Chemical Physics Letters 616–617 (2014) 20–24

Contents lists available at ScienceDirect

Chemical Physics Letters

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

# Multiscale enhanced sampling driven by multiple coarse-grained models

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

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

ARTICLE INFO

*Article history:*
Received 7 August 2014
In final form 6 October 2014
Available online 14 October 2014

ABSTRACT

Enhanced sampling has increased its importance in simulations of large-scale functional dynamics of proteins. "Multiscale enhanced sampling (MSES)" is an enhanced sampling method applicable to large protein systems, in which all-atom (MM) sampling is extended through a coupling with the accelerated dynamics of a coarse-grained (CG) model. Here, we show that the sampling efficiency of MSES can be further improved by using multiple CG copies instead of a single CG model and by introducing repulsive forces between the CG copies. An application to chignolin folding in explicit solvent has demonstrated a significant improvement in sampling efficiency due to the extension.



## 1. Introduction

Large-scale protein functional dynamics are usually described in the form of the free energy landscape derived from an all-atom molecular dynamics (MD) simulation. Since a straightforward MD simulation cannot cover the whole configurational space related to the functional dynamics, even for moderate-size globular proteins, the construction of the free-energy landscape requires acceleration of configurational sampling or "enhanced sampling". The temperature replica exchange and related methods [1,2] are the best known ways to enhance sampling. However, these methods do not have good scalability, or it is difficult to enhance the sampling of the whole protein molecule. There is another class of enhanced sampling method such as steered or targeted MD [3,4], umbrella sampling [5], conformational flooding [6], and metadynamics [7], which enhances the sampling along a pre-defined small dimensional "reaction coordinates" or "collective variables". The recent successes of enhanced sampling in large proteins have mostly relied on these methods together with proper choices of certain low-dimensional collective variables. However, it faces a difficulty in defining small dimensions when prior knowledge is not available. For these reasons, enhanced sampling remains a challenge for large proteins when the structural changes involve a considerable number of degrees of freedom.

To solve these problems, in our previous work, we proposed "multiscale enhanced sampling (MSES)", in which a sufficiently high dimensional space is sampled by steering the all-atom model through a coarse-grained (CG) model [8,9]. The MSES system contains an all-atom system composed of protein molecules and surrounding solvents (MM; the associated coordinate, $\mathbf{r}_{MM}$, and momentum, $\mathbf{p}_{MM}$) and the corresponding coarse-grained system ($\mathbf{r}_{CG}$ and $\mathbf{p}_{CG}$). The Hamiltonian, $H$, of this system is given by

$$H = V_{MM}(\mathbf{r}_{MM}) + K_{MM}(\mathbf{p}_{MM}) + V_{CG}(\mathbf{r}_{CG}) + K_{CG}(\mathbf{p}_{CG}) + V_{MMCG}(\mathbf{r}_{MM}, \mathbf{r}_{CG}), \quad (1)$$

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

where $V_{MM}$ and $V_{CG}$ ($K_{MM}$ and $K_{CG}$) are the potential (kinetic) energies for MM and CG, respectively, and the number of degrees of freedom in CG, $M$, is much smaller than that of MM, $N$. The CG model can be arbitrarily chosen with the help of prior knowledge or experimental information. The last term, $V_{MMCG}$, defines the coupling (harmonic constraint) for $K$ variables. Here, $\boldsymbol{\chi}_{CG}$ ($\mathbf{r}_{CG}$) is a $K$-dimensional projection of the CG coordinates, $k_{MMCG}$ is a force constant to drive the MM system using the fast dynamics of the CG system, and $\boldsymbol{\chi}_{MM}$ ($\mathbf{r}_{MM}$) is a $K$-dimensional vector that is a projection of $\mathbf{r}_{MM}$ onto the $K$-dimensional space. The coupled system (including both MM and CG) is at thermal equilibrium at temperature $T$.

Extrapolation of the Hamiltonian in Eq. (1) to $k_{MMCG} = 0$ eliminates the effects of biasing from $V_{MMCG}$ and yields an unbiased free energy surface originating from the intrinsic $V_{MM}$. To do this, the Hamiltonian replica exchange method [10] is used, in which a

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

http://dx.doi.org/10.1016/j.cplett.2014.10.009
0009-2614/

number of replicated systems are simulated for various $k_{MMCG}$ ranging from a large value to zero. The exchange probability between replicas $m$ and $n$ having different values of $k^m{}_{MMCG}$ and $k^n{}_{MMCG}$, satisfying the detailed balance condition, is given by

$$p_{mn} = \min(1, \exp(\Delta_{mn})), \tag{3}$$

with

$$\Delta_{mn} = \beta(k^m_{MMCG} - k^n_{MMCG})([\chi_{MM}(\mathbf{r}^m_{MM}) - \chi_{CG}(\mathbf{r}^m_{CG})]^2 - [\chi_{MM}(\mathbf{r}^n_{MM}) - \chi_{CG}(\mathbf{r}^n_{CG})]^2) \tag{4}$$

where $\beta$ is the inverse temperature (=$1/k_BT$; $k_B$ is the Boltzmann constant). Eq. (3) indicates that the probability is determined by the difference between $\chi_{MM}$ and $\chi_{CG}$ defined in the associated $K$-dimensional space. Because $K$ can be set much smaller than $N$, $\Delta_{mn}$ is kept small enough to provide a high exchange probability $p_{mn}$ irrespective of the size of the MM system $N$. This guarantees much higher scalability 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 $\Delta_{mn}$.

A major concern in the MSES method, especially when it is applied to large proteins in explicit solvent, is how to control the dynamics of the CG system: CG may be pulled and immobilized by the reaction from MM whose motion tends to stay in a stable basin. This situation may occur when the coupling force on CG from MM, $-\partial V_{MMCG}/\partial \mathbf{r}_{CG}$, dominates the CG intrinsic force, $-\partial V_{CG}/\partial \mathbf{r}_{CG}$. To solve this problem, the MSES method can be extended by coupling additional CG models with the MM model. This method is thus named "multiple-CG-driven MSES" or "mcMSES". The variants of the multiscale simulation using multiple MM and CG have also been adopted in [11,12]. The Hamiltonian of mcMSES is given by

$$H = V_{MM}(\mathbf{r}_{MM}) + K_{MM}(\mathbf{p}_{MM}) + \sum_{i=1}^{L}\{V_{CG,i}(\mathbf{r}_{CG,i}) + K_{CG,i}(\mathbf{p}_{CG,i}) + V_{MMCG,i}(\mathbf{r}_{MM}, \mathbf{r}_{CG,i})\} + \sum_{i\neq j}^{L} V_{CG,i/CG,j}(\mathbf{r}_{CG,i}, \mathbf{r}_{CG,j}), \tag{5}$$

where $L$ is the number of CG models, and $\mathbf{r}_{CG,i}$ and $\mathbf{p}_{CG,i}$ ($i$ = 1, 2,. . ., $L$) are the coordinate and momentum of the $i$th CG, respectively. $V_{MMCG,i}$ and $V_{CG,i/CG,j}$ respectively denote the couplings between MM and the $i$th CG and between the $i$th CG and the $j$th CG. The coupled system (including MM and the CGs) is also at thermal equilibrium at temperature $T$ (Figure 1). Note that the replica exchange procedure is controlled only by $V_{MMCG,i}$ and is independent of $V_{CG,i/CG,j}$. The force on the $i$th CG can be derived as,

$$f_{CG,i} = -\frac{\partial V_{CG,i}(\mathbf{r}_{CG,i})}{\partial \mathbf{r}_{CG,i}} - \frac{\partial V_{MMCG,i}(\mathbf{r}_{MM}, \mathbf{r}_{CG,i})}{\partial \mathbf{r}_{CG,i}} - \sum_{j\neq i}^{L} \frac{\partial V_{CG,i/CG,j}(\mathbf{r}_{CG,i}, \mathbf{r}_{CG,j})}{\partial \mathbf{r}_{CG,j}} \tag{6}$$

Eq. (6) indicates that $-\partial V_{MMCG,i}/\partial \mathbf{r}_{CG,i}$ can be counterbalanced by $L-1$ terms of $-\partial V_{CG,i/CG,j}/\partial \mathbf{r}_{CG,j}$, which in turn prevents CG from being trapped in a stable basin of MM. More significantly, as shown below, this perturbation will become much more enhanced by introducing a repulsive force between two CG models, as in, e.g.,

$$V_{CG,i/CG,j} = \frac{k_{CG,i/CG,j}}{1 + [\chi_{CG}(\mathbf{r}_{CG,i}) - \chi_{CG}(\mathbf{r}_{CG,j})]^2/\sigma^2} \tag{7}$$

where $k_{CG,i/CG,j}$ is a coupling constant and $\sigma^2$ is a measure of the distance correlation. In summary, mcMSES has the potential of increasing the sampling efficiency over that of the original MSES, and this owes to the increased dynamic perturbations via the interactions with MM and multiple CGs that are repulsive to each other. The forms of Eqs. (5) and (7) may not necessarily be optimal for all simulation systems, but many other choices are possible, e.g., using different $V_{CG,i}$ for each CG, different connectivities among MM and CG models, or different coupling functions for Eqs. (2) and (7).

As an illustrative application, we performed mcMSES for sampling the folding process of chignolin in explicit solvent. This system has been frequently used for validation of the enhanced sampling methods [8,13–18]. In our previous paper [8], the scalability of the MSES was shown to be better than the temperature replica exchange study [13].

Here, two CG copies were used in mcMSES, i.e., $L$ = 2. The Hamiltonian can be simply written as

$$H = V_{MM}(r_{MM}) + K_{MM}(\mathbf{p}_{MM}) + \sum_{i=1}^{2}\{V_{CG,i}(\mathbf{r}_{CG,i}) + K_{CG,i}(\mathbf{p}_{CG,i}) + k_{MMCG,i}[\chi_{MM}(r_{MM}) - \chi_{CG}(\mathbf{r}_{CG,i})]^2\}1 + [\chi_{CG}(\mathbf{r}_{CG,1}) - \chi_{CG}(\mathbf{r}_{CG,2})]^2/\sigma^2 \tag{8}$$

The Hamiltonian replica exchange was conducted using replicas with various values of $k_{MMCG}$ (here, the same value was used for two CG copies, i.e., $k_{MMCG} \equiv k_{MMCG,1} = k_{MMCG,2}$) by use of the exchange probability between replicas $m$ and $n$ with the following $\Delta_{mn}$.

$$\Delta_{mn} == \beta(k^m_{MMCG} - k^n_{MMCG})(\{[\chi_{MM}(\mathbf{r}^m_{MM}) - \chi_{CG}(\mathbf{r}^m_{CG,1})]^2 + [\chi_{MM}(\mathbf{r}^m_{MM}) - \chi_{CG}(\mathbf{r}^m_{CG,2})]^2\} - \{[\chi_{MM}(\mathbf{r}^n_{MM}) - \chi_{CG}(\mathbf{r}^n_{CG,1})]^2 + [\chi_{MM}(\mathbf{r}^n_{MM}) - \chi_{CG}(\mathbf{r}^n_{CG,2})]^2\}) \tag{9}$$

Note that $\Delta_{mn}$ in Eq. (9) contains two terms for $CG_1$ and $CG_2$, unlike a single term in Eq. (4), and this has the possibility of reducing the exchange probability. However, this problem can be avoided by using multiple CG models to improve the efficiency of the MM dynamics enhancement via multiple CG models, appearing in the use of much smaller values of $k_{MMCG}$ (see Section 2 for details).

In the following, we present the results of mcMSES simulations and demonstrate that the sampling ability of mcMSES is significantly better than that of the original method. It is evidenced in the

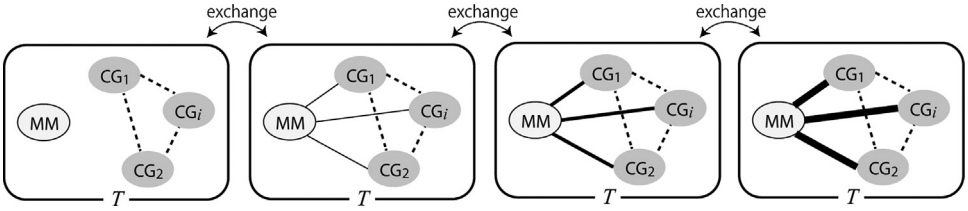


**Figure 1.** Schematic drawing of Hamiltonian replica exchange mcMSES. Each mcMSES system includes a MM model and multiple CG models, which is at thermal equilibrium at temperature $T$ (Eq. (5)). The solid lines indicate the coupling $V_{MMCG,i}$ between the MM and the $i$th CG (the thickness indicates the magnitude of $k_{MMCG}$), while dashed lines are the couplings $V_{CG,1/CG,2}$ between the $i$th and $j$th CGs. On the other hand, the original MSES system (Eq. (1)) consists of a single CG and thus has no $V_{CG,1/CG,2}$coupling.