Analyses of simulations of three-dimensional lattice proteins in comparison with a simplified statistical mechanical model of protein folding

H. Abe[1] and H. Wako[2]
[1]*Department of Natural Sciences, Nishinippon Institute of Technology, Fukuoka 800-0394, Japan*
[2]*School of Social Sciences, Waseda University, Tokyo 169-8050, Japan*
(Received 26 October 2005; published 18 July 2006)

Folding and unfolding simulations of three-dimensional lattice proteins were analyzed using a simplified statistical mechanical model in which their amino acid sequences and native conformations were incorporated explicitly. Using this statistical mechanical model, under the assumption that only interactions between amino acid residues within a local structure in a native state are considered, the partition function of the system can be calculated for a given native conformation without any adjustable parameter. The simulations were carried out for two different native conformations, for each of which two foldable amino acid sequences were considered. The native and non-native contacts between amino acid residues occurring in the simulations were examined in detail and compared with the results derived from the theoretical model. The equilibrium thermodynamic quantities (free energy, enthalpy, entropy, and the probability of each amino acid residue being in the native state) at various temperatures obtained from the simulations and the theoretical model were also examined in order to characterize the folding processes that depend on the native conformations and the amino acid sequences. Finally, the free energy landscapes were discussed based on these analyses.

DOI: 10.1103/PhysRevE.74.011913 PACS number(s): 87.15.Cc, 87.15.Aa, 87.15.−v

I. INTRODUCTION

Currently, elucidating the general principle governing protein folding is one of the most important open problems in modern biophysics. Recently, many papers have been published, and significant progress has been made toward solving this problem [1–5]. However, no clear solution has been found as yet.

In this paper, we focus on the following problems. First, from a physical perspective, protein folding is a cooperative phenomenon analogous to a first-order phase transition. The two-state nature (native and denatured states) is one of the characteristic features of protein folding to be discussed, although some proteins have intermediate states between these two states. The cooperativity of protein folding results from specific interactions between amino acid residues. In fact, some simulations considered only the interactions of the amino acid residues that are in contact with each other in a native conformation (referred to as native contact interaction or NC interaction), but they ignored non-NC interactions (interactions of amino acid residues not in contact with each other in the native conformation), even if such amino acid residues make contact during the simulation (Gō-type interaction [6]). However, non-NC interactions may possibly occur during the folding process and will affect folding in real proteins. Although the intermediate states around the transition temperature are important for understanding the two-state cooperativity [7,8], it is still an open problem. Then, the first question is the extent to which the non-NC interactions occur and the manner in which they affect protein folding, particularly around the transition temperature.

Second, Levinthal's paradox [9] indicates that a protein cannot fold spontaneously by randomly searching a conformational space; therefore, some folding pathways should exist. The concept of a funnel-shaped energy landscape [10,11] also states that an accessible area in the conformational space is considerably restricted and is directed toward the native conformation. The Gō model with the above-mentioned interactions realizes the main concept of the funnel-shaped energy landscape [12]. Besides, it is another scheme that simulates a restricted and directed conformational space in which a protein folds in a stepwise manner along the polypeptide chain. In this scheme, secondary structures such as α-helix, β-strand, and turn are formed in the first stage; in the next stage, they gradually grow through medium-range interactions; finally, they are associated together into the native structure by long-range interactions. However, the question as to what happens in the intermediate stage of the folding and/or unfolding in which several secondary structures coexist without being packed into the native structure arises again. The second question concerns the behavior of the secondary structures formed during an early stage of the folding in an intermediate stage. More specifically, to what extent they interact with each other and the manner in which the protein in the medium and/or intermediate stage proceeds to the final stage of folding.

In addition, in this paper, we also focus on the dependence of protein folding on native conformation and an amino acid sequence.

Since we are interested in protein folding as a statistical mechanical phenomenon, we approach this problem using two different methodologies: a Monte Carlo (MC) simulation of a three-dimensional (3D) lattice protein and a theoretical model for the folding of the lattice protein. We can understand such a complex system as real protein only by clarifying the essential aspects. The simulation and the theoretical models for the lattice protein involve a drastic and essential simplification of various aspects of the real proteins. Since no all-atom simulation is yet capable of folding real proteins, the lattice protein is a powerful tool for determining the basic characteristics of the protein folding process. The simulation and the theoretical models for the lattice protein serve as an

excellent basis for approximating real proteins.

The lattice protein was first introduced by Gō and coworkers [13,14] to reveal the mechanism of protein folding. Their MC simulations of lattice proteins are characterized by the Gō-type interactions described above. However, the amino acid types were not considered. These types of interactions not only ensure that the native structure has the lowest energy state but also result in a cooperative structural transition.

The Gō-type interactions were later modified to be more realistic by introducing the amino acid types. Lau and Dill [15] considered two types of amino acid residues, hydrophobic (H) and polar (P) residues, in 2D lattice proteins. Further, Shakhnovich and co-workers [16–19] considered 20 types of amino acid residues. These studies considered both NC and non-NC interactions that were assigned energy values depending on the amino acid types. In order to ensure protein folding in a given native conformation, amino acid sequences were carefully generated using the design method developed by Shakhnovich and co-workers. They then carried out the MC simulations and demonstrated that the lattice protein can be folded into its native conformation, even if both NC and non-NC interactions exist. They also showed that due to the non-NC interactions, different amino acid sequences that can be folded into the same native conformation undergo different structural transitions. On similar lines, we carried out the MC simulation.

Another approach to the concerned problem is the formulation of a simplified statistical mechanical model for protein folding. The theoretical model not only allows the rapid calculation of thermodynamic quantities of a given system, but also enables the analysis of the essential features of the folding problem. Wako and Saitô [20,21] first proposed a simplified statistical mechanical model (referred to as island model) for protein folding. This model is characterized by the formulation of a partition function that can explicitly take into account the specific native conformation and incorporate long-range interactions between amino acid residues. Subsequently, Gō and Abe [22] reproposed an improved and extended version of this model based on the molecular details. Further, they reformulated the partition function as a noninteracting local structure model (NILS model) in the form of a recurrent equation rather than the matrix form used by Wako and Saitô.

Abe and Gō [23] showed that the above-mentioned simplified statistical mechanical model can reproduce the computer simulation of 2D square-lattice proteins with Gō-type interactions remarkably well. Recently, Eaton and coworkers [24–26] applied the same model to the formation of a β hairpin and to the folding rates in real proteins. Bruscolini and Pelizzola [27] rediscovered a transfer matrix formalism to precisely evaluate the partition function and calculated several thermodynamic quantities for the given proteins under the above-mentioned assumption. Itoh and Sasai [28] developed further studies by analyzing the conformational dynamics related to the function of photoactive yellow protein.

However, in the simplified statistical mechanical model, it is generally difficult to estimate chain entropy with an excluded volume effect. Therefore, chain entropy is usually treated as an adjustable parameter so that the calculated thermodynamic quantities may best reproduce the experimental or simulation results. Fortunately, for the 3D lattice protein discussed in this study, chain entropy can be evaluated in a practical manner. In fact, we can count possible conformations for a small-sized segment by explicitly considering the excluded volume effect, and then extrapolate the approximated formula from the conformations for a larger-sized segment [29]. Consequently, we obtained the partition function by summing up the statistical weights of the local structures and the random-coil regions over all possible conformations without any adjustable parameter. Hereafter, we refer to this version of the simplified statistical mechanical model for the 3D cubic lattice protein as the Abe-Wako model (or the A-W model).

The main purpose of this paper is to examine in detail the NC and non-NC interactions and the interactions between secondary structures (redefined as local structures below) at various folding stages in the simulations in comparison with the results obtained from the A-W model. The equilibrium thermodynamic quantities (free energy, enthalpy, entropy, and the probability of each amino acid residue being in the native state) are calculated at various temperatures for the two native conformations and the two amino acid sequences for each of the two native conformations (as a result four amino acid sequences are examined) in order to reveal the dependence of folding on the native conformation and the amino acid sequence. The detailed comparisons of the simulation data with the A-W model will provide useful information for understanding the basic aspects of protein folding.

II. METHODS

A. Lattice proteins

We consider the conformation of protein molecules consisting of n monomers, each of which is regarded as an amino acid residue on a 3D cubic lattice. The two monomers that occupy the nearest-neighbor lattice points, but are not covalently bonded along the polypeptide chain, are considered to interact with each other in a conformation. Thus, the total energy of a conformation E is given as follows:

$$E = \sum_{1 \le i \le j \le n} U(\xi_i, \xi_j)\Delta_{ij}, \tag{1}$$

where $\Delta_{ij}=1$ if monomers i and j are lattice neighbors and $\Delta_{ij}=0$ otherwise. Now, ξ_i defines the type of amino acid residue in position i, and $U(\xi_i,\xi_j)$ is the magnitude of the contact interaction energy between the amino acid residues of types ξ_i and ξ_j. For $U(\xi_i,\xi_j)$, in accordance with Shakhnovich and co-workers [18,19], we used the values determined from the statistical distributions of contacts in real proteins by Miyazawa and Jernigan, as reported in Table 6 of Ref. [30]. Furthermore, the same energy parameter set was used in the A-W model.

We examined four proteins referred to as $a1$, $a2$, $b1$, and $b2$. The number of amino acid residues n is 36 for proteins $a1$ and $a2$ and 48 for proteins $b1$ and $b2$. Their amino acid sequences are given as follows:

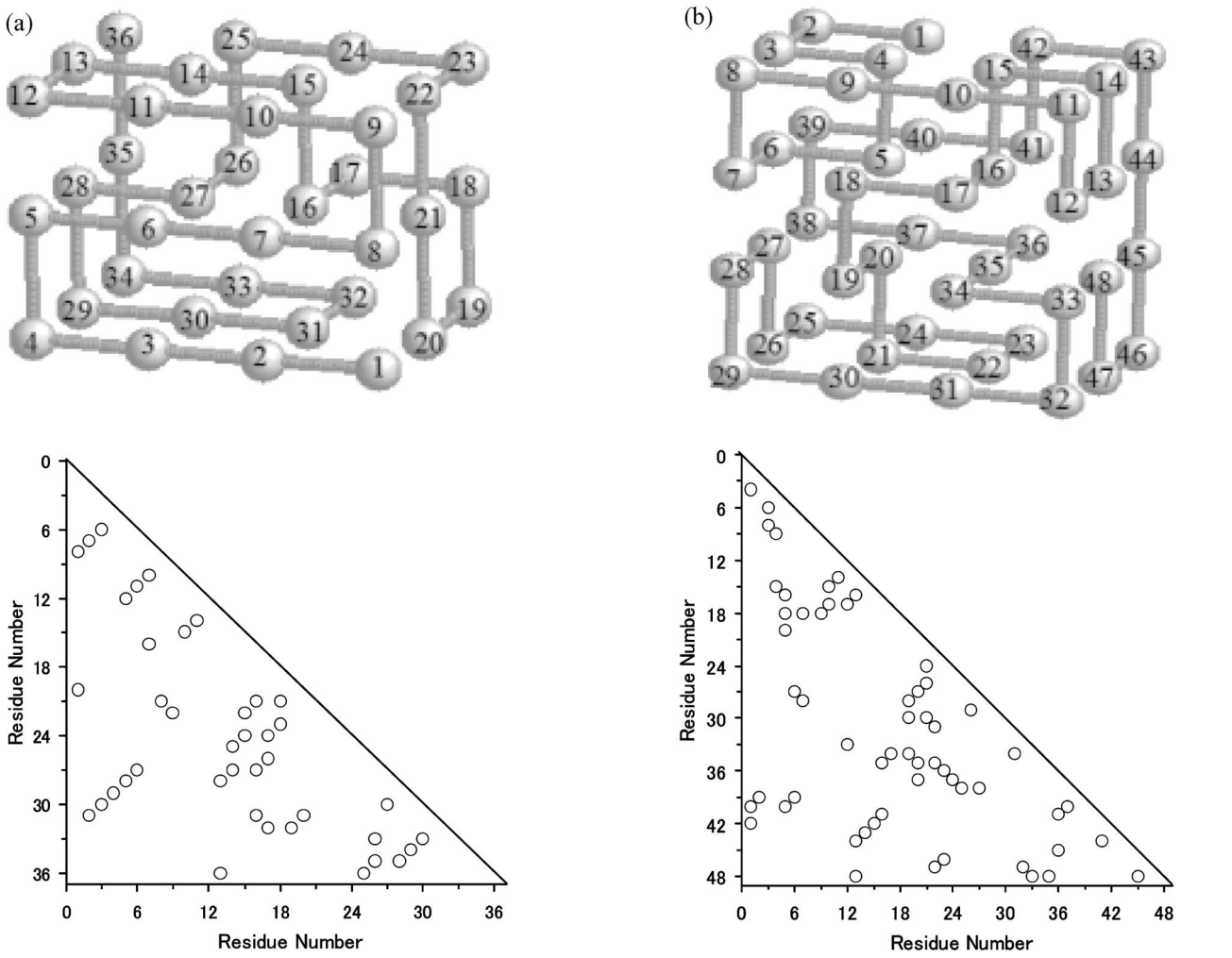

FIG. 1. Two native conformations of 3D cubic lattice proteins (top) and their contact maps (bottom). Two foldable amino acid sequences were designed for each conformation by Shakhnovich and co-workers. Proteins $a1$ and $a2$ [18] are folded into conformation (a) and $b1$ and $b2$ [19] into (b) (see text for their amino acid sequences). In the contact maps, the amino acid residue pairs in contact in the native conformations are indicated by an open circle. The total number of contacts (N_{total}) in these native conformations is 40 for proteins $a1$ and $a2$ and 57 for proteins $b1$ and $b2$.

$\langle a1\rangle$: NKTVVGEPWHCLLFPRRDKNQMSYLTGIAGEDSAAI;

$\langle a2\rangle$: SQKWLERGATRIADGDLPVNGTYFSCKIMENVHPLA;

$\langle b1\rangle$: TSKRQQPYPMSLGSPFIRIPMIGPRPRMRLLILLMGYPKRGRSGGGLF;

$\langle b2\rangle$: TEKGEEGYGGAAWTGPTSYKMAIYVWTTMWIYWAWAEAKKYGAYWAYM.

The amino acid sequences of $a1, a2$ and $b1, b2$ were designed to be folded into each native conformation as shown in Fig. 1(a) [18] and Fig. 1(b) [19], respectively, by Shakhnovich and co-workers. The native conformation of $a1$ and $a2$ is composed of two well-defined domains 1-12 and 13-36. The hairpin-type antiparallel β sheets of residues 1-12, 10-15, and 29-34 are characteristic. On the other hand, the native conformation of $b1$ and $b2$ is more complicated than $a1$ and $a2$, and their domains cannot be clearly defined. There are three extended strands 8-11, 29-32, and 43-46 that do not form a β sheet. The turns in $b1$ and $b2$ are greater than those in $a1$ and $a2$ (10 and 6, respectively). Therefore, it is expected that the folding of $b1$ and $b2$ is more complicated than that of $a1$ and $a2$.

We carried out the folding and unfolding simulations of these lattice proteins by using the MC method of Metropolis *et al.* [31]. Excluding very low temperatures, the simulations began from a fully extended conformation at various temperatures. At very low temperatures, we used the native conformation as the starting conformation because random conformations were easily trapped in the local minima from which it is difficult to escape in the simulations. We calculated various equilibrium thermodynamic quantities for folding and unfolding from the simulation records of 10^8 MC steps (10^7 MC steps for very low temperatures) [29].

B. Thermodynamic quantities characterizing folding

Several quantities that characterize protein folding are defined in this study. Since there is no difference between conformational energy and enthalpy, we simply regard energy as enthalpy. The first characteristic is the degree of folding $\theta_h^{\mathrm{sim}}(T)$ in the simulation at a temperature T. This degree of folding is defined as the normalized average enthalpy as follows:

$$\theta_h^{\mathrm{sim}}(T) = \langle E_h \rangle(T)/E_{\mathrm{native}}, \tag{2}$$

where E_h is the enthalpy, $\langle \cdots \rangle$ indicates the average over all conformations generated in the simulation, and E_{native} is the enthalpy of the native conformation. Further, $\theta_h^{\mathrm{sim}}(T)=1$ for the native conformation and $\theta_h^{\mathrm{sim}}(T)=0$ for the fully extended conformation.

The second characteristic is the fraction of the number N_c of both NCs and non-NCs, $C=\langle N_c \rangle/N_{\mathrm{total}}$. The third characteristic is the fraction of the number N_{native} of NCs, $Q=\langle N_{\mathrm{native}} \rangle/N_{\mathrm{total}}$, which are also averaged over all the conformations generated in the simulation at a given temperature T and normalized by the number of contacts in the native conformation N_{total}. Accordingly, $N_c - N_{\mathrm{native}}$ gives the number of non-NCs. In this study, $N_{\mathrm{total}}=40$ for proteins $a1$ and $a2$ and $N_{\mathrm{total}}=57$ for proteins $b1$ and $b2$ (see Fig. 1).

The fourth characteristic is the fraction of the number n_{res} of amino acid residues in the native state, $\eta=n_{\mathrm{res}}/(n-3)$. Since each amino acid residue is only a monomer unit in the 3D lattice protein, we consider an amino acid residue i ($i=2,\ldots,n-2$) to be in the native state if four continuous amino acid residues $(i-1)$ to $(i+2)$ form the same local conformation as in the native conformation, i.e., a local structure as defined below. Otherwise, the residue is considered to be in the random-coil state. Accordingly, the normalizing factor is $n-3$ rather than n. The characteristics C, Q, and η are regarded as progress variables for the protein folding reaction [2].

In addition, the probability p_i of an amino acid residue i being in the native state, enthalpy $H(\eta)$, entropy $S(\eta)$, free energy $F(\eta)$ for a given η value, and the free energy surfaces $F(Q,C)$ for given (Q,C) values are also useful characteristics for describing the protein folding.

C. Simplified statistical mechanical model for protein folding

In order to reveal the protein folding mechanism, it is convenient to use the simplified statistical mechanical model (the A-W model). Since the detailed description and formulation of the model are given elsewhere [29], we have summarized only the characteristic features of the model in this study (also see Fig. 2).

(1) Each amino acid residue is assumed to be in either of the two states, namely, native and random-coil states.

(2) Protein conformation at any stage of the folding process is represented by a sequence of two types of regions of various sizes, namely, a local structure and a random-coil region, arranged alternately along the chain. The local structure and the random-coil region are defined as continuous regions in which all amino acid residues are in the native state and the random-coil state, respectively.

(3) The key assumption of the model is that the interaction energy between two amino acid residues that are lattice neighbors is considered only when they are within the same local structure. The interaction energies for other amino acid residue contacts such as those between amino acid residues in different local structures and those between amino acid residues in random-coil regions are considered to be negligible. For example, in case of the conformation shown in Fig. 2, only the interaction energy for contacts within the same local structure (intra-N interactions) is taken into account while the other interactions (inter-N interactions) are ignored in the A-W model.

(4) The chain entropy for the 3D lattice protein is evaluated by explicitly counting possible conformations with the excluded volume effect for a small-sized segment. Then, the approximated formula for a segment of any size is extrapolated from the results of small-sized segments [29]. Ultimately, we can calculate the partition function without any adjustable parameter.

In the A-W model, the partition function for the present system can be finally represented as a function of temperature T by using a simple recurrent equation as follows [29]:

$$Z(T) = \sum_h W(E_h)\exp(-\beta E_h), \tag{3}$$

where $W(E_h)$ is the number of conformations at a given enthalpy E_h. Therefore, $k_B \ln W(E_h)$ is the entropy of a set of conformations with enthalpy E_h. For $\beta=1/k_BT$, we absorb the Boltzmann constant k_{B} into T, i.e., k_{B} is set to unity hereafter for convenience.

From the partition function, we can evaluate the probability p_i of an amino acid residue i in the native state. The degree of folding $\theta_h^{\mathrm{th}}(T)$ is defined as the normalized average enthalpy as follows:

$$\theta_h^{\mathrm{th}}(T) = \langle E_h \rangle(T)/E_{\mathrm{native}},$$

$$\langle E_h \rangle(T) = \sum_h E_h W(E_h)\exp(-E_h/T)/Z(T). \tag{4}$$

The partition function in the A-W model can also be rewritten as follows:

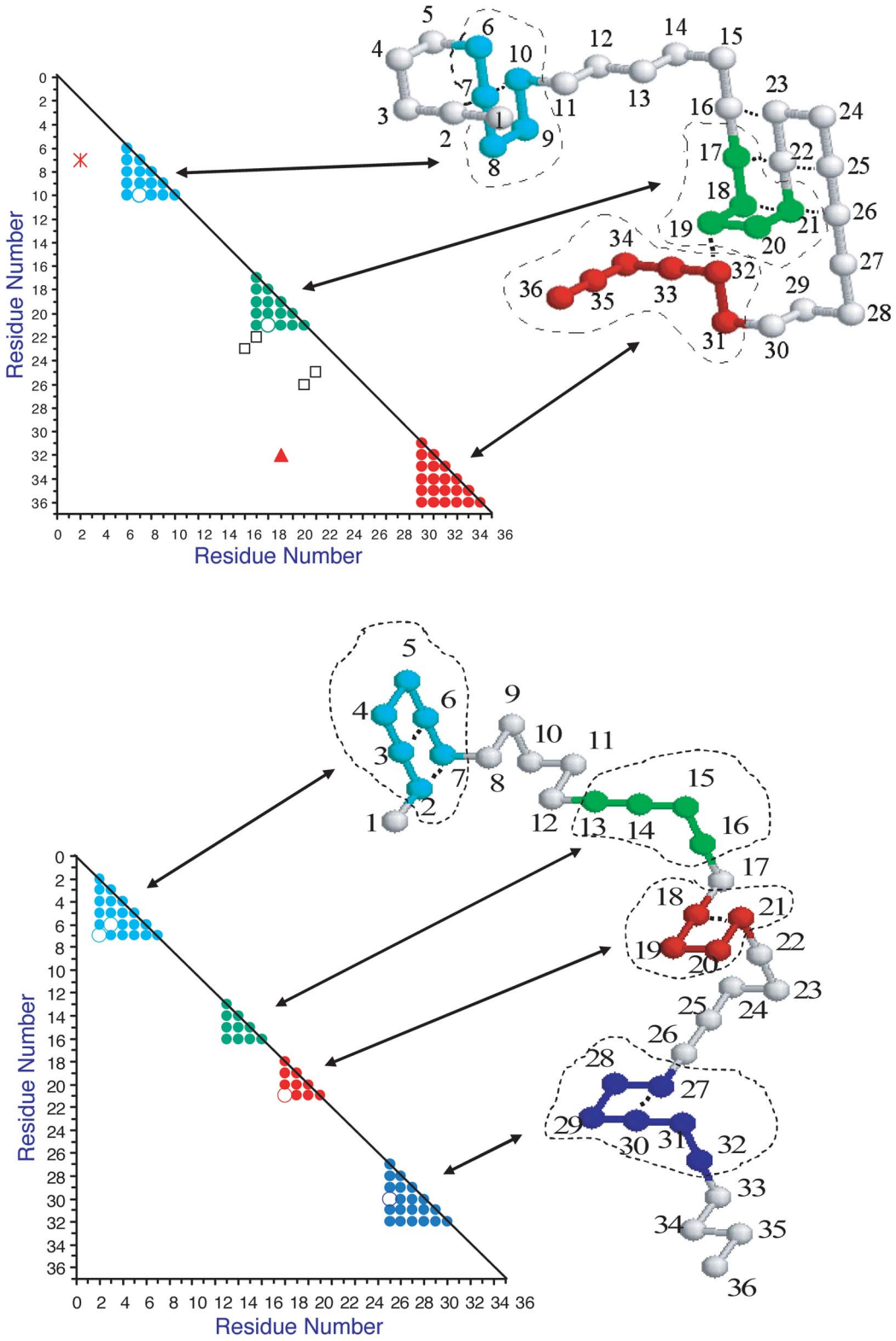

FIG. 2. (Color online) Two snapshots of the conformations generated in the simulation of protein $a1$ illustrate the existence of several local structures and various types of amino acid residue contacts. The local structures are the enclosed regions indicated by the dashed lines in the molecular models, and the triangular regions in the triangle maps correspond to them. The others are the random-coil regions. Top: A conformation with both native contact (NC) and non-native contact (non-NC) pairs. The NC pairs are classified into the following four classes: (1) intra-*N* (the amino acid residue pairs 7-10 and 18-21), (2) *N*-*N* (19-32), (3) *N*-*D* (2-7), and (4) *D*-*D* (no such pair found). In this case, the non-NC pairs are 16-23, 17-22, 21-26, and 22-25. In the *A*-*W* model, only the interaction energies for the intra-*N* are considered while the other interactions are neglected. Bottom: A conformation only with intra-*N* (2-7, 3-6, 18-21, and 27-30 in this case).

$$Z(T) = \sum_{\eta} z(\eta, T), \tag{5}$$

where $z(\eta,T)$ is the sum of the statistical weights for all states with a fraction η of the number of amino acid residues in the native state at a temperature T. By using Eq. (5), we can obtain the enthalpy $H(\eta)$, entropy $S(\eta)$, and free energy $F(\eta)$ as functions of η for the A-W model. The differences between these thermodynamic quantities in the simulations and the A-W model are discussed below.

D. Classification of contact energies

In order to analyze the records of the MC simulations, we classify the total interaction energy E into the following two types: E_{NC} and $E_{\text{non-NC}}$ for NC and non-NC, respectively (i.e., $E=E_{\mathrm{NC}}+E_{\text{non-NC}}$; see Fig. 2, for example). Furthermore, E_{NC} is divided into the following four types of energies: $\varepsilon_{\text{intra-}N}$ for amino acid residue pairs within the same local structure (intra-N), $\varepsilon_{N\text{-}N}$ for those in different local structures (N-N), $\varepsilon_{N\text{-}D}$ for one amino acid residue in the local structure and another in the random-coil region (N-D), and $\varepsilon_{D\text{-}D}$ for the other cases, i.e., for those in the same or different random-coil regions (D-D) (i.e., $E_{\mathrm{NC}}=\varepsilon_{\text{intra-}N}+\varepsilon_{N\text{-}N}+\varepsilon_{N\text{-}D}+\varepsilon_{D\text{-}D}$). Further, $\varepsilon_{N\text{-}N}$, $\varepsilon_{N\text{-}D}$, and $\varepsilon_{D\text{-}D}$ are collectively referred to as an inter-N interaction.

The same notation is used hereafter for representing various contact types between amino acids, i.e., there are two types of contacts: NC and non-NC. Further, NC is divided into the intra-N and inter-N; inter-N comprises three types of contacts, i.e., N-N, N-D, and D-D.

III. RESULTS

A. Folding transition and occurrence of non-NC interactions

In this section, the contributions of the NC and non-NC interactions at various folding stages in the simulations are examined in detail by comparing them with those in the A-W model.

Figure 3 shows the equilibrium transition curves $\theta_h^{\mathrm{sim}}(T)$ and $\theta_h^{\mathrm{th}}(T)$ for the four proteins $a1$, $a2$, $b1$, and $b2$. It should be emphasized here that no adjustable parameter was used in the A-W model to fit $\theta_h^{\mathrm{th}}(T)$ to $\theta_h^{\mathrm{sim}}(T)$. The transition temperatures T_m (defined in this study as the temperature at which $\theta_h^{\mathrm{sim}}=0.5$) obtained from the simulations are in good agreement with those obtained from the A-W model for proteins $a1$ and $b1$; however, they deviated slightly for $a2$ and $b2$. Although the transition curves are in good agreement with each other at lower and higher temperatures, their differences are significant at around T_m. The transition curves $\theta_h^{\mathrm{th}}(T)$ are steeper than $\theta_h^{\mathrm{sim}}(T)$. In particular, $\theta_h^{\mathrm{sim}}(T)$ for $a2$ and $b2$ is considerably more gradual than $\theta_h^{\mathrm{th}}(T)$ for these proteins; this shows that the transitions proceed through two phases.

In order to understand the reason why $\theta_h^{\mathrm{sim}}(T)$ and $\theta_h^{\mathrm{th}}(T)$ differed at around T_m, we examined the types of interaction energy between the amino acid residues that work in the simulation. The interaction energies for all the types of contacts were considered in the simulation, whereas the inter-N and non-NC interactions were not considered in the A-W model.

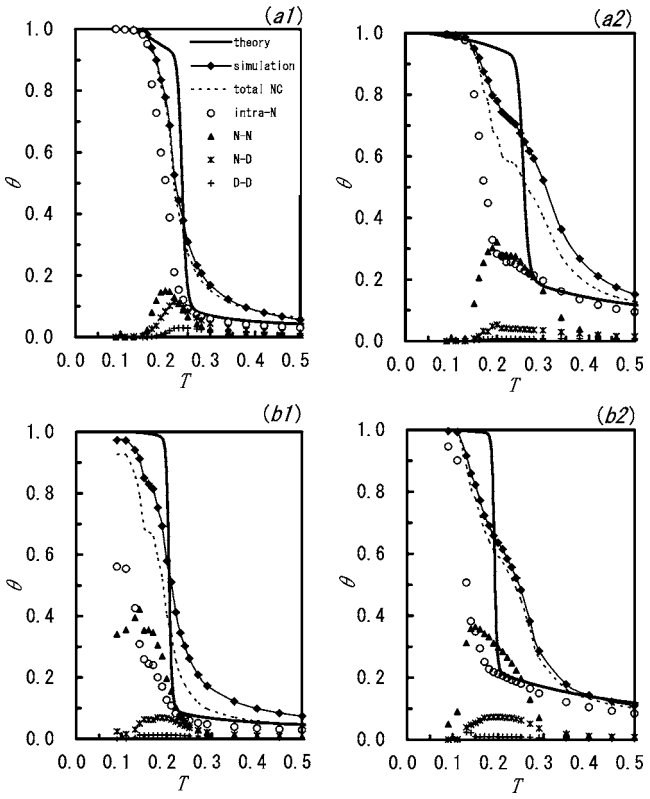

FIG. 3. Equilibrium transition curves $\theta_h^{\mathrm{th}}(T)$ and $\theta_h^{\mathrm{sim}}(T)$ obtained from the A-W model (theory) and the simulations (simulation), and the normalized average enthalpies $\theta_{\mathrm{NC}}^{\mathrm{sim}}(T)$, $\theta_{\text{intra-}N}^{\mathrm{sim}}(T)$, $\theta_{N\text{-}N}^{\mathrm{sim}}(T)$, $\theta_{N\text{-}D}^{\mathrm{sim}}(T)$, and $\theta_{D\text{-}D}^{\mathrm{sim}}(T)$ for all native contacts (total NC) and for the four types of native contacts, i.e., intra-N, N-N, N-D, and D-D, respectively, obtained from the simulations for the proteins $a1$, $a2$, $b1$, and $b2$ as functions of temperature T. The difference between $\theta_h^{\mathrm{sim}}(T)$ and $\theta_{\mathrm{NC}}^{\mathrm{sim}}(T)$ indicates the normalized average enthalpy for the non-NC $[\theta_{\text{non-NC}}^{\mathrm{sim}}(T)]$.

Figure 3 also shows the normalized average enthalpy for all native contacts (total NC) and the four types of native contacts (intra-N, N-N, N-D, and D-D) in the simulations. They are defined as follows:

$$\theta_{\mathrm{NC}}^{\mathrm{sim}}(T) = \langle E_{\mathrm{NC}} \rangle(T)/E_{\mathrm{native}}, \quad \theta_{\text{intra-}N}^{\mathrm{sim}}(T) = \langle \varepsilon_{\text{intra-}N} \rangle(T)/E_{\mathrm{native}},$$

$$\theta_{N\text{-}N}^{\mathrm{sim}}(T) = \langle \varepsilon_{N\text{-}N} \rangle(T)/E_{\mathrm{native}}, \quad \theta_{N\text{-}D}^{\mathrm{sim}}(T) = \langle \varepsilon_{N\text{-}D} \rangle(T)/E_{\mathrm{native}},$$

$$\theta_{D\text{-}D}^{\mathrm{sim}}(T) = \langle \varepsilon_{D\text{-}D} \rangle(T)/E_{\mathrm{native}}. \tag{6}$$

The normalized average enthalpy for the non-NCs $[\theta_{\text{non-NC}}^{\mathrm{sim}}(T)]$ is given by $\theta_h^{\mathrm{sim}}(T)-\theta_{\mathrm{NC}}^{\mathrm{sim}}(T)$; however, this is not shown explicitly in Fig. 3. Therefore, the very small differences between $\theta_h^{\mathrm{sim}}(T)$ and $\theta_{\mathrm{NC}}^{\mathrm{sim}}(T)$ for proteins $a1$ and $b2$ suggest very small $\theta_{\text{non-NC}}^{\mathrm{sim}}(T)$ values. This indicates that the non-NCs in the conformations generated in the simulations are very rare. In other words, the assumption of the Gō-type interactions holds true with regard to these proteins. On the contrary, for proteins $a2$ and $b1$, $\theta_{\text{non-NC}}^{\mathrm{sim}}(T)$ is not negligible for a wide temperature range.

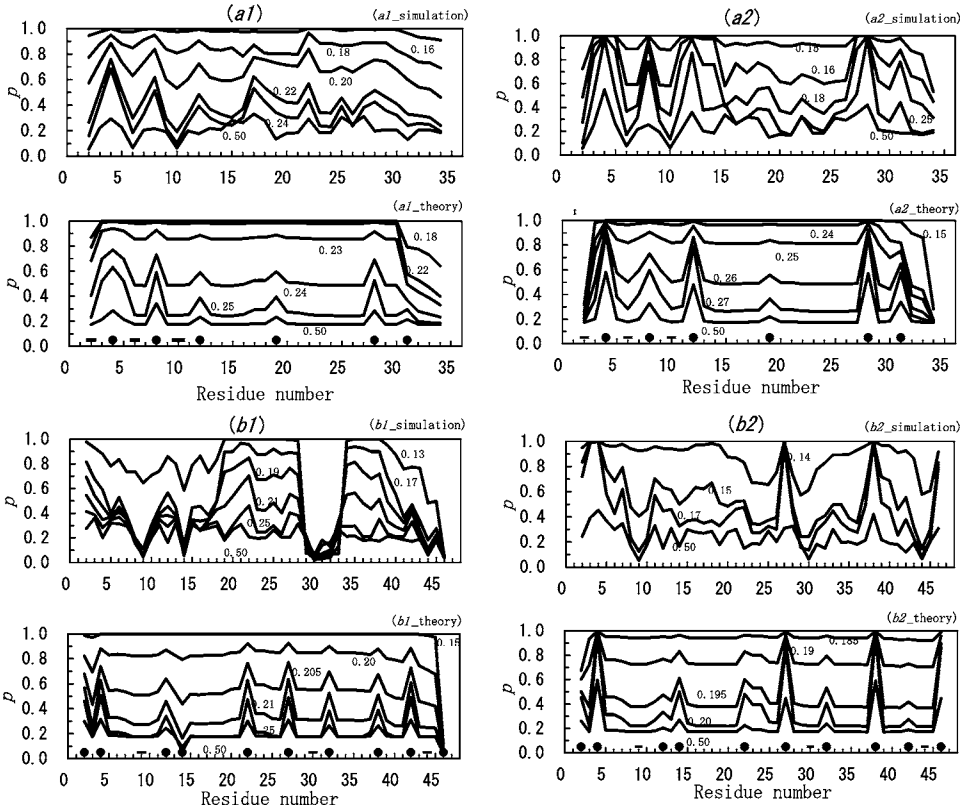

FIG. 4. Probability p_i of an amino acid residue i being in a native state obtained from the simulations and the A-W model at several temperatures for proteins $a1$, $a2$, $b1$, and $b2$. The temperatures are shown near the corresponding lines. The turns and extended regions in the native conformations are indicated by symbols ● and -, respectively (see also Fig. 1).

As shown in Fig. 3, $\theta_{\text{intra-}N}^{\text{sim}}$ agrees remarkably well with θ_h^{th} above T_m for any protein, whereas $\theta_{\text{intra-}N}^{\text{sim}}$ and θ_h^{th} differ considerably below T_m. These results imply that the intra-N interaction dominates above T_m. It should also be noted that $\theta_{\text{intra-}N}^{\text{sim}}$ and θ_h^{th} for proteins $a2$ and $b2$ are considerably larger than those for $a1$ and $b1$ above T_m. This fact indicates that some stable small local structures are formed in proteins $a2$ and $b2$ above T_m, whereas the chain is essentially in the random-coil state in $a1$ and $b1$.

As the temperature is lowered, $\theta_{N\text{-}N}^{\text{sim}}$ begins to increase immediately above T_m for any protein. In accordance with these changes in $\theta_{N\text{-}N}^{\text{sim}}$, the differences between $\theta_{\text{intra-}N}^{\text{sim}}$ and θ_h^{th} increase. It should be noted that $\theta_{N\text{-}N}^{\text{sim}}$ is much smaller for protein $a1$ than for proteins $a2$, $b1$, and $b2$. On the other hand, $\theta_{N\text{-}D}^{\text{sim}}$ and $\theta_{D\text{-}D}^{\text{sim}}$ are small for any protein. For proteins $a1$ and $b1$, the NCs between residues that are not within the same local structure increase below T_m; for $a2$ and $b2$, such contacts begin to form above T_m. This is most likely because some local structures are already formed at higher temperatures. Consequently, these analyses indicate that $\theta_{N\text{-}N}^{\text{sim}}$ is mainly responsible for the differences between $\theta_h^{\text{sim}}(T)$ and $\theta_h^{\text{th}}(T)$ below and immediately above T_m.

B. Probability of an amino acid residue being in the native state

In Fig. 4, the probability p_i of an amino acid residue i being in the native state is plotted against amino acid residue numbers at several temperatures for both the simulations and the A-W model for the four proteins. Since $\theta_h^{\text{sim}}(T)$ and $\theta_h^{\text{th}}(T)$ at the same temperature (at around T_m) differ significantly, as shown in Fig. 3, their p_i profiles also differ if they are compared at the same temperature. However, in this study, we intend to show that the p_i profiles for the simulations and the A-W model change with temperature. Accordingly, the selected p_i profiles are not at the same temperature, but possess similar features in Fig. 4. (Note: The regression correlation coefficients between p_i for the simulations and the A-W model at identical temperatures near T_m are 0.41, 0.80, 0.31, and 0.75 for proteins $a1$, $a2$, $b1$, and $b2$, respectively.)

The peaks in each profile indicate that at a given temperature the amino acid residues at the peaks are more likely to be in the native state than the other amino acid residues. In general, the locations of the peaks exhibit a good correspondence between the simulations and the A-W model. Closer agreements between the simulation and the A-W model are observed for the amino acid residues at the N and C terminals rather than for those in the middle region.

At high temperatures (lines with lower p_i values in Fig. 4), the turns are more likely to be in the native state. The peaks for $a2$ and $b2$ are remarkably higher than those for $a1$ and $b1$. This fact indicates that some small local structures, i.e., turns, are formed in proteins $a2$ and $b2$ at such high temperatures. The differences in p_i of these turns originate from the differences in their amino acid sequences. The favorable interactions between the residues $i-1$ and $i+2$ of turn i are obviously responsible for their higher p_i values (see Table I). This is consistent with the fact that $\theta_{\text{intra-}N}^{\text{sim}}$ and θ_h^{th} are larger for $a2$ and $b2$ and smaller for $a1$ and $b1$ at higher temperatures, as shown in Fig. 3.

On the other hand, the three extended structures 8-11, 29-32, and 43-46 in $b1$ and $b2$ exhibit very low p_i values as compared to 1-4, 5-8, and 9-12 in $a1$ and $a2$. While the latter

TABLE I. Interaction energy between amino acid residues forming a turn.

Residue pair in contact in a turn $i-1$ and $i+2$ (turn i)[a]	Interaction energy between $i-1$ and $i+2$	
	$a1$	$a2$
3, 6 (4)	−0.26	−0.97
7, 10 (8)	−0.45	−0.35
11, 14 (12)	−0.23	−0.72
18, 21 (19)	−0.17	−0.11
27, 30 (28)	−0.38	−0.97
30, 33 (31)	−0.16	−0.45
	$b1$	$b2$
1, 4 (2)	−0.35	−0.26
3, 6 (4)	−0.38	−0.97
11, 14 (12)	−0.20	−0.09
13, 16 (14)	0.38	−0.28
21, 24 (22)	−0.34	−0.13
26, 29 (27)	−0.38	−0.67
31, 34 (32)	−0.27	−0.22
37, 40 (38)	−0.25	−0.97
41, 44 (42)	−0.38	−0.06
45, 48 (46)	0.38	−0.67

[a] p_i in Fig. 4 is the probability that residues $i-1$ to $i+2$ form the same structure as the native one.

extended structures form a hairpin-type antiparallel β sheet, the former does not. Since the extended structure is not stable by itself, these regions in $b1$ and $b2$ tend to form other conformations rather than the extended structure. This makes the folding of $b1$ and $b2$ more complicated than that of $a1$ and $a2$.

As temperature decreases, the p_i values increase gradually and uniformly. At medium temperatures, the p_i profiles of the A-W model still show a good agreement with those of the simulations. However, in the simulations, the larger p_i values in the middle of the chain (e.g., the amino acid residues 17-25 in protein $a1$, 16-22 in $a2$, 19-28 in $b1$, and 11-23 in $b2$) indicate the formation of some local structures that are larger than the turns, whereas in the A-W model, these structures are still small in $a1$ and $a2$ and are localized only at the turn regions in $b1$ and $b2$. The latter fact may be related to the strong cooperativity (steeper transition curves in Fig. 3) in the A-W model for any protein.

The p_i profile of the simulation for $b1$ has attracted our attention. The p_i values in region 29-33 (this includes an extended region in the native conformation) are much smaller than those in other regions even at low temperatures, whereas those in regions 22-28 and 34-38 located on either side of the region 29-33 are larger. This indicates that a nonnative stable structure is formed in region 29-33, and it may be difficult to adjust to the correct conformation until the final stage of the folding. In fact, region 29-32 is an extended structure that does not form a β sheet similar to the three extended structures 1-4, 5-8, and 9-12 of proteins $a1$ and $a2$. Although the extended structure 29-32 can be stabilized in the native structure, it is unstable by itself. As a result, there is a high possibility that a wrong conformation will be formed. However, it should be noted that such a wrong conformation is formed in $b1$, but not in $b2$. It should be understood that this difference results from the difference in their amino acid sequences.

As the temperature is lowered further, every p_i value has a high probability of being in the native state; finally, protein folding is accomplished over all the regions for any protein.

C. Thermodynamic properties

Figure 5 shows the thermodynamic properties $H(\eta)$, $-TS(\eta)$, and $F(\eta)$ as functions of the progress variable η at several temperatures for the simulations and the A-W model for the four proteins. The enthalpy $H(\eta)$ decreases smoothly with η, indicating funnel-shaped energy landscapes toward the native conformation. The shapes of the curves for the simulations are essentially similar to those of the A-W model; however, $\Delta H(\eta)$ and $-T\Delta S(\eta)$ of the A-W model are considerably larger than those of the simulations, where Δ indicates the difference between the native and random-coil states.

In particular, the differences between the $F(\eta)$ curves in the simulations and the A-W model are considerably smaller than those between the $H(\eta)$ and $-TS(\eta)$ curves. This is because the differences in $H(\eta)$ are canceled out by the differences in $-TS(\eta)$. Further, $H(\eta)$ and $-TS(\eta)$ change drastically at $\eta > 0.7$. For any protein, the $F(\eta)$ curves of the A-W model have a notable peak at $\eta = 0.7-0.8$. These regions of η correspond to the transition state of the protein folding and unfolding, thereby indicating the two-state nature of these transitions. These peaks appear because $S(\eta)$ begins to decrease more drastically than $H(\eta)$ toward the native conformation. In addition, a metastable state exists at around $\eta = 0.6$ in the case of $b1$.

On the other hand, the $F(\eta)$ curves of the simulations do not have a distinct peak at around $\eta = 0.7-0.8$. However, small peaks are observed immediately below $\eta = 1.0$, and $F(\eta = 1.0)$ is significantly lower than $F(\eta)$ of the peaks. This is evident in Fig. 6 described below. However, the two-state nature of these transitions is ambiguous.

D. Free energy landscapes

Figure 6 shows the free energy surfaces $F(Q, C)$ plotted against the two characteristics of folding, Q and C, from the simulations of the four proteins at a temperature corresponding to $\theta_h^{\mathrm{sim}} = 0.6$ (immediately below T_m). The conformations on the diagonal region ($Q = C$) contain only native contacts, while those in the off-diagonal regions contain both NCs and non-NCs. The points $\partial F/\partial C = 0$ and $\partial F/\partial Q = 0$ (local minima) are indicated with different colors. The line of colored points indicates the most probable equilibrium path of folding. To a certain extent, we can visualize the folding process based on these folding paths. However, since a given

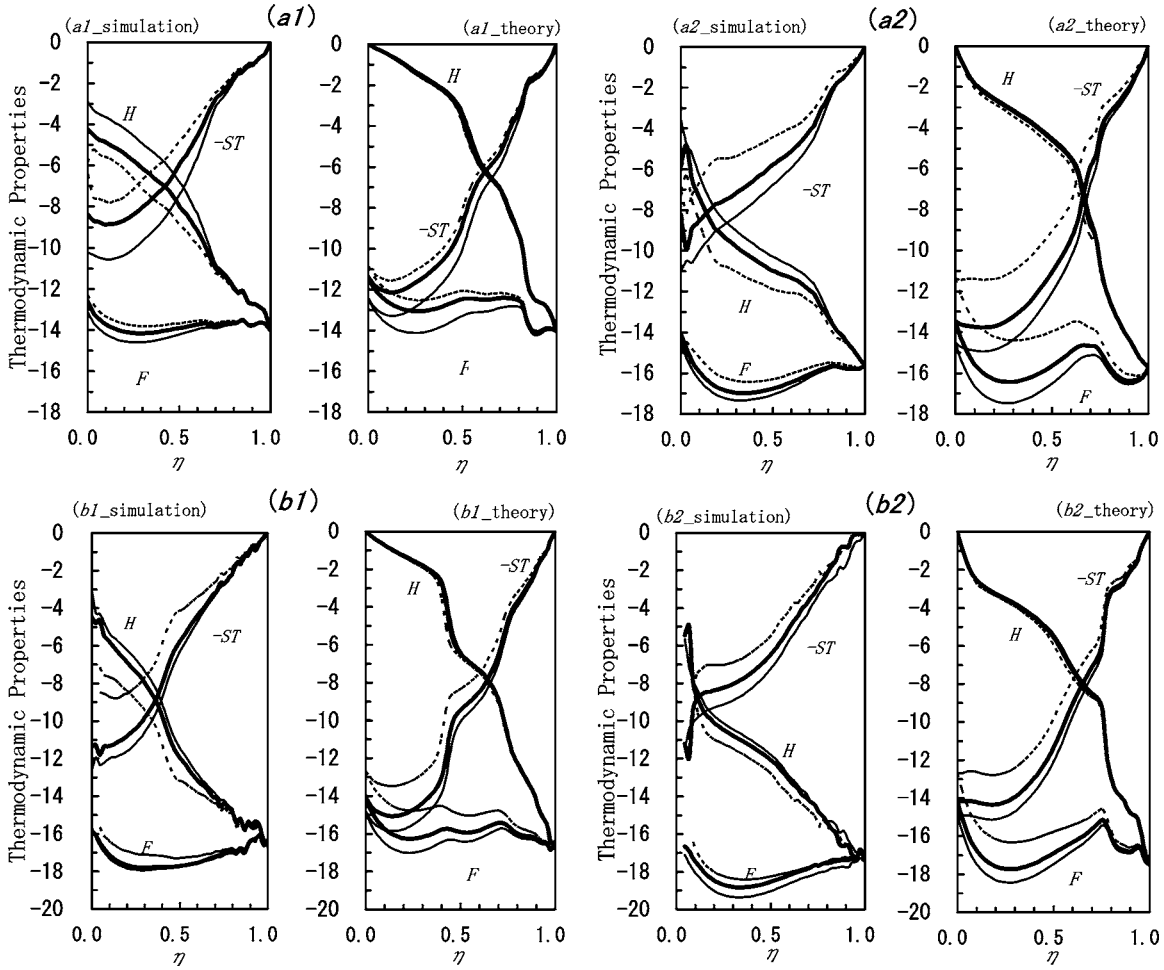

FIG. 5. Thermodynamic properties $H(\eta)$, $-TS(\eta)$, and $F(\eta)$ (in dimensionless k_BT unit) as functions of the progress variable η (fraction of the number of amino acid residues in the native state) obtained from the simulations and the A-W model for proteins $a1$, $a2$, $b1$, and $b2$. The thin solid, thick solid, and dotted lines are the profiles for T=0.24, 0.22, and 0.21 for $a1$; T=0.28, 0.26, and 0.22 for $a2$; and T =0.21, 0.20, and 0.18 for $b1$ and $b2$, respectively.

point (Q,C), as shown in Fig. 6, comprises many conformations classified according to the two parameters Q and C, it does not necessarily mean that only one unique pathway exists for the protein to proceed from the random-coil state to the native conformation. These paths should be understood in a statistical sense [2,4].

The valley bottoms are much deeper in $a2$ and $b1$ than in $a1$ and are intermediate in $b2$. These landscapes are consistent with the non-NCs indicated in Fig. 3 for the four proteins. By comparing Fig. 6 with Fig. 3, the contributions of the non-NCs were observed to be mainly entropic, because the latter (Fig. 3) indicates that the enthalpy of the non-NCs minorly contributes to the structural transitions even for proteins $a2$ and $b1$.

For proteins $a1$ and $b1$, the most probable equilibrium paths of folding lie in relatively flat regions, whereas for $a2$ and $b2$, they lie in the distinct minimum regions. For $a2$ and $b2$, the two paths of local minima cross each other, and their free energy landscapes are more complex than those of $a1$ and $b1$.

IV. DISCUSSION

A. NC and non-NC interactions

Gō [6] pointed out that various energy terms that contribute toward stabilizing the native state of globular proteins are consistent with each other in the native state. This is referred to as the consistency principle. Later, this principle was reproposed by Bryngelson and Wolynes [32] as the principle of minimum frustration with regard to a funnel-shaped energy landscape [10,11]. The most extreme case satisfying these principles is the Gō model for the simulation and the A-W model for the simplified theoretical model presented in this study, in which only the NC interactions are taken into account.

In this paper, we study lattice proteins with carefully designed amino acid sequences in order to allow their folding into unique native conformations. Since any pair of amino acid residues that are lattice neighbors can interact with each other, the occurrence of non-NC interactions is inevitable in the simulations. In such a situation, one of our interests was to determine the contribution of non-NC interactions to the

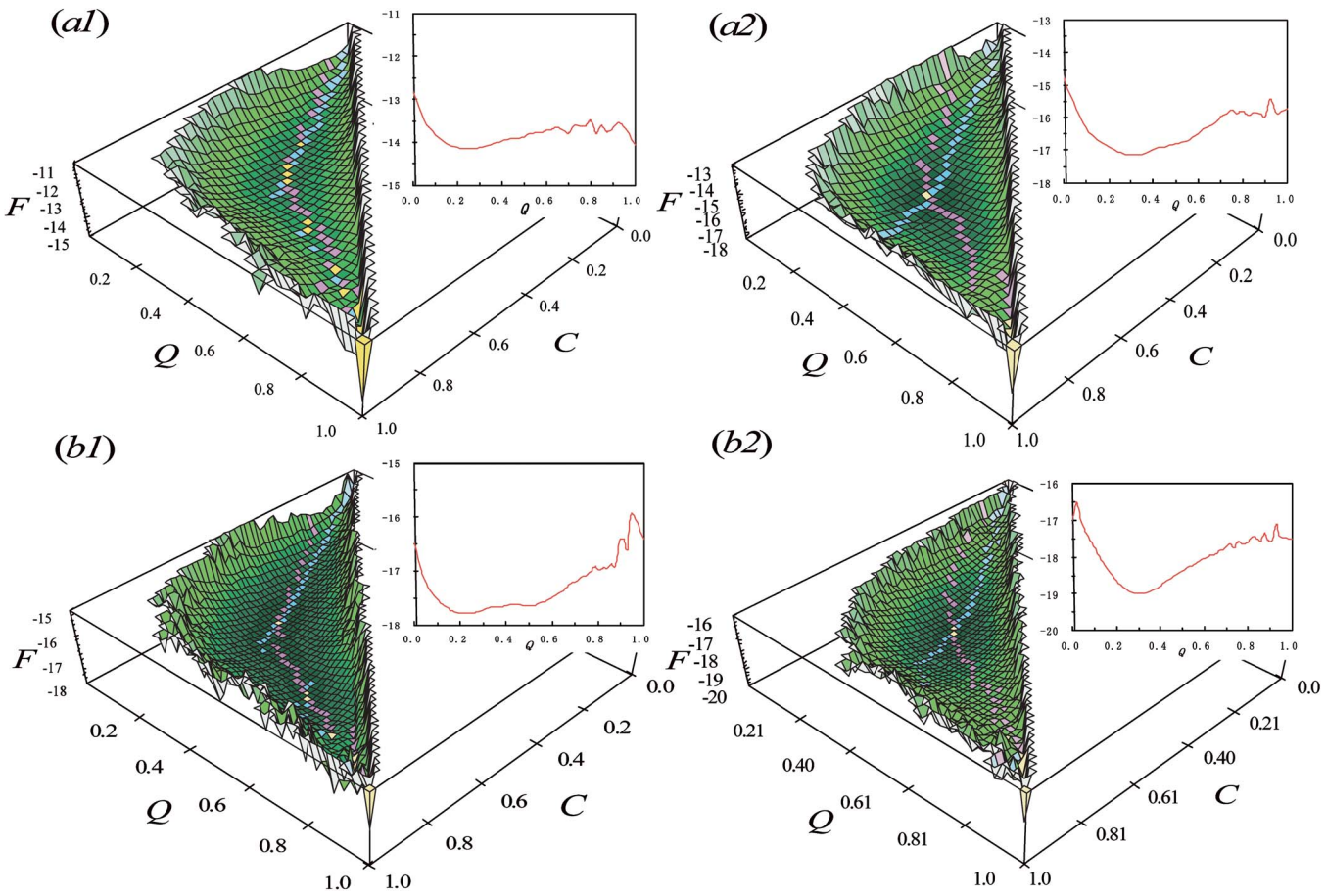

FIG. 6. (Color online) Free-energy $F(Q,C)$ (in dimensionless k_BT unit) plotted against the two characteristics of folding Q (fraction of the number of native contacts) and C (fraction of the number of both native and non-native contacts) obtained from the simulations for proteins $a1$, $a2$, $b1$, and $b2$ at the following temperatures that correspond to $\theta_h^{\mathrm{sim}}=0.6$: $T=0.22$ for $a1$, $T=0.26$ for $a2$, and $T=0.20$ for $b1$ and $b2$. For the native conformation, $(Q,C)=(1,1)$. The light blue and pink points indicate the local minima $\partial F/\partial C=0$ and $\partial F/\partial Q=0$, respectively. The yellow points show the coincident points of both the local minima. These points indicate the most probable equilibrium paths of the folding. In the inset boxes, free energies $F(Q)$ are plotted only as a function of Q. The swelling at $Q=0.9$ may be a lattice artifact due to the very small number of conformations [2].

folding process. Figure 3 shows that the non-NC interactions are negligible in proteins $a1$ and $b2$ and that although they occur in proteins $a2$ and $b1$, their (enthalpic) contributions are much smaller than those of the other interactions. In Fig. 6, however, the off-diagonal regions, which correspond to the conformations containing non-NC interactions, indicate that these contacts occur to a considerable extent. Therefore, it may be deduced that the contribution of non-NC interactions to the free energy is mainly entropic rather than enthalpic.

Furthermore, in addition to the intra-N interactions, since various interactions such as the non-NC and inter-N interactions exist between amino acid residues within the random-coil region and in the different local structures, the conformations in the denatured state may be more enthalpically favorable in the simulation than in the A-W model. These are the reasons why $\Delta H(\eta)$ and $-T\Delta S(\eta)$ in the A-W model are larger than those in the simulations shown in Fig. 5.

B. Intra-N and inter-N interactions

The assumption that the non-NC and inter-N interactions are negligible or in other words, the assumption that only the intra-N interactions exist enables the partition function of the A-W model to be calculated exactly; this is a remarkable characteristic of this model [20–23,29]. No *a priori* assumption is made for stabilizing specific local structures such as the α helix and β strand. Since all the possible local structures are considered, more stable local structures such as α helix and β strand are recognized in a statistical mechanical sense.

However, the A-W model will be unsuccessful in cases where the non-NC and inter-N interactions are dominant in many events of the folding process. The diffusion and collision mechanism by Karplus and Weaver [33] and the nucleation-condensation mechanism by Fersht [34], which involve various interactions such as non-NC and inter-N interactions, appear to contradict the basic assumption made in the A-W model. We are now concerned with the N-N interactions; this is important from the viewpoint of differences between the simulation and the A-W model. As shown in Fig. 3, the N-N interaction is not a rare occurrence as indicated by the large $\theta_{N\text{-}N}^{\mathrm{sim}}$ values below and near T_m for any protein. In fact, we have obtained better equilibrium transition curves $\theta_h^{\mathrm{th}}(T)$ for proteins $a1$ and $a2$ by modifying the A-W model

so as to introduce a part of the *N-N* interaction energies. The modified algorithm is so primitive and tentative that it requires a large amount of CPU time and is difficult to apply to proteins of any size. Therefore, the data are not shown here.

Since the NC interactions are usually favorable in the simulation, the *N-N* interactions are likely to form wrong conformations; this may consequently interfere with the correct folding. From the viewpoint of polymer physics, it is considered to be entropically difficult for two amino acid residues separated by many amino acid residues along the chain to accidentally make contact with each other unless all the amino acid residues intervening between the two amino acid residues form a part of the native conformation. It is necessary to carefully consider the possibility that such types of NC interactions may be exaggerated in the lattice protein. In a more realistic protein model, such types of NC interactions may be extremely rare owing to the much wider conformational space; this results in a much smaller contribution to the folding process. Thus, we expect the assumption that the NC interactions within the same local structure are dominant in the folding process is rationalized to the first approximation.

C. Transition temperature and chain entropy

In the *A-W* model, the chain entropy of the segment consisting of, for example, m amino acid residues is estimated by explicitly counting possible conformations of the segment for small m. The excluded volume effect is automatically taken into account. The good agreement of the transition temperatures of the *A-W* model with those of the simulations arises from this proper estimation of the chain entropy. In fact, we first used the α^n form for the chain entropy of n residues in the *A-W* model and tried to adjust the parameter α in order to fit the two transition curves of the *A-W* model and the simulation. However, the transition temperatures of the two transition curves never approached each other for any value of α. Therefore, an excluded volume effect is probably essential for estimating the chain entropy. In addition, the *A-W* model is correct at higher temperatures, where small local structures exist independently from each other. This fact also contributes to the good agreement of the transition temperatures.

However, this estimation of the chain entropy is not sufficiently correct. The interactions between amino acid residues attached to both terminals of the segment are not taken into account while counting conformations. In other words, the excluded volume effect of the attached amino acid residues is not considered; hence, the chain entropy obtained in this manner is overestimated.

Unfortunately, we cannot draw the free energy surface in the *A-W* model corresponding to Fig. 6. In the off-diagonal regions of this figure, the entropic contribution is dominant; however, the enthalpic contribution is not negligible in the simulation, as presented in the Sec. III. On the other hand, because it is included implicitly in the chain entropy term, only the entropic contribution to the free energy exists in these regions in the *A-W* model.

D. Transition states

The transition state of the protein folding and unfolding is located at around $\eta=0.7-0.8$, as shown by the free energy $F(\eta)$ as a function of progress variable η for the *A-W* model in Fig. 5. Therefore, the conformations in the transition state are rather close to those of the native state. The entropy and enthalpy change drastically around the transition state, and the conformations in the native state and in the denatured state can be distinctly characterized by these thermodynamic properties. These facts are obviously characteristic of the two-state nature of the structural transition of protein. However, in the case of $b1$, a metastable state is indicated at around $\eta=0.5$.

On the other hand, the transition state region in the simulations appears to be located very close to the native conformation, as seen in Figs. 5 and 6. However, it is not clear if the proteins in the simulations exhibit the two-state nature. It is probably because the proteins considered in this study are too small to exhibit the two-state nature. If a protein is larger, it becomes more two-state-like; thus, in case of large-sized proteins, thermodynamic quantities for the *A-W* model may approximate more those for the simulation.

E. Differences in amino acid sequences

In this study, two amino acid sequences that are foldable into the same native conformation have been examined. The difference in their amino acid sequences is reflected in the formations of local structures at high temperatures. According to Fig. 3, the stable local structures, i.e., turns, formed at high temperatures in $a2$, are characteristic as compared to those of $a1$. The *N-N* and non-NC interactions at around T_m are mainly responsible for the differences between $a1$ and $a2$. In other words, the formation of the more stable turns in $a2$ induces the *N-N* and non-NC interactions. Consequently, protein $a2$ shows a considerably different folding behavior than protein $a1$. For a comparison of the results of the simulations and the *A-W* model, protein $a1$ is better described by the *A-W* model.

The difference between proteins $b1$ and $b2$ is also reflected in the formations of local structures at high temperatures. Stable local structures (turns) are formed at high temperatures in $b2$. This is similar to $a2$. However, the formation of the turns in $b2$ induces the *N-N* interactions, but not the non-NC ones. On the other hand, in $b1$, non-native small structure formed incorrectly around the residues 29-34 may possibly interfere with the folding (see simulation of $b1$ in Fig. 4). If this interference is resolved by a proper amino acid substitution, the folding of protein $b1$ substituted may be described better by the *A-W* model.

Although we have discussed only the equilibrium properties in this study, the kinetics of protein folding and unfolding can be presumed from the MC simulation data to some extent. For example, protein $a1$ was folded more quickly than $a2$. This is most likely because the intra-*N* interaction is dominant in $a1$, whereas a considerable amount of *N-N* interactions occur in $a2$; the *N-N* interactions may interfere with the folding.

The better agreement between the simulation and the *A-W* model for $a1$ rather than $a2$ can be explained by the fact that $a1$ satisfies the consistency principle better than $a2$; this is stated from the viewpoint of physics. However, the ques-

tion of biological importance is whether $a1$ is more proteinlike than $a2$. If biological evolution exerts a pressure on a protein so that it folds rapidly, protein $a1$ would be more proteinlike. Further, such a protein will be well described by the A-W model.

V. CONCLUSION

The NC and non-NC interactions at various folding stages in the simulations of 3D lattice proteins have been examined by comparing them with the interactions in the simplified statistical mechanical model (the A-W model). In conclusion, the contribution of the non-NC interactions is small, and the A-W model (in spite of the simplified statistical mechanical model) yields equilibrium thermodynamic properties that are quite similar to those of one of the four proteins considered (protein $a1$). This is because the NC interactions, particularly the NC interaction within the same local structure (intra-N), are dominant in the folding process.

Furthermore, from the detailed comparisons of various equilibrium thermodynamic quantities obtained from the simulations with those from the A-W model described in this study, the degree of agreement between the results from the two methodologies is observed to be significant. This degree of agreement is important not only for the assessment of the A-W model as a simplified statistical mechanical model for describing a "general lattice polymer" but also for the assessment of the lattice polymer with a given amino acid sequence as a "protein" to be folded quickly and smoothly.

Although the MC simulation is a powerful method for the study of protein folding, it requires a large amount of computation time. On the other hand, by using a statistical mechanical model (the A-W model), we can more rapidly and systematically calculate various thermodynamic quantities characterizing folding using the available information on interacting amino acid pairs and their interaction energies in the native conformation. The number of operations for a protein of n residues is proportional to n^3, and the working space is proportional to n^2. Since the maximum value of n is usually several hundreds, the CPU time is not crucial for calculating the partition function and ultimately various thermodynamic quantities by using the A-W model.

Using this available information, we have also exhaustively examined the effects of various amino acid substitutions at arbitrary sites on the folding process by using the A-W model. The results will be published elsewhere.

As compared to general polymers, elucidating the characteristic feature of protein folding is an interesting problem to be solved. The study presented here will contribute to finding solutions to these types of questions.

ACKNOWLEDGMENTS

For this study, H.W. was partially supported by a Grant-in-Aid for the 21st century COE Program (Physics of Self-Organization Systems) at Waseda University from MEXT, Japan.

[1] K. A. Dill, S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan, Protein Sci. **4**, 561 (1995).
[2] C. M. Dobson, A. Šali, and M. Karplus, Angew. Chem. Int. Ed. **37**, 868 (1998).
[3] E. I. Shakhnovich, Curr. Opin. Struct. Biol. **7**, 29 (1997).
[4] A. R. Dinner, A. Šali, L. J. Smith, C. M. Dobson, and M. Karplus, TIBS **25**, 331 (2000).
[5] V. Daggett and A. R. Fersht, in *Mechanisms of Protein Folding*, edited by R. H. Pain (Oxford University Press, New York, 2000), p. 175.
[6] N. Gō, Annu. Rev. Biophys. Bioeng. **12**, 183 (1983).
[7] H. S. Chan, S. Bromberg, and K. A. Dill, Philos. Trans. R. Soc. London, Ser. B **348**, 61 (1995).
[8] H. S. Chan, Proteins **40**, 543 (2000).
[9] C. Levinthal, J. Chim. Phys. Phys.-Chim. Biol. **65**, 44 (1968).
[10] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, Proteins **21**, 167 (1995).
[11] K. A. Dill, Protein Sci. **8**, 1166 (1999).
[12] S. Takada, Proc. Natl. Acad. Sci. U.S.A. **96**, 11698 (1999).
[13] H. Taketomi, Y. Ueda, and N. Gō, Int. J. Pept. Protein Res. **7**, 445 (1975).
[14] Y. Ueda, H. Taketomi, and N. Gō, Biopolymers **17**, 1531 (1978).
[15] K. F. Lau and K. A. Dill, Macromolecules **22**, 3986 (1989).
[16] E. I. Shakhnovich and A. M. Gutin, Proc. Natl. Acad. Sci. U.S.A. **90**, 7195 (1993).
[17] E. I. Shakhnovich and A. M. Gutin, Protein Eng. **6**, 793 (1993).
[18] L. A. Mirny, V. Abkevich, and E. I. Shakhnovich, Folding Des. **1**, 103 (1996).
[19] V. I. Abkevich, A. M. Gutin, and E. I. Shakhnovich, Folding Des. **1**, 221 (1996).
[20] H. Wako and N. Saitô, J. Phys. Soc. Jpn. **44**, 1931 (1978).
[21] H. Wako and N. Saitô, J. Phys. Soc. Jpn. **44**, 1939 (1978).
[22] N. Gō and H. Abe, Biopolymers **20**, 991 (1981).
[23] H. Abe and N. Gō, Biopolymers **20**, 1013 (1981).
[24] V. Muñoz, E. R. Henry, J. Hofrichter, and W. A. Eaton, Proc. Natl. Acad. Sci. U.S.A. **95**, 5872 (1998).
[25] V. Muñoz and W. A. Eaton, Proc. Natl. Acad. Sci. U.S.A. **96**, 11311 (1999).
[26] E. R. Henry and W. A. Eaton, Chem. Phys. **307**, 163 (2004).
[27] P. Bruscolini and A. Pelizzola, Phys. Rev. Lett. **88**, 258101 (2002).
[28] K. Itoh and M. Sasai, Proc. Natl. Acad. Sci. U.S.A. **101**, 14736 (2004).
[29] H. Abe and H. Wako, J. Phys. Soc. Jpn. **73**, 1143 (2004).
[30] S. Miyazawa and R. L. Jernigan, Macromolecules **18**, 534 (1985).
[31] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[32] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **84**, 7524 (1987).
[33] M. Karplus and D. L. Weaver, Protein Sci. **3**, 650 (1994).
[34] A. R. Fersht, Curr. Opin. Struct. Biol. **7**, 3 (1997).