OUKA
The University of Osaka
Institutional Knowledge Archive

Title	Virtual system Coupled Adaptive Umbrella Sampling : An efficient method to compute potential of mean force along a reaction-coordinate
Author(s)	Dasgupta, Bhaskar
Citation	サイバーメディアHPCジャーナル. 2016, 6, p. 7-11
Version Type	VoR
URL	https://doi.org/10.18910/70506
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

Virtual system Coupled Adaptive Umbrella Sampling: An efficient method to compute potential of mean force along a reaction-coordinate

Bhaskar Dasgupta

Institute for Protein Research, Osaka University

1. Introduction

Polyatomic systems are complex because there exist many degrees of freedom to characterize the system. Biomolecular systems are particularly interesting polyatomic systems because of their medical importance. Such systems are characterized by an ensemble of structures (conformational ensemble), by which we want to realize the system in 3N dimensional space (where N is number of atoms present in the system). The free-energy landscape (3N dimensional) is rugged with many local minima (deep and shallow) and pinholes. This ruggedness of free-energy landscape is a crucial bottleneck to quantitatively characterize a system. Therefore, we need efficient methods to sample conformations of the system such that the system can be described with sufficient accuracy.

Any typical sampling method designed to obtain free-energy landscape suffer due to the ruggedness of energy landscape. In other words during the course of sampling the system may be trapped in a local minimum or basin. Therefore we need to come out of the basins and get in the basin not once, but many times if we are interested in computing thermodynamic quantities. Fortunately, a whole bunch of techniques have been proposed to overcome local basin problem, one of which is adaptive umbrella sampling (AUS) [1].

Suppose we want to compute an ensemble of configuration related to a physical process, e.g. binding between two molecules – a process that samples between two extreme scenarios including bound and unbound states. Because such a process can be a rare event kinetically, we need to bias the system to sample between two such states. In AUS, we add a bias to the potential energy of the system in such a way that the resulting energy landscape is free from ruggedness. The bias is function of collective reaction-coordinate (λ) realizing the process that we want to sample. For example, λ may indicate inter-molecular distance (Fig. 1). In AUS, more specifically we add a bias, which is negative of the potential of mean force (PMF) along λ, leading to modified energy,

$$E_m(\mathbf{r}) = E(\mathbf{r}) + RT \ln P_c\left(\boldsymbol{\lambda}(\mathbf{r}), T\right) \qquad (1)$$

where, **r** denotes a configuration (3N-dimensional vector), and E is potential energy calculated by empirical force-field equations, T is temperature, R is universal gas constant and P_c is canonical probability distribution of λ at T. If we sample our system using equation (1) for sufficiently long time and generate a trajectory (time-sequence) of λ, the observed probability distribution of λ ($P_{obs}(\lambda,T)$) is uniform [2]. However, note that, if we know the canonical probability distribution in advance, we could compute PMF directly ($PMF = -RT \ln P_c\left(\boldsymbol{\lambda}(\mathbf{r}), T\right)$), which has ample of information to understand conformational details of an ensemble. The word 'adaptive' in AUS refers that the method aims to iteratively update canonical distribution function from an initial guess such that observed probability distribution converges to a uniform function along λ. The necessary update formula is,

$$\ln P_c\left(\boldsymbol{\lambda}, T\right)^{(\nu+1)} = \ln P_c\left(\boldsymbol{\lambda}, T\right)^{(\nu)} + \ln P_{Obs}\left(\boldsymbol{\lambda}, T\right)^{(\nu)} \qquad (2),$$

where λ is the given iteration step.

One of the difficulty of AUS method is that it is still computationally expensive, as many iterations are required for convergence (uniformity of $P_{obs}(\lambda,T)$). Moreover, in many cases of simulation the method fails

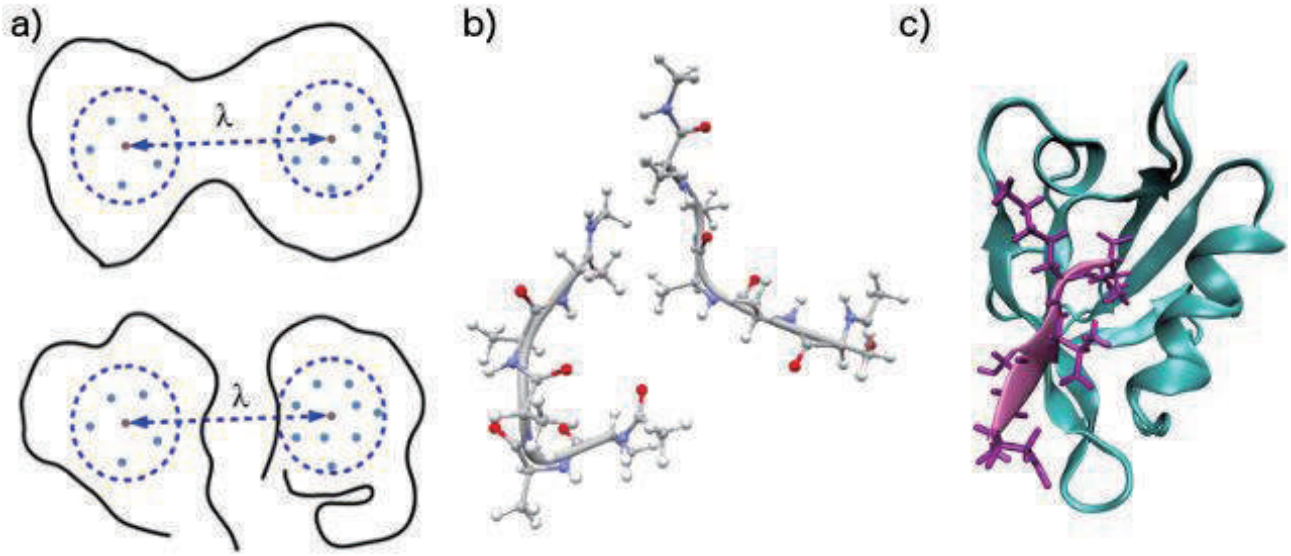

Fig. 1. a) Schematic diagram of reaction coordinate (λ) used in VAUS. Top cartoon suggests λ based on distances within one molecule, while bottom cartoon suggests λ between two molecules (currently used). b) Ace-Ala_5-NMe peptides system, dimerization of two peptides was studied. c) PDZ-GluR2 system as a more realistic example. PDZ is shown in green and GluR2 peptide is shown in magenta color.

because of the sensitivity of the biasing function, and artificial ruggedness is introduced in the landscape due to wrong guess of the canonical distribution function and insufficient sampling in an iterative step. To resolve this, method Virtual-system coupled adaptive umbrella sampling (VAUS) is proposed previously [3], and herein improvements of VAUS will be discussed.

2. Theory:

2.1. Virtual-system coupled adaptive umbrella sampling:

In VAUS, we simulate the molecular system (i.e., real system) by coupling it to a virtual system. The virtual system is defined by an array of virtual states. A virtual state (v_k) is defined along reaction coordinate λ and bounded by $\lambda_{k,min}$ and $\lambda_{k,max}$. Two virtual states are overlapping in λ if they are consecutive (i.e. v_k and $v_{k\pm1}$), but non-overlapping otherwise (e.g. v_k and $v_{k\pm2}$). A construction is shown in Fig. 2. Two extreme limits of λ ($\max\{\lambda_{k,max}\}$ and $\min\{\lambda_{k,min}\}$) denote the conformational space in which we are interested. During simulation the real molecular system is confined to a virtual state for a predetermined steps of integration (f_{int}), after which transition between two consecutive virtual states (v_l and v_k) are possible based on a probabilistic criteria. After f_{int} steps instantaneous value of λ of the real molecular system is checked to be in the overlapping region of virtual states. If this is true, then the total coupled system (real + virtual) transitions from current virtual state to the allowed virtual state (v_l to v_k). In this way, one effectively samples more in a given window of λ covering wider conformational variety. Therefore, the biasing function is more trained. The theory and actual procedures of VAUS is detailed elsewhere [2]. In summary, we use a modified version of equation 1 and 2 as given below,

$$E_m(\mathbf{r}) = E(\mathbf{r}) + RT \ln P_c\left(\boldsymbol{\lambda}(\mathbf{r}), v_k, T\right) \qquad (3), \text{ and}$$

$$\ln P_c\left(\boldsymbol{\lambda}, v_k, T\right)^{(\nu+1)} = \ln P_c\left(\boldsymbol{\lambda}, v_k, T\right)^{(\nu)} + \ln P_{Obs}\left(\boldsymbol{\lambda}, v_k, T\right)^{(\nu)} \quad (4),$$

where, probability distributions are jointly parameterized by λ and v_k.

Although VAUS was successful in a previous case

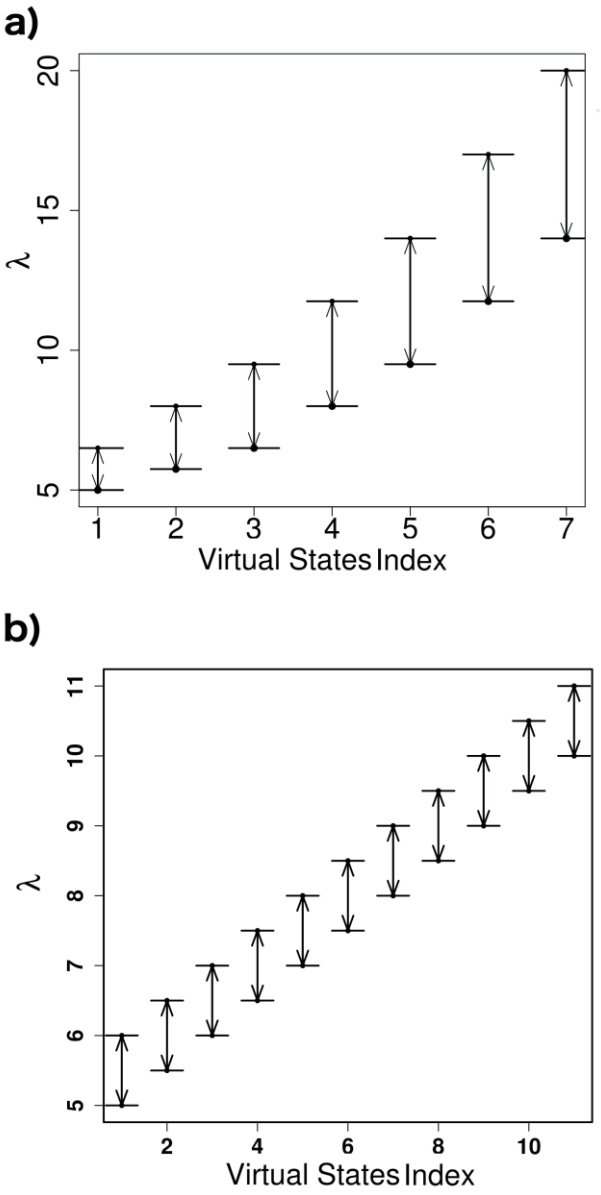

Fig. 2: Virtual-states used in a) Ala-pentapeptide simulation (see Fig. 1b) and b) PDZ-GluR2 simulation (see Fig. 1c).

aimed to sample between bound and unbound states of a dimerization process [3], we found that the method failsto converge in many cases. For example, when we aim to sample a dimerization process of two molecules with large conformational variety owing to their flexibility. One such generic system composed of two peptides where each one is Ace-$(Ala)_5$-NMe (where Ace and NMe are acetyl and N-methyl groups) with no significant conformational preference of the peptide chain. The simulations with VAUS method becomes very computationally demanding due to 1) sensitivity of the biasing forces, 2) finite sampling in each iteration, 3) transition between two virtual states may bring about abrupt change in biasing forces due to mismatch of probability distribution in the overlapping regions and 4) initial guess of bias is not trained over sufficient conformational space. To solve these problems, recently we improved VAUS method, which are briefly discussed below.

2.2. Markov approximated observed probability distribution calculation:

Typically observed probability distribution is calculated from the λ values recorded during the simulation by using number of counts in bins along λ. This disregards time-sequence of λ. To use time-sequence information we first discretize sampling space of λ by using definitions of the bins along λ, then we calculate bin-to-bin transition counts from the recorded trajectory of λ. This gives us a transition probability matrix for transitions between bins for a particular virtual state, which is referred as intra-virtual state transition probability matrix. In a similar way, one can look at trajectory of λ as a sequence of virtual states, because during the simulation total system transitions from one virtual state to other. Therefore, one can calculate transition probability matrix between virtual states, which is referred as inter-virtual state transition probability matrix. Such transition probability matrices can be used to calculate Markov approximated equilibrium counts in each bin, which is free from artifacts of short simulation trajectories. Correspondingly the biasing functions become better trained.

2.3. Iterative polynomial fitting to the observed probability distributions:

In VAUS we calculate observed probability distributions jointly parameterized by λ and v_k, which means that for each virtual state a certain piece of probability distributions is obtained. Typically such pieces of probability distributions are parameterized by n-order polynomials, which require considerable manual intervention. However, pieces of distributions may not

match smoothly in the overlapping regions. For this we invented iterative automatic polynomial fitting. In this approach, n-order polynomials are used to fit a piece of probability distribution with varying polynomial order from 2 to 10 and finding best order that minimizes fitting error. Such an initial fitting uses data from adjacent virtual states (±1 or ±2). Next these fitting parameters are used to reproduce distribution data, which again fitted by polynomials. This process is iterated until we get smooth matching of distributions in the overlapping region. Such technical improvement greatly alleviates abrupt change of forces due to virtual state transitions during simulations.

2.4 De-sensitization of biasing forces:

The update formula in equation 4 needs to be differentiated because we need forces to follow equations of motion. When each term is parameterized by a polynomial, the equation can be viewed as perturbing equation in which canonical distribution is modified by observed distribution in each iterative step. If there is error in current simulation then that will be reflected in polynomial parameterizing observed distribution, making simulation sensitive to the estimation of polynomial coefficients. To de-sensitize the bias a scaling factor is multiplied to those coefficients. We have used a scaling factor of 0.5.

2.5 Initial seed of conformations:

To initiate VAUS, the set of starting conformations should be exhaustively distributed in the configuration space. For this, the system box was split to many rectangular grids and one of the peptide chains was fixed in space and let the other peptide chain diffuse to different grids. We developed an algorithm to traverse all grids by a path most efficiently by combining enumerations of depth-first search. The diffusing peptide chain was allowed to follow the path by an advanced steered molecular dynamics. The break points in the path are approximately centers of grids, and the conformations that reach those break points were used to the initiate canonical simulation under VAUS. This way a large conformational ensemble was retrieved in the first canonical iterative step.

3. Results:

3.1. Binding between Ace-(Ala)$_5$-NMe peptides by VAUS:

We have used this improved version of VAUS procedure to simulate binding between two Ace-(Ala)$_5$-NMe peptides. The simulation involves a box with two peptide chains in explicit water. The system is equilibrated by usual NPT and NVT simulations. The use of grid-based initialization is not used in VAUS simulation. To compare the effectiveness of grid-based initialization in another iterative simulation of VAUS, 221 initial conformations were used which sample exhaustively binding and unbinding poses. The iterative simulations were rerun many times in order to improve our method. Namely, Markov approximated probability distribution calculation and automatic iterative fitting method need to be tested properly. For this, a large GPU-cluster machine was required. We have used Osaka University Cybermedia center GPU computing facility for such an advanced method development. We tested our method to compute potential of mean force along the reaction coordinate for the above Ala-pentapeptide binding (Fig. 3).

In our simulation designing a collective reaction coordinate is of prime importance. The current reaction coordinate properly envisions a binding process, because it is based on inter-peptide distance. The reaction coordinate is the distance between geometric centers of the sets of atoms from peptide chains. However, one must be careful in selecting atoms in peptide chains. After a number of trial-and-error we use Cα and carbonyl O-atom of each Ala-residue to define set of atoms in a peptide chain.

3.2. Active site remodeling during β-sheet augmentation reaction:

To test our method for binding of a larger biomolecular

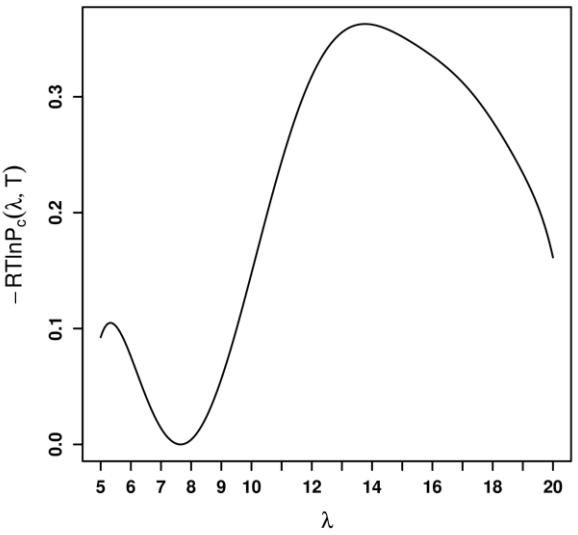

Fig. 3. Final PMF as obtained from Ala-pentapeptide dimerization study.

complex, we are currently simulating a β-sheet augmentation reaction occurring in a protein-peptide binding. The protein contains PDZ domain (87 amino-acid residues) that binds a small 5-residue peptide from C-terminal of GluR2 receptor protein. The interaction interface consists of β-sheet formed by two β-strands from PDZ domain and one β-strand from 5-residue peptide. We are interested in how the interactions remodel itself and forms a stable complex. The VAUS simulation is used with the reaction coordinate defining inter-molecular distance between two interfaces. The C-peptide fragment is allowed to diffuse out from the interface up to 10 Å. The whole system with water molecules and ions is composed of 18851 atoms, which is large compared to other systems treated by AUS or multicanonical molecular dynamics simulation.

The simulation is currently under operation. However, thanks to the computing facility of Osaka University Cybermedia center, we could parameterize initial simulation details and protocol of our method.

5. Conclusion

We developed a new method to simulate binding event between two molecules. Presently we performed Ala-pentapeptide dimerization to demonstrate capability of our method. We are performing the binding simulation in much larger biomolecular system. The ongoing simulation will provide valuable insight about binding mechanism and putative conformations related to loosely bound states for PDZ domain and C-terminal fragment from GluR2 receptor.

6. Acknowledgement

The author thanks Prof. Haruki Nakamura, Dr. Junichi Higo, and Mr. Gert-Jan Bekker in Osaka University and Dr. Narutoshi Kamiya in RIKEN for fruitful discussion. The author thanks Cyber Media Center, Osaka University, for using a VCC computer system as a HPCI project (hp150146).

References

[1] M. Mezei, Adaptive umbrella sampling: Self-consistent determination of the non-Boltzmann bias, J. Comput. Phys. 68 (1987) 237–248.

[2] J. Higo, J. Ikebe, N. Kamiya, H. Nakamura, Enhanced and effective conformational sampling of protein molecular systems for their free energy landscapes., Biophys. Rev. 4 (2012) 27–44.

[3] J. Higo, B. Dasgupta, T. Mashimo, K. Kasahara, Y. Fukunishi, H. Nakamura, Virtual-system-coupled adaptive umbrella sampling to compute free-energy landscape for flexible molecular docking., J. Comput. Chem. 36 (2015) 1489–1501.