Stochastic formulation of sampling dynamics in generalized ensemble methods

Jae Gil Kim,[1,*] Yoshifumi Fukunishi,[2] Akinori Kidera,[3] and Haruki Nakamura[4]
[1]*Japan Biological Information Research Center (JBIRC), Japan Biological Informatics Consortium, Aomi 2-41-6, Koto-ku, Tokyo 135-0064, Japan*
[2]*Biological Information Research Center (BIRC), National Institute of Advanced Industrial Science and Technology, Aomi 2-41-6, Koto-ku, Tokyo 135-0064, Japan*
[3]*Graduate School of Integrated Science, Yokohama City University, Yokohama 230-0045, Japan*
[4]*Laboratory of Protein Informatics, Research Center for Structural Biology Institute for Protein Research, Osaka University 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan*
(Received 11 July 2003; revised manuscript received 8 October 2003; published 10 February 2004)

The fundamental relations of the sampling process in the multicanonical ensemble and simulated tempering have been studied in the expanded ensembles formalism. The simulated tempering is identified as the multicanonical sampling with the generalized weight determined by a Laplace transform of the temperature weight. The characteristic dynamics of both sampling methods has been verified in the stochastic formulation of the sampling process. Our study gives a necessary and sufficient condition for the weights to realize a uniform sampling in the energy and temperature spaces. Based on the stochastic model, robust force biased iteration schemes have been proposed to allow automatic determinations of uniform sampling weights.

DOI: 10.1103/PhysRevE.69.021101 PACS number(s): 05.40.−a, 02.50.−r, 87.15.−v

I. INTRODUCTION

Potential energy surfaces of many important physical processes such as protein folding [1], cluster melting [2], and spin glasses [3] are dominated by a large number of local minima separated by high-energy barriers. Thus conventional Monte Carlo or molecular dynamics (MD) simulation will become trapped in one of local energy basins and fail to sample thermally accessible phase space. To overcome this quasiergodicity occurring in the simulation of rough energy landscapes, several sampling algorithms have been proposed during the past decade [4–10]. These include the parallel tempering or replica exchange method [4], umbrella sampling [5], generalized ensembles methods of multicanonical sampling [6,7] and simulated tempering (ST) [8,9], and random walk algorithm [10]. Among them, the generalized ensemble methods (GEMs) have been widely applied to biomolecules simulations as effective tools to alleviate multiple minima problem with substantial enhancements of conformational samplings [11].

The characteristic feature of these generalized ensemble methods is that they use non-Boltzmann sampling weights, which are adaptively modified to smooth the PES for a uniform sampling in the energy and temperature spaces. The multicanonical sampling (MUCA) [6] or an equivalent entropic sampling [7] generating a random walk on the energy space has been proved to be very effective in studying first-order phase transitions of spin systems [12] and folding problems of small peptide systems [13]. Another effective simulation technique of the GEMs is the ST [8] or expanded ensemble method [9], which samples states with a uniform distribution in the temperature space. The similarities and effectiveness of various generalized ensemble methods have been examined for protein folding simulations [14]. However, in spite of many impressive achievements, the fundamental dynamics of the GEMs has not been clearly understood since the weights, which play a critical role in the sampling, are not known *a priori*. Usually, the determination of the weights requires an iterative procedure, which is sometimes very tedious and time consuming in large size systems [15].

In the present study, we reveal that uniform sampling weights of the MUCA and ST are determined by the same functional relationship characterizing the average energy and temperature. Our finding relies on a unified formulation of the sampling process in the MUCA and ST within the expanded ensembles treating both energy and temperature as dynamical variables. The sampling process of the MUCA has been shown to be equivalent to the one of the ST by relating the temperature weight to the multicanonical weight. The simulated tempering is identified as the multicanonical sampling with the generalized weight determined by a Laplace transform of the temperature weight. Next, the characteristic sampling dynamics of the GEMs has been analyzed in terms of a Langevin stochastic process by considering the sampling process as a stochastic diffusion on the energy and temperature spaces. Our stochastic models reveal that the iterative refinements of the weights in the GEMs correspond to the dynamical processes approaching a free Brownian motion in the energy and temperature spaces via the cancellations of deterministic forces of governing Langevin equations. Based on this analysis, we propose robust force biased iteration schemes to allow an automatic determination of uniform sampling weights in the GEMs without any intervention in the simulation. The validation of our theory has been tested in a folding simulation of five residues peptide of Met-Enkephalin.

In Sec. II, a unified formulation of the sampling process in the MUCA and ST has been presented within the expanded ensembles. The formal equivalence of the MUCA and ST has been demonstrated by relating the temperature weight to the

*Corresponding author. Email address: jgkim@jbirc.aist.go.jp

multicanonical weight. In Sec. III, the characteristic sampling dynamics of the GEMs has been investigated in the stochastic formulation of the sampling process. Novel iteration schemes have been proposed for the determination of uniform sampling weights in the energy and temperature spaces. In Sec. IV, detailed numerical results and discussions are presented in a folding simulation of Met-Enkephalin. Finally, a conclusion and brief summary are added to Sec. V.

II. EXPANDED ENSEMBLES FORMULATION

One of the critical differences of the MUCA and ST is a dynamical variable in the sampling process. In the MUCA, the energy is a dynamical variable, while the temperature plays the same role in the ST. Thus, to formulate the MUCA and ST in a unified form, we need new ensembles treating both energy and temperature as dynamical variables simultaneously. For this purpose, we borrowed a key idea from Lyubartsev's expanded ensembles formalism. In the original paper [9], Lyubartsev *et al.*, have constructed the canonical expanded ensembles by considering the reciprocal temperature $\beta_i = 1/k_B T_i$ as a parameter for the subensembles, which gives the expanded partition function of

$$\Xi(N,V) = \sum_i Q(N,V,\beta_i)\mathcal{F}_i, \tag{1}$$

where $Q(N,V,\beta_i)$ denotes the canonical partition function at β_i. For each subensemble i, the positive weight $\mathcal{F}_i$ has been assigned to adjust a relative population. Usually, two types of MC moves occur in the expanded ensembles: (1) canonical sampling at a fixed value of β_i and (2) trial transition between neighboring β_i. The transitions between the subensembles take place with the acceptance ratio of

$$\min[1, e^{(\beta_i-\beta_j)E}\mathcal{F}_j/\mathcal{F}_i]. \tag{2}$$

Then, the occupation probability of $P_i = \exp\{-\beta_i A_i\}\mathcal{F}_i/\Xi$, $\beta_i A_i = -\ln Q(N,V,\beta_i)$, can be determined by accumulating a histogram of visits to β_i state. Once a broad sampling is achieved between the subensembles with a high statistics, free energy difference between the subensemble j and reference 0 can be calculated as $\beta_j A_j - \beta_0 A_0 = \ln P_0/P_j - \ln \mathcal{F}_0/\mathcal{F}_j$. However, the temperature weight $\mathcal{F}_i$ is not known *a priori* and has to be determined by preliminary iterative simulations before long production run.

In the present study, the expanded ensembles formalism has been generalized to a continuous limit of β_i by introducing the temperature weight function $\mathcal{F}(\beta)$ as

$$\Xi(N,V) = \int\int d\beta dE\ \Omega(E)e^{-\beta E}\mathcal{F}(\beta), \tag{3}$$

where we used the relation $Q(N,V,\beta) = \int dE \Omega(E) e^{-\beta E}$, $\Omega(E)$ being the density of state of the system. In the continuous expanded ensembles, the system with the configuration $(\mathbf{x},\beta)$ is populated by a product of the energy and temperature weights as $\mathcal{P}(\mathbf{x},\beta) = e^{-\beta E(\mathbf{x})}\mathcal{F}(\beta)/\Xi$. From a detailed balance condition, the transition between $(\mathbf{x},\beta)$ and $(\mathbf{x}',\beta')$ occurs with the probability of

$$\min[1, e^{-\beta' E(\mathbf{x}')+\beta E(\mathbf{x})}\mathcal{F}(\beta')/\mathcal{F}(\beta)]. \tag{4}$$

The acceptance ratio of Eq. (4) reduces to Eq. (2) when $E(\mathbf{x}) = E(\mathbf{x}')$, while it is the same as the Metropolis criterion with $\beta = \beta'$. The occupation probability of the system in (E,β) is determined by taking into account the density of state as $\mathcal{P}(E,\beta) = \Omega(E)e^{-\beta E}\mathcal{F}(\beta)/\Xi$.

Our basic idea is to regard the expanded ensembles of Eq. (3) as the multicanonical sampling associated with a non-Boltzmann weight $\mathcal{W}(E)$ as

$$\Xi(N,V) = \int dE\ \Omega(E)\mathcal{W}(E), \tag{5}$$

where $\mathcal{W}(E) = \mathcal{L}[\mathcal{F}]$ is a Laplace transform of $\mathcal{F}$ defined as $\int d\beta\, e^{-\beta E}\mathcal{F}(\beta)$. Then, the probability density function (PDF) for the energy is given by

$$P(E) = \Omega(E)\mathcal{L}[\mathcal{F}]/\Xi = e^{-\beta_0\Phi(E)}/\Xi, \tag{6}$$

where $\Phi(E) = \alpha(E) - T_0\mathcal{S}_m(E)$, $\mathcal{S}_m(E)$ being the microcanonical entropy defined by $k_B \ln \Omega(E)$. Here we defined the multicanonical potential $\alpha(E) = -1/\beta_0 \ln \mathcal{W}(E)$, β_0 being the arbitrarily defined temperature. The physical implication of $\Phi(E)$ is the effective free energy density associated with $\mathcal{W}(E)$, leading to the partition function of $\Xi(N,V) = \int dE e^{-\beta_0\Phi(E)}$.

On the other hand, the PDF for the inverse temperature is obtained by integrating $\mathcal{P}(E,\beta)$ with respect to an energy as

$$P(\beta) = \mathcal{L}[\Omega]\mathcal{F}(\beta)/\Xi = e^{-\Psi(\beta)}/\Xi, \tag{7}$$

where $\mathcal{L}[\Omega]$ is the Laplace transform of the density of state defined as $\int d\beta \Omega(E) e^{-\beta E}$, and $\Psi(\beta)$ is the effective potential of $\beta A(\beta) - \ln\mathcal{F}(\beta)$, $A(\beta)$ being the Helmholtz free energy. The interesting observation in Eq. (7) is that the Laplace transform of $\Omega(E)$ plays a role of the density of state in the temperature sampling. In a different way, Hansmann *et al.* also pointed out that the temperature weight is related to a Laplace transform of the density of state as $\mathcal{F}^{-1} = \mathcal{L}[\Omega]$ when the sampling in the temperature space becomes uniform [14].

The reformulation of the MUCA and ST in the expanded ensembles provides a clear picture on how does the temperature weight invoke a non-Boltzmann weight in the energy space. Conversely, an arbitrary energy weight can be decomposed to a mixture of the canonical ensembles parametrized by a temperature weight. The Gibbs-Boltzmann weight $\mathcal{W}_{GB} = e^{-\beta_0 E}$ is given by a δ function distribution $\mathcal{F}(\beta) = \delta(\beta - \beta_0)$ $(\beta_0 > 0)$. On the other hand, Lyubartsev's expanded ensembles are reconstructed by $\mathcal{F}(\beta) = \Sigma_i \mathcal{F}_i \delta(\beta - \beta_i)$. Recently, the Tsallis weight proposed in nonextensive statistical mechanics [16] has been applied to biomolecules simulations as one effective tool improving conformational sampling efficiency [17–20]. In contrast to the exponential suppression of $\mathcal{W}_{GB}$, the Tsallis weight shows an asymptotic algebraic decaying in the energy space as

$$\mathcal{W}_q(E) = [1 - (1-q)\beta_0 E)]^{1/(1-q)}, \tag{8}$$

where q is the Tsallis entropy index. Dynamical origin of the Tsallis weight has been related to the temperature fluctuations with χ^2 distribution [21]. This can be easily verified in our expanded ensembles formulation by denoting a following transformation of

$$\mathcal{W}_q(E)=A(a,p)\mathcal{L}[\beta^p\exp\{a\beta\}]=\mathcal{L}[\mathcal{F}_q] \tag{9}$$

with $p=1/(q-1)-1$ $(p>-1)$ and $1/a=(1-q)\beta_0$. Here $A(a,p)=[1/\Gamma(1+p)](-a)^{1+p}$, Γ being the Gamma function, and $\mathcal{F}_q$ is the Tsallis temperature weight. Note that the dynamical origin of the Tsallis weight is naturally involved with temperature fluctuations in our formulation via the transformation of $\mathcal{W}(E)=\mathcal{L}[\mathcal{F}(\beta)]$.

III. STOCHASTIC FORMULATION OF GENERALIZED ENSEMBLE METHODS

A. Multicanonical sampling

Since the MUCA can be considered as a canonical sampling associated with the multicanonical potential $\alpha(E)$ in Eq. (6), the energy trajectory can be generated to sample the ensemble of $\mathcal{W}(E)$ by constant temperature MD with a scaled Newton's equation [22]

$$\dot{\mathbf{p}}_i=-\frac{\partial\alpha(E)}{\partial\mathbf{q}_i}=\frac{\partial\alpha(E)}{\partial E}\mathbf{f}_i=\nu(E)\mathbf{f}_i, \tag{10}$$

where $\mathbf{q}_i$, $\mathbf{p}_i$, and $\mathbf{f}_i$ correspond to the coordinate, momentum, and force of the particle i on the potential energy E, respectively. Here, $\nu(E)$ is the force scaling function, which reduces to $\nu(E)=1$ in the canonical ensemble of $\alpha(E)=E$. The sampling dynamics of MUCA can be mapped into a one-dimensional stochastic diffusion on the external potential $\Phi(E)$ modeled by a Langevin equation of [23]

$$\partial_t E=\mathcal{G}(E)+\sqrt{k_BT_0}\ \zeta(t), \tag{11}$$

where

$$\mathcal{G}(E)=-\partial_E\Phi(E)=\nu_{\mathcal{S}}(E)-\nu(E) \tag{12}$$

and $\nu_{\mathcal{S}}(E)=T_0/T_{\mathcal{S}}(E)$, $T_{\mathcal{S}}(E)$ being the statistical temperature defined in microcanonical ensemble as $[\partial\mathcal{S}_m/\partial E]^{-1}$. In Eq. (11), thermal fluctuations are approximated by the unbiased δ correlated Gaussian white noise with $\langle\zeta(t)\zeta(t')\rangle=2\delta(t-t')$. The stationary PDF of Eq. (11) is obtained as $P(E)=e^{\beta_0\int\mathcal{G}(E')dE'}$ by considering a Fokker-Planck equation associated with a Langevin equation.

In Eq. (12), the deterministic force $\mathcal{G}(E)$ has two contributions, which are the system-dependent entropic force $\nu_{\mathcal{S}}(E)$ and weight-dependent force $\nu(E)$. The uniform energy sampling in the MUCA is realized from a generation of a random walk when $\mathcal{G}(E)=0$ by coinciding $\nu(E)$ with $\nu_{\mathcal{S}}(E)$. However, the force scaling function $\nu(E)$ has to be determined iteratively since the microcanonical entropy $\mathcal{S}_m(E)$ is not known *a priori*. The rearrangement of Eq. (12) gives an iteration scheme for the force scaling function as

$$\nu^{i+1}(E)=\nu^i(E)+\mathcal{G}^i(E), \tag{13}$$

where i means the iteration step [24]. Notifying that $\nu(E)$ is the external force driven by the weight in Eq. (12), the sampling in MUCA can be interpreted as the dynamical process approaching a free Brownian motion via the iterative cancellations of $\mathcal{G}(E)$. Conventional potential biasing iteration scheme [22] is obtained by integrating both sides of Eq. (13) as

$$\alpha^{i+1}(E)=\alpha^i(E)+\frac{1}{\beta_0}\ln P^i(E). \tag{14}$$

Since the trajectory in MD is determined directly by a force scaling, it is a more natural choice to bias the effective force $\nu(E)$ rather than potential $\alpha(E)$. Detailed application of the iteration scheme of Eq. (13) has been reported in our previous study for a helix-coil transition of $(\mathrm{Ala})_8$ system in a gas phase [24].

In the MUCA, overall sampling dynamics can be determined by identifying stationary points of $\Phi(E)$. When thermal fluctuations are ignored in Eq. (11), zeros of $\mathcal{G}(E)$ correspond to fixed points whose stabilities are analyzed by a linearization of $\mathcal{G}(E)$ as $\mathcal{G}(E)\simeq\mathcal{G}'(E^*)(E-E^*)+\cdots$, where

$$\mathcal{G}'(E)=\frac{\partial\mathcal{G}}{\partial E}=\frac{1}{\tilde{T}\tilde{T}_{\mathcal{S}}}\left(\frac{\partial\tilde{T}}{\partial E}-\frac{\partial\tilde{T}_{\mathcal{S}}}{\partial E}\right), \tag{15}$$

$\tilde{T}_{\mathcal{S}}=T_{\mathcal{S}}(E)/T_0$, and $\tilde{T}(E)=1/\nu(E)$. Here we defined the effective temperature $\tilde{T}(E)$ associated with the weight $\mathcal{W}(E)$, which becomes the exact statistical temperature when $\alpha(E)=\mathcal{S}_m(E)$ [24]. Then, the stabilities of fixed points are determined by

$$\kappa(E^*)=\left[\frac{\partial\tilde{T}}{\partial E}\right]\left[\frac{\partial\tilde{T}_{\mathcal{S}}}{\partial E}\right]^{-1}. \tag{16}$$

Since Φ is concave around E^* for $\kappa<1$, the sampling dynamics concentrates on stable fixed points, while it flows away from unstable fixed points with $\kappa>1$ due to a convexity of Φ. When $\kappa=1$, tangential points represents a marginal stability. The stability analysis of Eq. (16) can be used to characterize an essential dynamics of molecular dynamics simulation driven by Tsallis weight of Eq. (8). In the Appendix, we showed that fixed points determined by crossing points of the statistical temperature $T_{\mathcal{S}}(E)$ and Tsallis effective temperature play a critical role in overall sampling dynamics of MD driven by the Tsallis weight.

B. Simulated tempering

Previous analysis can be equally applied to the dynamics of β in the ST since β is a stochastic variable driven by MC simulation. The underlying stochastic differential equation (SDE) is derived to give the stationary solution $P(\beta)\sim e^{-\Psi(\beta)}$ as

$$\partial_t\beta=\Pi(\beta)+\zeta(t), \tag{17}$$

where the deterministic force $\Pi(\beta)$ becomes

$$\Pi(\beta) = -\partial_\beta \Psi(\beta) = -U(\beta) + \mu(\beta), \tag{18}$$

$U(\beta)$ being the average energy and $\mu(\beta) = \partial_\beta \ln \mathcal{F}(\beta)$. In Eq. (18), we used a canonical thermodynamic relation of $A = U - TS$, S being the canonical entropy given by $-(\partial A/\partial T)_V$. As noted in the MUCA, the deterministic force $\Pi(\beta)$ has also two kind of contributions; system-dependent force $U(\beta)$, which is not known *a priori*, and weight-contributed force $\mu(\beta)$, which is modulated by altering $\mathcal{F}(\beta)$.

The uniform sampling in β space is accomplished by a random walk condition of $\mu(\beta) = U(\beta)$. This can be also directly derived from the condition for a constant $P(\beta)$ in Eq. (7) i.e., $\mathcal{F}(\beta) = \mathcal{L}^{-1}[\Omega(E)]$. By differentiating both sides, we have

$$\mu = \frac{\partial \mathcal{F}(\beta)}{\partial \beta} = \frac{\int dE \Omega(E) e^{-\beta E} E}{\int dE \Omega(E) e^{-\beta E}} = U(\beta). \tag{19}$$

However, $\mu(\beta)$ has to be determined with an iterative manner since $U(\beta)$ cannot be determined before the simulation. Rearranging Eq. (18), the iteration scheme for $\mu(\beta)$ is derived as

$$\mu^{i+1}(\beta) = \mu^i(\beta) - \Pi^i(\beta), \tag{20}$$

where $\Pi^i(\beta) = \partial_\beta \ln P^i(\beta)$ for ith iteration step. The iteration scheme for the temperature weight is straightforward as

$$\ln \mathcal{F}^{i+1}(\beta) = \ln \mathcal{F}^i(\beta) - \ln P^i(\beta). \tag{21}$$

In terms of our stochastic model, the sampling process of ST can be also interpreted as the dynamical process converging to a free Brownian motion in β space by canceling the deterministic force iteratively.

The approximate form of the uniform sampling weight $\mathcal{F}_U(\beta)$ in β space can be simply calculated using various canonical samplings at different temperatures. By interpolating an average energy set $[\beta_i, U_i]$ of the canonical samplings, U_i being $U(\beta_i)$, $\mathcal{F}_U(\beta)$ is approximated as $\exp\{\int d\beta' U(\beta')\}$. With the weight $\mathcal{F}_U(\beta)$, the PDF of Eq. (6) can be rewritten as

$$P(E) = \int_0^\infty d\beta \; e^{\mathcal{D}(\beta,E)}/\Xi, \tag{22}$$

where $\mathcal{D}(\beta,E) = \beta[U - TS - \mathcal{A}(\beta,E)]$, $\mathcal{A}$ being the free energy density defined by $E - T\mathcal{S}_m(E)$. Since both U and S are proportional to the number of particle N, the integral of Eq. (22) receives contributions only from the neighborhood of the maximum integrand $U(\beta^*) = E$ in a thermodynamic limit. By using the expansion of $\mathcal{D}(\beta,E) = \mathcal{D}(\beta^*,E) - \frac{1}{2}\sigma(\beta - \beta^*)^2 + \cdots$, Eq. (22) can be rewritten as

$$P(E) \simeq e^{\mathcal{D}(\beta^*,E)} \int_{-\beta^*}^\infty d\beta e^{-\sigma\beta^2/2} = \sqrt{2\pi/\sigma}\; e^{\mathcal{D}(\beta^*,E)}, \tag{23}$$

where $\sigma = |\partial U/\partial\beta|_{\beta^*} = k_B T^{*2} C_V(T^*)$, C_V being the specific heat of the system. The lower limit of above integral has been replaced by $-\infty$ since $\sigma \propto N$. Denoting $\mathcal{D}(\beta^*,E) = \ln\Omega(E) - S[U(\beta^*)]/k_B = 0$, we conclude that the uniform sampling in β space brings about the uniform distribution in the energy space.

C. Uniform sampling weights of MUCA and ST

The condition for a uniform sampling in β space, i.e., $\mu = U(\beta)$, has a fundamental relationship with a uniform sampling condition in the energy space. Denoting $\alpha(E) = -\ln \mathcal{L}[\mathcal{F}]/\beta_0$, the force scaling function $\nu(E)$ can be rewritten as

$$\nu(E) = \frac{1}{\beta_0} \frac{\int d\beta \beta e^{-\beta E} \mathcal{F}(\beta)}{\int d\beta e^{-\beta E} \mathcal{F}(\beta)} = \frac{1}{\beta_0} \langle \beta \rangle_{\mathcal{F}}, \tag{24}$$

where $\langle \cdots \rangle_{\mathcal{F}}$ means the Boltzmann weighted average with respect to the distribution $\mathcal{F}$. The effective inverse temperature of Eq. (24) can be further identified as

$$\nu(E) = \beta^*/\beta_0 \tag{25}$$

by applying the steepest descent integral around the stationary point β^* as $\int_0^\infty d\beta \beta^n e^{-\beta E + \ln \mathcal{F}(\beta)} = e^{-\beta^* E + \ln \mathcal{F}(\beta^*)} I_n(\beta^*)$, where $I_n = \int_0^\infty d\beta \beta^n e^{-\sigma_\mu (\beta - \beta^*)^2/2}$ and $\sigma_\mu = |\partial\mu/\partial\beta|_{\beta^*}$. Here, β^* is determined by $E = \mu(\beta^*)$ corresponding to the stationary condition of the integrand of I_n. Applying the uniform sampling condition in β space of $\mu = U(\beta)$, the weight for a uniform energy sampling is obtained as

$$\nu(E) = U^{-1}(E)/\beta_0, \tag{26}$$

which can be approximated by interpolating average energy set $[U_i, \beta_i/\beta_0]$. Note that the uniform sampling weight in the energy space is complementary with the uniform β weight via the same functional relationship of β and $U(\beta)$. The condition of Eq. (26) has been also demonstrated in our previous study with the staircase temperature modulation in Eq. (11) [23]. Independently, Terada *et al.*, also reach the same conclusion that the multicanonical weight can be determined by interpolating maximum probability points of the canonical samplings at different temperatures [25]. Note that the maximum probability energy becomes identical to $U(\beta)$ in a thermodynamic limit.

IV. NUMERICAL COMPUTATIONS AND DISCUSSIONS

A. Implementation of simulated tempering into MD

The simulated tempering algorithm can be implemented into MD by considering the force scaling function as an effective inverse temperature parametrizing the expanded ensembles. The force scaling function $\nu(E)$ in Eq. (10) reduces to a temperature scaling factor when it is a constant as $\nu(E) = 1/\lambda$ $(\lambda > 0)$ [23,24]. In this case, the PDF of Eq. (6)

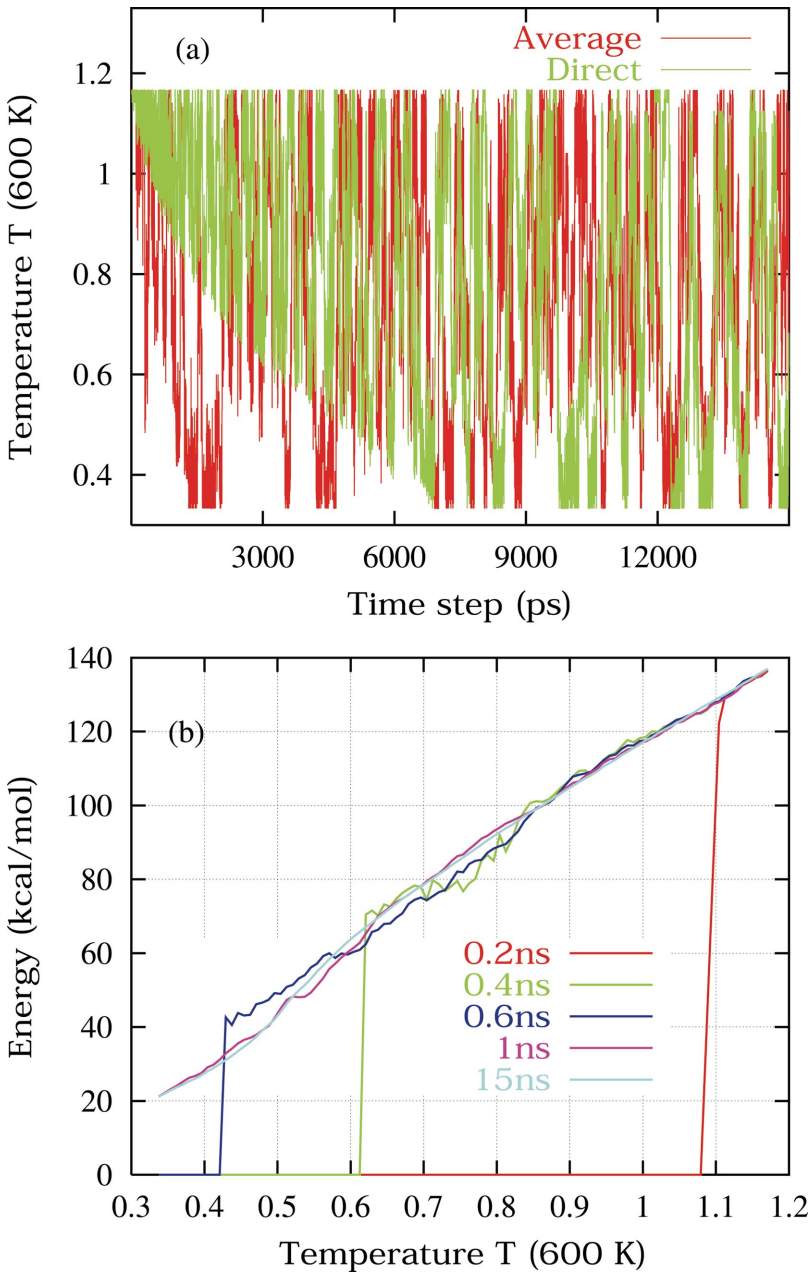

FIG. 1. (Color) (a) Temperature trajectories of force biased iteration of Eq. (28): direct (green) and average energy guided (red). (b) Approximate average energy $U_H^i(T)$ as a function of an iteration. The temperature weight $\tilde{\mu}(T)$ is determined as $U_H^i(T)/k_BT^2$.

reduces to a canonical one at a scaled temperature $T_\lambda = \lambda T_0$ as $P(E)=\Omega(E)e^{-E/k_BT_\lambda}$ denoting $\alpha(E)=E/\lambda$. Thus, the expanded ensembles can be constructed in the sampling process of the MUCA by considering $\nu(E)$ as a dynamical variable parametrizing the canonical subensembles of the temperature $T_0/\nu(E)$. However, we need a further consideration since the inverse relation of β and T makes a sampling biased toward low-temperature region as $P(T) = P(\beta)|\partial\beta/\partial T| \sim 1/T^2$ even though we can realize a uniform sampling in β space. To avoid an unfavorable bias, it is required to perform the simulation in the temperature space directly rather than β space.

This proceeds with a variable transformation from β to T in previous stochastic formulation. By defining a new effective potential $\tilde{\Psi}(T)=A(T)/k_BT-\ln\tilde{\mathcal{F}}(T)$, the governing SDE is transformed to

$$\partial_t T = \tilde{\Pi}(T) + \zeta(t), \tag{27}$$

where $\tilde{\Pi}(T)=\tilde{U}(T)-\tilde{\mu}(T)$, $\tilde{U}$ and $\tilde{\mu}$ being $U(T)/k_BT^2$ and $-\partial\ln\mathcal{F}(T)/\partial T$, respectively. The iteration scheme in the temperature space is obtained as

$$\tilde{\mu}^{i+1}(T) = \tilde{\mu}^i(T) + \tilde{\Pi}^i(T), \tag{28}$$

where $\tilde{\Pi}^i(T)=\partial\ln P^i(T)/\partial T$. The simulation proceeds with a combinatorial fashion: particle displacements are generated by the MUCA with a fixed $\nu=T_0/T$, while temperature transitions from T to T' have been tried with the acceptance ratio of

$$A(T\to T') = \min[1,\ e^{(\beta-\beta')E}\tilde{\mathcal{F}}(T')/\tilde{\mathcal{F}}(T)], \tag{29}$$

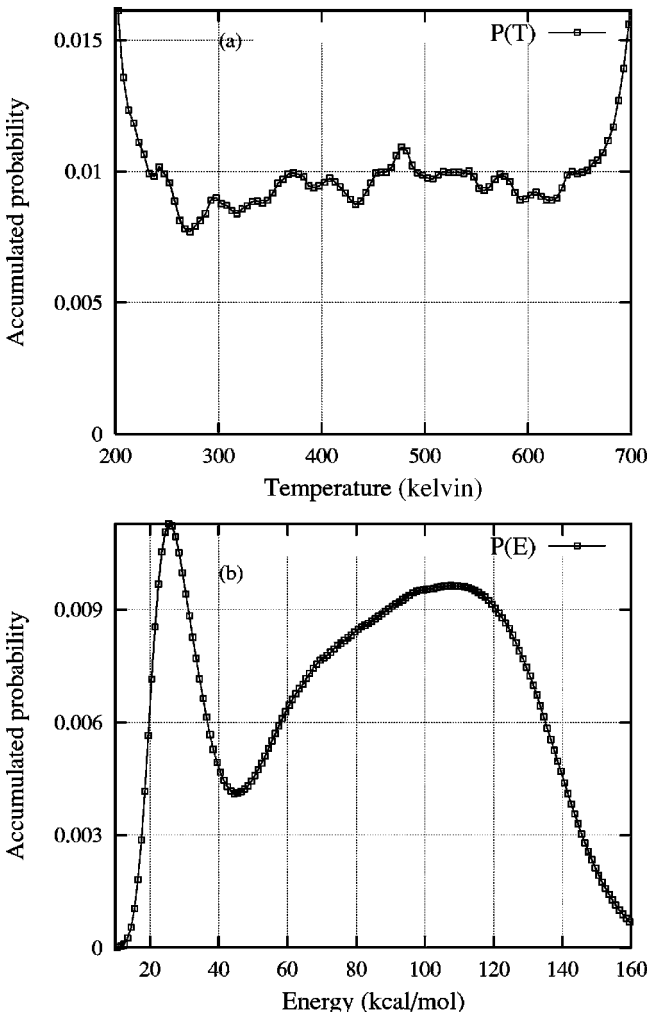

FIG. 2. Accumulated probability density functions for (a) the temperature and (b) energy.

where $\tilde{\mathcal{F}}(T)=\exp\{-\int^{T}\tilde{\mu}(T')dT'\}$.

Denoting a random walk condition in the temperature space i.e., $\tilde{\mu}=\tilde{U}(T)$, the convergence of Eq. (28) can be greatly accelerated by replacing $\overline{\mu^i}$ by an approximation of $\tilde{U}(T)$ i.e., $\tilde{U}_H^i=U_H^i(T)/k_BT^2$, in which $U_H^i(T)$ is an approximate internal energy. For each iteration, $U_H^i(T)$ is automatically updated by taking an average of all previous energy data in each temperature histogram. Since the weight is guided by the average energy $U_H^i(T)$ from the beginning of the simulation, preliminary simulations, which are essential in conventional ST, are not required in our iteration scheme. The detailed procedure of our simulation scheme is outlined as follows. (i) Perform the simulation at an arbitrary temperature T_0 with an initial weight $\mathcal{F}^0(E)=1$ $[\mu^0(E)=0]$. Since there is no restriction for a selection of trial values in temperature transitions, we used a discrete temperature by dividing a temperature space into grids with a size of $(T_{max}-T_{min})/N_G$, N_G being the number of the expanded ensembles. The temperature sampling has been restricted for $[T_{min},T_{max}]$ by applying the acceptance ratio as

$$A(T\rightarrow T')=\begin{cases} T_{min} & \text{for} \quad T'\leqslant T_{min} \\ \text{Eq. (29)} & \text{for} \quad T_{min}\leqslant T\leqslant T_{max} \\ T_{max} & \text{for} \quad T>T_{max}. \end{cases} \tag{30}$$

The temperature transitions are tried every predetermined time steps of N_T. (ii) By constructing the temperature histogram, calculate the deterministic force $\tilde{\Pi}^i(T)=\partial\ln P^i(T)/\partial T$ and approximate internal energy $U_H^i(T)$. (iii) Update the weight every N_I time steps using the scheme of

$$\tilde{\mu}^{i+1}(T)=\begin{cases} \tilde{U}_H^i(T)+\tilde{\Pi}^i(T) & \text{for} \quad T_1^i\leqslant T\leqslant T_2^i \\ \tilde{U}_H^i(T) & \text{for} \quad \text{otherwise}, \end{cases} \tag{31}$$

where T_1^i and T_2^i are determined by $P^i(T_1)\leqslant P_{th}$ and $P^i(T_2)\geqslant P_{th}$, respectively, P_{th} being the threshold value to maintain a statistical accuracy of the simulation. (iv) Repeat steps (i)–(iii) for a certain number of iterations until the broad distribution is obtained in the temperature space. For an automatic determination of the weight, four parameters N_G, N_T, N_I, and P_{th} have to be determined before starting the simulation.

B. Simulation of folding transition of Met-Enkephalin

The performance of our method has been examined in a folding simulation of Met-Enkephalin consisting of five residues Tyr-Gly-Gly-Phe-Met in a gas phase whose *N* and *C* termini were blocked with acetyl and *N*-methyl groups, respectively. Since Met-Enkephalin is very simple, but a nontrivial system, showing a folding transition even in a gas phase, it has been widely used to test a performance of many sampling algorithms [14,18,26]. The simulation was performed by the program PRESTO [22,27] and the force-field parameters were taken from all-atom version of AMBER [28] with 1 fs time step and no cutoff. The simulation has been performed with $N_T=100$ time steps, $N_G=100$, and $P_{th}=10^{-3}$. The sampling region in the temperature space has been restricted by boundary conditions of $T_{min}=200$ K and $T_{max}=700$ K. The update of weight has been done every $N_I=2\times10^5$ time steps. All hydrogen bonds have been fixed by SHAKE constraints.

As observed in Fig. 1(a), the temperature sampling shows a typical random walk covering an entire temperature space, which leads to a broad sampling in the energy space plotted in Fig. 2(b). The advantage of using the average energy guide of $\tilde{U}_H^i$ is clearly seen in convergence rates of the guided (red) and direct (green) iterations, in which the replacement of $\tilde{\mu}^i$ by $\tilde{U}_H$ enables a dramatic enhancement in an initial stage of the simulation. The approximate average energy $U_H^i(T)$ in the average guided simulation has been plotted in Fig. 1(b) as a function of the iteration. The weight $\tilde{\mu}^i(T)$ at each iteration corresponds to a scaled internal energy of $U_H^i(T)/k_BT^2$. The weight at 1 ns, in which an entire energy range has been sampled roughly, becomes almost identical to the final convergent one at 15 ns. It should be emphasized that the weight $\tilde{\mu}(T)$ has been determined fully

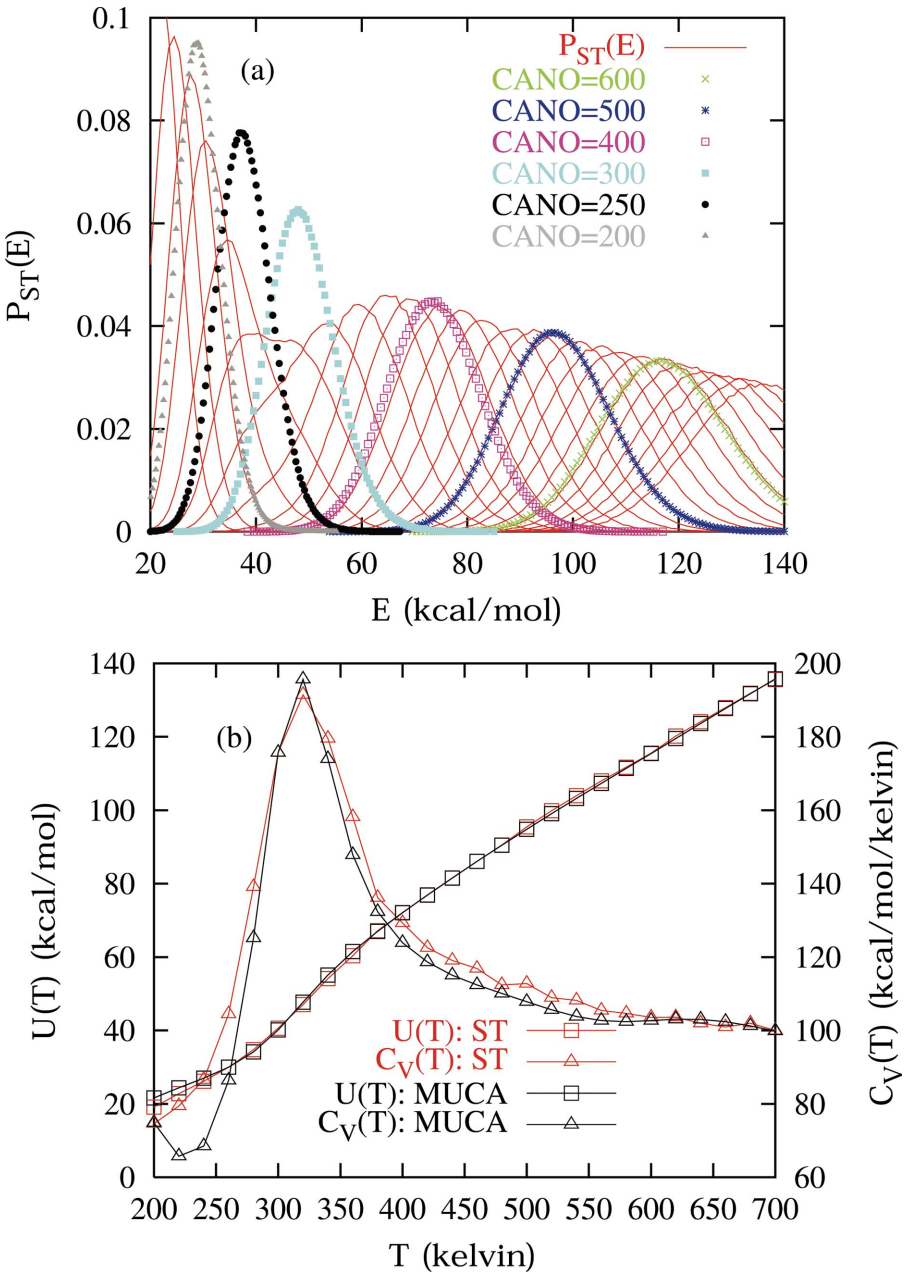

FIG. 3. (Color) (a) $P_{ST}(E)$: reconstructed canonical PDFs every 20 K from 200 K to 700 K. Above 400 K, conventional MD results coincide with P_{ST}, while they begin to depart from P_{ST} in the transition and low-temperature regions. (b) Average energy $U(T)$ and specific heat $C_V(T)$. The convergent weight $\tilde{\mu}(T)$ is identified with $U(T)/k_BT^2$. For a comparison, we plotted $U(T)$ and $C_V(T)$ obtained from the MUCA by the reweighing.

automatically by dividing a temperature space into grids with a size of $(T_{max}-T_{min})/N_G=5$ K. The attempts to automatize the iteration procedures in the MUCA have been tried in Monte Carlo [29] and molecular dynamics algorithm [24]. However, to the best of our knowledge, the same attempt has not been tried to the simulated tempering. Since the sampling energy range increases exponentially as a function of a size of the system, full automatic determination of the weight might play an important role in the applications of ST to large size systems.

The accumulated probability density functions for the temperature and energy are plotted in Figs. 2(a) and 2(b), respectively. Except for boundary regions where $P(T)$ shows unusual peaks due to an interruption of probability currents beyond T_{min} and T_{max}, $P(T)$ shows a broad distribution over entire temperature space, which also realizes a broad sampling in the energy space in Fig. 2(b). The less sampled region of $P(E)$ around 40 kcal/mol is due to a relative rare events of the folding transitions in Met-Enkephalin. Note that the sampling energy region of $P(E)$ around 40 kcal/mol in Fig. 2(b) exactly corresponds to the energy region of the transition temperature of 320 K by examining the profile of $U(T)$ in Fig. 3(b). Actually, a small bias in $P(E)$ does not affect a statistical accuracy of the simulation since temperature transitions obey a detailed balance condition.

The canonical distributions in the ST can be reconstructed by gathering energy data of the same temperature in the simulation trajectory. In Fig. 3(a), we plot the reconstructed canonical PDFs of P_{ST} (red line) and typical canonical PDFs (point) obtained from conventional canonical MD. The distributions of P_{ST} are nearly Gaussian at temperatures above and below $T_{trans}=320$ K in Fig. 3(a), implying the existence of well-defined conformational classes. Here, the transition temperature has been identified by the temperature corre-

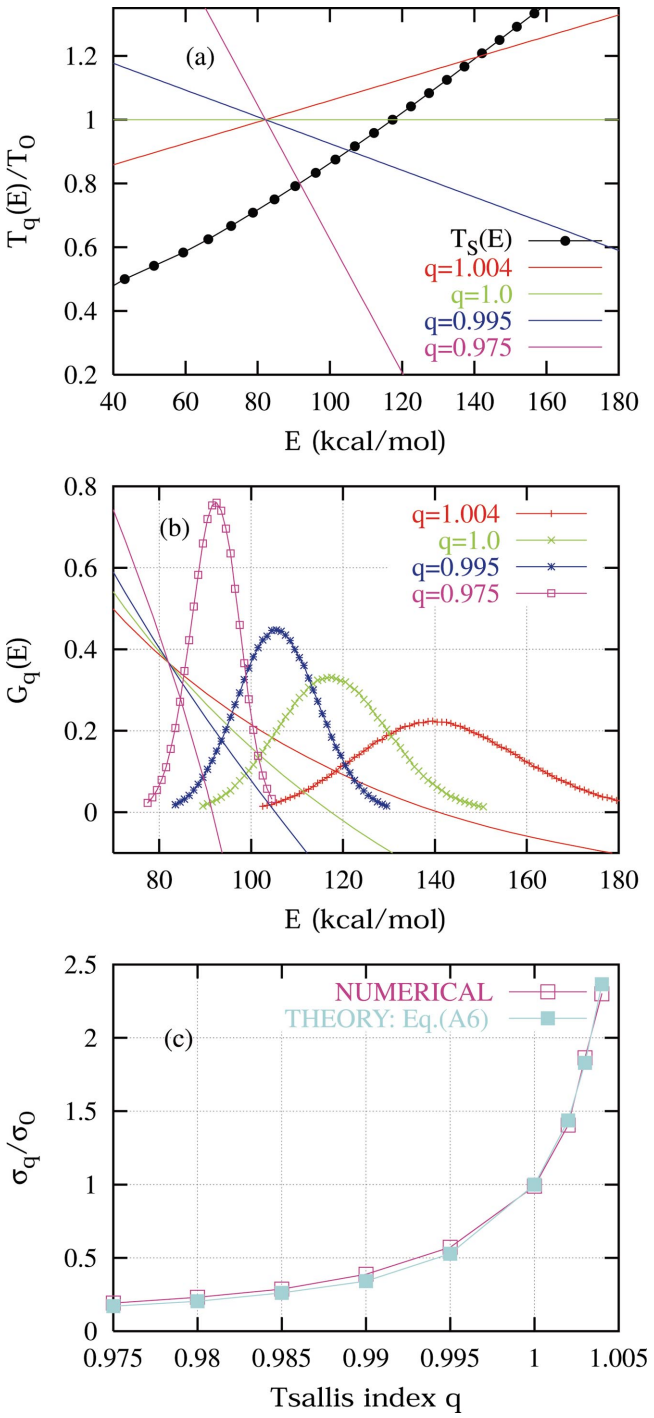

FIG. 4. (Color) (a) The statistical temperature $\tilde{T}_S(E)$ and Tsallis effective temperature $\tilde{T}_q(E)$ with $\epsilon=0.7E_0$, E_0 being the average energy of the canonical MD at T_0. (b) Deterministic forces $\mathcal{G}_q(E)$ (Solid lines). Fixed points E^* correspond to zeros of $\mathcal{G}_q(E)$. For a comparison, the Tsallis PDFs (Lines points) are magnified by ten times. (c) Ratio of σ_q/σ_0 with respect to q.

sponding to the maximum peak in the specific heat in Fig. 3(b). Above 400 K, typical canonical PDFs exactly coincide with $P_{ST}(E)$. However, due to a quasiergodicity, the PDFs of canonical MD show a significant deviation from $P_{ST}(E)$ in the transition and low-temperature regimes of Fig. 3(a). Note that $P_{ST}(E)$ from 280 K to 340 K shows a characteristic broadening of the distributions, showing the existence of two kinds of conformational states. The characteristic dynamics of the folding transition is clearly identified by the slope variation in $U(T)$ corresponding to the maximum peak of the specific heat $C_V(T)$ in Fig. 3(b). Similar maximum peaks in C_V have been also observed in multicanonical Monte Carlo simulation [18,29] and random walk simulation [26] with different force fields.

To check the accuracy of the ST, we compared thermodynamic quantities of $U(T)$ and $C_V(T)$ of the ST with those of the MUCA in Fig. 3(b). For this purpose, we performed total 10 ns MD simulation using repeated short time multicanonical samplings of 4×10^5 time steps with the update scheme of Eq. (13). Since the weights of each iteration step are slightly different, all simulation data has been joined to reconstruct the canonical ensembles via the multiple histogram technique [30]. Except for very low-temperature regions around 200 K, both results of the GEMs exhibit a good agreement for an entire temperature space, which validates our theory. However, the usefulness of our iteration scheme for the simulated tempering needs to be tested for larger systems to test a scalability of the algorithm with respect to the size of system.

V. CONCLUSIONS

In conclusion, the multicanonical sampling has been shown to be equivalent to the simulated tempering applying the expanded ensembles formalism. The simulated tempering is identified as the multicanonical sampling with the generalized weight determined by a Laplace transform of the temperature weight. The characteristic sampling dynamics of both generalized ensemble methods has been verified using the stochastic interpretation of the sampling process in the energy and temperature spaces modeled by a Langevin equation. We showed that the iterative refinements of the weights in the GEMs are equivalent to the dynamical processes converging to a free Brownian motion via the cancellations of deterministic forces in Langevin equations. Our study also reveals that the uniform sampling weight in the energy space is complementary with the uniform temperature weight via the same functional relationship of T and $U(T)$. From the stochastic models, we derive robust force biased iteration schemes, which enables full automatic determinations of uniform sampling weights.

ACKNOWLEDGMENTS

We would like to thank M.S. Yukihisa Watanabe, Yoshiaki Mikami, and Takashi Kurosawa for technical support. We acknowledge that this work was supported by NEDO and METI.

APPENDIX: ESSENTIAL DYNAMICS OF MOLECULAR DYNAMICS SIMULATION DRIVEN BY TSALLIS WEIGHT OF EQ. (8)

Molecular dynamics simulation driven by the Tsallis weight proceeds by introducing the Tsallis effective potential $\alpha_q(E)$ as

$$\alpha_q(E) = -\frac{1}{\beta_0}\ln \mathcal{W}_q(E)$$

$$= \frac{1}{\beta_0(q-1)}\ln[1-(1-q)\beta_0(E-\epsilon)], \quad \text{(A1)}$$

where the parameter ϵ is introduced to guarantee a positiveness of $\mathcal{W}_q$ [17]. Until now, the successful applications of the Tsallis weight for conformational samplings have been considered due to the smoothing of the Tsallis effective potential. Indeed, the transformation of Eq. (A1) with a proper q reduces the magnitude of barrier heights of the original potential E maintaining potential energy minima intact [19]. However, the detailed connection of the potential modification to the sampling dynamics can be clearly understood only by taking into account an influence of $\Omega(E)$ since the sampling depends not only the weight, but also the density of state $\Omega(E)$ of the system. This can be done by the fixed point analysis of the statistical temperature $T_{\mathcal{S}}(E)$ and the Tsallis effective temperature of

$$\tilde{T}_q = 1/\nu_q(E) = 1 + (q-1)\beta_0(E-\epsilon). \quad \text{(A2)}$$

Note that the Tsallis effective temperature $\tilde{T}_q(E)$ is a linear function of E with q-dependent slope and definite point of $(\epsilon,1)$.

When the Tsallis parameters q and ϵ are explicitly given, the fixed points are determined by $\tilde{T}_{\mathcal{S}}(E^*) = \tilde{T}_q(E^*)$. By using the Taylor expansion of

$$\tilde{T}_{\mathcal{S}}(E) = \eta + \xi(E-E^*) + \cdots, \quad \text{(A3)}$$

where η and ξ correspond to $\tilde{T}_{\mathcal{S}}(E^*)$ and $[T_0 C_V^*]^{-1}$, respectively, C_V^* being $[\partial T_{\mathcal{S}}/\partial E]^{-1}_{E^*}$, the deterministic force $\mathcal{G}_q$ in Eq. (12) reduces to

$$\mathcal{G}_q(E) \simeq \frac{(q-1)\beta_0 - \xi}{\eta^2}(E-E^*) + \cdots, \quad \text{(A4)}$$

where $\epsilon = E^* - (\eta-1)/(q-1)\beta_0$. By assuming that $\tilde{T}_{\mathcal{S}}$ is well approximated by a linear function around E^*, the first-order truncation in Eq. (A4) gives a stationary PDF of a Gaussian centered at E^* with q-dependent width σ_q of

$$P_q(E) \simeq \frac{1}{\sqrt{2\pi\sigma_q}}\exp\left\{-\frac{(E-E^*)^2}{2\sigma_q}\right\}, \quad \text{(A5)}$$

where

$$\sigma_q = \sigma_0 \eta^2/[1-(q-1)(\beta_0/\xi)], \quad \text{(A6)}$$

$\sigma_0 = k_B T_0^2 C_V^*$. When q approaches 1, the Tsallis effective temperature reduces to the Gibbs-Boltzmann limit of $\tilde{T}_q(E) = 1$ irrespective of ϵ. In the same limit, $P_q(E)$ becomes the canonical PDF for the temperature T_0 having the Gaussian width σ_0. Note that E^* and C_V^* are identified as the average energy and specific heat of the system denoting a thermodynamic relation of $\partial_E \mathcal{S}|_{E=E^*} = 1/T_0$.

To examine Eq. (A5), various samplings have been tried with different parameters set of q and ϵ for $T_0 = 600$ K. For the analysis, we approximated the statistical temperature $\tilde{T}_{\mathcal{S}}(E)$ by $U_{TS}(E)/T_0$, where U_{TS} is the average energy calculated from the ST in Fig. 3(b). In Fig. 4(a), the Tsallis effective temperatures have been plotted with $\tilde{T}_{\mathcal{S}}(E)$ for different values of q with maintaining $\epsilon = 0.7E_0$, E_0 being the average energy of the canonical MD at T_0. Since $\kappa(E^*) < 1$ for all cases of q, the Tsallis PDF shows a typical Guassian distribution centered at E^* in Fig. 4(a). As can be seen in Fig. 4(b), the deterministic force is greater (smaller) than zero for $E < E^*$ ($E > E^*$) biasing the sampling toward E^*. Note that maximum probability points of the Tsallis PDFs are exactly corresponding to zeros of the deterministic forces in Fig. 4(b). In present simulations, the fixed points are numerically identified as the average energies of $E^* = \int dE' P_q(E') E'$. The value of ξ has been determined as $(q-1)\beta_0 - \eta^2 p(E^*)$ by computing the slope $p(E^*)$ of the deterministic force in Eq. (A4) with $\eta = \tilde{T}_q(E^*)$. Numerically determined ratio of σ_q/σ_0 shows a good agreement with theoretical predictions of Eq. (A6) in Fig. 4(c), which validates our stochastic model.

[1] M. Karplus and G.A. Petsko, Nature (London) **347**, 631 (1990).
[2] J.P.K. Doye, D.J. Wales, and M.A. Miller, J. Chem. Phys. **109**, 8143 (1998).
[3] C. Maranas and C. Floudas, J. Chem. Phys. **97**, 7667 (1992).
[4] D.D. Frantz, D.L. Freeman, and J.D. Doll, J. Chem. Phys. **93**, 2769 (1990); K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).
[5] J.P. Valleau, J. Chem. Phys. **99**, 4718 (1993).
[6] B.A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249 (1991).
[7] J. Lee, Phys. Rev. Lett. **71**, 211 (1993).
[8] E. Marinari and G. Parisi, Europhys. Lett. **19**, 451 (1992).
[9] A.P. Lyubartsev, A.A. Martsinovski, S.V. Shevkunov, and P.N. Vorontsov-Velyaminov, J. Chem. Phys. **96**, 1776 (1992).
[10] F. Wang and D.P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
[11] U.H.E. Hansmann, Comp. Sci. Eng. **5**, 64 (2003).
[12] B.A. Berg and T. Celik, Phys. Rev. Lett. **69**, 2292 (1992).
[13] U.H.E. Hansmann and Y. Okamoto, J. Comput. Chem. **14**, 1333 (1993); N.A. Alves and U.H.E. Hansmann, J. Chem. Phys. **117**, 2337 (2002).
[14] U.H.E. Hansmann and Y. Okamoto, Phys. Rev. E **54**, 5863 (1996).
[15] S. Kumar, P. Payne, and M. Vasquez, J. Comput. Chem. **17**, 1269 (1996).
[16] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).
[17] I. Andricioaei and J.E. Straub, Phys. Rev. E **53**, 3055 (1996); J.

Chem. Phys. **107**, 9117 (1997).
[18] U.H.E. Hansmann and Y. Okamoto, Phys. Rev. E **56**, 2228 (1997).
[19] Y. Pak and S. Wang, J. Chem. Phys. **111**, 4359 (1999).
[20] I. Fukuda and H. Nakamura, Phys. Rev. E **65**, 026105 (2002).
[21] C. Beck, Phys. Rev. Lett. **87**, 180601 (2001).
[22] N. Nakajima, H. Nakamura, and A. Kidera, J. Phys. Chem. B **101**, 817 (1997); U.H.E. Hansman, Y. Okamoto, and F. Eisenmenger, Chem. Phys. Lett. **259**, 321 (1996).
[23] J.G. Kim, Y. Fukunishi, and H. Nakamura, Phys. Rev. E **67**, 011105 (2003).
[24] J.G. Kim, Y. Fukunishi, A. Kidera, and H. Nakamura, Phys. Rev. E **68**, 021110 (2003).
[25] T. Terada, Y. Matsuo, and A. Kidera, J. Chem. Phys. **118**, 4306 (2003).
[26] N. Rathore, T.A. Knotts IV, and J.J. de Pablo, J. Chem. Phys. **118**, 4285 (2003).
[27] K. Morikami, T. Nakai, A. Kidera, M. Saito, and H. Nakamura, J. Comput. Chem. **16**, 243 (1992).
[28] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, and P.A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
[29] F. Yasar, T. Celik, B.A. Berg, and H. Meirovitch, J. Comput. Chem. **21**, 1251 (2000); **23**, 1127 (2002).
[30] A.M. Ferrenberg and R.H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988); **63**, 1195 (1989).