REVIEW

Generalized-ensemble molecular dynamics and Monte Carlo algorithms beyond the limit of the multicanonical algorithm

Hisashi Okumura

Research Center for Computational Science, Institute for Molecular Science and Department of Structural Molecular Science, The Graduate University for Advanced Studies, Okazaki, Aichi 444-8585, Japan
E-mail: hokumura@ims.ac.jp

Received 7 July 2010
Accepted for publication 18 September 2010
Published 22 October 2010
Online at stacks.iop.org/ANSN/1/033002

Abstract
I review two new generalized-ensemble algorithms for molecular dynamics and Monte Carlo simulations of biomolecules, that is, the multibaric–multithermal algorithm and the partial multicanonical algorithm. In the multibaric–multithermal algorithm, two-dimensional random walks not only in the potential-energy space but also in the volume space are realized. One can discuss the temperature dependence and pressure dependence of biomolecules with this algorithm. The partial multicanonical simulation samples a wide range of only an important part of potential energy, so that one can concentrate the effort to determine a multicanonical weight factor only on the important energy terms. This algorithm has higher sampling efficiency than the multicanonical and canonical algorithms.

Keywords: molecular dynamics simulation, Monte Carlo simulation

Classification number: 3.02

1. Introduction

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 [1–4] and isobaric–isothermal [5, 6] ensemble, tend to get trapped in these local-minimum states. One of the powerful techniques to avoid this difficulty is generalized-ensemble algorithms [7–11], such as the multicanonical algorithm [12–15]. 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 the temperature dependence at certain pressure can be investigated as in experiments. In order to overcome this difficulty, I proposed recently the multibaric–multithermal algorithm [16–22]. In this ensemble, 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.

Another problem of the multicanonical algorithm is that the determination of the multicanonical weight factor to give a flat distribution becomes difficult in a large system. In order to alleviate this difficulty, I proposed the partial multicanonical algorithm [23, 24]. The partial multicanonical simulation samples a wide range of a part of the potential energy terms, which is necessary to sample the conformational space widely. Thus, one can concentrate the effort to determine a multicanonical weight factor only on the important energy terms, so that the partial multicanonical algorithm has a higher
sampling efficiency than the multicanonical and canonical algorithms.

In this review article, I explain the multibaric–multithermal algorithm and partial multicanonical algorithm. Applications of these algorithms to an alanine dipeptide in explicit water are also presented.

The multicanonical algorithm is briefly reviewed in section 2. In section 3, the multibaric–multithermal algorithm is explained. The partial multicanonical algorithm is described in section 4. Section 5 is devoted to conclusions.

2. The multicanonical algorithm

In the canonical ensemble, the distribution $P_{\text{NVT}}(E)$ of total potential energy $E$ is given by

$$P_{\text{NVT}}(E) = n(E) e^{-\beta_0 E},$$

where $n(E)$ is the density of states, $\beta_0$ is the inverse of the product of the Boltzmann constant $k_B$ and absolute temperature $T_0$, at which simulations are performed. This ensemble has a bell-shaped distribution in the potential-energy space.

In the multicanonical ensemble [12–15], every state is sampled with a weight factor $W_{\text{muca}}(E) = \exp\{\beta_0 E_{\text{muca}}(E)\}$ so that a uniform distribution of $E$ may be obtained,

$$P_{\text{muca}}(E) = n(E) W_{\text{muca}}(E) = \text{constant},$$

where $E_{\text{muca}}(E)$ and $W_{\text{muca}}(E)$ are referred to as the multicanonical energy and the multicanonical weight factor, respectively. The difference between $E_{\text{muca}}$ and $E$ is written as $\delta E(E) : E_{\text{muca}}(E) \equiv E + \delta E(E)$. The difference $\delta E(E)$ is characteristic of the multicanonical simulation. The case of $\delta E(E) = 0$ gives the regular canonical ensemble.

The multicanonical simulation covers a wide range of $E$ and escapes from local-minimum free-energy states. The expectation value of physical quantity $A$ at temperature $T$ is obtained by reweighting techniques [25] from

$$\langle A(E) \rangle_{\text{NVT}} \equiv \frac{\langle A(E) \exp[\beta_0(E + \delta E(E))] \exp(-\beta E) \rangle_{\text{muca}}}{\exp[\beta_0(E + \delta E(E))] \exp(-\beta E) \rangle_{\text{muca}},}$$

where $\langle \ldots \rangle_{\text{muca}}$ is the multicanonical ensemble average. Because of the random walks in the potential-energy space, we can calculate physical quantities in a wide range of $T$.

The multicanonical weight factor $W_{\text{muca}}(E)$ is not a priori known and has to be determined, for example, by the iterations of short simulations [26–28] and the Wang–Landau techniques [29].

3. The multibaric–multithermal algorithm

In the isobaric–isothermal ensemble [5, 6], the distribution $P_{\text{NPT}}(E, V)$ of potential energy $E$ and volume $V$ is given by

$$P_{\text{NPT}}(E, V) = n(E, V) e^{-\beta_0 H},$$

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 figure 1(a).

In the multibaric–multithermal ensemble [16–20], every state is sampled with a weight factor $W_{\text{mbt}}(E, V) = \exp\{-\beta_0 H_{\text{mbt}}(E, V)\}$ so that a uniform distribution of both $E$ and $V$, as shown in figure 1(b), may be obtained,

$$P_{\text{mbt}}(E, V) = n(E, V) W_{\text{mbt}}(E, V) = \text{constant}. \quad (5)$$

Here, $W_{\text{mbt}}(E, V)$ and $H_{\text{mbt}}$ are referred to as the multibaric–multithermal weight factor and the multibaric–multithermal enthalpy, respectively. The difference between $H_{\text{mbt}}$ and $H$ is written as $\delta H(E, V) : H_{\text{mbt}}(E, V) \equiv H + \delta H(E, V)$. 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 isobaric–isothermal Hamiltonian $H_{\text{mbt}}$ for a system consisting of $N$ atoms based on the Nosé thermostat [1, 2] and the Andersen barostat [5] is given by

$$H_{\text{mbt}} = \sum_{i=1}^{N} \frac{\bar{p}_i^2}{2m_i V^{2/3}s^2} + E(V^{1/3} r_i) + \frac{P_V^2}{2W} + P_0 V$$

$$+ \frac{P_s^2}{2 Q} + gk_B T_0 \log s + \delta H(E(V^{1/3} r_i), V).$$

(6)

The variable $\bar{r}_i$ is the coordinate scaled by the length of the simulation box: $\bar{r}_i = V^{-1/3} r_i$ ($r_i$ is the real coordinate). We have introduced a simplified notation for the set of the scaled coordinates, $\bar{r} = \{\bar{r}_1, \bar{r}_2, \ldots, \bar{r}_N\}$. The variable $\bar{p}_i$ is the conjugate momentum for $\bar{r}_i$. The real momentum $p_i$ is related to the virtual momentum $\bar{p}_i$ by $p_i = \bar{p}_i / V^{1/3}s$, where $s$ is the additional degree of freedom for the Nosé thermostat. The variable $P_V$ and $P_s$ are the conjugate momenta for $V$.
and $s$, respectively. The real-time interval $\Delta t$ is associated with the virtual time interval $\Delta t'$ by $\Delta t = \Delta t'/s$. 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. In the case of a system consisting of $N$ atoms, $g$ equals $3N$ if the time development is performed in real time $t$, or $g$ equals $3N + 1$ if the time development is performed in virtual time $t'$.

The equations of motion in real time are given from the Hamiltonian in equation (6) by

$$\dot{r}_i = \frac{p_i}{m_i} + \frac{\dot{V}}{3V} r_i, \quad (7)$$

$$p_i = \left(1 + \frac{\partial \delta H}{\partial E}\right) F_i - \left(\frac{\dot{V}}{3V} + \frac{s}{s}\right) p_i,$$  \quad (8)

$$\dot{V} = s \frac{p_V}{W}, \quad (9)$$

$$P_V = s \left[ \frac{1}{3V} \left( \sum_{i=1}^{N} \frac{p_i^2}{m_i} \right) + \left(1 + \frac{\partial \delta H}{\partial E}\right) \sum_{i=1}^{N} F_i \cdot r_i \right] - \left(P_0 + \frac{\partial \delta H}{\partial V}\right), \quad (10)$$

$$\dot{s} = \frac{P_s}{Q}, \quad (11)$$

$$P_s = \sum_{i=1}^{N} \frac{p_i^2}{m_i} - g k_B T_0, \quad (12)$$

where the dot (·) stands for the real-time derivative $d/dt$ and $F_i$ stands for the force acting on atom $i$. Performing the MD simulation by the equations of motion in equations (7)–(12), the multibaric–multithermal distribution $P_{\text{mbt}}(E, V)$ in equation (5) is obtained.

In the case of $\delta H(E, V) = \delta \Delta H(E, V)/\partial E = \delta \delta H(E, V)/\partial V = 0$, the Hamiltonian in equation (6) and the equations of motion in equations (7)–(12) become those for the regular isobaric–isothermal MD simulation of the Nosé–Andersen formulation.

In order to perform a multibaric–multithermal MC simulation, we employ the Metropolis criterion on $\tilde{r}_i$ and $V$. The trial moves from $\tilde{r}_i$ to $\tilde{r}_i'$ and is generated by uniform random numbers. The multibaric–multithermal enthalpy is consequently changed from $H_{\text{mbt}}(\tilde{r}, V)$ to $H'_{\text{mbt}} = H_{\text{mbt}}(\tilde{r}', V')$ by the trial moves. The trial moves are now accepted with the probability

$$\omega = \min(1, \exp[-\beta_0(H'_{\text{mbt}} - H_{\text{mbt}} - N k_B T_0 \ln(V'/V))]). \quad (13)$$

The multibaric–multithermal probability distribution $P_{\text{mbt}}(E, V)$ is obtained by this scheme.

After an optimal weight factor $W_{\text{mbt}}(E, V)$ is determined—for example, by the iterations of short simulations [26–28] and by the Wang–Landau techniques [29]—a long production run is performed for data collection. The reweighting techniques [25] 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 $T$ and $P$ is then obtained by

$$\langle A \rangle_{\text{NPT}} = \int dV \int dE A(E, V) \frac{P_{\text{NPT}}(E, V; T, P)}{\langle W_{\text{mbt}}^{-1}(E, V) e^{-\beta (E + PV)} \rangle_{\text{mbt}}}, \quad (14)$$

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

I introduce here an application of the multibaric–multithermal MD algorithm to a system consisting of one alanine dipeptide molecule ([S]-2-(acetylamino)-N-methylpropanamide) and 63 water molecules [22]. The AMBER parm96 force field [30] was used for the alanine dipeptide molecule and the TIP3P rigid-body [31] model was employed for the water molecules. The electrostatic potential was calculated by the Ewald method. The time step was taken to be $\Delta t = 0.5$ fs. The temperature was set as $T_0 = 600$ K and the pressure was set as $P_0 = 300$ MPa. The MD simulations were performed with the symplectic combination [32] of the Nosé–Poincaré thermostat [33, 34], the Andersen barostat [5] and the symplectic quaternions method [35]. In order to obtain an optimal multibaric–multithermal weight factor $W_{\text{mbt}}(E, V)$, two-dimensional versions of both the iterative method [26–28] and the Wang–Landau techniques [29] for multicanonical weight-factor determination were employed. An optimal weight factor was obtained after these preliminary simulations in total for 2.61 ns in the AMBER parm96 simulation. Two sets of the initial values of the alanine dipeptide backbone dihedral angles were prepared for the data-production runs. The initial condition 1 was $\phi = \psi = 180^\circ$ and the initial condition 2 was $\phi = 180^\circ$ and $\psi = 0^\circ$. The initial conformation 1 is shown in figure 2. We then performed a long multibaric–multithermal MD simulation for data collection. The simulation time was 20 ns for each initial condition. The isobaric–isothermal MD simulations of the same system were also performed for comparisons.

Figure 3 shows the contour maps of the distributions $P(\phi, \psi)$ at $T = 240$ K and $P = 0.1$ MPa by the MD simulations started from the initial condition 1. The distribution $P(\phi, \psi)$ by the multibaric–multithermal MD were obtained using the reweighting techniques. The isobaric–isothermal distribution $P(\phi, \psi)$ has only four peaks of the $m_0, m_0, P_0$ and $C_5$ states, as shown in figure 3(a). On the other hand, the multibaric–multithermal

![Figure 2](image-url)
MD simulation provides correct distributions $P(\phi, \psi)$ with all the six peaks of the $\alpha_R, \alpha_P, \alpha_L, C_5, C_\alpha$, and $C_\beta$ states, as shown in figure 3(b). These results mean that the multibaric–multithermal simulation can get over the free energy barriers and has a high sampling efficiency.

The volume under the surface $P(\phi, \psi)$ around each peak corresponds to the population $W$ of each state. The ratio of each state population $W$ against the $P_{II}$ state population was calculated as a function of the inverse of temperature $1/T$ at the constant pressure of $P = 0.1$ MPa, as shown in figure 4(a) and as a function of the inverse of pressure $P$ at the constant temperature of $T = 298\,\text{K}$, as shown in figure 4(b). The difference in partial molar enthalpy $\Delta H$ and the difference in partial molar volume $\Delta V$ can be obtained from the gradient in figure 4(a) and that in figure 4(b), respectively. Partial molar enthalpy and partial molar volume are important quantities in solution chemistry, which describes the change in the ratio of each state population due to temperature and pressure changes. As the temperature increases, the relative population $W_{C_5}/W_{P_{II}}$ of the $C_5$ state increases, as shown in figure 4(a). This indicates that the enthalpy for the $C_5$ state is higher than that of the $P_{II}$ state. The difference in partial molar enthalpy $\Delta H$ and the difference in partial molar volume $\Delta V$ of the $C_5$ state from that of the $P_{II}$ state, for example, are calculated from the derivative of $W_{C_5}/W_{P_{II}}$ by
\[
\Delta H = -R \left[ \frac{\partial \log(W_{C_5}/W_{P_{II}})}{\partial (1/T)} \right]_P,
\]
\[
\Delta V = -RT \left[ \frac{\partial \log(W_{C_5}/W_{P_{II}})}{\partial P} \right]_T,
\]
where $R$ is the gas constant: $R = 8.3145\,\text{J}\,(\text{mol}\,\text{K})^{-1}$. The values of $\Delta H$ and $\Delta V$ are listed in tables 1 and 2, respectively. The differences $\Delta H$ and $\Delta V$ of the $C_5$ state and $\Delta H$ of the $\alpha_R$ state are consistent in the multibaric–multithermal MD and the Raman spectroscopy experimental data [37]. The value of $\Delta V$ of the $\alpha_R$ state, however, is smaller than that of the Raman spectroscopy.

**Figure 3.** Contour maps of the logarithms of the probability distributions $P(\phi, \psi)$ of the backbone dihedral angles $\phi$ and $\psi$ at $T = 240\,\text{K}$ and $P = 0.1$ MPa. (a) $P(\phi, \psi)$ obtained from initial condition 1 by the conventional isobaric–isothermal MD simulation. (b) $P(\phi, \psi)$ obtained from initial condition 1 by the reweighting techniques from the results of the multibaric–multithermal MD simulation.

**Figure 4.** The population ratios $W/W_{P_{II}}$ against the $P_{II}$ (a), (a’) as functions of the inverse of temperature $1/T$ at constant pressure of $P = 0.1$ MPa and (b), (b’) as functions of pressure $P$ at constant temperature of $T = 298\,\text{K}$. These data were obtained by the reweighting techniques from the results of the multibaric–multithermal MD simulation of (a), (b) the $C_5$ state and (a’), (b’) the $\alpha_R$ state.

| Table 1 | Differences $\Delta H$/(kJ mol$^{-1}$) in partial molar enthalpy of the $C_5$ and $\alpha_R$ states from that of the $P_{II}$ state calculated by the multibaric–multithermal (MUBATH) MD simulations. Raman experimental data are taken from [37]. |
|---------|--------------------------------------------------------------------------------|
| State   | MUBATH MD | Raman exp. |
|---------|------------|------------|
| $C_5$   | 1.8 ± 0.5  | 2.5 ± 0.3  |
| $\alpha_R$ | 4.4 ± 1.1  | 4.4 ± 1.5  |

| Table 2 | Differences $\Delta V$/(cm$^3$ mol$^{-1}$) in partial molar volume of the $C_5$ and $\alpha_R$ states from that of the $P_{II}$ state calculated by the multibaric–multithermal (MUBATH) MD simulations. Raman experimental data are taken from [37]. |
|---------|--------------------------------------------------------------------------------|
| State   | MUBATH MD | Raman exp. |
|---------|------------|------------|
| $C_5$   | −0.3 ± 0.7 | 0.1 ± 0.3  |
| $\alpha_R$ | −2.3 ± 1.3 | 1.1 ± 0.2  |

The multicanonical simulation can escape from local-minimum free-energy states but cannot change volume. The conventional isobaric–isothermal MD simulation can
change volume but is trapped in a local free energy minimum. Therefore, $\Delta H$ and $\Delta V$ cannot be calculated by these simulation algorithms. The multibaric–multithermal algorithm has the advantages of both algorithms. This is the first simulational work to calculate $\Delta H$ and $\Delta V$ of biomolecules.

### 4. The partial multicanonical algorithm

Although the multicanonical algorithm is a powerful technique, the determination of the multicanonical weight factor is time consuming for a large system. In order to alleviate the effort to determine the weight factor, the total potential energy is divided into two terms in the partial multicanonical algorithm [23, 24],

$$E = E_1 + E_2$$ (16)

where $E_1$ is a part of the potential energy which is important to sample a wide range of the conformational space. The partial potential energy $E_1$ can be taken in any manner, in principle. The electrostatic, Lennard Jones and torsion-angle energy terms between solute atoms may be included in a biomolecular system. The second term $E_2$ is the rest of the potential energy. The multicanonical algorithm covers a wide range of $E_1$. The partial multicanonical simulation is performed with a weight factor $W_{\text{pmuca}}(E_1, E_2) = \exp \{-\beta_0 E_{\text{pmuca}}(E_1, E_2)\}$ so that a uniform distribution of $E_1$ may be obtained,

$$P_{\text{pmuca}}(E_1, E_2) = n(E_1, E_2) W_{\text{pmuca}}(E_1, E_2) = \text{constant on } E_1,$$

where $E_{\text{pmuca}}(E_1, E_2)$ and $W_{\text{pmuca}}(E_1, E_2)$ are referred to as the partial multicanonical energy and the partial multicanonical weight factor, respectively. The difference between $E_{\text{pmuca}}$ and $E$ depends only on $E_1$ and is written as $\delta E_1(E_1)$,

$$E_{\text{pmuca}}(E_1, E_2) \equiv E_1 + E_2 + \delta E_1(E_1) = E + \delta E_1(E_1)$$ (18)

The difference $\delta E_1(E_1)$ is characteristic of the partial multicanonical simulation. If we define here a distribution $P_{\text{pmuca}}(E_1)$ of $E_1$ and the density of states $n(E_1)$ of $E_1$ as

$$P_{\text{pmuca}}(E_1) = \int dE_2 \, P_{\text{pmuca}}(E_1, E_2),$$

$$n(E_1) = \int dE_2 \, n(E_1, E_2) \exp \{-\beta_0 E_2\},$$

equation (17) is written as

$$P_{\text{pmuca}}(E_1) = n(E_1) \exp \{-\beta_0 (E_1 + \delta E_1(E_1))\} = \text{constant.}$$ (21)

The expectation value of a physical quantity $A$ is obtained by the reweighting techniques [25] from

$$\langle A \rangle_{\text{NVT}} = \frac{\langle A(E) \exp[\beta_0 (E + \delta E_1(E_1))] \exp \{-\beta E\} \rangle_{\text{pmuca}}}{\langle \exp[\beta_0 (E + \delta E_1(E_1))] \exp \{-\beta E\} \rangle_{\text{pmuca}}}.$$

where $\langle \cdot \cdot \cdot \rangle_{\text{pmuca}}$ is the partial multicanonical ensemble average. Note that $\langle A \rangle_{\text{NVT}}$ can be calculated correctly at a temperature $T$ only in the vicinity of $T_0$, although $\langle A \rangle_{\text{NVT}}$ can be calculated in a wide range of $T$ in the multicanonical algorithm. This is because a limited range of $E_2$ is sampled with the normal Boltzmann weight factor $\exp(-\beta_0 E_2)$ in partial multicanonical algorithm.

The scheme of the partial multicanonical algorithm includes the canonical and the multicanonical algorithms. In the case of $E_1 = 0$ and $E_2 = E$, this algorithm becomes the regular canonical algorithm. In the case of $E_1 = E$ and $E_2 = 0$, the multicanonical algorithm is obtained.

In order to perform a partial multicanonical MD simulation, a constant-temperature MD method is employed. The partial multicanonical Hamiltonian based on the Nosé thermostat [1, 2] is given by

$$H_{\text{pmuca}} = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + E(r) + \frac{p_i^2}{2Q} + gk_B T_0 \log s + \delta E_1(E_1(r))$$ (23)

where $p_i$ is the conjugate momentum for $r_i$. The real momentum $p_i$ is related to the virtual momentum $p_i'$ by $p_i = p_i'/s$. In the Nosé–Hoover formalism [1–3], a variable $\zeta \equiv s/s_0$ is introduced and the equations of motion are given from the Hamiltonian in equation (23) by

$$\dot{r}_i = \frac{p_i}{m_i}$$ (24)

$$\dot{p}_i = \left\{ \left( 1 + \frac{\partial \delta E_1}{\partial E_1} \right) \frac{\partial E_1}{\partial r_i} + \frac{\partial E_2}{\partial r_i} \right\} - \zeta \, p_i$$ (25)

where $\zeta = 1$ and

$$\dot{\zeta} = \frac{1}{Q} \sum_{i=1}^{N} \frac{p_i^2}{m_i} - gk_B T_0.$$

Equation (25) is obtained from regular canonical-ensemble equations of motion by modifying force from

$$F_i = -\frac{\partial E}{\partial r_i}$$ (27)

to

$$F_i = \left\{ \left( 1 + \frac{\partial \delta E_1}{\partial E_1} \right) \frac{\partial E_1}{\partial r_i} + \frac{\partial E_2}{\partial r_i} \right\}. $$ (28)

For a partial multicanonical MC simulation, we perform Metropolis sampling on the coordinates $r$. The trial moves from $r_i$ to $r'_i$ are generated by uniform random numbers. The partial multicanonical energy is consequently changed from $E_{\text{pmuca}}(r)$ to $E_{\text{pmuca}}' = E_{\text{pmuca}}(r')$ by these trial moves. The trial moves are accepted by the Metropolis criterion [4] with the probability

$$\omega = \min(1, \exp \{-\beta_0 (E_{\text{pmuca}}' - E_{\text{pmuca}})\}).$$ (29)

The partial multicanonical probability distribution $P_{\text{pmuca}}(E_1)$ is obtained by this scheme.

The partial multicanonical MD algorithm was applied to a system consisting of one alanine dipeptide molecule and 67 water molecules [23]. The total potential energy is the sum of the potential energy between solute atoms $E_{\text{slit}}$ and that of the
solute–solvent interaction $E_{\text{slt-svn}}$, and that between solvent atoms $E_{\text{svn-svn}}$, as follows:

$$E = E_{\text{slt-slt}} + E_{\text{slt-svn}} + E_{\text{svn-svn}}. \quad (30)$$

Each term for the AMBER force field [30] is given by

$$E_{\text{slt-slt}} = E_{\text{bond}}^{\text{slt}} + E_{\text{angle}}^{\text{slt}} + E_{\text{torsion}}^{\text{slt}} + E_{\text{imp}}^{\text{slt}} + E_{\text{elec}}^{\text{slt}} + E_{\text{LJ}}^{\text{slt}} \quad (31)$$

$$E_{\text{slt-svn}} = E_{\text{elec}}^{\text{slt}} + E_{\text{LJ}}^{\text{slt}} \quad (32)$$

$$E_{\text{svn-svn}} = E_{\text{svn-svn}}^{\text{elec}} + E_{\text{LJ}}^{\text{svn}} \quad (33)$$

where the subscripts bond, angle, torsion, imp, elec and LJ stand for the bond-length, bond-angle, torsion-angle, improper torsion-angle, electrostatic and Lennard Jones potential energy, respectively. Although these terms may vary for different force fields, the partial multicanonical algorithm can be applied in essentially the same way. Even if we sample wide ranges of bond-length, bond-angle and improper torsion-angle energy in $E_{\text{slt-slt}}$, the structure of the solute molecule is not expected to be changed very much. Sampling a wide range of the solvent energy $E_{\text{svn-svn}}$ would not change the solute structure very much, either. Sampling a wide range of the solute–solvent potential energy $E_{\text{slt-svn}}$ may sample a wide range of the interface area between the solute molecule and solvent molecules. The solute molecule of the alanine dipeptide in this simulation, however, is small and this effect was not taken into account in this simulation. Finally, the partial potential energy $E_1$ was here taken to be the sum of the torsion-angle, electrostatic and Lennard Johnes energy between the solute molecules as follows:

$$E_1 = E_{\text{slt-slt}}^{\text{elec}} + E_{\text{slt-slt}} + E_{\text{LJ}} \quad (34)$$

and $E_2$ was the rest of the potential energy,

$$E_2 = E_{\text{slt-slt}}^{\text{bond}} + E_{\text{slt-slt}}^{\text{angle}} + E_{\text{imp}}^{\text{slt}} + E_{\text{slt-svn}}^{\text{elec}} + E_{\text{svn-svn}}^{\text{slt}} \quad (35)$$

Note that this choice of $E_1$ is not always the best. The best choice of $E_1$ should depend on the system and the force field.

In order to obtain an optimal partial multicanonical weight factor $W_{\text{pmuc}}(E_1, E_2)$, both the iterative method [26–28] and the Wang-Landau techniques [29] were employed. An optimal weight factor was obtained after these preliminary simulations for 2 ns in total. I then performed a long partial multicanonical MD simulation for data collection for 6 ns at $T_0 = 300$ K.

The probability distributions of $E$ and $E_1$ are shown in figure 5. The distribution of $E$ in the multicanonical simulation was much wider than those in the other two simulations. The partial multicanonical algorithm sampled a much wider range of $E_1$. These generalized-ensemble MD simulations actually sampled a wide range of the potential energy, which and only which are predesignated to be sampled widely.

The time series of $\phi$ is shown in figure 6. The canonical MD simulation did not make a complete rotation, as shown in figure 6(a). It covered only 65.6% of the whole range of $\phi$. There seem to be potential energy barriers around $\phi = 0^\circ$ and $\phi = 120^\circ$, which cannot be overcome by the canonical simulation (these energy barriers correspond to the boundaries between the local-minimum states). This means that the canonical simulation was trapped in local-minimum free-energy states. The multicanonical MD simulation also did not make a complete rotation and sampled 82.8% of the whole range of $\phi$, as shown in figure 6(b). It overcame the potential energy barrier around $\phi = 0^\circ$ but did not overcome that around $\phi = 120^\circ$. The partial multicanonical MD simulation, on the other hand, overcame both energy barriers and covered the whole range of $\phi$, as shown in figure 6(c). The backbone dihedral angle $\phi$ underwent a complete rotation (covering 360°) six times (six turns) in this MD simulation for 6 ns.

Note that these results do not mean that the multicanonical algorithm cannot sample the whole range of $\phi$. If a better weight factor $W_{\text{pmuc}}(E)$ were obtained with the longer preliminary simulations, the whole range of $\phi$ could be sampled also in the multicanonical ensemble. These results mean that the partial multicanonical algorithm shows higher sampling efficiency than the multicanonical algorithm if the weight factor is determined with the same effort.

The probability distributions $P_{\text{NVT}}(\phi, \psi)$ of $\phi$ and $\psi$ are shown in figure 7. In order to obtain the distribution $P_{\text{NVT}}(\phi, \psi)$ in the canonical ensemble from the multicanonical simulation and that from the partial
multicanonical algorithm will thus be of great use for the protein folding problem. This idea, in which a simulation covers a wide range of only important potential energy, which is necessary to investigate a wide range of the conformational space. Thus, one can concentrate the effort for the weight factor determination on the important energy terms. The MD simulation of an alanine dipeptide showed that the multicanonical simulation showed higher sampling efficiency than the canonical simulation. The partial multicanonical algorithm will thus be of great use for the protein folding problem. This idea, in which a simulation covers a wide range of only important potential energy terms, can be utilized not only for the multicanonical algorithm but also for multidimensional extensions of the multicanonical algorithm, such as the multibaric–multithermal [16–22] and multicanonical–multioverlap [38] algorithms.

5. Conclusions

I have reviewed two new generalized-ensemble MD and MC algorithms: the multibaric–multithermal and partial multicanonical algorithms. These generalized-ensemble algorithms allow the simulation to escape from local-minimum free-energy states and sample many conformations.

The multibaric–multithermal 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. From the temperature and pressure dependences of an alanine dipeptide in explicit water, the differences in the partial molar enthalpy $\Delta H$ and the partial molar volume $\Delta V$ were calculated between the $P_{II}$ state and other states. This is the first work to calculate $\Delta H$ and $\Delta V$ by molecular dynamics simulations. The multibaric–multithermal algorithm will thus be a powerful simulation technique to study the temperature and pressure dependences of biomolecules like proteins.

In the partial multicanonical algorithm, the simulation covers a wide range of only important potential energy, which is necessary to investigate a wide range of the conformational space. Thus, one can concentrate the effort for the weight factor determination on the important energy terms. The MD simulation of an alanine dipeptide showed that the multicanonical simulation showed higher sampling efficiency than the canonical simulation. The partial multicanonical algorithm will thus be of great use for the protein folding problem. This idea, in which a simulation covers a wide range of only important potential energy terms, can be utilized not only for the multicanonical algorithm but also for multidimensional extensions of the multicanonical algorithm, such as the multibaric–multithermal [16–22] and multicanonical–multioverlap [38] algorithms.

![Figure 6](image.png)

Figure 6. The time series of the backbone dihedral angle $\phi$ obtained by (a) the canonical MD simulation, (b) the multicanonical MD simulation and (c) the partial multicanonical MD simulation.

![Figure 7](image.png)

Figure 7. Contour maps of the logarithms of the probability distributions $\log P_{NVT}(\phi, \psi)$ at $T = 300$ K obtained (a) by the canonical MD simulation, (b) by the reweighting techniques from the results of the multicanonical MD simulation and (c) by the reweighting techniques from the results of the partial multicanonical MD simulation.
References

[1] Nosé S 1984 Mol. Phys. 52 255
[2] Nosé S 1984 J. Chem. Phys. 81 511
[3] Hoover W G 1985 Phys. Rev. A 31 1695
[4] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 J. Chem. Phys. 21 1087
[5] Andersen H C 1980 J. Chem. Phys. 72 2384
[6] McDonald I R 1972 Mol. Phys. 23 41
[7] Hansmann U H E and Okamoto Y 1999 Ann. Rev. Comput. Phys. VI ed D Stauffer (Singapore: World Scientific) p 129
[8] Mitsutake A, Sugita Y and Okamoto Y 2001 Biopolymers (Pept. Sci.) 60 96
[9] Berg B A 2002 Comput. Phys. Commun. 104 52
[10] Okamoto Y 2004 J. Mol. Graph. Mod. 22 425
[11] Itoh S G, Okamura H and Okamoto Y 2007 Mol. Simul. 33 47
[12] Berg B A and Neuhaus T 1991 Phys. Lett. B 267 249
[13] Berg B A and Neuhaus T 1992 Phys. Rev. Lett. 68 9
[14] Hansmann U H E, Okamoto Y and Eisenmenger F 1996 Chem. Phys. Lett. 259 321
[15] Nakajima N, Nakamura H and Kidara A 1997 J. Phys. Chem. B 101 817
[16] Okumura H and Okamoto Y 2004 Chem. Phys. Lett. 383 391
[17] Okumura H and Okamoto Y 2004 Phys. Rev. E 70 026702
[18] Okumura H and Okamoto Y 2004 J. Phys. Soc. Japan 73 3304
[19] Okumura H and Okamoto Y 2004 Chem. Phys. Lett. 391 248
[20] Okumura H and Okamoto Y 2006 J. Comput. Chem. 27 379
[21] Okumura H and Okamoto Y 2007 Bull. Chem. Soc. Japan 80 1114
[22] Okumura H and Okamoto Y 2008 J. Phys. Chem. B 112 12038
[23] Okumura H 2008 J. Chem. Phys. 129 124116
[24] Okumura H 2010 Phys. Chem. Chem. Phys. submitted
[25] Ferrenberg A M and Swenden R H 1988 Phys. Rev. Lett. 61 2635
[26] Berg B A and Celik T 1992 Phys. Rev. Lett. 69 2292
[27] Lee J 1993 Phys. Rev. Lett. 71 211
[28] Lee J 1993 Phys. Rev. Lett. 71 2353
[29] Okamoto Y and Hansmann U H E 1995 J. Phys. Chem. 99 11276
[30] Wang F and Landau D P 2001 Phys. Rev. Lett. 86 2050
[31] Kollman P A, Dixon R, Cornell W, Fox T, Chipot C and Pohorille A 1997 Computer Simulation of Biomolecular Systems vol 3 ed A Wilkinson, P Weiner and W F van Gunsteren (Dordrecht: Elsevier) p 83
[32] Jorgensen W L, Chandradekhar J, Madura J D, Impey R W and Klein M L 1983 J. Chem. Phys. 79 926
[33] Okumura H, Itoh S G and Okamoto Y 2007 J. Chem. Phys. 126 084103
[34] Bond S D, Leimkuhler B J and Laird B B 1999 J. Comput. Phys. 151 114
[35] Nosé S 2001 J. Phys. Soc. Japan 70 75
[36] Miller T F, Elefttheriou M, Pattnaik P, Ndiriango A, Newns D and Martyna G J 2002 J. Chem. Phys. 116 8649
[37] Sayle R A and Milner-White E J 1995 Trends Biochem. Sci. 20 374
[38] Takekiyo T, Imai T, Kato M and Taniguchi Y 2004 Biopolymers 73 283
[39] Itoh S G and Okamoto Y 2007 Phys. Rev. E 76 026705