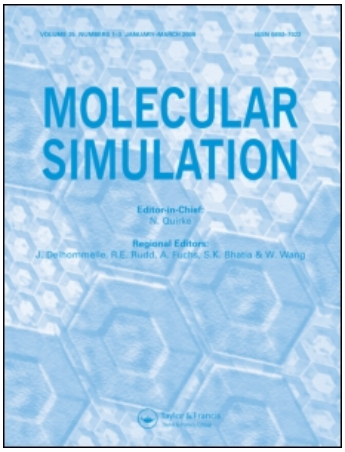

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Generalized-ensemble algorithms for molecular dynamics simulations

Satoru G. Itoh[a]; Hisashi Okumura[a]; Yuko Okamoto[a]
[a] Department of Physics, School of Science, Nagoya University, Nagoya, Aichi, Japan

To cite this Article Itoh, Satoru G. , Okumura, Hisashi and Okamoto, Yuko(2007) 'Generalized-ensemble algorithms for molecular dynamics simulations', Molecular Simulation, 33: 1, 47 — 56
To link to this Article: DOI: 10.1080/08927020601096812
URL: http://dx.doi.org/10.1080/08927020601096812

Molecular Simulation, Vol. 33, Nos. 1–2, 15 January–15 February 2007, 47–56

Generalized-ensemble algorithms for molecular dynamics simulations

SATORU G. ITOH†, HISASHI OKUMURA‡ and YUKO OKAMOTO*

Department of Physics, School of Science, Nagoya University, Nagoya, Aichi 464 8602, Japan

(Received October 2006; in final form November 2006)

In complex systems with many degrees of freedom such as biomolecular systems, conventional Monte Carlo and molecular dynamics simulations in canonical ensemble or isobaric–isothermal ensemble suffer from the multiple-minima problem, resulting in entrapment in states of energy local minima. A simulation in generalized ensemble performs a random walk in specified variables and overcomes this difficulty. In this article we review the generalized-ensemble algorithms. Multicanonical algorithm is described first. In this method, a random walk in potential energy space is realized and the simulation can avoid the multiple-minima problem. We then present two new generalized-ensemble algorithms, namely multioverlap algorithm and multibaric–multithermal algorithm, which are multi-variable/multi-dimensional extensions of the multicanonical algorithm. In the former method, a random walk in overlap space is realized, and in the latter that in both potential energy space and volume space is obtained. Emphasis is laid in the description of the molecular dynamics versions of these algorithms.

Keywords: Molecular dynamics simulation; Generalized-ensemble algorithms; Multicanonical algorithm; Multioverlap algorithm; Multibaric–multithermal algorithm

1. Introduction

Since the pioneering work of Nosé about two decades ago [1,2], molecular dynamics (MD) simulations that yield realistic statistical mechanical ensembles have been widely used. Examples of realistic physical ensembles are the *canonical ensemble* where the number of particles, volume, and temperature of the systems are constant (NVT ensemble) and the *isobaric–isothermal ensemble* where the number of particles, pressure, and temperature are constant (NPT ensemble). While by far the most popular Monte Carlo (MC) method that reproduces the canonical ensemble is the Metropolis algorithm [3], there are several widely employed MD methods for this ensemble [1,2,4–8]. Likewise, there are a MC method [9] and several MD methods [1,2,4,10–13] that reproduce the isobaric–isothermal ensemble. These simulation algorithms have been extensively applied to many fields of computational science.

In complex systems such as spin systems and protein systems, however, conventional simulations at low temperatures in the canonical ensemble and those at low temperatures or high pressures in the isobaric–isothermal ensemble tend to get trapped in states of energy local minima, which are separated by high energy barriers. In order to overcome this difficulty, simulations in artificial, generalized ensembles are commonly employed. This class of simulation methods are often referred to as the *generalized-ensemble algorithms* (for recent reviews, see, e.g. Refs. [14,15]). In a generalized-ensemble simulation, each state is weighted by a non-Boltzmann probability weight factor so that a random walk in potential energy space may be realized. The random walk allows the simulation to escape from any energy barrier and to sample much wider configurational space than by conventional methods.

One of the most well-known generalized-ensemble methods is perhaps *multicanonical algorithm* (MUCA) [16,17] (for a textbook see, e.g. Ref. [18]). The probability weight factor, which is referred to as the multicanonical weight factor, is defined to be inversely proportional to the density of states so that a flat

*Corresponding author. Email: okamoto@phys.nagoya-u.ac.jp
†Email: itoh@tb.phys.nagoya-u.ac.jp
‡Email: hokumura@tb.phys.nagoya-u.ac.jp

Molecular Simulation
ISSN 0892-7022 print/ISSN 1029-0435 online © 2007 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/08927020601096812

distribution in potential energy may be obtained. The uniform distribution induces a free random walk in the potential energy space, and the multiple-minima problem is overcome. MUCA was first applied to spin systems (see, e.g. Refs. [19–24]). MUCA was also introduced to the molecular simulation field [25]. Molecular dynamics version of MUCA was then developed [26–28]. MUCA and its generalizations have been extensively used in many applications in protein and related systems [14,15].

Soon after its development, the idea of MUCA was extended so that a flat distribution in magnetization instead of potential energy may be obtained (hence, it was referred to as multimagnetic algorithm) [20]. MUCA has further been extended so that flat distributions in other parameters and/or multidimensional parameter space may be realized [21,29–33].

In this article, we first describe both MC and MD versions of MUCA. We then present two of newly-developed generalized-ensemble algorithms that are multidimensional/multicomponent extensions of MUCA. Emphasis will be laid in the description of the MD versions of these algorithms. One is the *multioverlap algorithm* (MUOV), which performs a random walk in the overlap space instead of the potential energy space (see Ref. [34] for the MC version and Refs. [35,36] for the MD version). A similar method was previously used in spin glass simulations [22,23]. The difference is: While in Refs. [22,23] the weighting is done for the self-overlap of two replicas of the system, in the present method we deal with the overlap with a predefined configuration. The other method that we present here is the *multibaric–multithermal algorithm* (MUBATH), which realizes a random walk both in the potential energy space and in the volume space (see Refs. [37–39] for the MC version and Refs. [40,41] for the MD version),

In Section 2 we give the methodological details of the three generalized-ensemble algorithms, namely, MUCA, MUOV, and MUBATH. In Section 3 we present the results of applications of MUOV MD and MUBATH MD simulations. Section 4 is devoted to conclusions.

2. Generalized-ensemble algorithms

2.1 Multicanonical algorithm

Let us consider a system of N atoms of mass m_k ($k = 1, \ldots, N$) with their coordinate vectors and momentum vectors denoted by $q \equiv \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N\}$ and $p \equiv \{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N\}$, respectively. The Hamiltonian $\mathcal{H}(q,p)$ of the system is the sum of the kinetic energy $K(p)$ and the potential energy $E(q)$:

$$\mathcal{H}(q,p) = K(p) + E(q), \tag{1}$$

where

$$K(p) = \sum_{k=1}^{N} \frac{\boldsymbol{p}_k^2}{2m_k}. \tag{2}$$

In the canonical ensemble at temperature T_0 each state $x \equiv (q,p)$ with the Hamiltonian $\mathcal{H}(q,p)$ is weighted by the Boltzmann factor:

$$W_{\mathrm{B}}(x; T_0) = \exp\left(-\beta_0 \mathcal{H}(q,p)\right), \tag{3}$$

where the inverse temperature β_0 is defined by $\beta_0 = 1/k_{\mathrm{B}}T_0$ (k_{B} is the Boltzmann constant) and after the semicolon (;) we have written explicitly some of the fixed parameters in this ensemble (temperature here). The average kinetic energy at temperature T_0 is given by

$$\langle K(p)\rangle_{T_0} = \left\langle \sum_{k=1}^{N} \frac{\boldsymbol{p}_k^2}{2m_k} \right\rangle_{T_0} = \frac{3}{2} N k_{\mathrm{B}} T_0. \tag{4}$$

Because the coordinates q and momenta p are decoupled in equation (1), we can suppress the kinetic energy part and can write the Boltzmann factor as

$$W_{\mathrm{B}}(x; T_0) = W_{\mathrm{B}}(E; T_0) = \exp\left(-\beta_0 E\right). \tag{5}$$

The canonical probability distribution of potential energy $P_{\mathrm{B}}(E; T_0)$ at temperature T_0 is then given by the product of the density of states $n(E)$ and the Boltzmann weight factor $W_{\mathrm{B}}(E; T_0)$:

$$P_{\mathrm{B}}(E; T_0) \propto n(E) W_{\mathrm{B}}(E; T_0). \tag{6}$$

Because $n(E)$ is a rapidly increasing function and $W_{\mathrm{B}}(E; T_0)$ decreases exponentially, the canonical ensemble yields a bell-shaped potential energy distribution which has a maximum around the average energy at temperature T_0. The conventional MC or MD simulations at constant temperature are expected to yield $P_{\mathrm{B}}(E; T_0)$. A MC simulation based on the Metropolis algorithm [3] is performed with the following transition probability from a state x of potential energy E to a state x' of potential energy E':

$$w(x \to x') = \min\left(1, \frac{W_{\mathrm{B}}(E'; T_0)}{W_{\mathrm{B}}(E; T_0)}\right) = \min\left(1, \exp\left[-\beta_0 (E' - E)\right]\right). \tag{7}$$

A MD simulation, on the other hand, is based on the following Newton equations of motion:

$$\dot{\boldsymbol{q}}_k = \frac{\boldsymbol{p}_k}{m_k}, \tag{8}$$

$$\dot{\boldsymbol{p}}_k = -\frac{\partial E}{\partial \boldsymbol{q}_k} = \boldsymbol{F}_k, \tag{9}$$

where F_k is the force acting on the kth atom ($k = 1, \ldots, N$). This set of equations actually yield the microcanonical ensemble, and we have to add a thermostat in order to obtain the canonical ensemble at temperature T_0. Here, we just follow Nosé's prescription

[1,2], and we have

$$\dot{\boldsymbol{q}}_k = \frac{\boldsymbol{p}_k}{m_k}, \tag{10}$$

$$\dot{\boldsymbol{p}}_k = -\frac{\partial E}{\partial \boldsymbol{q}_k} - \frac{\dot{s}}{s}\boldsymbol{p}_k = \boldsymbol{F}_k - \frac{\dot{s}}{s}\boldsymbol{p}_k, \tag{11}$$

$$\dot{s} = s\frac{p_s}{Q}, \tag{12}$$

$$\dot{P}_s = \sum_{i=1}^{N} \frac{\boldsymbol{p}_i^2}{m_i} - 3Nk_{\mathrm{B}}T_0 = 3Nk_{\mathrm{B}}(T(t) - T_0), \tag{13}$$

where s is Nosé's scaling parameter, Q is its mass, P_s is its conjugate momentum, and the "instantaneous temperature" $T(t)$ is defined by

$$T(t) = \frac{1}{3Nk_{\mathrm{B}}} \sum_{i=1}^{N} \frac{p_i(t)^2}{m_i}. \tag{14}$$

Given the prescriptions for MC or MD simulations, it is still very difficult to obtain accurate canonical distributions of complex systems at low temperatures. This is because simulations at low temperatures tend to get trapped in one or a few of local-minimum-energy states.

In the *multicanonical algorithm* (MUCA) [16,17], on the other hand, each state is weighted by a non-Boltzmann weight factor $W_{\mathrm{muca}}(E)$ (which we refer to as the *multicanonical weight factor*) so that a uniform potential energy distribution $P_{\mathrm{muca}}(E)$ is obtained:

$$P_{\mathrm{muca}}(E) \propto n(E)W_{\mathrm{muca}}(E) \equiv \text{constant}. \tag{15}$$

The flat distribution implies that a free random walk in the potential energy space is realized in this ensemble. This allows the simulation to escape from any local minimum-energy states and to sample the configurational space much more widely than the conventional canonical MC or MD methods.

The definition in equation (15) implies that the multicanonical weight factor is inversely proportional to the density of states, and we can write it as follows:

$$W_{\mathrm{muca}}(E) \equiv \exp\left[-\beta_0 E_{\mathrm{muca}}(E; T_0)\right] = \frac{1}{n(E)}, \tag{16}$$

where we have chosen an arbitrary reference temperature, $T_0 = 1/k_{\mathrm{B}}\beta_0$, and the "*multicanonical potential energy*" is defined by

$$E_{\mathrm{muca}}(E; T_0) \equiv k_{\mathrm{B}}T_0 \ln n(E) = T_0 S(E) \ . \tag{17}$$

Here, $S(E)$ is the entropy in the microcanonical ensemble.

A multicanonical MC simulation is performed, for instance, with the usual Metropolis criterion [3]: the transition probability of state x with potential energy E to state x' with potential energy E' is given by

$$\begin{aligned} w(x \to x') &= \min\left(1, \frac{W_{\mathrm{muca}}(E')}{W_{\mathrm{muca}}(E)}\right) \\ &= \min\left(1, \frac{n(E)}{n(E')}\right) \\ &= \min\left(1, \exp\left(-\beta_0 \Delta E_{\mathrm{muca}}\right)\right), \end{aligned} \tag{18}$$

where

$$\Delta E_{\mathrm{muca}} = E_{\mathrm{muca}}(E'; T_0) - E_{\mathrm{muca}}(E; T_0). \tag{19}$$

The MD algorithm in the multicanonical ensemble also naturally follows from equation (16), in which the regular constant temperature MD simulation (with $T = T_0$) is performed by replacing E by E_{muca} in equation (11) [26,27]:

$$\begin{aligned} \dot{\boldsymbol{p}}_k &= -\frac{\partial E_{\mathrm{muca}}(E; T_0)}{\partial \boldsymbol{q}_k} - \frac{\dot{s}}{s}\boldsymbol{p}_k \\ &= \frac{\partial E_{\mathrm{muca}}(E; T_0)}{\partial E}\boldsymbol{F}_k - \frac{\dot{s}}{s}\boldsymbol{p}_k. \end{aligned} \tag{20}$$

In general, the multicanonical weight factor $W_{\mathrm{muca}}(E)$, or the density of states $n(E)$, is not *a priori* known, and one needs its estimator for a numerical simulation. This estimator is usually obtained from iterations of short trial simulations. The details of this process are described, for instance, in Refs. [14,15,18].

After an optimal multicanonical weight factor is obtained, we perform with this weight factor a multicanonical simulation with high statistics (production run). Let $N_{\mathrm{muca}}(E)$ be the histogram of potential energy distribution $P_{\mathrm{muca}}(E)$ obtained by this production run. The best estimate of the density of states can then be given by the reweighting techniques [42] as follows (see the proportionality relation in equation (15)):

$$n(E) = \frac{N_{\mathrm{muca}}(E)}{W_{\mathrm{muca}}(E)}. \tag{21}$$

By substituting this quantity into the following equation, one can calculate the ensemble averages of a physical quantity A at any temperature $T(= 1/k_{\mathrm{B}}\beta)$:

$$\langle A \rangle_T = \frac{\sum_E A(E)P_{\mathrm{B}}(E; T)}{\sum_E P_{\mathrm{B}}(E; T)} = \frac{\sum_E A(E)n(E)\exp\left(-\beta E\right)}{\sum_E n(E)\exp\left(-\beta E\right)}. \tag{22}$$

2.2 Extensions of multicanonical algorithm

While MUCA yields a flat distribution in potential energy and performs a random walk in potential energy space, we can, in principle, choose any other variable and induce a random walk in that variable. One such example is the *multioverlap algorithm* (MUOV) [34–36]. Here, we choose a protein system and define the overlap in the space

of dihedral angles by [43]

$$O = 1 - d, \tag{23}$$

where d is the dihedral-angle distance given by

$$d = \frac{1}{n\pi}\sum_i d_a\left(\theta_i, \theta_i^0\right). \tag{24}$$

θ_i is the dihedral angle i, and θ_i^0 is the dihedral angle i of the reference configuration. The distance $d_a\left(\theta_i, \theta_i^0\right)$ between two dihedral angles is defined by

$$d_a\left(\theta_i, \theta_i^0\right) = \min\left(|\theta_i - \theta_i^0|, 2\pi - |\theta_i - \theta_i^0|\right). \tag{25}$$

The dihedral-angle distance d in equation (24) takes a value in the range $0 \le d \le 1$. If $d = 0$, all dihedral angles are coincident with those of the reference configuration. The dihedral-angle distance is thus an indicator of how similar the conformation is to the reference conformation. As one can see in equation (23), the dihedral-angle distance d is equivalent to the overlap O. We will deal with the dihedral-angle distance instead of the overlap hereafter.

In the multioverlap ensemble at a constant temperature T_0, the probability distribution is given by the following non-Boltzmann weight factor, which we refer to as the multioverlap weight factor:

$$W_{\mathrm{muov}}(d, E; T_0) = e^{-\beta_0 E_{\mathrm{muov}}}, \tag{26}$$

where E_{muov} is the "multioverlap potential energy" defined by

$$E_{\mathrm{muov}}(d, E; T_0) = E - k_{\mathrm{B}} T_0 f(d; T_0). \tag{27}$$

The function $f(d; T_0)$ is the dimensionless free energy at dihedral-angle distance d.

The generalization to the multi-dimensional dihedral-angle distance space is straightforward, and the multioverlap weight factor is given by

$$\begin{aligned} W_{\mathrm{muov}}(d_1, \ldots, d_L, E; T_0) &= e^{-\beta_0 E_{\mathrm{muov}}} \\ &\equiv e^{-\beta_0 E + f(d_1, \ldots, d_L; T_0)}, \end{aligned} \tag{28}$$

where L is the number of the reference configurations and d_i is the dihedral-angle distance with respect to reference configuration $i\,(i = 1, \ldots, L)$. The function $f(d_1, \ldots, d_L; T_0)$ is the dimensionless free energy with the fixed value of dihedral-angle distances $d_1, \ldots, d_L$. The dimensionless free energy $f(d_1, \ldots, d_L; T_0)$ is defined so that the probability distribution of dihedral-angle distances $P_{\mathrm{muov}}(d_1, \ldots, d_L; T_0)$ is flat:

$$\begin{aligned} P_{\mathrm{muov}}(d_1, \ldots, d_L; T_0) &= \int \mathrm{d}E\ P_{\mathrm{muov}}(d_1, \ldots, d_L, E; T_0) \\ &\propto \int \mathrm{d}E\ n(d_1, \ldots, d_L, E) W_{\mathrm{muov}}(d_1, \ldots, d_L, E; T_0) \\ &= \int \mathrm{d}E\ n(d_1, \ldots, d_L, E) e^{-\beta_0 E + f(d_1, \ldots, d_L; T_0)} \\ &\equiv \text{constant}, \end{aligned} \tag{29}$$

where $P_{\mathrm{muov}}(d_1, \ldots, d_L, E; T_0)$ is the probability distribution of potential energy and dihedral-angle distances, and $n(d_1, \ldots, d_L, E)$ is its density of states.

The MD algorithm in the multioverlap ensemble also naturally follows from equation (28), in which the regular constant temperature MD simulation (with $T = T_0$) is performed by replacing E by E_{muov} in equation (11) [35,36]:

$$\begin{aligned} \dot{\boldsymbol{p}}_k &= -\frac{\partial E_{\mathrm{muov}}}{\partial \boldsymbol{q}_k}(d_1, \ldots, d_L, E; T_0) - \frac{\dot{s}}{s}\boldsymbol{p}_k \\ &= \boldsymbol{F}_k + k_{\mathrm{B}} T_0 \frac{\partial f}{\partial \boldsymbol{q}_k}(d_1, \ldots, d_L; T_0) - \frac{\dot{s}}{s}\boldsymbol{p}_k. \end{aligned} \tag{30}$$

The multioverlap weight factor, or the dimensionless free energy, is not *a priori* known and has to be determined by the usual iterations of short simulations [14,15,18] (see also Refs. [34–36]). Suppose that we have determined an appropriate dimensionless free energy $f(d_1, \ldots, d_L; T_0)$ at temperature T_0 and that we have made a production run at this temperature. The results of the multioverlap production run can then be analyzed by the reweighting techniques [42]. Namely, the expectation value of a physical quantity A at any temperature T is given by

$$\begin{aligned} \langle A \rangle_T &= \frac{\sum_{d_1, \cdots, d_L, E} A(d_1, \cdots, d_L, E) N_{\mathrm{muov}}(d_1, \cdots, d_L, E) W_{\mathrm{muov}}(d_1, \cdots, d_L, E; T_0)^{-1} e^{-\beta E}}{\sum_{d_1, \cdots, d_L, E} N_{\mathrm{muov}}(d_1, \cdots, d_L, E) W_{\mathrm{muov}}(d_1, \cdots, d_L, E; T_0)^{-1} e^{-\beta E}} \\ &= \frac{\sum_{d_1, \ldots, d_L, E} A(d_1, \ldots, d_L, E) N_{\mathrm{muov}}(d_1, \ldots, d_L, E) e^{-(\beta - \beta_0)E - f(d_1, \ldots, d_L; T_0)}}{\sum_{d_1, \ldots, d_L, E} N_{\mathrm{muov}}(d_1, \ldots, d_L, E) e^{-(\beta - \beta_0)E - f(d_1, \ldots, d_L; T_0)}}, \end{aligned} \tag{31}$$

where $N_{\mathrm{muov}}(d_1, \ldots, d_L, E)$ is the histogram of the probability distribution $P_{\mathrm{muov}}(d_1, \ldots, d_L, E; T_0)$ of potential energy and dihedral-angle distances that was obtained by the multioverlap production run.

The multioverlap algorithm can further be combined with the multicanonical algorithm as follows (this method is referred to as the *multicanonical–multioverlap algorithm*) [44]. In analogy with the multicanonical ensemble in equation (15) or the multioverlap ensemble in equation (29), by employing the non-Boltzmann weight factor $W_{\mathrm{mcmo}}(d_1, \ldots, d_L, E)$, which we refer to as the multicanonical–multioverlap weight factor, a uniform probability distribution with respect to the potential

energy and dihedral-angle distances is obtained:

$$P_{\mathrm{mcmo}}(d_1, \ldots, d_L, E) \propto n(d_1, \ldots, d_L, E) W_{\mathrm{mcmo}}(d_1, \ldots, d_L, E) \equiv \text{constant}. \quad (32)$$

In this method we obtain a random walk not only in the potential energy space but also in the dihedral-angle distance space.

Besides the canonical ensemble, molecular simulations in the isobaric–isothermal ensemble [1,2,4,9–13] are also commonly used. This is because most experiments are carried out under the constant pressure and constant temperature conditions. The canonical probability distribution $P_{\mathrm{B}}(E; T_0)$ in equation (6) is here replaced by the isobaric–isothermal distribution $P_{\mathrm{NPT}}(E, V; T_0, P_0)$ for potential energy E and volume V:

$$P_{\mathrm{NPT}}(E, V; T_0, P_0) \equiv n(E, V) e^{-\beta_0 H}. \quad (33)$$

Here, the density of states $n(E, V)$ is given as a function of both E and V, and H is the "enthalpy" (without the kinetic energy contributions):

$$H = E + P_0 V, \quad (34)$$

where P_0 is the pressure at which simulations are performed. This weight factor produces an isobaric–isothermal ensemble at constant temperature (T_0) and constant pressure (P_0). This ensemble has bell-shaped distributions in both E and V.

As for the MD methods in this ensemble, we just present the Nosé–Andersen algorithm [1,2,4]. The equations of motion in equations (10)–(13) are now generalized as follows:

$$\dot{\boldsymbol{q}}_k = \frac{\boldsymbol{p}_k}{m_k} + \frac{\dot{V}}{3V}\boldsymbol{q}_k, \quad (35)$$

$$\dot{\boldsymbol{p}}_k = -\frac{\partial H}{\partial \boldsymbol{q}_k} - \left(\frac{\dot{s}}{s} + \frac{\dot{V}}{3V}\right)\boldsymbol{p}_k = \boldsymbol{F}_k - \left(\frac{\dot{s}}{s} + \frac{\dot{V}}{3V}\right)\boldsymbol{p}_k, \quad (36)$$

$$\dot{s} = s\frac{P_s}{Q}, \quad (37)$$

$$P_s = \sum_{i=1}^{N} \frac{\boldsymbol{p}_i^2}{m_i} - 3Nk_{\mathrm{B}}T_0 = 3Nk_{\mathrm{B}}(T(t) - T_0), \quad (38)$$

$$\dot{V} = s\frac{P_V}{M}, \quad (39)$$

$$\dot{P}_V = \frac{1}{3V}\left(\sum_{i=1}^{N} \frac{\boldsymbol{p}_i^2}{m_i} - \sum_{i=1}^{N} \boldsymbol{q}_i \cdot \frac{\partial H}{\partial \boldsymbol{q}_i}\right) - \frac{\partial H}{\partial V} = P(t) - P_0, \quad (40)$$

where M is the artificial mass associated with the volume, P_V is the conjugate momentum for the volume, and the "instantaneous pressure" $P(t)$ is defined by

$$P(t) = \frac{1}{3V}\left(\sum_{i=1}^{N} \frac{\boldsymbol{p}_i(t)^2}{m_i} - \sum_{i=1}^{N} \boldsymbol{q}_i(t) \cdot \frac{\partial H}{\partial \boldsymbol{q}_i}(t)\right) = \frac{1}{3V}\left(\sum_{i=1}^{N} \frac{\boldsymbol{p}_i(t)^2}{m_i} + \sum_{i=1}^{N} \boldsymbol{q}_i(t) \cdot \boldsymbol{F}_i(t)\right). \quad (41)$$

We now introduce the idea of the multicanonical technique into the isobaric–isothermal ensemble method and refer to this generalized-ensemble algorithm as the *multibaric–multithermal algorithm* (MUBATH) [37–41]. The molecular simulations in this generalized ensemble perform random walks both in the potential energy space and in the volume space.

In the multibaric–multithermal ensemble, each state is sampled by the multibaric–multithermal weight factor $W_{\mathrm{mbt}}(E, V) \equiv \exp\{-\beta_0 H_{\mathrm{mbt}}(E, V)\}$ (H_{mbt} is referred to as the multibaric–multithermal enthalpy) so that a uniform distribution in both potential energy and volume is obtained [37]:

$$P_{\mathrm{mbt}}(E, V) \propto n(E, V) W_{\mathrm{mbt}}(E, V) = n(E, V)\exp\{-\beta_0 H_{\mathrm{mbt}}(E, V)\} \equiv \text{constant}. \quad (42)$$

In order to perform the multibaric–multithermal MD simulation, we just solve the above equations of motion (equations (35)–(40)) for the regular isobaric–isothermal ensemble (with $T = T_0$ and $P = P_0$), where the enthalpy H is replaced by the multibaric–multithermal enthalpy H_{mbt} in equations (36) and (40) [40].

The multibaric–multithermal weight factor is, however, not *a priori* known and has to be determined by the usual iterations of short simulations [14,15,18] (see also Refs. [37–41]). After an optimal weight factor $W_{\mathrm{mbt}}(E, V)$ is obtained, a long production simulation is performed for data collection. We employ the reweighting techniques [42] for the results of the production run to calculate the isobaric–isothermal-ensemble averages. The probability distribution $P_{\mathrm{NPT}}(E, V; T, P)$ of potential energy and volume in the isobaric–isothermal ensemble at the desired temperature T and pressure P is given by

$$P_{\mathrm{NPT}}(E, V; T, P) = \frac{N_{\mathrm{mbt}}(E, V) W_{\mathrm{mbt}}(E, V)^{-1} \mathrm{e}^{-\beta(E+PV)}}{\sum_{E,V} N_{\mathrm{mbt}}(E, V) W_{\mathrm{mbt}}(E, V)^{-1} \mathrm{e}^{-\beta(E+PV)}}, \quad (43)$$

where $N_{\mathrm{mbt}}(E, V)$ is the histogram of the probability distribution $P_{\mathrm{mbt}}(E, V)$ of potential energy and volume that was obtained by the multibaric–multithermal production run. The expectation value of a physical quantity A at T and P is then obtained from

$$\langle A \rangle_{T,P} = \sum_{E,V} A(E, V) P_{\mathrm{NPT}}(E, V; T, P). \quad (44)$$

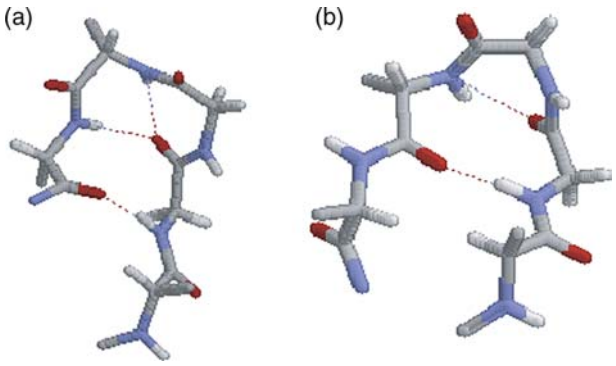

Figure 1. (a) Reference configuration 1 and (b) reference configuration 2. The side chains are suppressed and only backbone structures are shown. The dotted lines denote the hydrogen bonds. The N-terminus and the C-terminus are on the right-hand side and on the left-hand side, respectively. The figures were created with RasMol [47].

3. Results

We now present the results of our simulations based on the algorithms described in the previous section. The first example is a multioverlap MD simulation of the system of a penta peptide, Met-enkephalin, in vacuum [35]. The amino-acid sequence is Tyr–Gly–Gly–Phe–Met. The N-terminus and the C-terminus were blocked with the acetyl group and the *N*-methyl group, respectively. The force field that we adopted is the CHARMM param 22 parameter set [45]. Our multioverlap MD simulations were performed by implementing the method in the CHARMM macromolecular mechanics program [46].

We considered two energy local-minimum states of Met-enkephalin as reference configurations. In figure 1 we show these two reference configurations. We then set $L = 2$ in equation (28) and the dimensionless free energy is expressed as $f(d_1, d_2; T_0)$. The multioverlap MD simulation was carried out at $T_0 = 300\,\mathrm{K}$ with a time step of 0.5 fs.

Figure 2 shows the time series of the dihedral-angle distances with respect to each of the two reference configurations. While figure 2(a),(b) are the results of the conventional canonical MD simulation at $T_0 = 300\,\mathrm{K}$, figure 2(c),(d) are the results of the multioverlap MD simulation at the same temperature. When $d_1 = 0$, the values of dihedral angles of backbone completely coincide with those of reference configuration 1 and $d_2 = 0.122$. Conversely, when $d_2 = 0$, $d_1 = 0.122$. When d_1 (d_2) is near zero, the conformation is similar to reference

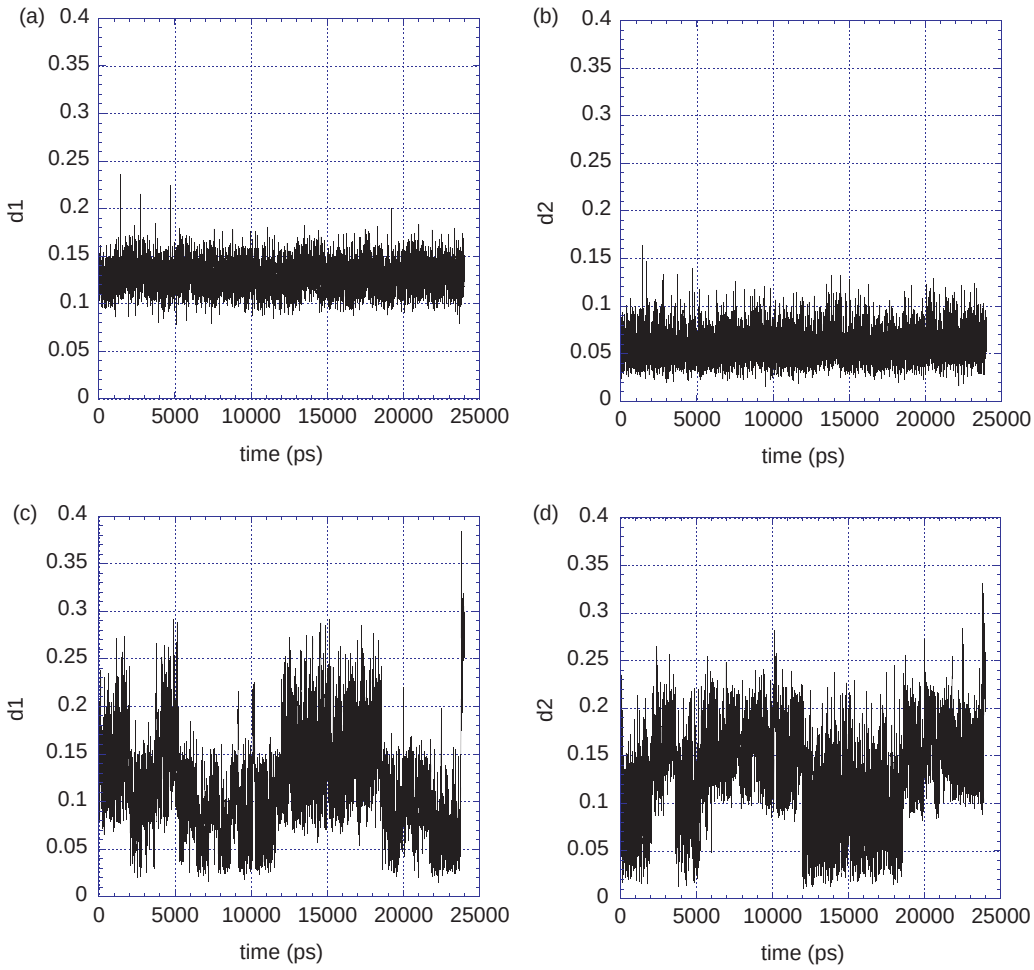

Figure 2. The time series of the dihedral-angle distances d_1 and d_2. (a) and (b) are from the conventional canonical MD simulation, and (c) and (d) are from the multioverlap MD simulation at $T_0 = 300\,\mathrm{K}$.

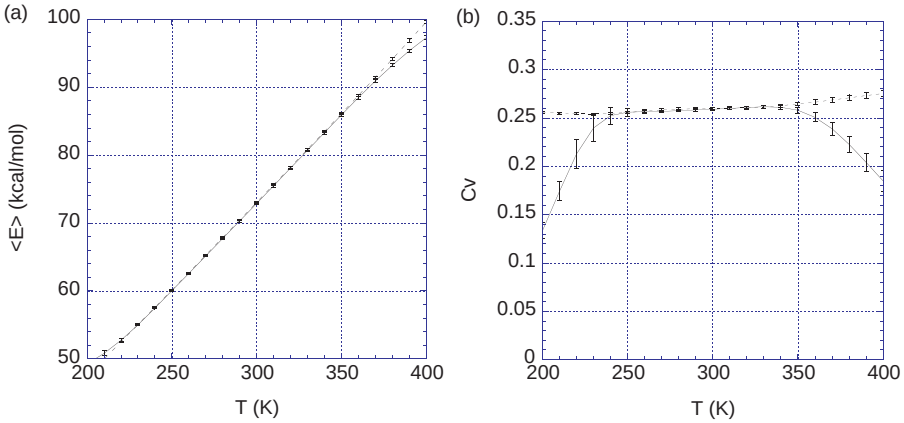

Figure 3. (a) Average potential energy and (b) specific heat calculated from the multioverlap MD simulation (solid line) and the multicanonical MD simulation (dashed line) by the reweighting techniques.

conformation 1 (2). Therefore, figure 2 implies that the multioverlap MD simulation performed a random walk in the dihedral-angle distance space between reference configurations 1 and 2, whereas the usual canonical MD simulation got trapped in a local-minimum state near configuration 2.

In figure 3 the average potential energy and specific heat as functions of temperature are shown. They were calculated from the results of the multioverlap MD simulation by the reweighting techniques in equation (31). The results from a multicanonical MD simulation are also shown as a reference in the figure, because the multicanonical algorithm is well-known for giving accurate expectation values for a wide range of temperature [18]. These results from the multioverlap MD simulation well coincide with those from the multicanonical MD simulation between 250 and 350 K. In the region under 250 K and above 350 K, however, we see deviations between the results of the two simulations. The reason is that the multioverlap algorithm samples conformations in the dihedral-angle distance space but not in the potential energy space. In other words, the multioverlap simulation does not necessarily give an accurate estimate of the density of states in wide energy range. Thus, the expectation value calculated by the reweighting techniques in equation (31) is correct only in the neighborhood of temperature T_0 at which simulations were performed. This problem is now solved by introducing the multicanonical–multioverlap algorithm defined in equation (32) [44].

The free energy $F(d_1, d_2; T)$ (or, the potential of mean force) at temperture T is defined by

$$F(d_1, d_2; T) = -k_B T \ \ln P_B(d_1, d_2; T), \qquad (45)$$

where $P_B(d_1, d_2; T)$ is the reweighted canonical probability distribution of d_1 and d_2 at T and given by (see equation (31))

$$P_B(d_1, d_2; T) = \frac{\sum_E N_{muov}(d_1, d_2, E) e^{-(\beta-\beta_0)E - f(d_1,d_2;T_0)}}{\sum_{d_1,d_2,E} N_{muov}(d_1, d_2, E) e^{-(\beta-\beta_0)E - f(d_1,d_2;T_0)}}. \qquad (46)$$

In figure 4 we illustrate the free-energy landscapes with respect to the dihedral-angle distances that were calculated from the results of the conventional canonical

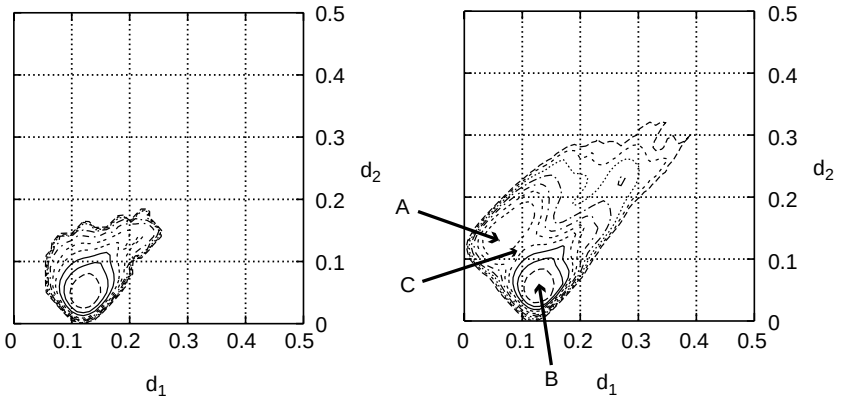

Figure 4. The free-energy landscape obtained from (a) the conventional canonical MD simulation and (b) the multioverlap MD simulation at $T_0 = 300$ K. Contour lines are drawn every 1 kcal/mol. The labels A and B locate the local-minimum states. The label C stands for the saddle point, or the transition state, between these two local-minimum states.

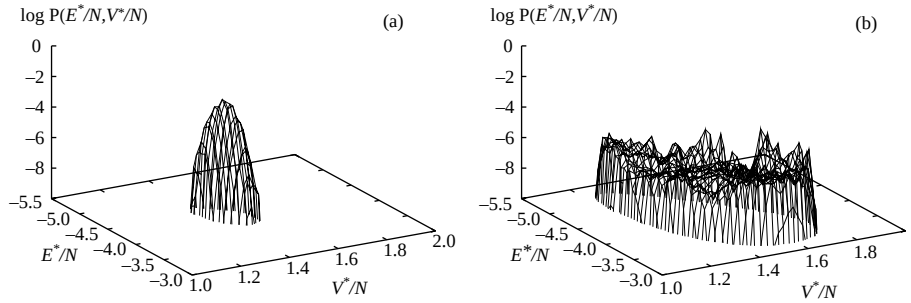

Figure 5. (a) The probability distribution $P_{\mathrm{NPT}}(E^*/N, V^*/N)$ in the isobaric–isothermal MD simulation at $(T_0^*, P_0^*) = (2.0, 3.0)$ and (b) the probability distribution $P_{\mathrm{mbt}}(E^*/N, V^*/N)$ in the multibaric–multithermal MD simulation.

MD simulation and those of the multioverlap MD simulation. While in figure 4(a) only one local-minimum state exists near reference configuration 2, in figure 4(b) we find a local-minimum state A and a local-minimum state B near reference configurations 1 and 2, respectively. This result again implies that the canonical MD simulation got trapped in the latter local-minimum state. The local-minimum state B near reference configuration 2 corresponds to the global-minimum state at 300 K. The local-minimum state A near reference configuration 1 is another local-minimum state at 300 K. The free-energy difference between the global-minimum state (B) and the local-minimum state (A) is about 3 kcal/mol.

The saddle point C in figure 4(b) corresponds to the transition state between the global-minimum state (B) and the local-minimum state (A). The free-energy difference between B and C is about 5 kcal/mol and that between A and C is 2 kcal/mol. Because $k_{\mathrm{B}}T \approx 0.6\,\mathrm{kcal/mol}$ at $T = 300\,\mathrm{K}$, these barrier heights are rather high. This is why the conventional canonical MD simulation got trapped in the vicinity of the global-minimum state B.

We now present the results of a multibaric–multithermal MD simulation [40]. We considered a Lennard–Jones 12–6 potential system. The length and the energy are scaled in units of the Lennard–Jones diameter σ and the depth of the potential ϵ, respectively. We use an asterisk (*) for quantities reduced by σ and ϵ.

We used 500 particles ($N = 500$) in a cubic unit cell with periodic boundary conditions. We started the multibaric–multithermal weight factor determination from a regular isobaric–isothermal simulation at $T_0^* = 2.0$ and $P_0^* = 3.0$ (the multibaric–multithermal production run was also performed at this set of temperature and pressure values). These temperature and pressure values are respectively higher than the critical temperature T_c^* and the critical pressure P_c^* [48,49]. Recent reliable data are $T_c^* = 1.3207(4)$ and $P_c^* = 0.1288(5)$ [49]. The cutoff radius r_c^* was taken to be $r_c^* = 4.0$. A cut-off correction was added for the pressure and the potential energy.

In order to carry out the multibaric–multithermal MD simulation in equations (35)–(40) with the replacement of H by H_{mbt}, we employed the Nosé–Poincaré formalism [50,51]. This gives the same equations of motion as the Nosé thermostat and provides a symplectic integrator. Therefore, it has an advantage that the secular deviation of the Hamiltonian is suppressed. We have recently shown that this integrator is also very effective for rigid-body molecules [52]. We performed a long production run of 10^6 MD steps.

In figure 5(a) we show the probability distribution $P_{\mathrm{NPT}}(E^*/N, V^*/N)$ from the isobaric–isothermal simulation that was carried out first. It is a bell-shaped distribution. As the iteration of the multibaric–multithermal weight factor determination proceeds, $P_{\mathrm{mbt}}(E^*/N, V^*/N)$ will become flat and broad gradually. Figure 5(b) depicts the probability distribution $P_{\mathrm{mbt}}(E^*/N, V^*/N)$ from the multibaric–multithermal

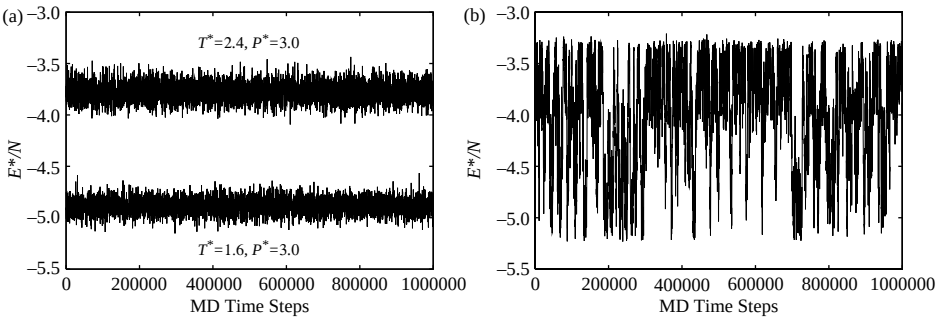

Figure 6. The time series of E^*/N from (a) the conventional isobaric–isothermal MD simulations at $(T_0^*, P_0^*) = (2.4, 3.0)$ and at $(T_0^*, P_0^*) = (1.6, 3.0)$ and (b) the multibaric–multithermal MD simulation.

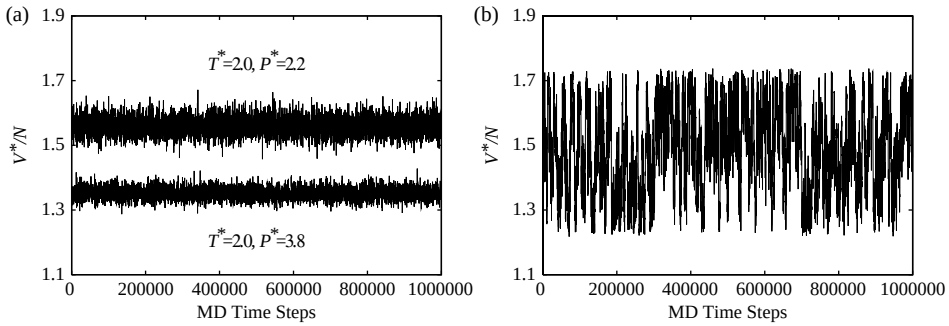

Figure 7. The time series of V^*/N from (a) the conventional isobaric–isothermal MD simulations at $(T_0^*, P_0^*) = (2.0, 2.2)$ and at $(T_0^*, P_0^*) = (2.0, 3.8)$ and (b) the multibaric–multithermal MD simulation.

simulation that was finally performed. It shows a flat distribution, and the multibaric–multithermal MD simulation indeed sampled the configurational space in wider ranges of E^*/N and V^*/N than the conventional isobaric–isothermal MD simulation.

The time series of E^*/N from two conventional isobaric–isothermal MD simulations at $(T_0^*, P_0^*) = (1.6, 3.0)$ and (2.4, 3.0) are given in figure 6(a). The potential energy fluctuates in narrow ranges of $E^*/N = -4.0 \sim -3.5$ at the higher temperature of $T_0^* = 2.4$ and in the ranges of $E^*/N = -5.1 \sim -4.7$ and at the lower temperature of $T_0^* = 1.6$. On the other hand, figure 6(b) shows that the multibaric–multithermal MD simulation realizes a random walk in the potential-energy space and covers a wide energy range.

A similar situation is observed in V^*/N. In figure 7(a) the time series of two conventional isobaric–isothermal MD simulations at $(T_0^*, P_0^*) = (2.0, 2.2)$ and (2.0, 3.8) are shown. The volume fluctuations are only in the range of $V^*/N = 1.3 \sim 1.4$ and $V^*/N = 1.5 \sim 1.6$ at $P_0^* = 3.8$ and at $P_0^* = 2.2$, respectively. On the other hand, the multibaric–multithermal MD simulation performs a random walk that covers even a wider volume range as shown in figure 7(b).

4. Conclusions

In this article we first described one of well-known generalized-ensemble algorithms, namely, multicanonical algorithm. We then introduced two new generalized-ensemble algorithms, which we refer to as multioverlap algorithm and multibaric–multithermal algorithm, as extensions of the multicanonical algorithm. The former is useful for studyiing the transition states among specified configurations, and the latter for investigating pressure induced transitions of protein and other molecular systems.

With these new methods available, we believe that we have working molecular dynamics simulation algorithms that can yield accurate physical ensembles such as the canonical and isobaric–isothermal ensembles.

Acknowledgements

We would like to thank our co-workers for useful discussions. In particular, we are grateful to Drs B. A. Berg of Florida State University and H. Noguchi of Forschungszentrum Juelich for collaborations that led to the multioverlap algorithm. This work was supported, in part, by the Grants-in-Aid for the Next Generation Super Computing Project, Nanoscience Program and for Scientific Research in Priority Areas, "Water and Biomolecules", from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

[1] S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, **52**, 255 (1984).
[2] S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, **81**, 511 (1984).
[3] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087 (1953).
[4] H.C. Andersen. Molecular dyanmics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, **72**, 2384 (1980).
[5] W.G. Hoover, A.J.C. Ladd, B. Moran. High-strain-rate plastic flow studied via nonequilibrium molecular dynamics. *Phys. Rev. Lett.*, **48**, 1818 (1982).
[6] D.J. Evans. Computer "experiment" for nonlinear thermodynamics of Couette flow. *J. Chem. Phys.*, **78**, 3297 (1983).
[7] W.G. Hoover. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A*, **31**, 1695 (1985).
[8] G.J. Martyna, M.L. Klein, M. Tuckerman. Nosé–Hoover chains: the canonical ensemble via continuous dynamics. *J. Chem. Phys.*, **97**, 2635 (1992).
[9] I.R. McDonald. NPT-ensemble Monte Carlo calculations for binary liquid mixtures. *Mol. Phys.*, **23**, 41 (1972).
[10] M. Parrinello, A. Rahman. Crystal structure and pair potentials: a molecular-dynamics study. *Phys. Rev. Lett.*, **45**, 1199 (1980).
[11] D.M. Heyes. Molecular dynamics at constant pressure and temperature. *Chem. Phys.*, **82**, 285 (1983).
[12] W.G. Hoover. Constant-pressure equation of motion. *Phys. Rev. A*, **34**, 2499 (1986).
[13] G.J. Martyna, D.J. Tobias, M.L. Klein. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.*, **101**, 4177 (1994).
[14] A. Mitsutake, Y. Sugita, Y. Okamoto. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers (Peptide Science)*, **60**, 96 (2001).
[15] Y. Okamoto. Generalized-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. *J. Mol. Graphics Mod.*, **22**, 425 (2004).

[16] B.A. Berg, T. Neuhaus. Multicanonical algorithms for first order phase transitions. *Phys. Lett.*, **B267**, 249 (1991).
[17] B.A. Berg, T. Neuhaus. Multicanonical ensemble: a new approach to simulate first-order phase transitions. *Phys. Rev. Lett.*, **68**, 9 (1992).
[18] B.A. Berg. *Markov Chain Monte Carlo Simulations and Their Statistical Analysis*, World Scientific, Singapore (2004).
[19] B.A. Berg, T. Celik. New approach to spin-glass simulations. *Phys. Rev. Lett.*, **69**, 2292 (1992).
[20] B.A. Berg, U.H.E. Hansmann, T. Neuhaus. Simulation of an ensemble with varying magnetic field: a nemerical determination of the first-order interface tension in the $D = 2$ Ising model. *Phys. Rev. B*, **47**, 497 (1993).
[21] W. Janke, S. Kappler. Multibondic cluster algorithm for Monte Carlo simulations of first-order phase transitions. *Phys. Rev. Lett.*, **74**, 212 (1995).
[22] B.A. Berg, W. Janke. Multioverlap simulations of the 3D Edwards–Anderson Ising spin glass. *Phys. Rev. Lett.*, **80**, 4771 (1998).
[23] B.A. Berg, A. Billoire, W. Janke. Spin-glass overlap barriers in three and four dimensions. *Phys. Rev. B*, **61**, 12143 (2000).
[24] N. Hatano, J.E. Gubernatis. A multicanonical Monte Carlo study of the 3D (J spin glass. *Prog. Theor. Phys. (Suppl.)*, **138**, 442 (2000).
[25] U.H.E. Hansmann, Y. Okamoto. Prediction of peptide conformation by multicanonical algorithms: new approach to the multiple-minima problem. *J. Comput. Chem.*, **14**, 1333 (1993).
[26] U.H.E. Hansmann, Y. Okamoto, F. Eisenmenger. Molecular dynamics Langevin and hybrid Monte Carlo simulations in a multicanonical ensemble. *Chem. Phys. Lett.*, **259**, 321 (1996).
[27] N. Nakajima, H. Nakamura, A. Kidera. Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides. *J. Phys. Chem. B*, **101**, 817 (1997).
[28] C. Bartels, M. Karplus. Probability distributions for complex systems: adaptive umbrella sampling of the potential energy. *J. Phys. Chem. B*, **102**, 865 (1998).
[29] J. Lee, M.A. Novotny, P.A. Rikvold. Method to study relaxation of metastable phases: macroscopic mean-field dynamics. *Phys. Rev. E*, **52**, 356 (1995).
[30] S. Kumar, P. Payne, M. Vásquez. Method for free-energy calculations using iterative techniques. *J. Comput. Chem.*, **17**, 1269 (1996).
[31] C. Bartels, M. Karplus. Multidimensional adaptive umbrella sampling: applications to main chain and side chain peptide conformations. *J. Comput. Chem.*, **18**, 1450 (1997).
[32] J. Higo, N. Nakajima, H. Shirai, A. Kidera, H. Nakamura. Two-component multicanonical Monte Carlo method for effective conformational sampling. *J. Comput. Chem.*, **18**, 2086 (1997).
[33] Y. Iba, G. Chikenji, M. Kikuchi. Simulation of lattice polymers with multi-self-overlap ensemble. *J. Phys. Soc. Jpn.*, **67**, 3327 (1998).
[34] B.A. Berg, H. Noguchi, Y. Okamoto. Multioverlap simulations for transitions between reference configurations. *Phys. Rev. E*, **68**, 036126 (2003).
[35] S.G. Itoh, Y. Okamoto. Multi-overlap molecular dynamics methods for biomolecular systems. *Chem. Phys. Lett.*, **400**, 308 (2004).
[36] S.G. Itoh, Y. Okamoto. Theoretical studies of transition states by the multioverlap molecular dynamics methods. *J. Chem. Phys.*, **124**, 104103 (2006).
[37] H. Okumura, Y. Okamoto. Monte Carlo simulations in multibaric–multithermal ensemble. *Chem. Phys. Lett.*, **383**, 391 (2004).
[38] H. Okumura, Y. Okamoto. Monte Carlo simulations in generalized isobaric–isothermal ensembles. *Phys. Rev. E*, **70**, 026702 (2004).
[39] H. Okumura, Y. Okamoto. Liquid–gas phase transitions studied by multibaric–multithermal Monte Carlo simulations. *J. Phys. Soc. Jpn.*, **73**, 3304 (2004).
[40] H. Okumura, Y. Okamoto. Molecular dynamics simulation in the multibaric–multithermal ensemble. *Chem. Phys. Lett.*, **391**, 248 (2004).
[41] H. Okumura, Y. Okamoto. Multibaric–multithermal ensemble molecular dynamics simulations. *J. Comput. Chem.*, **27**, 379 (2006).
[42] A.M. Ferrenberg, R.H. Swendsen. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.*, **61**, 2635 (1988); ibid. **63**, 1658 (1989).
[43] U.H.E. Hansmann, M. Masuya, Y. Okamoto. Characteristic temperatures of folding of a small peptide. *Proc. Natl. Acad. Sci. USA*, **97**, 10652 (1997).
[44] S.G. Itoh, Y. Okamoto. Effective sampling in the configurational space by the multicanonical–multioverlap algorithm, e-print: cond-mat/0610169, submitted for publication.
[45] A.D. MacKerell Jr, D. Bashford, M. Bellott, R.L. Dunbrack Jr, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher III, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586 (1998).
[46] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, **4**, 187 (1983).
[47] R.A. Sayle, E.J. Milner-White. RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.*, **20**, 374 (1995).
[48] H. Okumura, F. Yonezawa. Liquid–vapor coexistence curves of several interatomic model potentials. *J. Chem. Phys.*, **113**, 9162 (2000).
[49] H. Okumura, F. Yonezawa. Reliable determination of the liquid–vapor critical point by the NVT plus test particle method. *J. Phys. Soc. Jpn.*, **70**, 1990 (2001).
[50] S.D. Bond, B.J. Leimkuhler, B.B. Laird. The Nosé–Poincaré method for constant temperture molecular dynamics. *J. Comput. Phys.*, **151**, 114 (1999).
[51] S. Nosé. An improved symplectic integrator for Nosé–Poincaré thermostat. *J. Phys. Soc. Jpn.*, **70**, 75 (2001).
[52] H. Okumura, S.G. Itoh, Y. Okamoto. Explicit symplectic integrators of molecular dynamics algorithms for rigid-body molecules in the canonical, isothermal–isobaric, and related ensembles, e-print: cond-mat/0610382, submitted for publication.