論文の内容の要旨

Development of molecular simulation methods for efficient conformational sampling of peptides and proteins
(ペプチドやタンパク質の効率的立体構造サンプリングを目指した分子シミュレーション法開発)

山守　優

Biological processes are often associated with significant conformational changes of biomolecules, which are relevant to their functions. The free energy landscape (FEL) along appropriate reaction coordinates provides us essential information to characterize the mechanism of conformational changes and functions. In this sense, efficient and accurate sampling of the conformational space to calculate the FEL is a major topic for Molecular Dynamics (MD). However, sampling of large conformational space within a limited simulation time is still a challenging problem, because of complexity of FEL and the large gap between simulation time step and time scale of biological process. In this thesis, I proposed three methods to tackle this problem. Firstly, the direct method of enlarging the time step of MD in order to make the time gap of simulation and phenomena close is proposed. Secondly, a new multi-scale efficient conformational sampling method, MuSTAR MD (Multi-scale Sampling using Temperature Accelerated and Replica-exchange MD) is proposed. Finally, an analysis method for FEL is proposed.

The common goal is to calculate and analyze the FEL efficiently. Let $\boldsymbol{\theta}(\boldsymbol{x})$ be Collective Variables (CVs) defined as functions of $\boldsymbol{x}$, where $\boldsymbol{x}$ is the Cartesian coordinate of atom. Supposing that $\boldsymbol{z}$ is a particular realization of these CVs, the FEL $F(\boldsymbol{z})$ is defined as:

$$F(\boldsymbol{z}) = -\frac{1}{\beta}\ln\left\{Q^{-1}\int e^{-\beta V(\boldsymbol{x})}\prod_{l=1}^{L}\delta(\theta_l(\boldsymbol{x}) - z_l)\,d\boldsymbol{x}\right\}, \qquad (1)$$

where $Q = \int e^{-\beta V(\boldsymbol{x})}d\boldsymbol{x}$, L is the number of CVs, $V(\boldsymbol{x})$ is the potential energy, δ is Dirac's delta function, and $\beta = 1/k_B T$, where k_B is the Boltzmann constant and T is the temperature. The ideal analysis of FEL is to compute the Minimum Free Energy Path (MFEP) that is defined as the maximum likelihood pathway connecting two stable states. Denote γ is a pathway on $F(\boldsymbol{z})$. The MFEP satisfies the condition,

$$\nabla F(\boldsymbol{z})^{\perp}(\gamma) = 0, \qquad (2)$$

where $\perp$ indicates the perpendicular component of $F(\boldsymbol{z})$.

1) Enlargement of time step using Internal Coordinate MD

The overall fold of protein is mainly determined by torsion angles. The time scale torsion angle fluctuations are significantly slower than other internal motions. MD simulation in torsion angle space can enlarge the time step of integration to solve equation of motion without loss of significant information; however, directly solving of the equation in torsion angle space requires the calculation of inverse matrix of $n \times n$ mass matrix where n indicates the number of degree of freedom, and it costs calculation of $O(n^3)$. In robotics research, Articulated-Body Algorithm was proposed to reduce the cost of computation of inverse matrix significantly from $O(n^3)$ to $O(n)$. In this research, ABA was further extended to treat non-fixed end systems. The performance of this method, freed-end ABA, was demonstrated for twenty dipeptides and two proteins, villin head-piece subdomain and ubuqutin. The results showed that time step can be elongated up to 6 fs, if the modified force field is used.

2) MuSTAR MD

MuSTAR MD is an extension of temperature accelerated MD (TAMD)[1] and can also be considered as a variation of replica-exchange MD (REMD)[2]. In the MuSTAR MD, each replica contains an all-atom model, at least one coarse-grained model, and a CVs that interacts with the other models through coupling terms. The coarse-grained model is introduced to drive efficient sampling of large conformational space and the all-atom model can serve to conduct accurate conformational sampling. Equation of motion for MuSTAR MD consists of those for those of Cartesian space and CV space,

$$\begin{aligned} m_i^\alpha \ddot{\boldsymbol{x}}_i^\alpha &= -\frac{\partial V^\alpha(\boldsymbol{x}^\alpha)}{\partial \boldsymbol{x}_i^\alpha} - \frac{\partial W^\alpha(\boldsymbol{x}^\alpha, \boldsymbol{z})}{\partial \boldsymbol{x}_i^\alpha} + (\mathrm{thr}, \beta^\alpha) \\ &= -\frac{\partial U^\alpha(\boldsymbol{x}^\alpha, \boldsymbol{z})}{\partial \boldsymbol{x}_i^\alpha} + (\mathrm{thr}, \beta^\alpha), \end{aligned} \tag{3}$$

$$m^z \ddot{z}_l = -\sum_\alpha^N \frac{\partial W^\alpha(\boldsymbol{x}^\alpha, \boldsymbol{z})}{\partial z_l} + (thr, \beta^z). \tag{4}$$

where $\boldsymbol{x}_i^\alpha$ and $V^\alpha(\boldsymbol{x}^\alpha)$ are the position vector in Cartesian coordinate for the i-th atom of model α and potential energy functions in each model, respectively, and $W^\alpha(\boldsymbol{x}^\alpha, \boldsymbol{z})$ is a coupling terms between model α and the CV system. α can represent either fine-grained or coarse-grained model. The thermostat parameters are related to temperatures as $\beta^\alpha = 1/k_B T^\alpha$ and $\beta^z = 1/k_B T^z$. The coupling terms are defined as:

$$W^\alpha(\boldsymbol{x}^\alpha, \boldsymbol{z}) = \frac{K^\alpha}{2} \sum_{l=1}^{L} (\theta_l(\boldsymbol{x}^\alpha) - z_l)^2, \tag{5}$$

$$U^\alpha(\boldsymbol{x}^\alpha, \boldsymbol{z}) = V^\alpha(\boldsymbol{x}^\alpha) + W^\alpha(\boldsymbol{x}^\alpha, \boldsymbol{z}), \tag{6}$$

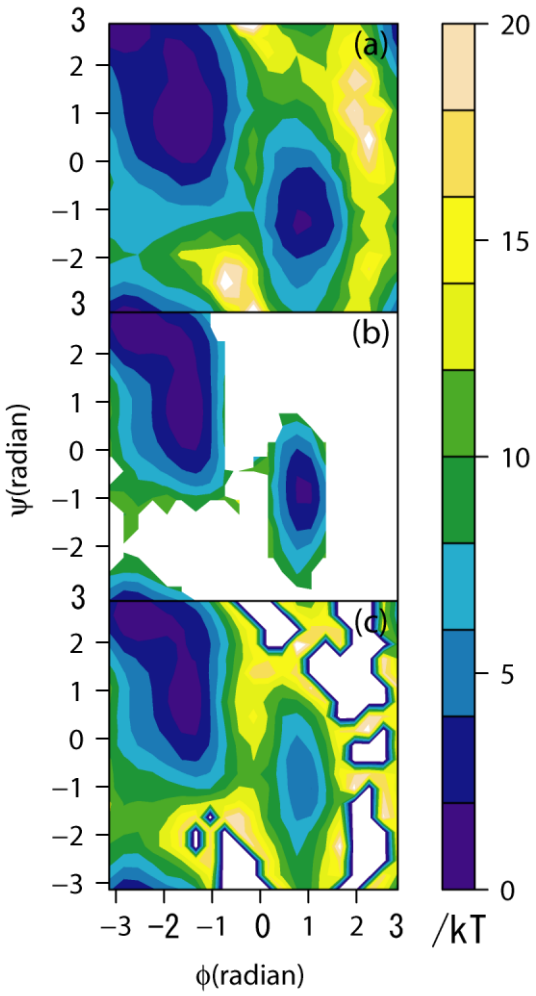

Figure 1. FEL of Ala-dipeptide in vacuum using the AMBER force field parm99SB with respect to backbone dihedral angles ϕ and ψ. The results from (a) MuSTAR MD (10 ns×8 replicas), (b) REMD (10 ns×8 replicas), (c) 80 ns TAMD .

where K^{α} is the coupling-strength between model α and the CV system. The parameters are exchanged between neighboring replicas in some interval obeying the Metropolis method.

I applied MuSTAR MD for typical test cases. Fig.1 (a)-(c) is FEL at 300 K calculated from the results of MuSTAR MD, REMD, and TAMD, respectively. The sampled conformational space is broader in the order of MuSTAR MD>TAMD> REMD. Comparison the result of MuSTAR MD with that of umbrella sampling shows the accuracy of MuSTAR MD.

3) Path Search Method

The path search method proposed is designed to determine the MFEP on a given CVs space after the FEL calculation. Because analytical derivative of the FEL is useful for finding of the MFEP, the FEL is approximated using combination of Gaussian distributions with the expectation-maximization algorithm, as follows.

$$F_{app}(\{\boldsymbol{z}\}) = -\frac{1}{\beta}\log \rho_{app}(\{\boldsymbol{z}\}) = -\frac{1}{\beta}\log \sum_{k=1}^{K} \pi_{\mathrm{MLH}} N(\boldsymbol{z}|\boldsymbol{\mu}_{\mathrm{MLH}}\boldsymbol{\Sigma}_{\mathrm{MLH}}), \qquad (7)$$

where K is the number of Gaussian distribution, π_k is the fraction of the k-th Gaussian distribution. $N(\boldsymbol{z}|\boldsymbol{\mu}_k\boldsymbol{\Sigma}_k)$ is a L-dimensional Gaussian distribution whose average and variance are $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$. The method originally proposed to calculate the path on potential energy surface, zero-temperature string method[3], was applied to calculation of MFEP in this thesis.

The method was applied to the transition of Ala-dipeptide between two stable states. 19 Gaussians were used. The convergence is confirmed by Kullbacl-Leiblear divergence

between approximated and original FEL. The FEL cross sectio along the MFEP (in Fig. 2 (a)) suggested the transition state is the corresponding structure at (ϕ, ψ)=(*-1.09˚,-57.9˚*) (Fig. 2(b)). Committer test supported that the structure is corresponding to transition state.

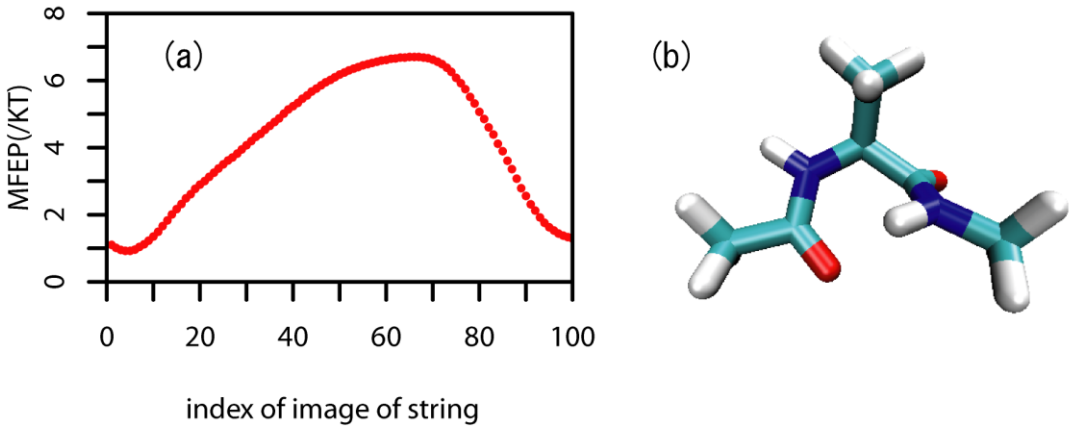

Figure 2. (a) The cross section of the FEL along the minimum free energy path calculated by zero-temperature string method, the x-axis indicates the index of the point on the string, and (b) corresponding structure of Ala-dipeptide o at the energy barrier (index 67, and dihedral angles are (ϕ, ψ)=(-1.09˚,-57.9˚)) .

Conclusions and Perspectives

I proposed three methods to calculate and analyze the FEL. Two of them are the methods to calculate FEL by make simulation efficient, i.e. enlargement of time step in torsion angle MD, and MuSTAR MD. The other is the method to analyze the calculated FEL by MFEP search. The Freed-end ABA program developed in this work can enlarge the time step up to 6 fs. MuSTAR MD shows the high performance in sampling efficiency and accuracy compared to established enhanced sampling methods. The advantageous features of MuSTAR MD are: 1) High temperature of the CV system enhances conformational sampling. 2)Multiple coarse-grained models can be introduced to guide the system to move to multiple structures. 3) Applications to larger system are expected to be relatively easy. It was demonstrated that the proposed MFEP searching method can detect the maximum-likelihood path and the transition state. The proposed methods are expected to be applied to the further computational studies of biological phenomena including large conformational change, such as allosteric transition.

[1] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**. 141 (1999).
[2] L. Maragliano and E. Vanden-Eijnden, Chem. Phys. Lett. **426**. 168 (2006)
[3] W. Ren, E. Vanden-Eijnden, P. Maragakis, W.N. E, Journal of Chemical Physics 123 (2005).