FREE-ENERGY CALCULATION OF A PROTEIN AS A FUNCTION OF TEMPERATURE AND PRESSURE: MULTIBARIC-MULTITHERMAL MOLECULAR DYNAMICS SIMULATIONS

Hisashi Okumura*,**

*Institute for Molecular Science, Okazaki, Aichi 444-8585, Japan, E-mail: hokumura@ims.ac.jp
**The Graduate University for Advanced Studies, Okazaki, Aichi 444-8585, Japan

ABSTRACT

Free-energy calculation is one of the main topics in thermodynamics. However, biomolecules such as proteins have complicated free energy surfaces with many local minima. Conventional molecular dynamics (MD) and Monte Carlo (MC) simulations in physical ensembles, such as the canonical and isobaric-isothermal ensemble, tend to get trapped in these local-minimum states and cannot give the correct free-energy difference between different-conformational states. In order to avoid this difficulty, generalized-ensemble algorithms such as the multicanonical algorithm are frequently employed. However, because the multicanonical simulation is performed in a fixed volume, neither the pressure dependence nor temperature dependence at certain pressure can be investigated as in experiments. To overcome this difficulty, the author recently proposed multibaric-multithermal MD and MC algorithms. In this ensemble, two-dimensional random walks in the potential-energy space and in the volume space are realized. In this paper, the multibaric-multithermal molecular dynamics algorithm is reviewed and application to a 10-residue protein, chignolin is presented.

INTRODUCTION

Molecular dynamics (MD) simulation is a standard tool to calculate free-energy landscape and to investigate the conformational changes for biomolecules such as proteins at atomic level. However, biomolecules have complicated free energy surfaces with many local minima. Thus, conventional MD simulations in physical ensembles, such as the canonical [1; 2; 3] and isobaric-isothermal [4] ensemble, tend to get trapped in these local-minimum states. One of the powerful techniques to avoid this difficulty is generalized-ensemble algorithms [5; 6; 7; 8; 9] such as the multicanonical algorithm [10; 11; 12; 13]. In the multicanonical ensemble, a free one-dimensional random walk is realized in the potential-energy space and a simulation does not get trapped in free-energy-minimum states.

However, because the multicanonical simulation is performed in a fixed volume, neither the pressure dependence nor temperature dependence at certain pressure can be investigated as in experiments. To overcome this difficulty, the author have proposed the multibaric-multithermal algorithm [14; 15; 16; 17; 18; 19; 20]. In this algorithm, two-dimensional random walks are realized both in the potential-energy space and in the volume space, so that the temperature and pressure dependence can be discussed. This algorithm can be also used for Monte Carlo (MC) simulations. In this paper, the multibaric-multithermal algorithm is reviewed and applications of this algorithm to chignolin in explicit water is presented.

METHOD

In the isobaric-isothermal ensemble [4], the distribution $\mathcal{P}_{NPT}(E,V)$ of potential energy E and volume V is given by

$$\mathcal{P}_{NPT}(E,V) = n(E,V)\mathrm{e}^{-\beta_0 H} , \qquad (1)$$

where $n(E,V)$ is the density of states as a function of E and V and H is the "enthalpy" (without the kinetic energy contributions): $H = E + P_0 V$. Here, P_0 is the pressure at which simulations are performed. This ensemble has bell-shaped distributions both in the potential-energy space and in the volume space, as shown in Fig. 1(a). In order to obtain the isobaric-isothermal ensemble, the combination of the Nosé thermostat [1; 2] and the Andersen barostat [4] is frequently employed.

In the multibaric-multithermal ensemble [14; 15; 16; 17; 18], every state is sampled with a weight factor $W_{\mathrm{mbt}}(E,V) \equiv \exp\{-\beta_0 H_{\mathrm{mbt}}(E,V)\}$ so that a uniform distribution of both E and V, as shown in Fig. 1(b), may be obtained:

$$\mathcal{P}_{\mathrm{mbt}}(E,V) = n(E,V)W_{\mathrm{mbt}}(E,V) = \text{constant} . \qquad (2)$$

Here, $W_{\mathrm{mbt}}(E,V)$ and H_{mbt} are referred to as the multibaric-multithermal weight factor and the multibaric-multithermal enthalpy, respectively. The difference between H_{mbt} and H is written as $\delta H(E,V)$: $H_{\mathrm{mbt}}(E,V) = H + \delta H(E,V)$. The difference $\delta H(E,V)$ is therefore characteristic of the multibaric-multithermal simulation. The case of $\delta H(E,V) = 0$ gives the regular isobaric-isothermal ensemble.

The equations of motion in the multibaric-multithermal ensemble based on the Nosé thermostat [1; 2] and the Andersen barostat [4] are given by

$$\dot{r}_i = \frac{p_i}{m_i} + \frac{\dot{V}}{3V} r_i , \qquad (3)$$

$$\dot{p}_i = \left(1 + \frac{\partial \delta H}{\partial E}\right) F_i - \left(\frac{\dot{V}}{3V} + \frac{\dot{s}}{s}\right) p_i , \qquad (4)$$

$$\dot{V} = s\frac{P_V}{W} , \qquad (5)$$

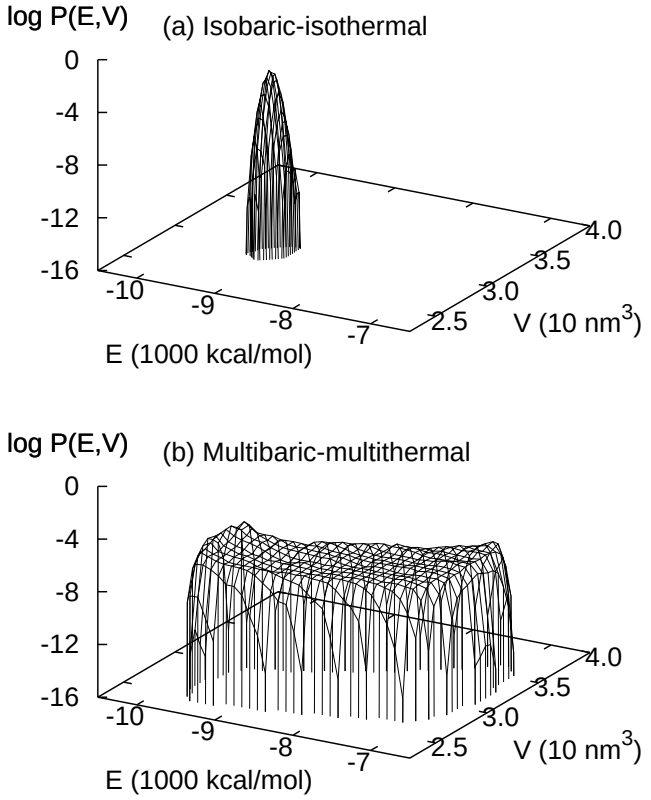

Figure 1. Distributions $\mathcal{P}(E,V)$ of potential energy E and volume V of an chignolin in explicit water (a) by the isobaric-isothermal MD simulation at $T = 298$ K and $P = 0.1$ MPa and (b) by the multibaric-multithermal MD simulation.

$$\dot{P_V} = s\left[\frac{1}{3V}\left\{\sum_{i=1}^{N}\frac{p_i^2}{m_i} + \left(1+\frac{\partial\delta H}{\partial E}\right)\sum_{i=1}^{N}F_i\cdot r_i\right\} - \left(P_0+\frac{\partial\delta H}{\partial V}\right)\right], \tag{6}$$

$$\dot{s} = s\frac{P_s}{Q}, \tag{7}$$

$$\dot{P_s} = \sum_{i=1}^{N+M}\frac{p_i^2}{m_i} - gk_{\mathrm{B}}T_0, \tag{8}$$

where r_i is the coordinate, p_i is the momentum, s is the additional degree of freedom for the Nosé thermostat, the dot ($\dot{}$) stands for the real time derivative d/dt, and F_i stands for the force acting on atom i. The variables P_V and P_s are the conjugate momenta for V and s, respectively. The constant m_i is the mass of atom i. The constants W and Q are the artificial "mass" related to V and s, respectively. Performing the MD simulation by the equations of motion, the multibaric-multithermal distribution $\mathcal{P}_{\mathrm{mbt}}(E,V)$ in Eq. (2) is obtained.

After an optimal weight factor $W_{\mathrm{mbt}}(E,V)$ is determined, for example, by the iterations of short simulations [21; 22] or by the Wang-Landau techniques [23], a long production run is performed for data collection. The reweighting techniques [24] are used for the results of the production run to calculate the isobaric-isothermal-ensemble averages. The expectation value of a physical quantity A at the desired temperature T and pressure P is given by

$$\langle A\rangle_{NPT} = \frac{\langle A(r,V)W_{\mathrm{mbt}}^{-1}(E(r,V),V)\mathrm{e}^{-\beta\{E(r,V)+PV\}}\rangle_{\mathrm{mbt}}}{\langle W_{\mathrm{mbt}}^{-1}(E(r,V),V)\mathrm{e}^{-\beta\{E(r,V)+PV\}}\rangle_{\mathrm{mbt}}}, \tag{9}$$

where $\langle\cdots\rangle_{\mathrm{mbt}}$ is the multibaric-multithermal ensemble average. Because of the random walks both in the potential-energy space and in the volume space, physical quantities can be calculated in wide ranges of T and P.

In order to calculate free-energy landscape, we should calculate first an unnormalized histogram as a function of reaction coordinates $(\xi_1,\xi_2,\cdots)$ by the reweighting techniques. In the case of two reaction coordinates (ξ_1,ξ_2), it is given by

$$\langle N(\xi_1,\xi_2)\rangle_{NVT} = \frac{\langle N(\xi_1(r),\xi_2(r))W_{\mathrm{mbt}}^{-1}(E(r,V),V)\mathrm{e}^{-\beta\{E(r,V)+PV\}}\rangle_{\mathrm{mbt}}}{\langle W_{\mathrm{mbt}}^{-1}(E(r,V),V)\mathrm{e}^{-\beta\{E(r,V)+PV\}}\rangle_{\mathrm{mbt}}}. \tag{10}$$

Probability distribution $P(\xi_1,\xi_2)$ is calculated by normalizing $\langle N(\xi_1,\xi_2)\rangle_{NVT}$:

$$P(\xi_1,\xi_2) = \frac{\langle N(\xi_1,\xi_2)\rangle_{NVT}}{\int d\xi_1 d\xi_2\langle N(\xi_1,\xi_2)\rangle_{NVT}} \tag{11}$$

The free-energy landscape then can be calculated by

$$F(\xi_1,\xi_2) = -k_{\mathrm{B}}T\log P(\xi_1,\xi_2). \tag{12}$$

APPLICATION TO CHIGNOLIN

Application to chignolin in explicit water solvent [25] is now presented. Chignolin is a 10 residue protein (GYDPETGTWG) of which the native state is a β-hairpin structure [26]. The N and C termini of the protein were left uncapped as in the experiment by Honda et al. [26]. That is, the N terminus and C terminus have a positive and negative electric charge, respectively. The system is consisting of one chignolin molecule, 902 water molecules, and two sodium ions Na^+ as counter ions. The initial values of the chignolin backbone dihedral angles were $\phi = \psi = 180^\circ$ except for proline. The initial dihedral-angle values of PRO4 were set to $\phi = -60^\circ$ and $\psi = 180^\circ$.

AMBER parm99SB force field [27] was used for the chignolin molecule and the TIP3P [28] rigid-body model was used for the water molecules. A cubic unit cell was employed with periodic boundary conditions. The electrostatic potential was calculated by the Ewald method. Cutoff distance was $r_{\mathrm{c}} = 12$ Å for both electrostatic and Lennard-Jone potential. The combination [29] of the Nosé-Hoover thermostat [1; 2; 3], the Andersen barostat [4], and the symplectic quaternion scheme [30; 31] was used for the rigid-body water molecules. Reversible multiple time scale molecular dynamics techniques [32] were also applied. The time step was taken to be Δt = 0.5 fs for the protein atoms and Δt = 4.0 fs for the water molecules. Because the symplectic rigid-body algorithm was used for the water molecules here, Δt was able to be taken as long as 4.0 fs [29].

During this MD simulation, the root mean square deviation (RMSD) decreased and increased repeatedly. It mean that folding and unfolding events occurred. the unfolding events occurred four times.

Temperature and pressure dependences of the fraction of the folded chignolin f_{fold} are shown in Fig. 2. The fraction of the folded protein f_{fold} decreases as temperature and /or pressure

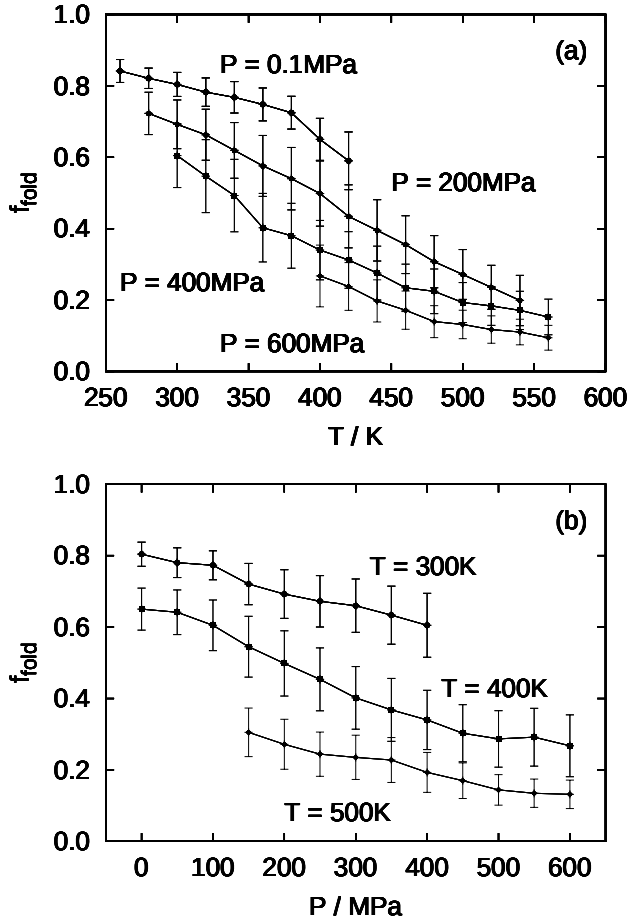

Figure 2. (a) Temperature dependence of fraction of the folded protein $f_{\rm fold}$ at $P = 0.1$, 200, 400, and 600 MPa and (b) pressure dependence of $f_{\rm fold}$ at $T = 300$, 400, and 500 K obtained by the reweighting techniques from the results of the multibaric-multithermal MD simulation.

increases. The partial molar enthalpy change ΔH and the partial molar volume change ΔV on unfolding is calculated by

$$\Delta H = R\left[\frac{\partial \log\{f_{\rm fold}/(1-f_{\rm fold})\}}{\partial(1/T)}\right]_P , \qquad (13)$$

$$\Delta V = RT\left[\frac{\partial \log\{f_{\rm fold}/(1-f_{\rm fold})\}}{\partial P}\right]_T . \qquad (14)$$

Interpolating or extrapolating the temperature dependence of $f_{\rm fold}$, the derivatives in Eqs. (13) and (14) were calculated. The partial molar enthalpy change ΔH was determined that $\Delta H = 24.1 \pm 4.9$ kJ/mol at $P = 0.1$ MPa, as listed in Table 1. Honda et al. determined ΔH by the CD spectroscopy and NMR experiments and their data distribute between 25.9 kJ/mol and 32.2 kJ/mol. The partial molar enthalpy change ΔH by the present multibaric-multithermal MD simulation is slightly lower than the experimental data, but shows a reasonably good agreement.

The partial molar volume change ΔV was determined that $\Delta V = -5.6 \pm 1.5$ cm^3/mol at $T = 298$ K, as listed in Table 1. Imamura and Kato obtained $\Delta V = -8.8$ cm^3/mol by their Fourier transform infraredspectrometer (FT-IR) experiments [33]. The absolute value of the present multibaric-multithermal MD result is slightly lower than their experimental data, but still agrees well.

The lower absolute values of ΔH and ΔV mean that chignolin

Table 1. Difference in the Gibbs free energy ΔG between the unfolded state and the folded state, difference in the partial molar enthalpy ΔH at $P = 0.1$ MPa, and difference in the partial molar volume ΔV at $T = 298$ K calculated by the multibaric-multithermal (MUBATH) MD simulation. Experimental data are taken from Refs. [26; 33].

Method	ΔG/(kJ/mol)	ΔH/(kJ/mol)	ΔV/(cm^3/mol)
MUBATH	3.5 ± 0.5	24.1 ± 4.9	-5.6 ± 1.5
Exp.	1.07 – 1.87	25.9 – 32.2	-8.8

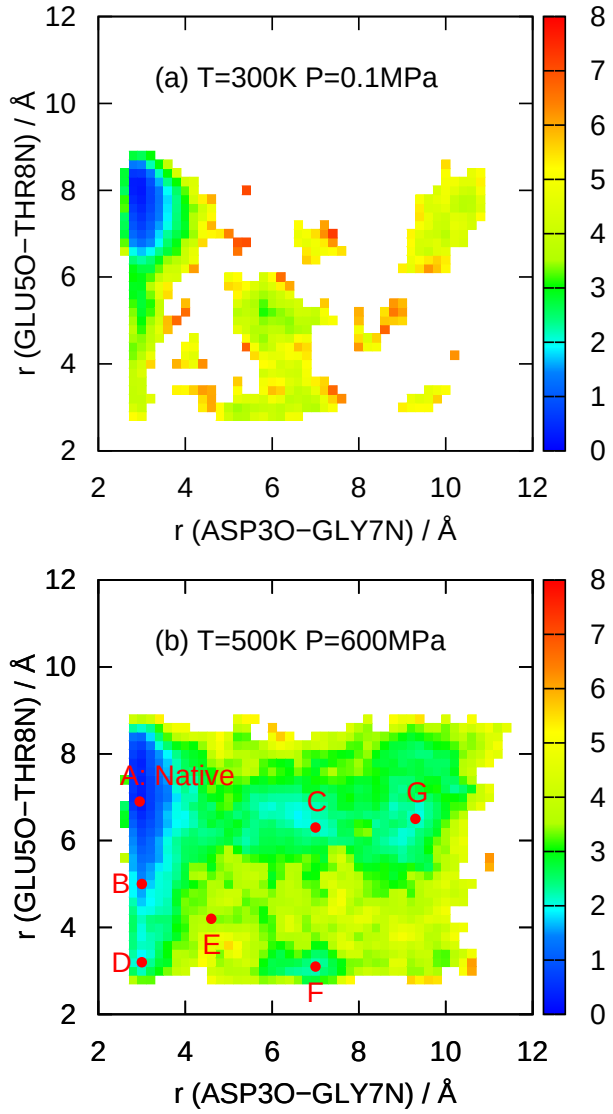

Figure 3. Potential of mean force (PMF) as a function of $r(\mathrm{ASP3O-GLY7N})$ and $r(\mathrm{GLU5O-THR8N})$ (a) at $T = 300$ K and $P = 0.1$ MPa and (b) at $T = 500$ K and $P = 600$ MPa obtained by the reweighting techniques from the results of the multibaric-multithermal MD simulation.

does not unfold with the increasing temperature or pressure in the MD simulation as fast as in the experiment. A possible reason is that the Coulomb potential parameter or the electrostatic charge in the classical force field is too large. Reducing the absolute value of the electrostatic charge may evaluate the hydrogen bonds weaker so that the simulational temperature/pressure

dependence of $f_{\rm fold}$ and ΔG may agree better with the experiment.

The present multibaric-multithermal MD simulation sampled not only the β-hairpin structure but also a 3_{10}-helix structure. Figure 3 shows potential of mean force as a function of $r(\mathrm{ASP3O}-\mathrm{GLY7N})$ and $r(\mathrm{GLU5O}-\mathrm{THR8N})$, which are the distance between ASP3O and GLY7N atoms and that between GLU5O and THR8N atoms, respectively. ASP3O and GLY7N atoms make a hydrogen bond at the β-hairpin structure and GLU5O and THR8N atoms make a hydrogen bond at the 3_{10}-helix structure. The potential of mean force has a global minimum state at $r(\mathrm{ASP3O}-\mathrm{GLY7N}) = 3.0$ Å and $r(\mathrm{GLU5O}-\mathrm{THR8N}) = 7.8$ Å at $T = 300$ K and $P = 0.1$ MPa, as shown in Fig. 3(a). This state is the native state, at which the chignolin folds into the β-hairpin structure. There also exists other states, which correspond to unfolded states. We can see blue and green distribution in a wide area at $T = 500$ K and $P = 600$ MPa, which means that the unfolded states increase, as shown in Fig. 3(b). There are several local-minimum free-energy states besides the global-minimum state. The state E is not a local-minimum states but a transition state between the states B and F or the states D and C.

Typical conformations obtained by the multibaric-multithermal MD simulation at these states are illustrated in Fig. 4. At the global-minimum state A, chignolin forms the native β-hairpin structure. The local-minimum state B is obtained from the state A by bending the C terminus of the β-hairpin structure. The local-minimum state C is obtained from either state A or state B by bending both N and C terminus of the β-hairpin structure. This conformation looks like a Greek letter "Ω". If the N terminus of the state B does not bent like the state C, but make a turn, an α-helix structure is obtained at the state D. The transition sate E is obtained by breaking some of the hydrogen bonds of the α-helix structure at the state D. It can be obtained also from the state B or C. The local-minimum state F, a 3_{10}-helix structure, is obtained from the state E by forming hydrogen bonds in a different way from the α-helix structure at the state D. This 3_{10}-helix structure can also be attained from the Ω-like structure at the state C by making turn structures. Finally an extended structure is obtained at the state G if all hydrogen bonds are broken from the state C or F. Following these pathways, the native β-hairpin structure at the state A unfolds to the extended structure at the state G. When chignolin folds, it follows the reverse process to the native structure.

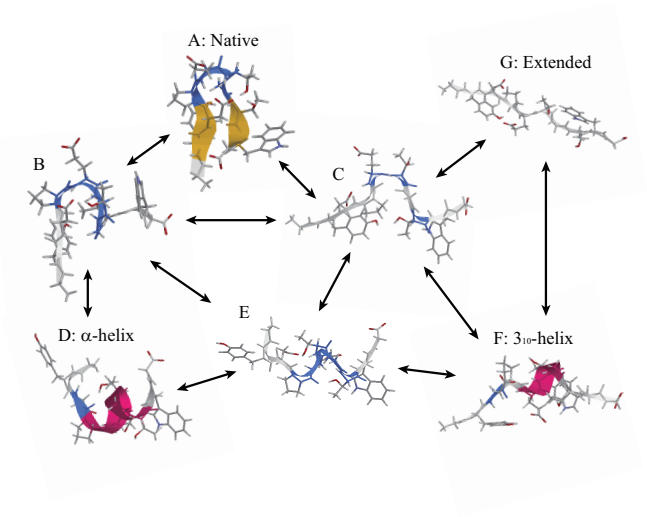

Figure 4. Folding and unfolding pathways of chignolin from the state A to G. These conformations are obtained by the multibaric-multithermal MD simulation. The N terminus and the C terminus are on the left-hand side and on the right-hand side, respectively. The figures were created with RasMol [34].

CONCLUSION

The multibaric-multithermal ensemble algorithm is reviewed. The multibaric-multithermal MD or MC simulation performs a two-dimensional random walk both in the potential-energy space and in the volume space so that one can obtain various isobaric-isothermal ensemble averages at different temperatures and pressures from only one simulation run. The multibaric-multithermal algorithm will thus be a powerful simulation technique to study the temperature and pressure dependences of biomolecules like proteins.

REFERENCES

[1] S. Nosé, A molecular-dynamics method for simulations in the canonical ensemble, *Mol. Phys.* vol. 52, pp. 255-268, 1984.

[2] S. Nosé, A unified formulation of the constant temperature molecular-dynamics methods, *J. Chem. Phys.* vol. 81, pp. 511-519, 1984.

[3] W.G. Hoover, Canonical dynamics: Equilibrium phase-space distributions, *Phys. Rev. A* vol. 31, pp. 1695-1697, 1985.

[4] H.C. Andersen, Molecular dynamics simulations at constant pressure and/or temperature, *J. Chem. Phys.* vol. 72, pp. 2384-2393, 1980.

[5] U. H. E. Hansmann and Y. Okamoto, The generalized-ensemble approach for protein folding simulations, in D. Stauffer (ed.), *Ann. Rev. Comput. Phys. VI*, pp. 129-158, World Scientific, Singapore, 1999.

[6] A. Mitsutake, Y. Sugita, Y. Okamoto, Generalized-ensemble algorithms for molecular simulations of biopolymers, *Biopolymers (Pept Sci)* vol. 60, pp. 96-123, 2001.

[7] B. A. Berg, Biased metropolis sampling for rugged free energy landscapes, *Comp. Phys. Commun.* vol. 104, pp. 52-57, 2002.

[8] H. Okumura, Partial multicanonical algorithm for molecular dynamics and Monte Carlo simulations, *J. Chem. Phys.* vol. 129, 124116, 2008.

[9] H. Okumura, Optimization of partial multicanonical molecular dynamics simulations applied to an alanine dipeptide in explicit water solvent, *Phys. Chem. Chem. Phys.* vol. 13, pp. 114-126, 2011.

[10] B. A. Berg, T. Neuhaus, Multicanonical algorithms for first order phase transitions, *Phys. Lett. B* vol. 267, pp. 249-253, 1991.

[11] B. A. Berg, T. Neuhaus, Multicanonical ensemble: a new approach to simulate first-order phase transitions, *Phys. Rev. Lett.* vol. 68, pp. 9-12, 1992.

[12] U. H. E. Hansmann, Y. Okamoto, and F. Eisenmenger, Molecular dynamics, Langevin and hybrid Monte Carlo simulations in a multicanonical ensemble, *Chem. Phys. Lett.* vol. 259, pp. 321-330, 1996.

[13] N. Nakajima, H. Nakamura, and A. Kidera, Multicanonical ensemble generated by molecular dynamics simula-

tion for enhanced conformational sampling of peptides, *J. Phys. Chem. B* vol. 101, pp. 817-824, 1997.
[14] H. Okumura and Y. Okamoto, Monte Carlo simulations in multibaric-multithermal ensemble, *Chem. Phys. Lett.* vol. 383, pp. 391-396, 2004.
[15] H. Okumura and Y. Okamoto, Monte Carlo simulations in generalized isobaric-isothermal ensembles, *Phys. Rev. E* vol. 70, 026702, 2004.
[16] H. Okumura and Y. Okamoto, Liquid-gas phase transitions studied by multibaric-multithermal Monte Carlo simulations, *J. Phys. Soc. Jpn.* vol. 73, pp. 3304-3311, 2004.
[17] H. Okumura and Y. Okamoto, Molecular dynamics simulations in the multibaric-multithermal ensemble, *Chem. Phys. Lett.* vol. 391, pp. 248-253, 2004.
[18] H. Okumura and Y. Okamoto, Multibaric-multithermal ensemble molecular dynamics simulations, *J. Comput. Chem.* vol. 27, pp. 379-395, 2006.
[19] H. Okumura and Y. Okamoto, Multibaric-multithermal molecular dynamics simulation of alanine dipeptide in explicit water, *Bull. Chem. Soc. Jpn.* vol. 80 pp. 1114-1123, 2007.
[20] H. Okumura and Y. Okamoto, Temperature and pressure dependence of alanine dipeptide studied by multibaric-multithermal molecular dynamics simulations, *J. Phys. Chem. B* vol. 112 pp. 12038-12049, 2008.
[21] B. A. Berg and T. Celik, New approach to spin-glass simulations, *Phys. Rev. Lett.* vol. 69, pp. 2292-2295, 1992.
[22] U. H. E. Hansmann and Y. Okamoto, Thermodynamics of helix-coil transitions studied by multicanonical algorithms, *J. Phys. Chem.* vol. 99, pp. 11276-11287, 1995.
[23] F. Wang and D. P. Landau, Efficient, multiple-range random walk algorithm to calculate the density of states, *Phys. Rev. Lett.* vol. 86, pp. 2050-2053, 2001.
[24] A. M. Ferrenberg and R. H. Swendsen, Optimized Monte Carlo data analysis, *Phys. Rev. Lett.* vol. 61, pp. 2635-2638, 1988. ibid. vol. 63, p. 1658(E), 1989.
[25] H. Okumura, Temperature and pressure denaturation of chignolin: Folding and unfolding simulation by multibaric-multithermal molecular dynamics method, *Proteins* vol. 80, pp. 2397-2416, 2012.
[26] S. Honda, K. Yamasaki, Y. Sawada, H. Morii, 10 residue folded peptide designed by segment statistics, *Structure* vol. 12, pp. 1507-1518, 2004.
[27] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Comparison of multiple amber force fields and development of improved protein backbone parameters, *Proteins* vol. 65, pp.712-725, 2006.
[28] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* vol. 79, pp. 926-925, 1983.
[29] H. Okumura, S. G. Itoh, and Y. Okamoto, Explicit symplectic integrators of molecular dynamics algorithms for rigid-body molecules in the canonical, isobaric-isothermal, and related ensembles, *J. Chem. Phys.* vol. 126, 084103, 2007.
[30] H. Yoshida, Construction of higher order symplectic integrators, *Phys. Lett. A* vol. 150, pp. 262-268, 1990.
[31] T. F. Miller, M. Eleftheriou, P. Pattnaik, A. Ndirango, D. Newns, and G. J. Martyna, Symplectic quaternion scheme for biophysical molecular dynamics, *J. Chem. Phys.* vol. 116, pp. 8649-8659, 2002.
[32] M. Tuckerman, B. J. Berne, and G. J. Martyna, Reversible multiple time scale molecular dynamics, *J. Chem. Phys.* vol. 97, pp. 1990-2001, 1992.
[33] H. Imamura and M. Kato, Unfolding of β-hairpin peptides by pressure : FT-IR and FRET studies, *The 48th Annual Meeting of the Biophysical Society of Japan*, 3P066, 2010.
[34] R. A. Sayle and E. J. Milner-White, RASMOL: biomolecular graphics for all, *Trends Biochem. Sci.* vol. 20, pp. 374-376, 1995.