, leaving a hole in the simulation cell. In contrast, the system without chain connections tends to be more uniformly distributed at the same density and temperature. As the results of particle segregation, the pair correlation function $g(r)$ of the single-bead model at low temperature (0.83) as shown in Fig. 7(b) exhibits a higher first-neighbor peak (hence larger coordination number) in comparison with that of the unconnected system.

At higher temperatures of $T=1.25$ and 2.5, the distribution of the particles in single-bead system becomes more uniform. The first peak of $g(r)$ is now lower for single-bead chain as compared to that of the corresponding unconnected $(k=0.0)$ system, as shown in Fig. 7(b). The average coordination numbers $n(r)$ of the system are estimated by integrating the $g(r)[n=\int_{r=0}^{r_1} g(r)4\pi r^2 dr]$ over residue-residue distance from the origin to $r_1=8.22$ Å, which corresponds approximately to the distance of the first minimum of $g(r)$ for the unconnected LJ system. The results show that at the average density $(4.8/\mathrm{nm}^3)$ the coordination number of the single-bead system is reduced by about 6% and 5% at $T=1.25$ and 2.5, respectively, as compared to that of the corresponding unconnected LJ systems. We note that the temperature dependence of the chain effect on coordination numbers is not significant in this temperature range.

At the lower density, the simulations show that the particles in the single-bead model have a stronger tendency to form an inhomogeneous liquid than they do at the average density. The aggregation phenomena for low-density systems have been well studied [46–48]. The inhomogeneous structures are also observed in our low-density simulations at both $T=0.83$ and 1.25. At higher temperature of $T=2.5$, the particle distribution appears to be uniform and the first peak of $g(r)$ is lower than that of the corresponding unconnected system, as can be seen from Fig. 7(a). At $T=2.5$, the coordination number of the single-bead model is estimated (using $r_1=8.43$ Å) to decrease by about 6% with respect to that of the unconnected system, similar to that at an average density of $4.8/\mathrm{nm}^3$. In comparison, the distribution of the particles at

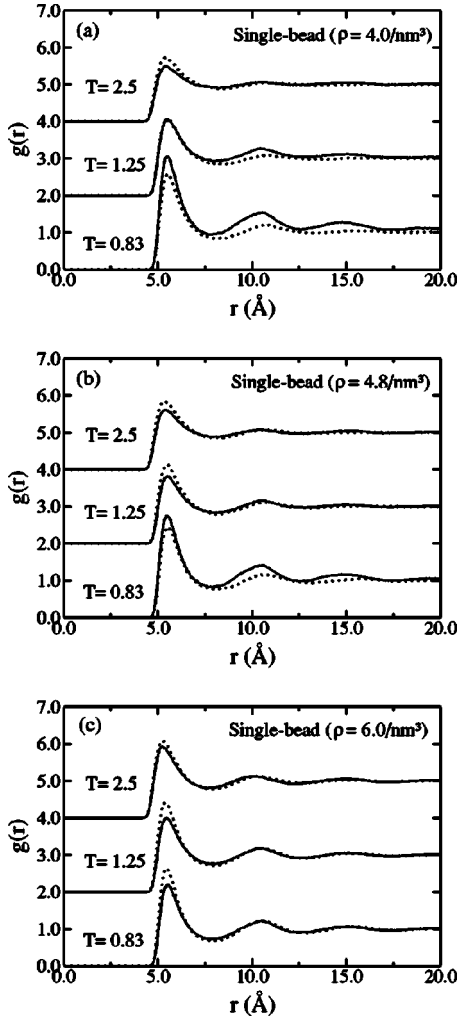

FIG. 7. Simulation results of the pair correlation function, with two-neighbor corrections, of the single-bead model ($k=1.0$) at the various temperatures and densities (dotted line indicates the unconnected system).

the high-density regime ($\rho=6.0/\mathrm{nm}^3$) is much more uniform at all three temperatures used in the simulations. As shown in Fig. 7(c), the peaks of $g(r)$ are found to be sharper at this density than those at average and low densities. The first peak of $g(r)$ is lower than that of the corresponding unconnected LJ system at all three temperatures. The coordination numbers at $\rho=6.0/\mathrm{nm}^3$ are calculated (using $r_1=7.92$ Å) to be reduced by about 6%, 6%, and 4% at $T=0.83$, 1.25, and 2.5, respectively, with respect to unconnected systems.

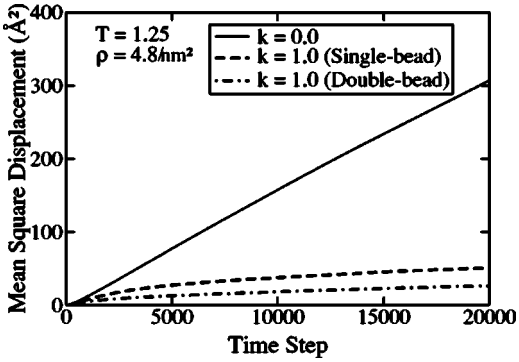

FIG. 8. Mean-square displacement of the residue particles as a function of time steps in the unconnected, single-bead, and double-bead systems.

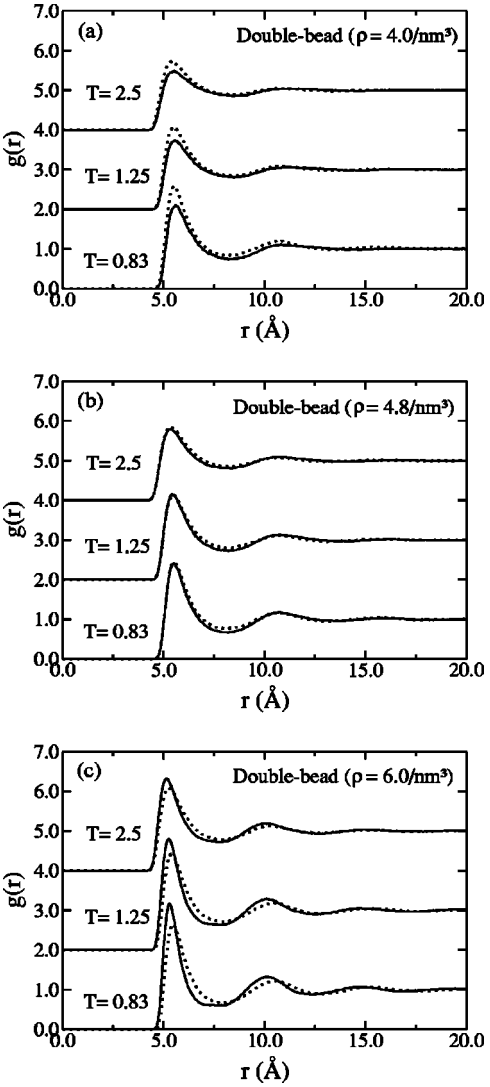

FIG. 9. Simulation results of the pair correlation function, with one-neighbor corrections, of the double-bead model ($k=1.0$) at the various temperatures and densities (dotted line indicates the unconnected system).

The dynamical properties of the single-bead model are studied by calculating the mean-square displacement of the particles as a function of time. The result of $\langle R(t)^2\rangle$ obtained from the simulation is plotted in Fig. 8 and compared to that of the unconnected system. It is found that both the value and the slope of the mean-square displacement of the particles in the single-bead model are much smaller than those of the corresponding unconnected system. This result indicates that the diffusion processes of chain-connected particles are significantly hindered by chain connectivity.

B. Double-bead model

The pair correlation function $g(r)$ of the side-chain residues in the double-bead model from our simulations is pre-

sented in Fig. 9. In contrast to the single-bead model, we do not find an inhomogeneous state in the double-bead model at the temperature and density regime we studied, although the inhomogeneous structures may appear at lower temperature or lower density.

At average density ($4.8/\mathrm{nm}^3$), the corrected $g(r)$ of the side-chain residues in the double-bead model is again found to be almost independent of the spring force constant. The first peak of $g(r)$ in the chain-connected system has a similar height but narrower width as compared to that of the free-particle system. The intensity between the first and second peaks in the $g(r)$ is lower for the chain connected system as shown in Fig. 9(b). At different temperatures of $T=0.83$, 1.25, and 2.5, the coordination numbers of the side-chain residues are reduced by almost the same amount, 7%, 5%, and 5%, respectively. The reductions in the coordination numbers are similar to those in the single-bead model.

At low density ($4.0/\mathrm{nm}^3$), the chain effect on $g(r)$ is larger than that at the average density ($4.8/\mathrm{nm}^3$), as shown in Fig. 9(a). The first peak of $g(r)$ is lowered by a larger amount, and the coordination numbers of the side-chain residues are reduced by 13%, 9%, and 8% at three different temperatures with respect to those of unconnected systems. Therefore, the chain effects in the low-density systems are more than 1.5 times larger in comparison with the corresponding systems at the average density. It can be seen from Figs. 9(a) and 9(b) that at low and average densities, the location of the first peak of $g(r)$ changes very little no matter whether chain connectivity exists or not, indicating that the packing of the side-chain residues is dominated by the hydrophobic interactions, which in our case is the LJ potential with an equilibrium distance of 5.45 Å.

It can be seen from Fig. 9(c) that at high density ($6.0/\mathrm{nm}^3$), the first peak of $g(r)$ is higher and much sharper than that of the unconnected system. At $T=0.83$, 1.25, and 2.5, the coordination numbers of the side-chain residues are decreased by 6%, 5%, and 4% with respect to the reference system, similar to those at the average density.

The mean-square displacement of the side-chain residues versus time at $T=1.25$ and $\rho=4.8/\mathrm{nm}^3$ is plotted in Fig. 8 in comparison with that of the single-bead model as well as that of the unconnected system. The results show that the mean-square displacement of the residues in the double-bead model is smaller than that of the single-bead model. Thus the resistance to the motions of the residues due to the chain connectivity is slightly larger in the double-bead model as compared to the single-bead model.

V. CONCLUSIONS

In this work, the effects of chain connectivity on the pair correlation function $g(r)$ of a LJ liquid are studied. The simulation results show that $g(r)$ of the chain-connected beads is different from that of the unconnected ones, leading to smaller coordination numbers for both directly (single-bead model) and indirectly (double-bead model) chain-connected residues. By ignoring the chain connectivity, the error in counting coordination numbers or contact numbers would be large for low-density and low-temperature systems (e.g., 13% for $\rho=4.0/\mathrm{nm}^3$, $T=0.83$). But the error becomes much smaller when the density reaches the average value in the protein density profile (see Fig. 3), i.e., 5% for $\rho=4.8/\mathrm{nm}^3$, $T=1.25$. In recent years, there has been a controversy about how much error is introduced in statistical contact potentials for proteins by neglecting the chain connectivity when the potentials are constructed. From our simulation results, we find that the coordination number is indeed affected at a moderate level by the chain connectivity in the double-bead model.

ACKNOWLEDGMENTS

Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under Contract No. W-7405-Eng-82. This work was supported by the Director for Energy Research, Office of Basic Energy Sciences, including a grant of computer time at the National Energy Research Supercomputing Center (NERSC) in Berkeley. We are also grateful to Professor S. Miyazawa for his critical reading of our manuscript.

[1] S. Miyazawa and R. Jernigan, Macromolecules **18**, 534 (1985).
[2] S. Miyazawa and R. Jernigan, J. Mol. Biol. **256**, 623 (1996).
[3] M. Levitt and A. Warshel, Nature (London) **253**, 694 (1975).
[4] A. Torda, Curr. Opin. Struct. Biol. **7**, 200 (1997).
[5] L. A. Miruy and E. I. Shakhnovich, J. Mol. Biol. **283**, 507 (1998).
[6] H. Cao, Y. Ihm, C. Z. Wang, J. R. Morris, M. Su, D. Dobbs, and K. M. Ho, Polymer **45**, 687 (2004).
[7] J. Skolnick and A. Kolinski, Science **250**, 1121 (1990).
[8] C. Wilson and S. Doniach, Proteins: Struct., Funct., Genet. **6**, 193 (1989).
[9] A. Kolinski and J. Skolick, Proteins: Struct., Funct., Genet. **18**, 338 (1994).
[10] M. Qin, J. Wang, Y. Tang, and W. Wang, Phys. Rev. E **67**, 061905 (2003).
[11] M. Pellegrini and S. Doniach, Proteins: Struct., Funct., Genet. **15**, 436 (1993).
[12] S. Vajda, M. Sippl, and J. Novotny, Curr. Opin. Struct. Biol. **7**, 222 (1997).
[13] H. Li, R. Helling, C. Tang, and N. Wingreen, Science **273**, 666 (1996).
[14] H. Li, C. Tang, and N. S. Wingreen, Proteins **49**, 403 (2002).
[15] J. Miller, C. Zeng, N. S. Wingreen, and C. Tang, Proteins **47**, 506 (2002).
[16] S. Tanaka and H. A. Scheraga, Macromolecules **9**, 945 (1976).
[17] M. J. Sippl, J. Mol. Biol. **213**, 859 (1990).
[18] A. Kolinski and J. Skolnick, J. Chem. Phys. **98**, 7420 (1993).
[19] A. Kolinski and J. Skolnick, Proteins: Struct., Funct., Genet. **18**, 338 (1994).

[20] S. Shimizu and H. S. Chan, Proteins **58**, 15 (2002).
[21] M. Rooman, J. A. Kocher, and S. Wodak, Biochemistry **32**, 10 226 (1992).
[22] R. Dewitte and E. Shakhnovich, Protein Sci. **3**, 1570 (1994).
[23] S. Miyazawa and R. L. Jernigan, Proteins **36**, 347 (1999).
[24] L. A. Mirny and E. I. Shakhnovich, J. Mol. Biol. **264**, 1164 (1996).
[25] P. D. Thomas and K. A. Dill, J. Mol. Biol. **257**, 457 (1996).
[26] P. D. Thomas and K. A. Dill, Proc. Natl. Acad. Sci. U.S.A. **93**, 11 628 (1996).
[27] J. P. Kocher, M. J. Rooman, and S. J. Wodak, J. Mol. Biol. **235**, 1598 (1994).
[28] A. Godzik, A. Kolinski, and J. Skolnick, Protein Sci. **4**, 2101 (1995).
[29] E. Furuichi and P. Koebl, Proteins **31**, 139 (1998).
[30] L. Zhang and J. Skolnich, Protein Sci. **7**, 112 (1998).
[31] M. Vendruscolo and E. Domany, J. Chem. Phys. **109**, 11 101 (1998).
[32] S. R. Sunyaev, F. Eisenhaber, P. Argos, E. N. Kuznetsov, and V. G. Tumanyan, Proteins **31**, 225 (1998).
[33] P. G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, NY, 1979).
[34] K. Kremer and G. S. Grest, J. Chem. Phys. **92**, 5057 (1990), and references therein.
[35] K. Binder, J. Baschnagel and W. Paul, Prog. Polym. Sci. **28**, 115 (2003), and references therein.
[36] P. E. Rouse, J. Chem. Phys. **21**, 1272 (1953).
[37] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Clarendon, New York, 1986).
[38] We find that the protein data set used in Liang's paper [39] has six proteins (1eca, 1arb, 1plc, 1ycc, 3sdh, and 5p2l) in common with our data set. If we take the van der Waals volumes of these six proteins from the VOLUME methods [39] and divide them by a packing density of 0.75 [40] , we get the values for the envelope volumes as 24 987, 43 515, 16 774, 20 217, 54 990, and 31 449 $Å^3$, respectively, which are very close to the results of 25 872, 42 413, 17 444, 20 212, 58 269, and 31 130 $Å^3$ obtained from our method.
[39] J. Liang, H. Edelsbrunner, P. Fu, P. V. Sudhakar, and S. Subramaniam, Proteins: Struct., Funct., Genet. **33**, 1 (1998).
[40] J. Zhang, R. Chen, C. Tang, and J. Liang, J. Chem. Phys. **118**, 6102 (2003); J. Liang and K. A. Dill, Biophys. J. **81**, 751 (2001).
[41] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1987).
[42] F. F. Abraham, Adv. Phys. **35**, 1 (1985).
[43] A. Rahman, in *NATO Advanced Study Institute Series: Correlation Functions and Quasiparticle Interactions in Condensed Matter*, edited by J. W. Halley (Plenum, New York, 1978).
[44] J. P. Hansen and L. Verlet, Phys. Rev. **184**, 151 (1969).
[45] J. Q. Broughton and G. H. Gilmer, J. Chem. Phys. **79**, 5095 (1983).
[46] F. H. Stillinger, Phys. Rev. E **63**, 011110 (2000).
[47] D. L. Malandro and D. J. Lacks, J. Chem. Phys. **107**, 5804 (1997).
[48] A. D. Mackie, A. Z. Panagiotopoulos, and I. Szleifer, Langmuir **13**, 5022 (1997).