Chemical Physics Letters 661 (2016) 279–283

Contents lists available at ScienceDirect

Chemical Physics Letters

journal homepage: www.elsevier.com/locate/cplett

Editor's choice

Multiscale enhanced sampling for protein systems: An extension via adiabatic separation

Kei Moritsugu [a,*], Tohru Terada [b], Akinori Kidera [a]

[a] Graduate School of Medical Life Science, Yokohama City University, 1-7-29 Tsurumi, Yokohama 230-0045, Japan
[b] Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo, Tokyo 113-8657, Japan

ARTICLE INFO

Article history:
Received 3 August 2016
In final form 31 August 2016
Available online 1 September 2016

Keywords:
Multiscale enhanced sampling
Adiabatic separation
All-atom model
Chignolin

ABSTRACT

Multiscale enhanced sampling (MSES) calculates the configurational ensemble of all-atom (MM) protein systems with the help of coupling to a coarse-grained (CG) model. Here, for further improvement of the sampling efficiency, the approximation of adiabatic separation was introduced to the original MSES, by adopting a high CG temperature limit. An application to the folding of chignolin in explicit solvent demonstrated that the MSES formula based on adiabatic separation correctly sampled the canonical ensemble with excellent efficiency and robustness against the parameter selection, and thus MSES successfully achieved the scalability for applications to large protein systems.

1. Introduction

In the computational point of view, comprehensive understanding of protein functional dynamics requires the calculation of the associated free energy landscape, or the well-converged structural ensemble in the space of related dynamical variables. Due to the large computational burden, however, the evaluation of a free energy landscape still remains challenging for large protein systems. Large-scale structural sampling of protein systems has frequently been attained by the help of "enhanced sampling". The temperature replica exchange and related methods [1–4] are the best known methods to enhance sampling, because of their broad applicability. However, since these methods simply attempt to enhance the sampling by increasing the temperature of the *whole* system, one inevitably encounters difficulties in applications to large systems. The other class of enhanced sampling methods, including umbrella sampling [5], conformational flooding [6], metadynamics [7] and temperature accelerated molecular dynamics [8], enhances the sampling along pre-defined *small dimensional* "reaction coordinates" or "collective variables (CV)". The difficulty in the applications of these methods is that the knowledge of the proper reaction coordinates is prerequisite for successful sampling.

To alleviate these problems in the two types of sampling methods, we have developed "multiscale enhanced sampling (MSES)", in which the sampling of the all-atom model is enhanced by the coupling with the associated coarse-grained (CG) model, moving on a substantially smoothed potential energy surface with reduced dimensionality [9–12]. The multiscale system is composed of an all-atom system containing protein molecules and surrounding solvents (MM; the coordinates, $\mathbf{r}_{MM}$), and the corresponding coarse-grained system ($\mathbf{r}_{CG}$) that is defined flexibly, according to prior knowledge or experimental information. The potential energy of the multiscale system, V, is given by

$$V(\mathbf{r}_{MM}, \mathbf{r}_{CG}, k) = V_{MM}(\mathbf{r}_{MM}) + V_{CG}(\mathbf{r}_{CG}) + V_{MMCG}, \quad (1)$$

with the coupling term between MM and CG, V_{MMCG}, as follows:

$$V_{MMCG} = k_{MMCG}[\chi_{MM}(\mathbf{r}_{MM}) - \chi_{CG}(\mathbf{r}_{CG})]^2. \quad (2)$$

V_{MM} and V_{CG} are the potential energy functions for MM and CG, respectively. In this scheme, the number of degrees of freedom in CG, M, is by definition much smaller than that in MM, N. V_{MMCG} is described by harmonic constraints with a force constant k_{MMCG} for K variables, χ_{CG}, that are defined from the CG coordinates. The K-dimensional vector χ_{MM} is a projection of $\mathbf{r}_{MM}$ onto the K-dimensional space. The dimension K is then taken so as to satisfy the condition, $K < M \ll N$, and also can be set to be much larger than the number of CVs in the second class of the sampling methods [5–8] described above.

The target quantity calculated in the MSES is the unbiased free energy surface originating solely from the MM potential, V_{MM}, unlike the second class of the sampling methods that attempt to obtain the free energy surfaces along the pre-defined CVs by simulating at a sufficiently large value of k_{MMCG}. For this purpose,

* Corresponding author.
E-mail address: moritugu@tsurumi.yokohama-cu.ac.jp (K. Moritsugu).

http://dx.doi.org/10.1016/j.cplett.2016.08.075
0009-2614/

it is necessary to derive the joint distribution related not only to the configuration ($\mathbf{r}_{MM}$, $\mathbf{r}_{CG}$) but also to the coupling constant k_{MMCG} as an extended-system variable; i.e., $\rho(\mathbf{r}_{MM}, \mathbf{r}_{CG}, k_{MMCG}) \propto \exp(-\beta V)$, which allows the system to be extrapolated to $k_{MMCG} = 0$. Here, to explain the MSES method and its extension, we follow the formulation of "Gibbs sampling" [13,14]. Gibbs sampling is an algorithm to generate the probability distribution of a multivariate system, by sampling each variable separately in an iterative manner by the use of the conditional probability. The Gibbs sampling process of MSES proceeds as follows. In the first step of the iteration, the coordinates of the system, ($\mathbf{r}_{MM}$, $\mathbf{r}_{CG}$), are sampled by an MD simulation with a fixed value of k_{MMCG} under the conditional probability of

$$\rho(\mathbf{r}_{MM}, \mathbf{r}_{CG}|k_{MMCG}) = \exp(-\beta V)/Z(k_{MMCG}), \tag{3}$$

where β is the inverse temperature, and $Z(k_{MMCG}) \equiv \int d\mathbf{r}_{MM} d\mathbf{r}_{CG} \exp(-\beta V)$ is treated as the state-dependent (the k_{MMCG}-dependent) weighting factor in Gibbs sampling [14]. In the MD simulation of ($\mathbf{r}_{MM}$, $\mathbf{r}_{CG}$), the value of k_{MMCG} is fixed, and thus Z is taken as a constant. In the second step, the sampling in terms of k_{MMCG} is carried out by Markov chain Monte Carlo (MCMC) simulations for discretized values of k_{MMCG}, by introducing many replicated systems with various values of k_{MMCG} ranging from a large value to zero, and by exchanging the k_{MMCG} values between neighboring replicas. This sampling scheme is called Hamiltonian replica exchange [15]. Suppose the exchange between the neighboring replicas of system m with the conditional probability of $\rho(k^m_{MMCG}|\mathbf{r}^m_{MM}, \mathbf{r}^m_{CG})$ and of system n with $\rho(k^n_{MMCG}|\mathbf{r}^n_{MM}, \mathbf{r}^n_{CG})$, which produces the exchanged systems of $(\mathbf{r}^m_{MM}, \mathbf{r}^m_{CG}, k^n_{MMCG})$ and $(\mathbf{r}^n_{MM}, \mathbf{r}^n_{CG}, k^m_{MMCG})$. The exchange probability satisfying the detailed balance condition is then given by

$$p_{mn} = \min(1, \exp(\Delta_{mn})), \tag{4}$$

with

$$\begin{aligned}\Delta_{mn} &= \ln\left[\frac{\rho(k^n_{MMCG}|\mathbf{r}^m_{MM}, \mathbf{r}^m_{CG})\rho(k^m_{MMCG}|\mathbf{r}^n_{MM}, \mathbf{r}^n_{CG})}{\rho(k^m_{MMCG}|\mathbf{r}^m_{MM}, \mathbf{r}^m_{CG})\rho(k^n_{MMCG}|\mathbf{r}^n_{MM}, \mathbf{r}^n_{CG})}\right] \\ &= \beta(k^m_{MMCG} - k^n_{MMCG})\Big([\chi_{MM}(\mathbf{r}^m_{MM}) - \chi_{CG}(\mathbf{r}^m_{CG})]^2 \\ &\quad - [\chi_{MM}(\mathbf{r}^n_{MM}) - \chi_{CG}(\mathbf{r}^n_{CG})]^2\Big),\end{aligned} \tag{5}$$

and

$$\rho(k_{MMCG}|\mathbf{r}_{MM}, \mathbf{r}_{CG}) = \exp(-\beta V)/Z(\mathbf{r}_{MM}, \mathbf{r}_{CG}), \tag{6}$$

where $Z(\mathbf{r}_{MM}, \mathbf{r}_{CG}) \equiv \int dk_{MMCG} \exp(-\beta V)$. In Eq. (5), the weighting factors, $Z(\mathbf{r}^m_{MM}, \mathbf{r}^m_{CG})$ and $Z(\mathbf{r}^n_{MM}, \mathbf{r}^n_{CG})$, are canceled out between the numerator and the denominator. Eq. (5) shows that the probability is determined by the difference between χ_{MM} and χ_{CG}, defined in the K-dimensional space. Since $K \ll N$, Δ_{mn} is kept small enough to provide a high exchange probability p_{mn}, irrespective of the size of the MM system. This guarantees high scalability as compared with the conventional temperature replica exchange method, where the difference in the potential energy of MM (scaling up to N^2) determines the exchange probability Δ_{mn}. An extension of MSES using multiple CG copies for a single MM system was also proposed, and was found to improve the sampling efficiency [12].

In the present study, another extension is proposed for the further improvement of the sampling efficiency, by strengthening the CG's driving force for MM. Since CG has to accelerate MM with large degrees of freedom and a complicated interaction energy, reinforcing the driving force for MM is a key factor to improve the sampling power. For this purpose, we now introduce the approximation of "adiabatic separation" that has been widely used in various systems, including proteins [8,16–18]. Adiabatic separation imposes the condition that CG moves much slower than MM or the mass of CG is sufficiently heavier than that of MM. To improve sampling efficiency, the temperature of CG, T_{CG}, is set to be much higher than that of MM, or β' ($=1/k_B T_{CG}$) $\ll \beta$, so that CG may easily drive MM. Under these conditions, the CG simulation at β' with the coupling to MM generates the conditional probability for CG, given in the form of the potential of mean force of MM [8,16],

$$\begin{aligned}\rho(\mathbf{r}_{CG}|k_{MMCG}) &= \left[\int d\mathbf{r}_{MM} \exp(-\beta V)\right]^{\beta'/\beta} \Big/ Z(k_{MMCG}) \\ &= \exp[-\beta' V_{CG}(\mathbf{r}_{CG})] Z(\mathbf{r}_{CG}, k_{MMCG})^{\beta'/\beta} \Big/ Z(k_{MMCG}),\end{aligned} \tag{7}$$

with

$$Z(\mathbf{r}_{CG}, k_{MMCG}) \equiv \int d\mathbf{r}_{MM} \exp(-\beta[V_{MM}(\mathbf{r}_{MM}) + V_{MMCG}(\mathbf{r}_{MM}, \mathbf{r}_{CG}, k_{MMCG})]), \tag{8}$$

$$Z(k_{MMCG}) \equiv \int d\mathbf{r}_{CG} \exp[-\beta' V_{CG}(\mathbf{r}_{CG})] Z(\mathbf{r}_{CG}, k_{MMCG})^{\beta'/\beta}. \tag{9}$$

Meanwhile, MM evolves with a given CG on the conditional probability [8],

$$\begin{aligned}\rho(\mathbf{r}_{MM}|\mathbf{r}_{CG}, k_{MMCG}) &= \exp(-\beta[V_{MM}(\mathbf{r}_{MM}) \\ &\quad + V_{MMCG}(\mathbf{r}_{MM}, \mathbf{r}_{CG}, k_{MMCG})])/Z(\mathbf{r}_{CG}, k_{MMCG}).\end{aligned} \tag{10}$$

The conditional probability for MM and CG at a given value of k_{MMCG} is therefore,

$$\begin{aligned}\rho(\mathbf{r}_{MM}, \mathbf{r}_{CG}|k_{MMCG}) &= \rho(\mathbf{r}_{MM}|\mathbf{r}_{CG}, k_{MMCG})\rho(\mathbf{r}_{CG}|k_{MMCG}) \\ &= \exp(-\beta[V_{MM}(\mathbf{r}_{MM}) + V_{MMCG}(\mathbf{r}_{MM}, \mathbf{r}_{CG}, k_{MMCG})] \\ &\quad - \beta' V_{CG}(\mathbf{r}_{CG}))Z(\mathbf{r}_{CG}, k_{MMCG})^{\frac{\beta'}{\beta}-1}/Z(k_{MMCG}).\end{aligned} \tag{11}$$

Note that Eq. (11) reduces back to the original form of Eq. (3) when $\beta' = \beta$. $\rho(k_{MMCG}|\mathbf{r}_{MM}, \mathbf{r}_{CG})$ is derived by changing the denominator of Eq. (11) from $Z(k_{MMCG})$ to $Z(\mathbf{r}_{MM}, \mathbf{r}_{CG})$, and then we have the exchange probability,

$$p_{mn} = \min(1, \exp(\Delta'_{mn})) \tag{12}$$

$$\begin{aligned}\Delta'_{mn} &= \ln\left[\frac{\rho(k^n_{MMCG}|\mathbf{r}^m_{MM}, \mathbf{r}^m_{CG})\rho(k^m_{MMCG}|\mathbf{r}^n_{MM}, \mathbf{r}^n_{CG})}{\rho(k^m_{MMCG}|\mathbf{r}^m_{MM}, \mathbf{r}^m_{CG})\rho(k^n_{MMCG}|\mathbf{r}^n_{MM}, \mathbf{r}^n_{CG})}\right] \\ &= \Delta_{mn} + \left(\frac{\beta'}{\beta} - 1\right)\ln\left[\frac{Z(\mathbf{r}^m_{CG}, k^n_{MMCG})Z(\mathbf{r}^n_{CG}, k^m_{MMCG})}{Z(\mathbf{r}^m_{CG}, k^m_{MMCG})Z(\mathbf{r}^n_{CG}, k^n_{MMCG})}\right],\end{aligned} \tag{13}$$

where Δ_{mn} is the term defined in Eq. (5). The assumption of adiabatic separation turns out to be the addition of the second term of Eq. (13), which is not easy to evaluate.

To circumvent the difficulty in the evaluation of the second term of Eq. (13), we consider the high temperature limit of adiabatic separation, or $\beta'/\beta \to 0$. At this limit, the conditional probability of Eq. (7) becomes independent of the influence from the potential of mean force of MM, and CG behaves freely from MM; that is,

$$\rho(\mathbf{r}_{CG}|k_{MMCG}) \sim \exp[-\beta' V_{CG}(\mathbf{r}_{CG})]. \tag{14}$$

When the CG motion is generated by Eq. (14), it is now possible to set the replicated systems of the MCMC simulation consisting of many MMs driven by a *single copy* of CG, or $\mathbf{r}^m_{CG} = \mathbf{r}^n_{CG}$ (Fig. 1). Under these conditions, the exchange probability can be reduced from Eq. (13) back to Eq. (5). The present extension of MSES produces the largest driving force of MM, because CG can move freely without feeling the counter force from the potential of mean force of MM. In summary, the simulation process consists of the iterated pro-