Quasiequilibrium unfolding thermodynamics of a small protein studied by molecular dynamics simulation with an explicit water model

Jihua Wang,[1,2,*] Zhiyong Zhang,[2] Haiyan Liu,[2] and Yunyu Shi[2,†]
[1]*Department of Physics, De zhou University, De zhou, Shandong 253023, People's Republic of China*
[2]*Key Laboratory of Structural Biology, University of Science and Technology of China (USTC), Chinese Academy of Sciences, Hefei, Anhui 230026, People's Republic of China*
and School of Life Sciences, USTC, Hefei, Anhui, 230026, People's Republic of China
(Received 23 January 2003; published 13 June 2003)

The 124 independent molecular dynamics simulations are completed with total time of 196.8 ns. The calculated unfolding quasiequilibrium thermodynamics of G-IgG-binding domain B_1 (GB1) shows the experimentally observed protein transitions: a coil to disordered globule transition, a disordered globule to molten globule transition, a molten globule to nativelike transition, and a nativelike to solidlike state transition. The first protein unfolding phase diagram has been constructed from molecular dynamics simulations with an explicit water model. The calculated melting temperature of GB1 agrees with early experiment. The results also agree with the recent experiment result in which GB1 has more than one intermediate.

DOI: 10.1103/PhysRevE.67.061903 PACS number(s): 87.14.Ee

I. INTRODUCTION

Protein folding problem remains one of the most challenging problems in modern molecular biology and biophysics. Thermodynamic properties of proteins are closely related to the mechanism of protein folding. Much attention has been paid to it. Privalov and co-workers have done a lot of work to illustrate the thermodynamic properties of protein during folding and unfolding processes by microcalorimetric [1] and circular dichroism spectroscopy, spectrofluorimetry and heat capacity scanning calorimetry methods [2]. Many people have shown that protein folding and unfolding is similar to the phase transition in condense physics. Wolynes and co-workers have studied the general phase behavior of proteins using statistical mechanics of spin glass and associative memories [3–7]. They have shown that the phase diagram of proteins consists of random coil, collapsed but fluid state, collapsed frozen state, and folded state [6]. Based on the lattice model, Shakhnovich and Karplus have obtained its common diagram [8]. Zhou and Karplus have investigated the phase behavior of a protein with a three-helix bundle using off-lattice model [9]. Phase transition of folding and unfolding of the protein can be determined by a simple parameter **g**. This parameter is the difference of the strength of contacts between native and non-native state [9,10].

Although simple models such as lattice model have revealed certain fundamental aspects of thermodynamic properties of proteins, it is necessary for us to study these properties by a more accurate model at atomic level with explicit water. However, few studies have been done on the subject because of current computer power.

In this paper, the unfolding process of a small protein with 56 amino acid residues (i.e., G-IgG-binding domain B_1(GB1), see Fig. 1) is studied by molecular dynamics (MD) simulations. X-ray and NMR structures of GB1 have been solved [11,12]. The thermodynamic properties of GB1 unfolding process have been studied by circular dichroism spectroscopy (CD) and differential scanning calorimetry (DSC) [13]. Experiment indicates that the melting temperature is 87.5 °C and the protein folding and unfolding is reversible. Two-state transition happened at 87.5 °C on pH $=5.4$. However, the recent experiment using fluorescence spectroscopy gave different results. It has shown that GB1 has at least one intermediate along the folding pathway [14]. This intermediate represents an ensemble of early intermediates [14,15]. In this work, we performed independent MD simulations of 196.8 ns for GB1 at atomic level with explicit water. Excess heat capacity as a function of temperature was calculated. The GB1 unfolding phase diagram (Fig. 8) was constructed. To our knowledge, it is the first phase diagram calculated by the MD simulation with molecular force field at atomic level in explicit water, which reveals the complex phase behavior of the protein unfolding process. The results agree with the recent experiment result in which GB1 has more than one intermediate [14]. The calculated melting temperature of GB1 also agrees with the early experiment [13].

II. METIODS

A. Molecular dynamics simulation

At different temperatures and four pressures, the 98.4-ns MD simulations (which are called the MD simulations here-

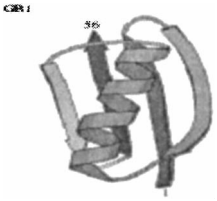

FIG. 1. Ribbon structure of the B1 domain of protein G (Gronenborn *et al.* [11]) drawn using the program Molscript.

*Corresponding author. FAX: +86-551-3603754. Email address: jhwyh@yahoo.com.cn
†Email address: yyshi@ustc.edu.cn

inafter for convenience) are performed for various time periods (Table I) using the GROMACS2.0 MD software package [16] with GROMOS96 force field 43a1. At the same temperatures (from 280 K to 540 K, the same as in Table I) and the same pressures (i.e., 1bar, 2000 bars, 5000 bars, and 8000 bars, respectively, the same as in Table I), we added 98.4-ns control simulations (which are called the control simulations hereinafter) in order to validate the quasiequilibrium condition. The 98.4-ns control simulations, which are the same as the MD simulations in Table I but only with different atomic initial velocities of all atoms of the system, were completed using the GROMACS3.1.4 software package with parallel computing [17]. The atomic initial velocities were taken from the Maxwellian distribution under some random number generator seeds. So selecting different random number generator seeds specified the atomic initial velocities of the system. Here, the random number generator seed was set to one number (gen_seed1) for the 98.4-ns MD simulations in Table I, and it was set to another number (gen_seed2) for the added control simulations of 98.4 ns.

The LINCS was used to constrain covalent bond length [18]. The SPC/E water model was used [19] for water molecules. The initial velocities of all atoms were given from a Maxwellian distribution at the chosen temperature and pressure.

Weakly coupling solute and solvent to an external temperature bath separately at the reference temperature maintained the temperature of the system [20]. The temperature coupling constant was 0.1 ps. The pressure was maintained by weakly coupling the system to an external pressure bath at reference pressure with a coupling constant of 0.5 ps. The cut-off radius was set to 1.4 nm for van der Waals interactions and to 1.4 nm for Coulomb interactions, and the cut-off distance in the short-range neighborlist or rlist was set to 0.8 nm.

The system was prepared in the following way. First, the initial structure was obtained from NMR of GB1 [18]. Then, a rectangular box (its dimension is $5.45\times4.342\times3.932$ nm^3, respectively) with equilibrated SPC/E water molecules was created. Before being inserted into the water, the N and C termini of the protein molecule were capped with neutral acetyl and methyl amide groups. The protein was inserted into the box (distance between the solute and the box is 0.9 nm), and water molecules overlapping with the inserted proteins were removed. Then four Na^+ ions were inserted into the box and the four water molecules overlapping with the four Na^+ ions were removed, so the system gives a total charge of zero.

The resulting system was composed of the protein in the four Na^+ ions and 2656 water molecules. The total system contained 8522 atoms. After that, 200 steps of energy minimization were performed with a steepest descent method until the maximum force of the system is smaller than 100 KJ mol^{-l} nm^{-1}. Then, 100 ps of the equilibration dynamics of the whole system were performed before the initiation of the production run that was used for analysis. The duration of time steps of simulation is 2fs.

B. Quasiequilibrium simulation

In order to study the phase transition of GB1, we must perform many independent molecular dynamics simulations (62 independent unfolding simulations in this paper) from 1.1 ns to 2.6 ns at different temperatures and at four pressures (Table I). If a full equilibrium for each simulation were reached, the total simulation time would be over 3 μs according to the criterion of full equilibrium simulation [21,22]. However, the longest simulation by far was a computational tour involving a 1 μs "refolding" simulation of HP-36 reported by Duan and Kollman [23]. Obviously, the 3 μs simulation of the 56-residue GB1 is beyond our computational power. Giving consideration to both computational power and a meaning calculated result, it may be a helpful exploration for us to employ a concept of quasiequilibrium simulation, which is an advisable approach to simulate protein folding or unfolding using an atomic level model in explicit water at present. The quasiequilibrium simulation is defined as the one in which the average change of positional root-mean-square deviation of denatured protein is less than a given value during certain simulation time. The quantity R_d is defined as follows:

$$R_d=\frac{1}{\langle R_{msd}\rangle}\sqrt{\frac{1}{N_C}\sum_i R_{msd}i^2-\langle R_{msd}\rangle^2}. \tag{1}$$

Here,

$$\langle R_{msd}\rangle=\frac{1}{N_C}\sum_i R_{msdi},$$

where R_{msdi} is the positional root-mean-square deviation of conformation i and N_C is the number of conformation. R_d describes equilibrium degree of a simulation during certain simulation period to some extent. It is considered to be quasiequilibrium simulation when R_d is less than a given value (here, the R_d value is set to be less than 10%, see below).

C. Heat capacity of protein

Because heat capacity is a sensitive indictor of different transitions [24], we study the phase transition mainly by the analysis of heat capacity of protein. The system to be studied is an isothermal-isobaric ensemble (fixed N, P, and T). The heat capacity of protein is defined as follows [25]:

$$C_P=\left(\frac{\partial H}{\partial T}\right)_P=\left(\frac{\partial(U+PV)}{\partial T}\right)_P=\left(\frac{\partial U}{\partial T}\right)_P+P\left(\frac{\partial V}{\partial T}\right)_P. \tag{2}$$

Here, U is the internal energy of protein, V is the volume of protein, and P is the pressure. The heat capacity defined by Eq. (2) is the molar heat capacity of protein because it is the multi-ingredient system composed of protein, water, and ions.

One of the main goals of this paper is to study the phase transition of protein unfolding. So we focus on the transition excess heat capacity C_{Pex} defined as follows [26,27]:

TABLE I. The MD simulation scheme at different temperatures and pressures. T (K), temperature; P (bar), pressure; MD (ps), MD simulation time. QSP (ps), quasiequilibrium simulation period.

T (K) / P (bar)	280	300	320	340	360	380	400	420	440	460	480	500	520	540
1	S01	S11	S21	S31	S41	S51	S6	S71	S81	S91	S101	S111	S121	S131
MD (ps)	1100	2100	1100	1100	2100	2100	2600	2100	2100	2100	2100	2100	2100	2100
QSP (ps)	600–1100	500–2100	700–1100	800–1100	1500–2100	1500–2100	2200–2600	1500–2100	1500–2100	1700–2100	1500–2100	1700–2100	1400–2100	1600–2100
2000	S02	S12	S22	S32	S42	S52	S62	S72	S82	S92	S102	S112	S122	S132
MD (ps)	1100	1100	1100	1100	1100	1600	1100	1600	1600	1600	1600	2500	1600	1600
QSP (ps)	700–1100	700–1100	800–1100	700–1100	900–1100	1400–1600	500-1100	1200–1600	1000–1600	1000–1600	1300–1600	2000–2500	1300-1600	1300-1600
5000	S03	S13	S23	S33	S43	S53	S63	S73	S83	S93	S103	S113	S123	S133
MD (ps)	1100	1100	1100	1100	1100	1100	1100	1600	1100	2600	1600	2100	1600	2000
QSP (ps)	840–1100	600–1100	600–1100	800–1100	900–1100	900–1100	650–1100	1200-1600	800–1100	2200–2600	1300–1600	1500–2100	1100–1600	1500–2000
8000	S04	S14	S24	S34	S44	S54	S64	S74	S84	S94	S104	S114	S124	S134
MD (ps)	1100	1100	1100	1100	1600	1600	1100	1600	1600	1600	2100	2100	2100	2500
QSP (ps)	700–1100	600–1100	600–1100	800–1100	1300–1600	1200–1600	700–1100	1400–1600	1100–1600	1000-1600	1500–2100	1800–2100	1500–2100	2100–2500

T (K) / P (bar)	290	310	330	350	370	376
1	S101	S102	S103	S104	S105	S106
MD (ps)	1100	1100	1100	2100	2100	2100
QSP (ps)	900–1100	900–1100	800–1100	1600–2100	1700–2100	1700–2100

$$C_{Pex}=\frac{\partial\langle\Delta H(T)\rangle}{\partial T}, \tag{3}$$

$$\langle\Delta H(T)\rangle=\langle H(T)\rangle-H_N. \tag{4}$$

Here, H_N is the enthalpy of the native state. The $\langle H(T)\rangle$ is the mean enthalpy of a system at temperature T during the quasiequilibrium simulation time period.

The heat capacity for protein unfolding can be calculated using three methods. They are a finite difference second derivative of the free energy, a finite first derivative of the enthalpy, and the fluctuations in the enthalpy [25], but the third method was not used because of the large numerical uncertainties in the fluctuations of enthalpy [28]. Here, we chose the second method to calculate C_{Pex} which is shown as follows:

$$C_{Pex}=\frac{\partial\langle\Delta H(T)\rangle}{\partial T}\approx\frac{H(P,T+\Delta T)-H(P,T-\Delta T)}{2\Delta T}, \tag{5}$$

$$H(P,T)=U(P,T)+PV, \tag{6}$$

$$C_{Pex}\approx\frac{\langle U(P,T+\Delta T)\rangle-\langle U(P,T-\Delta T)\rangle}{2\Delta T}+P\frac{\langle V(P,T+\Delta T)\rangle-\langle V(P,T-\Delta T)\rangle}{2\Delta T}, \tag{7}$$

$$\langle U(P,T)\rangle=\langle U_{p-p}(P,T)\rangle+\langle U_{p-s}(P,T)\rangle+\langle U_{p-Na^+}(P,T)\rangle+\langle U_{bond}(P,T)\rangle, \tag{8}$$

where $\langle U_{p-p}(P,T)\rangle$ is mean interaction energy within protein; $\langle U_{p-s}(P,T)\rangle$ is mean interaction energy between protein and water; $\langle U_{p-N}(P,T)\rangle$ is the mean interaction energy between protein and Na^+ ions; $\langle U_{bond}(P,T)\rangle$ is mean bond energy of protein; $\langle V(P,T)\rangle$ is mean volume of protein. All the quantities above are averaged during the quasiequilibrium simulation time period.

The contribution of interaction energy within solvent in the process of protein unfolding may be neglected because most of the bulk water around the protein can be removed without altering the heat capacity [29].

D. Contact and the Lindermann criterion analysis

Contact analysis was used to calculate the number of contacts and reaction coordinates of the simulation system. For a protein, a contact was defined as being present if the centers of geometry of side chains of two residues (for residue pairs i, j with $j>i+1$) are within 0.65 nm [30]. During the reference state simulation, all contacts that satisfied the definition are recorded as ''native contacts.'' Reaction coordinates are defined based on the number of native contacts, which has been a straightforward and computationally inexpensive way to discriminate the native state from non-native conformation of a protein.

The Lindemann ratio for protein can be expressed [31] as

$$L(P,T)=\frac{r_{ms}}{a}, \tag{9}$$

$$r_{ms}=\sqrt{\frac{\langle(r_i-\langle r_i\rangle)^2\rangle}{N}}. \tag{10}$$

Here, r_{ms} is the root-mean-squared flunctuation, N is the number of atoms, and ''a'' is the most probable nonbonded near-neighbor distance. Here we chose that a=0.45 nm, the same as the distance used by Zhou and Karplus [32]. The $L(P,T)$ is used as melting and freezing criterion for protein solids. Generally, when $L(P,T)<0.14$, the system is solidlike, while $L(P,T)>0.14$, it is considered to be liquidlike [31].

III. RESULTS AND DISCUSSIONS

The results are discussed in four parts. First, the 62 unfolding quasiequilibrium MD simulations and the 62 control simulations are completed. Second, the melting temperature of GB1 unfolding at 1 bar is discussed, and the result is compared with that of the experiment. Third, the structural characteristic of the intermediates presented on GB1 unfolding process is analyzed. Fourth, we discuss the phase behavior of GB1 unfolding process. Finally, the quasiequilibrium condition in the molecular dynamics simulations is discussed.

A. Unfolding quasiequilibrium simulation

Full equilibrium simulation is too expensive for the current computer power [21,22]. To draw unfolding phase diagram of GB1 at atomic level, it is necessary for us to complete the multiple MD quasiequilibrium simulations. We have computed 62 unfolding trajectories at temperatures ranging from 280 K to 540 K and pressures from 1 bar to 8000 bars as shown in Table I. In order to validate the quasiequilibrium condition, we have added the other 62 control simulations with different atomic initial velocities. According to our experience, the quasiequilibrium simulation is defined as $R_d\leqslant 10\%$. That is to say, an MD simulation is considered to be quasiequilibrium simulation if R_d is less than 10% during certain time period. R_d less than 10% determines the MD simulation time in Table I, and each quasiequilibrium simulation period (QSP) was selected according to the principle that R_d is less than 10% (Table I). Table I indicates that the MD simulation time is longer at high temperatures than at low temperatures at the same pressure in general [Figs. 2(b) and 2(c)]. It is shorter at higher pressures than at lower pressures [Figs. 2(c) and 2(d)] in general. For example, the MD simulation time is 1.1 ns at 320 K and at 1 bar [Fig. 2(b)], 2.1 ns at 520 K and at 1 bar [Fig. 2(c)], and 1.6 ns at 520 K and at 5000 bars [Fig. 2(d)], respectively. Their corresponding quasiequilibrium simulation period is 400 ps (from 700 ps to 1100 ps) [Fig. 2(b)], 700 ps (from 1400 ps to 2100 ps) [Fig. 2(c)] and 500 ps (from 1100 ps to 1600 ps) [Fig. 2(d)], respectively. During these quasiequilibrium

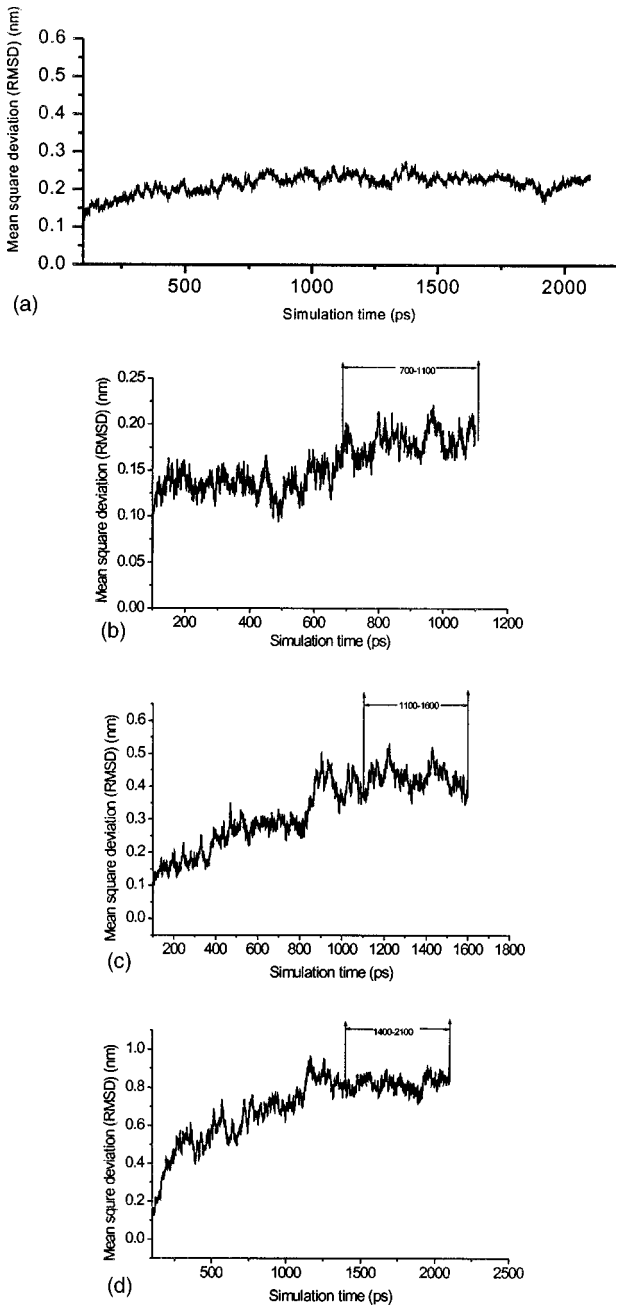

FIG. 2. The positional root-mean-square deviation (RMSD) of GB1 molecular dynamics (MD) simulation as a function of time at different temperatures and pressures: (a) at 300 K and 1 bar; (b) at 320 K and 1 bar; QSP equal to 400 ps; (c) at 520 K and 1 bar; QSP equal to 500 ps; (d) at 520 K and 5000 bars; QSP equal to 700 ps. QSP: quasiequilibrium simulation period.

simulation periods their corresponding R_d is 7.2%, 3.3%, 7.4%, respectively, as shown in Figs. 2(b), 2(c), and 2(d). The atomic-positional root-mean-square deviations (RMSD) of C_α atoms of the protein at the room temperature (300 K) and at the pressure of 1 bar is shown in Fig. 2(a).

B. The melting temperature of GB1

The melting temperature of GB1 is obtained as follows. The $U(P,T)$ of GB1 [see Eq. (8)] was averaged during each

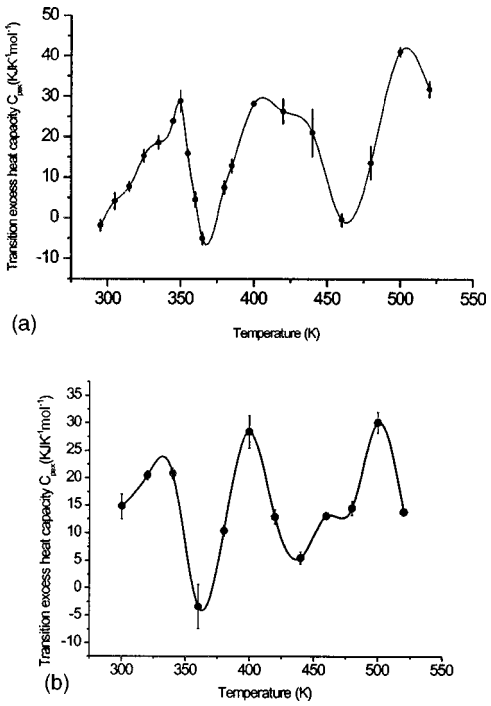

FIG. 3. The transition excess heat capacity C_{Pex} as a function of temperature at different pressures. (a) C_{Pex} as a function of temperature at 1 bar; (b) C_{Pex} as a function of temperature at 2000 bars.

quasiequilibrium simulation period (Table I). And the transition excess heat capacities were computed at different temperatures and different pressures by Eq. (7). Then we drew the curve of transition excess heat capacity C_{Pex} as a function of temperature at 1 bar [Fig. 3(a)] by the MD simulations. The shape of C_{pex} curve agrees with the experimental result [13] in the range of experiment (below 100 °C). However, the maximum melting temperature of GB1 is 87.5 °C at pH of 5.4 in the experiment [13], while the GB1's melting temperature obtained from the MD simulations is 77 °C according to Fig. 3(a). The difference is 10.5 °C. The reason can be attributed to pH. The pH value of the experiment is 5.4, while the pH value in the MD simulations is 7.0 since they are performed in neutral water. Considering this factor, the melting temperature of GB1 obtained by the MD simulations is basically in conformity with that in the experiment [13].

C. The intermediate of GB1

The successful simulation results discussed above encourage us to run further the GB1 unfolding of MD simulation at higher temperatures and at higher pressures. The 62 unfolding trajectories are computed and the total MD simulation time is 98.4 ns (Table I).

The heat capacity curves as a function of temperature at 1 bar and 2000 bars are plotted from MD simulation results shown in Fig. 3. The heat capacity curves at 5000 bars and 8000 bars are similar to the curves at 1 bar and 2000 bars (omitted here). Figure 3(a) indicates that in the heat capacity curve at 1 bar, in addition to the first peak, there are two other peaks. The result has shown that there is at least three

phase transitions at 350 K, 410 K, and 504 K, respectively. Therefore, we can conclude that there may be two intermediates on GB1 unfolding process. This shows that the so-called two-state small proteins may have intermediates on their unfolding or folding processes. So some small proteins might unfold by multiple-state mechanism [14], which supports the recent arguments in which refolding from the unfolded state proceeds along a sequential pathway with at least one obligatory intermediate, which lies along a direct path toward the native state. This result agrees with the recent experimental result, GB1 has more than one intermediate [14].

D. Phase transition of GB1

We study phase transition of GB1 based on heat capacity because it is a sensitive indicator of phase transition [24]. Figure 3(a) indicates that the curve of C_{Pex} has three peaks at 1 bar. Starting from low temperature, the first peak is at 350 K, which corresponds to the phase transition in the range of experiment [13]. The second peak at 410 K may coincide with a strong collapse transition [Fig. 5(a)]. At the second peak, square radius of gyration (R_g^2) changes dramatically around the second peak. The R_g^2 increases from 1.0 nm^2 at 400 K to 1.5 nm^2 at 440 K. The corresponding fraction of contact number decreases from 0.5 to 0.1 [Fig. 4(a)], and the contact number within α helix is almost disappearing at 440 K [Fig. 4(a)]. Figure 6(a) further indicates that area of solvent accessible surface and hydrophobic surface changes drastically around 410 K. These results show that the second peak may correspond to collapse transition. And the phase transition at 410 K may be of first order. The third peak is at 505 K. At this peak, the α helix content disappears completely and the contact number within α helix tends to be zero. And the area of solvent accessible hydrophobic surface increases obviously, which shows that the hydrophobic core is further exposed to water.

At 2000 bars, phase transitions occur at 332 K, 400 K, and 500 K, respectively [Fig. 3(b)]. Starting from high temperature, the first peak at 500 K may coincide with the strong collapse transition as shown in Fig. 5(b). The collapse transition moves to higher temperature (500 K corresponding to the first peak starting from high temperature) at 2000 bars compared with that (410 K corresponding to the second peak starting from high temperature) at 1 bar, and the collapse transition is stronger than that at 1 bar [Figs. 5(a) and 5(b)]. From Fig. 6(b), we know that the areas of solvent accessible surface and hydrophobic surface are maximum at 500 K (corresponding to collapse transition), which shows that most part of the hydrophobic core is exposed to water. The collapse transition may be of first order. The second peak at 400 K and the third peak at 332 K move to lower temperature compared with those corresponding peaks at 1 bar, which originates from pressure effect [33,34]. Figure 4(b) shows that the number of contacts decreases according to gradient function approximately with increasing temperature, and the three peaks correspond to three minimum points. The collapse transition corresponds to the most obvious decreasing of contact number [Fig. 4(b)].

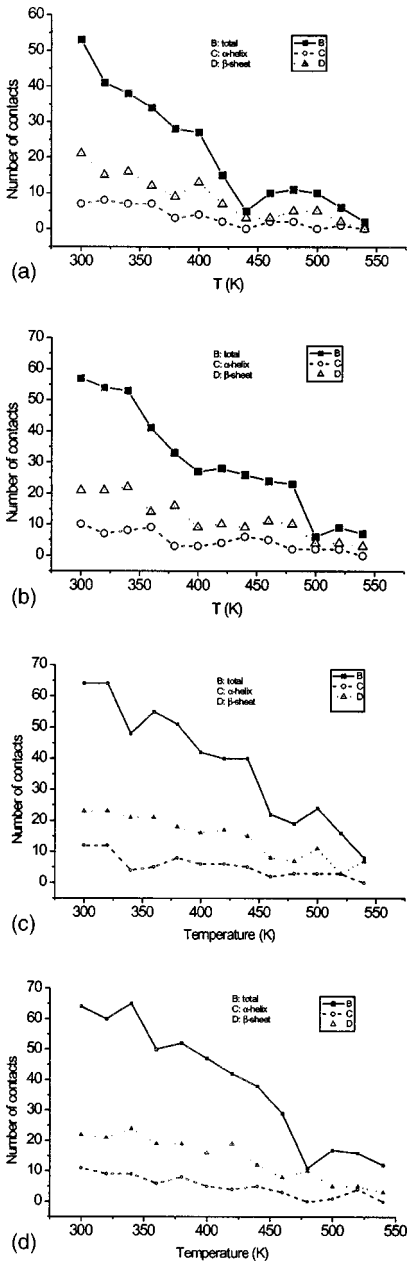

FIG. 4. The number of native contacts internal GB1, internal α helix, and internal β strand as a function of temperature at 1 bar (a), 2000 bars (b), 5000 bars (c), and 8000 bars (d), respectively.

Using the methods similar to those applied at 1 bar and 2000 bars, we analyze phase transitions of GB1 at 5000 bars and 8000 bars. At 5000 bars, the phase transitions occur at 345 K, 420 K, and 480 K, respectively. At 8000 bars, the phase transition of GB1 occurs at 320 K, 425 K, and 488 K, respectively. Figure 4(a) reveals that the number of contacts becomes almost zero at 540 K and 1 bar, while the number of native contacts remains above 10% up to 540 K at 2000 bars, 5000 bars, and 8000 bars [Figs. 4(b), 4(c), and 4(d)]. The strong collapse transitions at 440 K and 1 bar and 500 K and 2000 bars are replaced by several much weaker transitions at 5000 bars and 8000 bars, respectively [Figs. 5(c) and 5(d)].

The phase transition of solidlike and nativelike is judged by the Lindermann ratio. From Eqs. (9) and (10), we com-

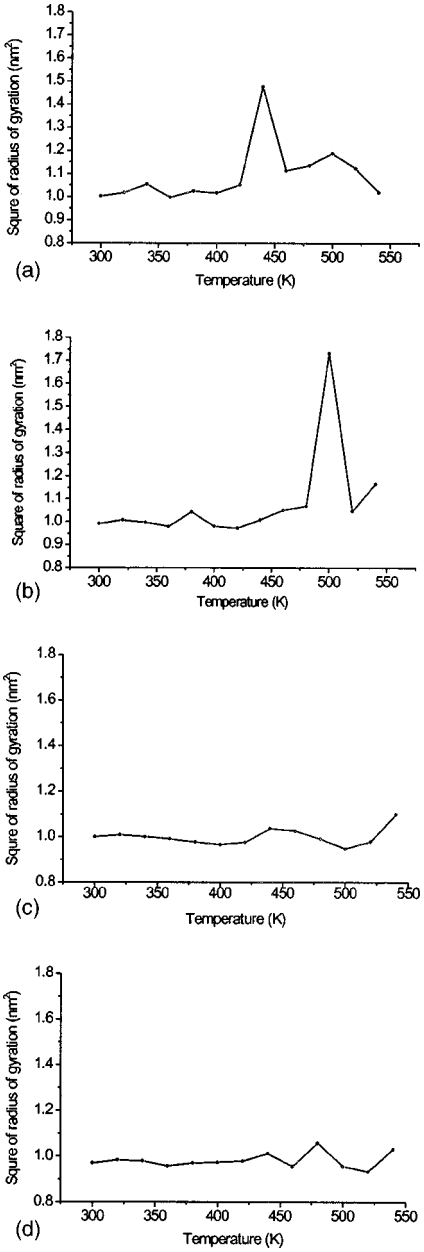

FIG. 5. The radius of gyration as a function of temperature at pressures of 1 bar (a), 2000 bars (b), 5000 bars (c), and 8000 bars (d), respectively.

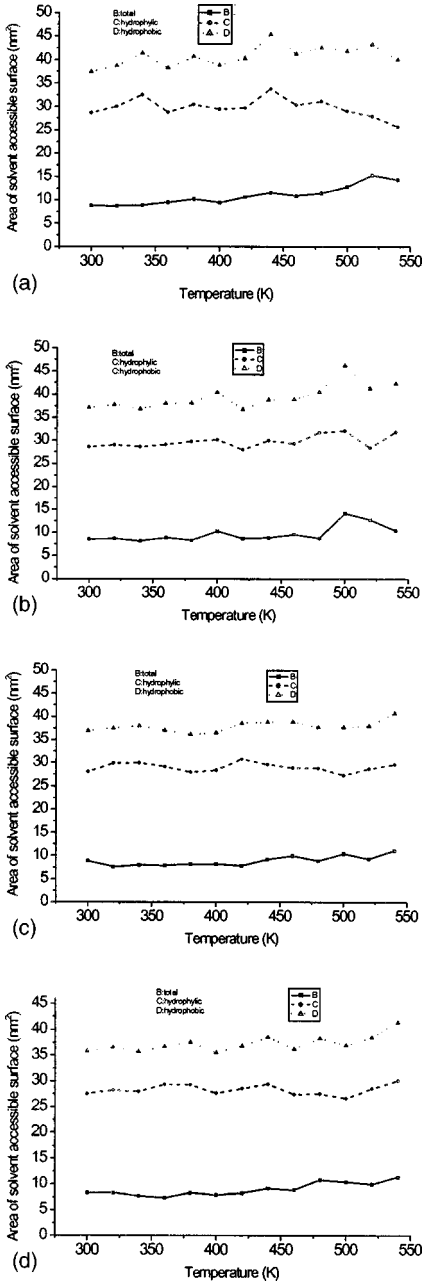

FIG. 6. Areas of solvent accessible surface, hydrophobic and hydrophilic solvent accessible surfaces as a function of temperature at pressures of 1 bar (a) and 2000 bars (b).

pute the Lindermann ratios at different temperatures (from 280 K to 540 K) and pressures (from 1 bar to 8000 bars) (Fig. 7). From Fig. 8, we obtain that $L(P,T)>0.14$ at temperatures ranging from 280 K to 540 K and pressures ranging from 1 bar to 8000 bars, which indicates that the protein in the range of temperatures (from 280 K to 540 K) and pressures (from 1 bar to 8000 bars) is not solidlike [31]. Let $L(P,T)=0.14$, we induce the phase transition temperatures between solidlike and nativelike phase at different pressures by linear extrapolation. The phase transition temperature between solidlike and nativelike is 262.2 K at 1 bar, 262.4 K at 2000 bars, 269.9 K at 5000 bars, and 272.7 K at 8000 bars, respectively. The phase transition curve between solidlike and nativelike phases is approximately a beeline (Fig. 8). If we extend the transition line by linear extrapolation, we can obtain that the frozen pressure (the pressure that protein is solidlike) is about 25 kbar at room temperature (300 K) that is in qualitative agreement with experiment [35].

On the basis of the phase transition temperatures obtained from the results above, a protein phase diagram as a function of temperature and pressure can be constructed as shown in Fig. 8. The phase diagram consists of five phases: solidlike, nativelike, molten globule, disordered globule, and coil, which agree with experiment [36] and simple models [8,9]. These phase transitions are not very sensitive to pressure below 8000 bars (Fig. 8). The phase diagram presented here may have two advantages. On the one hand, the diagram is

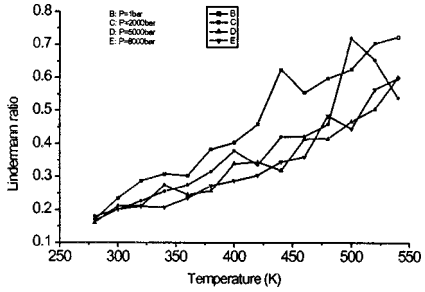

FIG. 7. The Lindermann ratio as a function of temperature at 1 bar, 2000 bars, 5000 bars, and 8000 bars, respectively, on protein GB1 unfolding.

constructed by molecular dynamics simulation. To our knowledge, it is the first time that the phase diagram on protein unfolding process was obtained at atomic level, which is more close to actual phase transition of protein unfolding process. On other hand, the phase diagram of the protein unfolding is obtained first as a function of temperature and pressure, which is the form that most substances usually adopt in physics and chemistry. It is convenient for us to compare the protein phase diagram with other phase diagrams of any substance in the same formation in physics and chemistry. However, the protein phase transition above 8000 bars is unclear in the phase diagram (Fig. 8).

E. Structural characteristic of the phases

The structural characteristic of all kinds of phases in the protein phase diagram (Fig. 8) is summed up as follows.

Coil. More than 70% native contacts have been vanished. The α helix is almost fading away. Only few residues of β strand remain. The most part of residues of the protein has become random coil, so we call it the Coil phase. However, there is a very small amount of native structure in the phase, which agrees with experimental result [37].

Disordered globule. The α helix and β strand are partly vanished. R_g^2 is bigger than that of native state. The fraction of native contacts is between 30% and 65%.

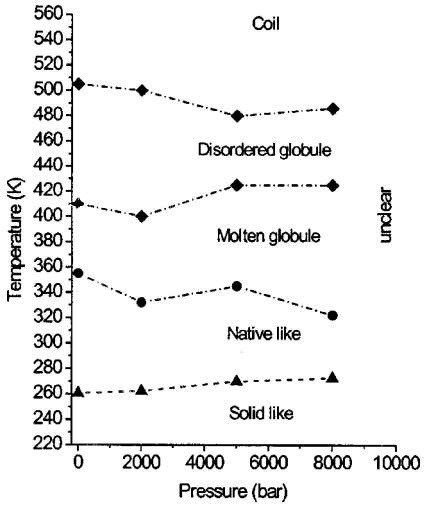

FIG. 8. Phase diagram of protein GB1 on protein unfolding at atomic atom level. The horizontal axis is the pressure. The vertical axis is the temperature. ''Unclear'' expresses that it remains to be studied in the scope of the unclear region.

Molten globule. R_g^2 is close to that of native state. The fraction of native contacts is between 50–90 %. The slight denaturation in α helix and β strand occurs.

Nativelike phase. R_g^2 almost equals that of native state. The α helix remains unchanged. A very small amount of β strand has denatured. The secondary structure of the protein is similar to that of native state. The fraction of native contacts exceeds 80%.

F. The quasiequilibrium condition in MD simulation

Based on the added 98.4-ns control simulations, we study the thermodynamic properties of GB1 unfolding process and discuss the quasiequilibrium condition. First, we compute the mean enthalpy [Fig. 9(a)], the transition excess heat capacity C_{Pex} [Fig. 9(b)], the Linderman ratio [Fig. 9(c)], the number of contacts [Figs. 9(d), 9(e), and 9(f)], the number of hydrogen bonds [Figs. 9(g) and 9(h)], and the area of solvent accessible surface [Fig. 9(i)] under the quasiequilibrium condition (here the R_d is also set to be less than 10%) using 38.7-ns control simulations at 20 temperatures (from 280 K to 540 K) and 1 bar with the gen_seed 2 (the new random number generator seed). And they are compared with the corresponding qualities [Figs. 9(a)–9(i)] obtained by the MD simulations (Table I) with the gen_seed1 (the old random number generator seed). The enthalpy has the same changing tendency with temperature as that computed by the MD simulations [Fig. 9(a)]. Based on the Eqs. (5), (6), (7), and (8), we compute the transition excess heat capacity C_{Pex} using the first derivative method [Fig. 9(b)]. Figure 9(b) indicates that there are three peaks in the heat capacity curve at 1 bar. So we obtain that there are at least three phase transitions at 355K, 403 K, and 499 K, respectively, which are almost the same as that the three phase transitions occurred at 350 K, 410 K, and 505 K, respectively [Fig. 3(a)]. We compute the Lindermann ratio based on Eq. (10) shown in Fig. 9(c). From Fig. 9(c), we obtain that $L(T)=0.14$ at temperatures ranging from 280 K to 540 K and at pressure 1 bar, which indicates that the protein in the range of temperatures (from 280 K to 540 K) and pressure at 1 bar is not solidlike. Let $L(T)=0.14$, we induce the phase transition temperatures between solidlike phase and nativelike phase at different pressures by linear extrapolation. The phase transition temperature between solidlike and nativelike is 267 K that agrees with the corresponding temperature 262.2 K computed from the MD simulations with gen_seed1. Second, the other 59.7-ns control simulations at 2000 bars, 5000 bars, and 8000 bars and at 14 temperatures (from 280 K to 540 K) with gen_seed2 are also studied by the same method (omitted). We run the total control simulation 98.4 ns. The phase transition temperatures are obtained under the pressures 2000 bars, 5000 bars, and 8000 bars, respectively. They are 268.6 K, 340 K, 393 K, and 496 K at 2000 bars, 275.0 K, 354 K, 424 K, and 485 K at 5000 bar, 292.6 K, 328 K, 404 K, and 476 K at 8000 bar, respectively. The phase transition temperatures obtained by MD simulations in Table I are 262.4 K, 332 K, 400 K, and 480 K at 2000 bar, 269.9 K, 345 K, 420 K, and 480 K at 5000 bars, and 272.7 K, 320 K, 425 K, and

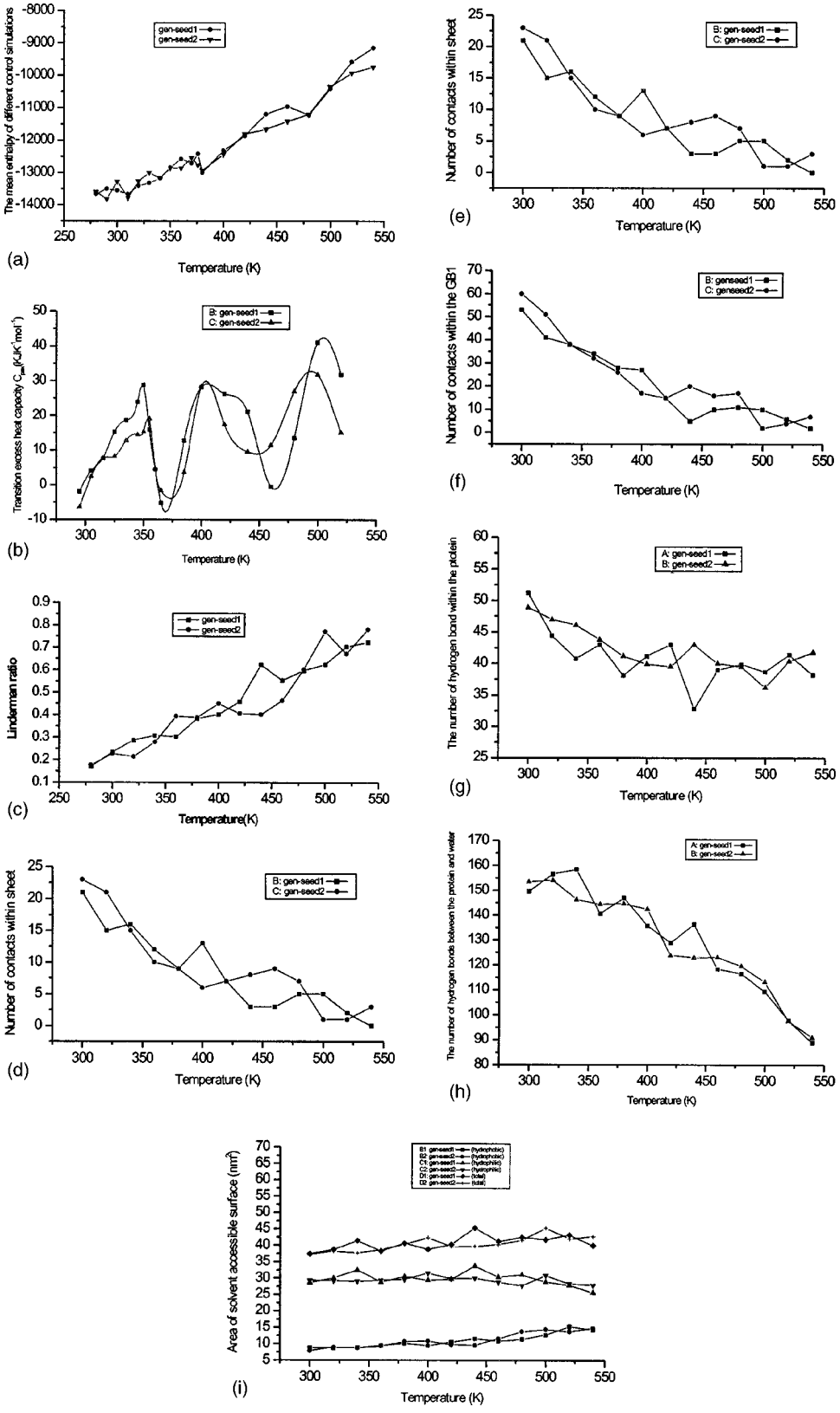

FIG. 9. Comparison of the mean enthalpies (a), the transition excess heat capacities C_{Pex} (b), the Linderman ratios (c), the number of contacts [(d), (e), and (f)], the number of hydrogen bonds [(g) and (h)], and the area of solvent accessible surface (i) between the MD simulations (with gen_seed1) and the control simulations (with gen_seed2) under the quasiequilibrium condition.

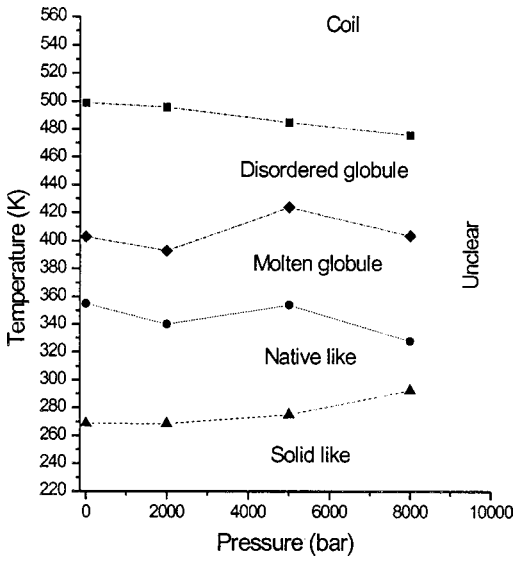

FIG. 10. The phase diagram constructed by the control simulations with gen_seed2.

488 K at 8000 bars, respectively. The two group corresponding transition temperatures computed by the control simulations and the MD simulations in Table I are almost the same. Third, the protein phase diagram as a function of temperature and pressure has been constructed again using the transition temperatures computed by the added 98.4-ns control simulations shown in Fig. 10, which is the same as the phase diagram of Fig. 8. Furthermore, the contact number [Figs. 9(d), 9(e), and 9(f)], the number of hydrogen bond, and the area of solvent accessible simulations [Figs. 9, 9(h), and 9(i)] obtained by the control simulations are about the same as those of the old MD simulations. Of course, there are also small diversities for the results between the MD simulations and the added control simulations [Figs. 9, 10, and 8]. The main reason is that these MD simulations, which are the quasiequilibrium simulations, do not reach the full equilibrium.

In summary, we have obtained almost the same results from the MD simulations and the control simulations under quasiequilibrium condition, which shows that the quasiequilibrium condition is at least valid in this work. The quasiequilibrium simulations can provide some useful information (such as phase transition) on protein folding or unfolding process. So it is worth employing the quasiequilibrium simulations to study the thermodynamic properties of protein unfolding or folding process at present.

IV. CONCLUSIONS

Within the experimental range, the calculated results presented here agree with the experimental results. Beyond the experiments, the MD simulations results have provided us with more useful information.

The calculated melting temperature and the curve of heat capacity of GB1 obtained from the MD simulations are in agreement with the experimental result [13]. The results presented here support the argument that there is at least one intermediate for the so-called two-state small proteins [38,14]. We think that intermediates of small proteins may be observed by more experiments in the future with improvement of experimental accuracy.

The protein phase diagram (Fig. 8) has been constructed. To our knowledge, this is the first time that the phase diagram of protein unfolding process was obtained by the MD simulations. At 1 bar, the phase transition from molten globule to disordered globule for GB1 is found to be of first order, and the phase transition from disordered globule to coil at 2000 bars is found to be also first order. The phase transitions are insensitive to pressure at least below 8000 bars. The phase diagram is composed of five phases that correspond to coil, disordered globule, molten globule, nativelike, and solidlike, respectively, which agrees with the results obtained from the simple models [8,9], the analysis theory [5] and the experiment [36]. The characteristics of the secondary structure, the native contact, and the radius of gyration (Figs. 4 and 5) corresponding to all kinds of phases have been revealed. The general phase transition tendency of the protein GB1 has been displayed in the phase diagram (Fig. 8) though it is a rough one for two reasons. One is that the MD simulations (Table I) at high temperatures and high pressures are not long enough to reach full equilibrium for the current computer power limit. The MD simulations are the quasiequilibrium simulations. However, it is shown that the phase diagram almost holds the line under the quasiequilibrium condition by the control simulations (Figs. 8 and 10). Another is that the molecular field at high temperatures may lead to some artifacts [17]. In Figs. 8 and 10, we have not shown phase behavior in the region beyond 8000 bars. The region is called the ''unclear.'' The unclear region remains to be further studied.

The experiment and computer simulation have shown that BPT1 starts denaturation at 10 kbar. And a further increase of the pressure such as beyond 15 kbar results in a freezing of the protein [39,35]. In this work, we deduced that the frozen pressure of GB1 was 25 kbar by extrapolation. So GB1 and BPT1 have approximately the same frozen pressure.

It is at least now useful for us to study thermodynamic properties of protein unfolding (or folding) under the quasiequilibrium condition considering the current computer power. The control simulations with different atomic initial velocities of the system show that the quasiequilibrium condition is valid (Figs. 9, 8, and 10) at least in this work. With the development of computer power, a further accurate phase diagram of proteins can be constructed by MD simulation, which will provide more rich and accurate informations for us to understand protein folding and unfolding mechanism intuitively. Recently, Simmerling and co-workers have correctly predicted the Trp-cage's final shape purely from its genetic code using computer simulations [40]. The successful prediction further shows that computer simulations have a broad prospect for the protein study.

ACKNOWLEDGMENTS

We gratefully thank Professor H.J.C. Berendsen (Department of Biophysical Chemistry, University of Groningen) for providing us with the GROMACS2.0 and GROMACS3.1.4 pack-

ages and W.F. van Gunsteren (Department of Chemistry, ETHZ) for the GROMOS96 force field. Dr J. Zhu in our group has provided us with the algorithms of contact analysis. J.H. Wang gratefully thanks Dr. J. Zhu, Mr. X.H. Dou, Dr. H. Fan, Dr. Y.D. Yang, and Dr. J.B. He in our group for their helpful discussions. This work was supported by the Chinese National Fundamental Research Project (Grant Nos. G19990756 and G90103032), the Chinese National Natural Science Foundation (Grant No. 399906000) and granted from the Chinese Academy of Science.

[1] P.L. Privalov, Annu. Rev. Biophys. Biophys. Chem. **18**, 47 (1989).
[2] A.I. Dragan and P.L. Privalov, J. Mol. Biol. **321**, 891 (2002).
[3] J.D. Bryngelson and P.G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **84**, 7524 (1987).
[4] J.D. Bryngelson and P.G. Wolynes, J. Phys. Chem. **93**, 6902 (1989).
[5] M. Sasai and P.G. Wolynes, Phys. Rev. Lett. **65**, 2740 (1990).
[6] J.D. Bryngelson, J.N. Onuchic, N.D. Socci, and P.G. Wolynes, Proteins **21**, 167 (1995).
[7] S.S. Plotkin, J. Wang, and P.G. Wolynes, J. Chem. Phys. **106**, 2932 (1997).
[8] A. Dinner, A. Sali, M. Karplus, and E.I. Shakhnovich, J. Chem. Phys. **101**, 1444 (1994).
[9] Y. Zhou and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **94**, 14429 (1997).
[10] Y. Zhou and M. Karplus, Nature (London) **401**, 400 (1999).
[11] A.M. Gronenborn, D.R. Filpula, N.Z. Essig, A. Achari, M. Whitlow, P.T. Wingfield, and G.M. Clore, Science **253**, 657 (1991).
[12] T. Gallagher, P. Alexander, P. Bryan, and G.L. Gilliland, Biochemistry **33**, 4721 (1994).
[13] P. Alexander, S. Fahnestock, T. Lee, J. Orban, and P. Bryan, Biochemistry **31**, 3597 (1992).
[14] S.H. Park, K.T. O' Neil, and H. Roder, Biochemistry **36**, 14277 (1997).
[15] A. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (Freeman, New York, 1999).
[16] D. Van der Spoel, R. Van Drunen, and H.J.C. Berendsen, Bioson Research Institute Nijenborgh Report No. 4 NL-9717 AG Groningen, 1999 (unpublished).
[17] E. Lindahal, B. Hess, and D. Van der Spoel, J. Mol. Model. [Electronic Publication] **7**, 306 (2001).
[18] B. Hess, H. Bekker, H.J.C. Berendsen, and J.G.E.M. Fraaije, J. Phys. Chem. **18**, 1463 (1997).
[19] H.J.C. Berendsen, J.R. Grigera, and T.P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
[20] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A.D. Nola, and J.R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[21] X. Daura, B. Jaun, D. Seebach, W.F. van Gunsteren, and A.E. Mark, J. Mol. Biol. **280**, 925 (1998).
[22] X. Daura, W.F. van Gunsteren, and A.E. Mark, Proteins **34**, 269 (1999).
[23] Y. Duan and P.A. Kollman, Science **282**, 740 (1998).
[24] Y. Zhou and M. Karplus, J. Chem. Phys. **107**, 3684 (1997).
[25] S.W. Rick, J. Phys. Chem. B **104**, 6884 (2000).
[26] H.S. Chan, Proteins **40**, 543 (2000).
[27] H. Kaya and H.S. Chan, Proteins **40**, 637 (2000).
[28] M.P. Allen and D.J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1987).
[29] P.H. Yang and J.A. Rupley, Biochemistry **18**, 2654 (1979).
[30] F.B. Sheinerman and C.L. Brooks, J. Mol. Biol. **278**, 439 (1998).
[31] F.H. Stillinger, Science **267**, 1935 (1995).
[32] Y. Zhou, D. Vitkup, and M. Karplus, J. Mol. Biol. **285**, 1371 (1999).
[33] H. Li, H. Yamada, and K. Akasaka, Biophys. J. **77**, 2801 (1999).
[34] J.L. Silva and G. Weber, Annu. Rev. Phys. Chem. **44**, 89 (1993).
[35] B. Wroblowski, J.F. Diaz, K. Heremans, and Y. Engelborghs, Proteins **25**, 446 (1996).
[36] O.B. Ptitsyn, Adv. Protein Chem. **47**, 83 (1995).
[37] R.L. Baldwin and B.H. Zimm, Proc. Natl. Acad. Sci. U.S.A. **97**, 12391 (2000).
[38] T. Lazaridis and M. Karplus, Science **278**, 1928 (1997).
[39] K. Goossens, L. Smeller, J. Frank, and K. Heremans, Eur. J. Biochem. **236**, 254 (1996).
[40] C. Simmerling, S. Strockbine, and A.E. Roitberg, J. Am. Chem. Soc. **124**, 11258 (2002).