Chemical Physics Letters 662 (2016) 327–332

Contents lists available at ScienceDirect

Chemical Physics Letters

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

Research paper

Flexible binding simulation by a novel and improved version of virtual-system coupled adaptive umbrella sampling

Bhaskar Dasgupta [a,b,*], Haruki Nakamura [a], Junichi Higo [a]

[a] Institute for Protein Research, Osaka University, 3-2, Yamadaoka, Suita, Osaka 565-0871, Japan
[b] Technology Research Association for Next Generation Natural Products Chemistry, 2-3-26 Aomi, Koto-ku, Tokyo 135-0064, Japan

ARTICLE INFO

Article history:
Received 4 July 2016
In final form 20 September 2016
Available online 21 September 2016

Keywords:
Adaptive umbrella sampling (AUS)
Virtual-system coupled AUS
Markov approximated probability distribution
Force-bias parameterization
Binding free-energy landscape

ABSTRACT

Virtual-system coupled adaptive umbrella sampling (VAUS) enhances sampling along a reaction coordinate by using a virtual degree of freedom. However, VAUS and regular adaptive umbrella sampling (AUS) methods are yet computationally expensive. To decrease the computational burden further, improvements of VAUS for all-atom explicit solvent simulation are presented here. The improvements include probability distribution calculation by a Markov approximation; parameterization of biasing forces by iterative polynomial fitting; and force scaling. These when applied to study Ala-pentapeptide dimerization in explicit solvent showed advantage over regular AUS. By using improved VAUS larger biological systems are amenable.

1. Introduction

Molecular dynamics simulation (MD) is useful to understand the complexities of a polyatomic system, however inefficient because of the complicated force-field equations and explicit treatment of solvents with the system of interest [1]. This makes conformational sampling computationally expensive and sometimes inadequate to simulate long-timescale events [2]. Therefore, understanding microscopic details of those rare events require special generalized ensemble techniques [3–8], for example adaptive umbrella sampling (AUS) [9–12].

Understanding details of rare events are particularly crucial for biomolecular systems, because many of the molecular functions related to the survival of an organism are slow processes (e.g. folding, binding, conformational change) [13,14]. Yet, those microscopic details are useful in medical research [15]. In this context, binding is one of the central functions of biomolecules, which can be regarded as a slow process of sampling between two extremes – bound and unbound states [2]. For such a slow process thermodynamic details may include large statistical error. Supposing a free-energy difference of 3 kcal/mol between bound (major basin) and unbound (minor basin) states, one samples 1.7×10^7 snapshots of bound state per snapshot of unbound state [16], which means large statistical error may be included in binding free-energy landscape. Additionally, when the bound state includes multiple basins the minor basins may also be insignificantly sampled. To overcome this sampling problem AUS or similar techniques are used.

AUS is one of the flat histogram techniques, quite akin to sampling by multicanonical MD (McMD) [2,9]. In AUS, energy of the simulated system is biased by potential of mean force (PMF) along a chosen reaction coordinate (λ). This bias is applied at a given temperature (T), at which the simulation is performed; therefore multiple basins are sampled at that temperature. By iterative MD initially unknown PMF is updated till convergence. The converged PMF corresponds to an uniform probability distribution of the reaction coordinate ($P_{obs}(\lambda, T)$). However, we observed that many cases of iterations fail to converge because of hysteresis of $P_{obs}(\lambda, T)$. One reason for the hysteresis is that there is theoretically infinite number of microstates in a given λ-slice $(\lambda, \lambda + d\lambda)$, therefore in a given iteration only a tiny subset of the microstates may be sampled, and in the successive iteration a different subset is sampled. This can be improved by sampling a λ region exhaustively. This motivated us to design virtual-system coupled AUS or VAUS [16].

The method VAUS has a precursor named virtual-system coupled multicanonical MD (VMcMD) [4,17–20]. In VAUS we simulate the biomolecular system of interest coupled to an arbitrary virtual system composed of virtual states (see Section 2). Moreover, VMcMD can be generalized by VAUS, i.e., the energy is adopted as the reaction coordinate in VMcMD [21].

* Corresponding author at: Institute for Protein Research, Osaka University, 3-2, Yamadaoka, Suita, Osaka 565-0871, Japan.
E-mail address: bhaskar@protein.osaka-u.ac.jp (B. Dasgupta).

http://dx.doi.org/10.1016/j.cplett.2016.09.059
0009-2614/

Previously VMcMD and VAUS were successfully used for dimerization of a endothelin-1 derivative [19] and Aβ-peptide [16], respectively. However, there were a few theoretical limitations on sampling during iterative process. Moreover, we are also interested in qualitative difference between ensembles obtained by VAUS and AUS.

In VAUS we have used a reaction coordinate defined by the distance between geometric centers of a pair of clusters of atoms. The clusters of atoms are included in two different polypeptide chains (Fig. 1), thereby λ can be regarded as a inter-molecular distance. This λ is simple to realize, but sensitive to the variation in biasing potential. This sensitivity requires that one should be careful in estimating the biasing potential. In AUS and VAUS this estimation is crucial to the convergence of simulation and is one of the primary focuses of the current work. To this end, we have improved the calculation of bias by (1) a novel calculation method of probability distribution by using Markov state modeling, (2) refinement of fitting to better parameterize the probability distribution, and (3) de-sensitizing biasing force. The result was compared with that from a regular AUS procedure.

2. Theory

2.1. VAUS

First we briefly introduce VAUS method, whose details are explained in supplementary material S1. In VAUS a virtual degree of freedom is introduced to the real biomolecular system (r) [2,16]. We assume that the virtual degree of freedom is a property of a virtual system (v) and takes discrete states v_k, where $k = 1, \ldots, m$ and m is the number of virtual states. In VAUS along a reaction-coordinate λ, the state of the entire system is designated by λ and v_k. Virtual state v_k is specified by a window $[\lambda_{k,min}, \lambda_{k,max}]$ along the λ axis (supplementary table ST1), and the configuration is restrained in the window if the system is in v_k. VAUS is a sampling method where the system moves along the λ axis continuously and hops among the virtual states (see supplementary material S1 and supplementary Fig. SF1a). The virtual and molecular systems interact to each other [19,22] because potential energy function involves v_k and molecular configurations x. The effective potential energy of the biomolecular system is

$$E_{VAUS}(x, v_k) = E(x, v_k) + RT \ln [P_c(\lambda, v_k, T)/g(v_k, \lambda)], \quad (1)$$

where E is the original potential energy of the molecular system at configuration x using an empirical force-field, R is the gas constant. The term P_c is a canonical probability distribution as a function of λ in virtual state v_k, and g is a function introduced to restrain the biomolecular system in the window of virtual state v_k.

If P_c is given in advance, E_{VAUS} is calculated by Eq. (1), and VAUS simulation is performed using E_{VAUS}. However, P_c is unknown *a priori* and therefore, it is determined by an iterative procedure, where P_c is updated by using the following equation [2]:

$$\ln P_c(\lambda, v_k, T)^{(\gamma+1)} = \ln P_c(\lambda, v_k, T)^{(\gamma)} + \ln P_{VAUS}(\lambda, v_k, T)^{(\gamma)}, \quad (2)$$

where $P_{VAUS}(\lambda, v_k, T)$ is observed probability distribution of λ in virtual state v_k. The detail of this update procedure is outlined in supplementary material S1.

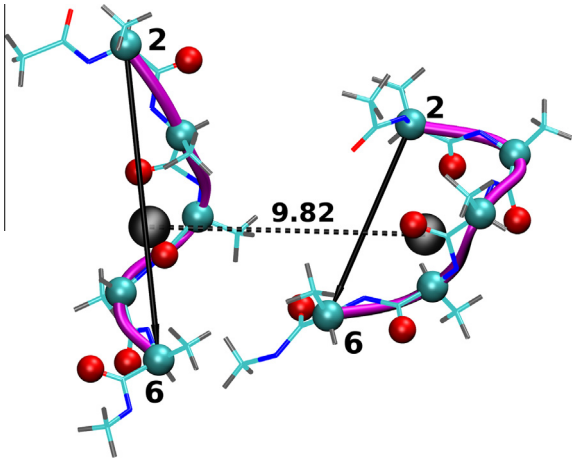

Fig. 1. Equilibrated two Ace-$(Ala)_5$-Nme peptides system (without solvent) used for initial conformation of simulation. Traces of peptides are shown in magenta, and atoms are in stick representation with CPK color scheme. Cyan and red spheres are Cα- and carbonyl O-atoms, respectively. Geometric center of cyan and red-spheres for each peptide is shown by black sphere, and inter-peptide distance (λ), shown by broken line, is the distance between the geometric centers (in Å). Thick black arrows define inter-Cα vector from the first (residue 2) to fifth (residue 6) Ala residues. Angle between these two vectors is referred by θ.

2.2. Improvements of VAUS

Although VAUS is a powerful sampling method [16], some improvements are required. The improvements are aimed at maintenance of detailed balance [2] and better approximation of the distribution functions. In Monte Carlo sampling the detailed balance is automatically satisfied. In MD, however, it is assumed that the detailed balance is satisfied if a trajectory is significantly long. In enhanced sampling schemes, such as VAUS, additional attentions are required.

2.2.1. Markov approximated probability distribution

The trajectory (sequence of recorded λ) obtained from an iteration γ may not be significantly long, and then, the observed distribution may be different from the equilibrium distribution. To obtain the equilibrated distribution, a method (e.g. Markov state model; MSM) that imposes the detailed balance to the trajectory is required. For this purpose, we assume that even for a short trajectory a configuration propagator follows Markov rules [23–25]. In fact MSM is one of the successful methods to understand equilibrium dynamics [26–29]. In short, $P_{VAUS}(\lambda, v_k, T)$ is converted to $P^{MA}_{VAUS}(\lambda, v_k, T)$ in which detailed balance is imposed (supplementary Fig. SF2).

Next we explain details of the technique. In VAUS, the configuration moves along the λ axis by keeping its virtual state at v_k for f_{int} steps of simulation (supplementary information S1, [22]), and the system can hop to a different virtual state only at the f_{int}-th step. Hence, a complete VAUS trajectory can be broken to a set of subtrajectories of length f_{int} and each subtrajectory can be assigned to a virtual state. Accordingly, MSM consists of two layers; one layer is composed of transitions among bins along the λ axis for each virtual state (within a subtrajectory), and the other layer is composed of transitions among virtual states along virtual degree of freedom. A virtual-state transition probability matrix (Ω) can be calculated from the virtual-state sequence, whose (l, k)-th element is the virtual-state transition probability $\Omega(v_l|v_k)$ from v_k to v_l (Fig. SF2b). In the other layer, by collecting subtrajectories belonging to v_k, we calculate an intra-state transition probability matrix Ω^k_{intra} whose matrix element $\Omega^k_{intra}(b_j|b_i)$ provides a transition probability from bin b_i to b_j (Fig. SF2b), where b_i is i-th bin in v_k.

To compute $\Omega(v_l|v_k)$ from the virtual-state sequence, a count-matrix (C) was defined, where $C(v_k, v_l)$ denotes transition counts from v_k to v_l. If the simulation is long enough, an equality stands theoretically: $C(v_k, v_l) = C(v_l, v_k)$. It is known that the use of multiple parallel runs increases sampling statistics [16,30,31].