Efficient Umbrella Sampling for Biomolecular System

Shun Sakuraba

Student ID: 47-56907
Department of Compuational Biology
Graduate School of Frontier Sciences
the University of Tokyo

Supervisor: Akio Kitao

Abstract

A variation of umbrella sampling is proposed. I have developed an efficient umbrella sampling method based on the concept of Jumping-Among-Minima model and quasi harmonic approximation. I demonstrated that the new method is capable of finding both the meta-stable state structure and the transition state structure of 5-residue peptide Met-enkephalin. I also showed that the same method can search the transition state structure of 160-residue protein T4 lysozyme.

0 Preface

> "Computers are incredibly fast, accurate, and stupid; humans are incredibly slow, inaccurate and brilliant; together they are powerful beyond imagination." – Albert Einstein

Although computers have changed, science has evolved, and so many things are different from the era he lived, I think this quote still holds truth. I am dreaming, working, and studying, for the day that I can see, or hopefully I can open by myself, "beyond imagination."

I would like to thank my supervisor, Prof. Dr. Akio Kitao. He had led my research through many insightful discussions, even before I have a clear goal in my mind, and provided me some critical ideas necessary for completing this work. I would also like to thank Prof. Dr. Yasumasa Joti, for daily discussion, conversation, and education. He gave me plenty of valuable comments, including critic ones, to push my research toward the goal. I thank all Sosei lab members: Dr. Furuta, Dr. Ramaswamy, Dr. Löfller, Dr. Takemura, Dr. Yang, Mr. Chng, Mr. Nishima, Mr. Arima, Mr. Harada, Ms. Joti, Ms. Eitai and Ms. Uemura for giving me confortable and joyful life in the laboratory. Additionally, I would also like to thank friends of mine, Yu Sugawara, Takahiro Horiuchi, Atsushi Seki and Hidehiko Abe, for valueable discussions.

As a final statement, I would like to thank my both parents, who financially supported me in all these two years.

1 Introduction

Molecular dynamics (MD) simulation has become a powerful tool for analyzing macromolecules such as proteins, in cooperation with various experimental methods. Recent development of MD simulation techniques, combined with faster computers, opens up a detailed observation of macromolecules. Although the simulation is an established technique to investigate dynamic structures, it can hardly find non-stable states such as transition states during a limited simulation time. The gap between the time scale of simulations and that of functionally important events still exists even now. This is the sampling problem in protein dynamics.

A possible way to overcome is disposing time-dependent information. First we carry out a simulation, using a modified Hamiltonian to obtain modified probability distribution. After that, the original probability distribution is reproduced by a certain procedure. Modified Hamiltonian is designed to sample a larger configurational space within a limited simulation time. Two ideas are successful in obtaining the modified Hamiltonian: one is the *iterative* procedure, and the other is the *parallel* procedure.

In the former, we optimize an additional energy term iteratively. In multicanonical [1] method, for example, the additional energy term is optimized until a flat probability distribution as a function of energy is obtained, while in adaptive umbrella sampling [2] as a function of coordinate space. These methods can be applied without *a priori* knowledge on the target molecules, but they require much computation time, and are difficult to parallelize. It should also be noted that they are reported to have problems with complex macromolecules [3].

In the latter idea we first run simulations with different Hamiltonian in parallel. After that, we perform a reweighting calculation to reproduce distribution. This idea is commonly used in umbrella sampling [4, 5]. It requires, however, *a priori* knowledge on the system such as reaction coordinates, thus it is not straightforward to apply them to the molecules whose behavior is not well investigated.

Is it possible to combine the advantages of both ideas? One such successful example is replica exchange multicanonical MD method [6]. It combines both two ideas, replica exchange method and multicanonical method, to generate good ensemble over phase space. This method first optimize the modified Hamiltonian by iterative procedure, then applies parallel procedure for detailed analysis. Although this method still requires rather long simulation time for setting up, results on small molecules are quite promising. One lesson from this method is that even if our knowledge on the target system is imperfect, multiple simulations can minimize the error, and realize the exploration of a large phase space. This implies that we only need approximate information on the target system.

In this thesis, I present a method to exploit both two, as a natural extension to chemical flooding [7] combined with JAM model [8]. This method exploits approximate information on protin dynamics based on currently existing structures. By the suggested method, we are now able to sample larger phase space, with only a short preparation time, without *a priori* knowledge on reaction coordinates.

2 Methods

In this section, first I review the umbrella sampling method. After that, I will describe Chemical flooding and JAM flooding method, which is based umbrella sampling method. To consult the basics of umbrella sampling and molecular dynamics, please see Appendix B.

2.1 Umbrella sampling

The detail of umbrella sampling is described everywhere [5]. I briefly review umbrella sampling method below.

If the system obeys canonical ensemble, the density probability is given as:

$$\rho(\mathbf{\Gamma}) = \frac{\exp(-\beta V(\mathbf{\Gamma}))}{\int \mathrm{d}\mathbf{\Gamma}' \exp(-\beta V(\mathbf{\Gamma}'))}, \tag{1}$$

where, $\beta = 1/k_B T$, and $\mathbf{\Gamma}$, $V(\mathbf{\Gamma})$, k_B, and T being generalied coordinate, potential function, Boltzmann constant, and temperature, respectively. In umbrella sampling method, the original poteintial $V(\mathbf{\Gamma})$ is changed to the modified potential $V'(\mathbf{\Gamma})$, to generate a weighted sampling. Changes to the potential function is described as,

$$V'(\mathbf{\Gamma}) = V(\mathbf{\Gamma}) + V_{\mathrm{umb}}(\mathbf{\Gamma}). \tag{2}$$

The additional term V_{umb} is called umbrella potential. Performing simulations with the umbrella potential, we can obtain new probability density ρ', with the following relation to ρ:

$$\rho(\mathbf{\Gamma}) = \frac{\rho'(\mathbf{\Gamma})\exp(\beta V_{\mathrm{umb}}(\mathbf{\Gamma}))}{\int \mathrm{d}\mathbf{\Gamma}'\rho'(\mathbf{\Gamma}')\exp(\beta V_{\mathrm{umb}}(\mathbf{\Gamma}'))} \tag{3}$$

Using this relation, we can reconstruct ρ from ρ', which can be obtained as the result of simulation on the modified potential. To obtain an accurate probability density map ρ, we have to set proper V_{umb} so that we can sample ρ' sufficiently. This requires *a priori* knowledge on simulated system, thus it was one of the largest problem in applying umbrella sampling to large molecules.

2.2 JAM flooding

I propose Jumping-Among-Minima flooding (JAM flooding), a natural specialization to multiple potential chemical flooding method [7, 9]. This method depends on the concept of principal component analysis (PCA) [10, 11] and Jumping-Among-Minima (JAM) model [8]. Here is their brief review.

Given a set of structure $\{\mathbf{\Gamma}\}$, variance-covariance matrix A can be calculated as follows:

$$A = \{a_{ij}\} = \left\langle Q_i - \bar{Q}_i \right\rangle \left\langle Q_j - \bar{Q}_j \right\rangle. \tag{4}$$

$$Q_i = m_i^{1/2}\Gamma_i. \tag{5}$$

In this equation, m_i is the mass of atom that correspond to i-th coordinate, $\boldsymbol{Q}$ being square root mass weighted atomic coordinate, $\{\boldsymbol{Q}\}$ being a set of ones, and $\bar{\boldsymbol{Q}}$ is the averaged value of $\{\boldsymbol{Q}\}$. Diagonalizing A, we can obtain k-th eigenvector $\boldsymbol{X}_k$, that is called *principal component axis* (PC axis),

$$A\boldsymbol{X}_k = \lambda_k \boldsymbol{X}_k, \tag{6}$$

$$\boldsymbol{X}_k^T \boldsymbol{X}_k = 1. \tag{7}$$

$\boldsymbol{X}_k$ denotes the direction along which trajectories have strong correlation. Using $\boldsymbol{X}_k$ as new basis, we can introduce the linear combinations of atomic coordinates, *PC coordinate* $\boldsymbol{\sigma}$:

$$\sigma_k = \boldsymbol{X}_k^T \boldsymbol{Q}. \tag{8}$$

Jumping-Among-Minima model denotes that protein energy landscape is mostly harmonic or quasi-harmonic if projected on PC coordinate space, and only a few coordinates have hierarchical multiple minima. Based on this model, I approximated potential function to be harmonic around each local minimum structure.

$$V(\mathbf{\Gamma}) \simeq \sum_m^{\text{minima}} \Delta\boldsymbol{\sigma}_m^T S_m^{-1} \Delta\boldsymbol{\sigma}_m. \tag{9}$$

$$\Delta\boldsymbol{\sigma}_m = \boldsymbol{\sigma} - \bar{\boldsymbol{\sigma}}_m. \tag{10}$$

Here $\bar{\boldsymbol{\sigma}}_m$ is the coordinate of m-th minimum structure, S_m is the variance-covariance matrix of trajectory around each local minimum structure. Additionally, I assumed potential function to have no strong correlation between different axis.

$$S_m = \{s_{ab}^m\} = \delta_{ab} s_a^m. \tag{11}$$

The desired shape of umbrella potential should "flood" potential well of each minimum. That is, umbrella function should also have harmonic property around some minimum structure. For practical reason, I propose to employ Gaussian function for flooding out from stable structures,

$$V_{\mathrm{umb}}(\mathbf{\Gamma}) = \sum_m^{\text{minima}} E_m^{\mathrm{eff}} \exp\left(-\Delta\boldsymbol{\sigma}_m^T K_m^{-1} \Delta\boldsymbol{\sigma}_m\right). \tag{12}$$

$$K_m = \{k_{ab}^m\} = \delta_{ab} k_a^m. \tag{13}$$

Here E_m^{eff} is an adjustable parameter that determines the height of umbrella function, and K is the diagnoal matrix that determines fluctuation. Comparing V and V_{umb} upto the second order, E_m^{eff} and K_m are required to have the following relation,

$$E_m^{\mathrm{eff}} K_m^{-1} = S_m^{-1}. \tag{14}$$

Thus, only E_m^{eff} should be determined to run this method. This value should be determined empirically, to reflect the real simulation time scale.

To perform MD simulation with this potential, forces to all atoms should also be calculated. Further details are described in Appendix A.

ρ' obtained with umbrella potential can be reweighted to original probability density, using eq. (3) for single trajectory. Free energy f can be obtained by,

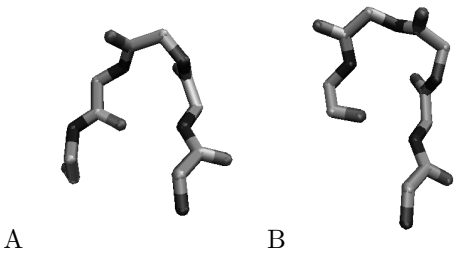

Figure 1: Reference structure A (left) and reference structure B (right). The former corresponds to the structure **2**, and the latter to the structure **1** of [13], respectively. This figure is generated by VMD [14].

$$f = -\frac{1}{\beta} \ln \rho + \text{const.} \tag{15}$$

For multiple ones, WHAM method [12] is applicable.

3 Results and discussion

3.1 Met-enkephalin test

Preliminary simulations For the test of new method, I choose Met-enkephalin, 5-residue peptide (sequenced YGGFM) as a model case. Previous results on this molecule is discussed in [13]. I modeled Met-enkephalin as an all-atom model in vacuum, where acetyl group and N-methyl group were added at the N- and C- terminus, respectively.

Two reference structures, denoted in figure 1, were chosen from local minima structures, as represented in [13]. All simulations were performed with molecular dynamics package NAMD [15], using CHARMM22 force field [16]. Two structures are modeled with NAMD to reconstruct the hydrogen bonding pattern described in the literature. They were then equilibrated at 300K, with standard molecular dynamics procedure. Langevin heat bath was used to establish NVT ensemble. First I performed two preliminary simulation on Met-enkephalin's known two stable structure (structure A and B in figure 1), 5 ns each, plus 5 ns equilibration time. PCA was performed on these two trajectories with an equal weighting to each set of coordinates. Only the coordinates of 23 main chain atoms were used for this analysis.

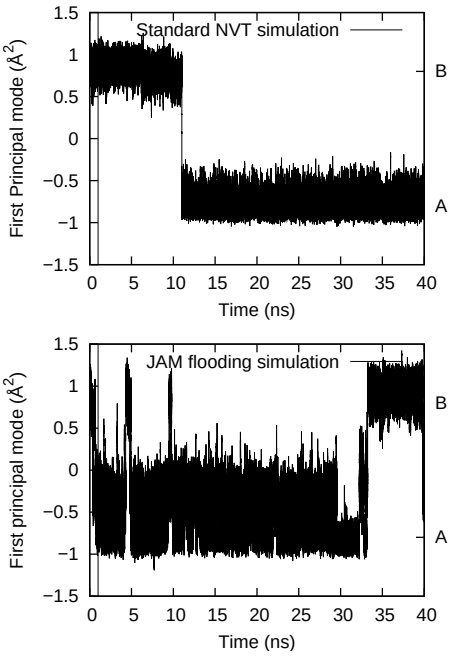

Figure 2: Typical trajectory of 40 ns simulation on M-enkephalin, projected on first and second PC modes. In the reweighting procedure, I dropped first 1 ns trajectoy for eqilibration. (top) Trajectory from canonical ensemble. Trajectory only switch once between two stable states. (bottom) Trajectory from JAM flooding. With JAM flooding, the system achieves frequent switching between two stable states.

JAM flooding simulation After obtaining PC vector, 30 simulations of 10-ns run were performed, with JAM flooding potential enabled. I applied JAM flooding on this molecule with the number of minima equals to 2 in eq. (12). Flooding heights E_m^{eff} are set to 3.0, 4.5, 6.0, 7.5, 9.0, 10.5 kcal/mol for all m. 5 distinct simulations were performed with each parameter. First 1-ns trajectory of all simulation was dropped as equilibration time. For the comparison of trajectory, I also have performed 40-ns simulations on both canonical ensemble and JAM flooding condition. Typical trajectory of JAM flooding is shown in figure 2. While standard canonical sampling could not accomplish switching between two minimum states, JAM flooding simulation demonstrates a good sampling over PC coordinate.

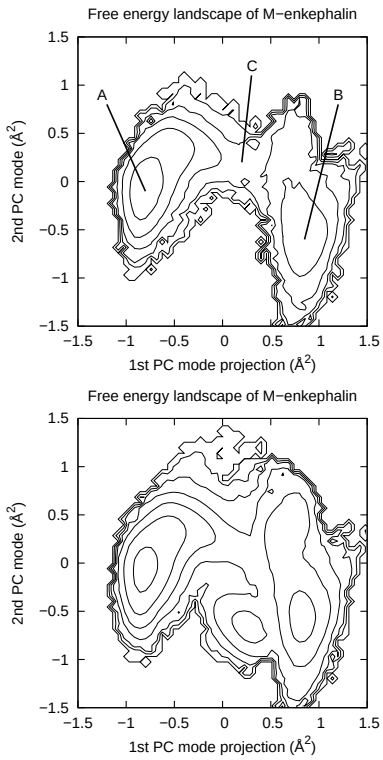

Figure 3: Free energy maps, reconstructed from JAM flooding simulation runs. Each contour represents 2 kcal/mol difference on free energy. Two axis represents first and second PC axis. (top) Map was reconstructed from 6 trajectories. A and B is reference conformation, where C is transition state found by this algorithm. (bottom) Map was reconstructed from all 30 trajectories.

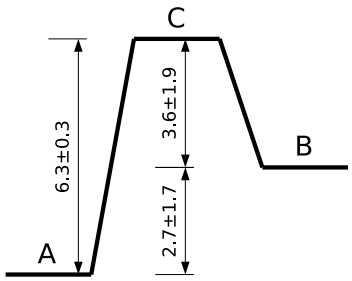

Figure 4: Free energy difference of M-enkephalin's stable structures and a transition structure, in figure 3. The unit of each value is kcal/mol. Values shown are the averaged values and standard deviations of 5 distinct set of simulations.

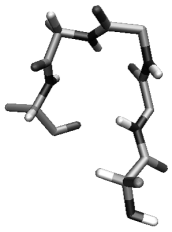

Figure 5: Presumed transition state structure, reconstructed from C of figure 3. This figure is generated by VMD [14].

Reweighting and free energy analysis Free energy map was reconstructed from the trajectories of JAM flooding, according to eq. (3) and WHAM analysis [12]. Figure 3 shows the diagram, representing free energy map of M-enkephalin reconstructed from simulation result. From the short simulation result, we can clearly observe both two stable minima structure. Additionally, we can also clearly see transition structure coordinate on the map. Energy difference between three states are calculated and shown in figure 4. The result demonstrates the capability of obtaining accurate energy on small molecules.

Presumed transition state structure of free energy map in figure 3 was reconstructed. Figure 5 shows the resulted structure, which has a hydrogen bonding between CO of Gly-2 and NH of Phe-4. The structure and hydrogen bonding pattern showed an excellent agreement with previously proposed transition state conformation in [13].

From these analysis, it is shown that the new method is capable to sample multiple minima of small molecules in a reasonable time. It also shows that the method can be used to estimate the energy difference between stable state and transition structure.

3.2 Application to macromolecules: T4 lysozyme

Preliminary simulations As the example of application to macromolecule, I have used JAM flooding method to T4 lysozyme, 18 kDa hydrolysis enzyme, which is known to make large conformational change with only one sequence. T4 lysozyme structures were taken from PDB-ID 150L (open) and

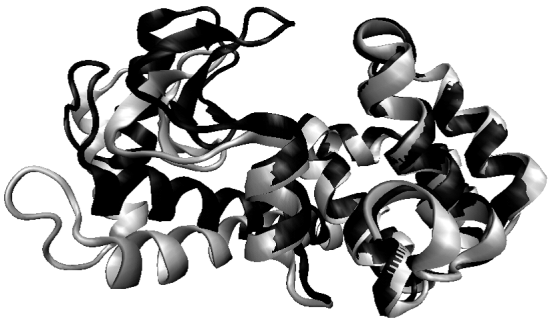

Figure 6: Reference structure of two T4 lysozyme structure. (light gray) "Open" structure, taken from PDB-ID 150L. (black) "Closed" structure from 2LZM.

2LZM (close), respectively. Figure 6 shows two reference structure used for this simulation. Missing terminal loop was constructed by MODELLER [17]. Four residues from C terminus were dropped off, thus 160 amino acid residues were used for the simulation. Models were solvated by 10 Å TIP3P water box and equilibrated on 1 atom, 300K condition, with a standard molecular dynamics procedure. Langevin piston algorithm was used for achieving constant pressure controls. CHARMM22 force field [16] was used for all simulation. All simulations were performed on molecular dynamics package NAMD [15], with integration time step set to 2 fs, along with a SHAKE algorithm.

After the 10-ns unconstrained equilibration run, two 5-ns MD simulations were performed for each starting structure. Snapshots were taken and saved for every 1 ps of simulations. PCA was performed on trajectory from both 150L and 2LZM with equal weighting factor, where only main-chain heavy atoms were used in all three analysis.

JAM flooding simulation I applied JAM flooding for this system, with the number of minima set to 2 on eq. (12). Four PC axes were chosen for JAM flooding so that they cover more than 80 % of the coordinate space. Intensity of flooding umbrella (E_m^{eff} in eq. (12)) was chosen from 4.0-8.0, 5.0-7.0, 6.0-6.0, 7.0-5.0, 8.0-4.0, 9.0-3.0, 10.0-2.0 kcal/mol, where the former number is the energy on the "closed" structure A, and the latter is on "open" structure B. Two simulations were performed on each 7 conditions, 14 simulations in total. Each trajectory was taken for 5 ns, after 1-ns equilibration time.

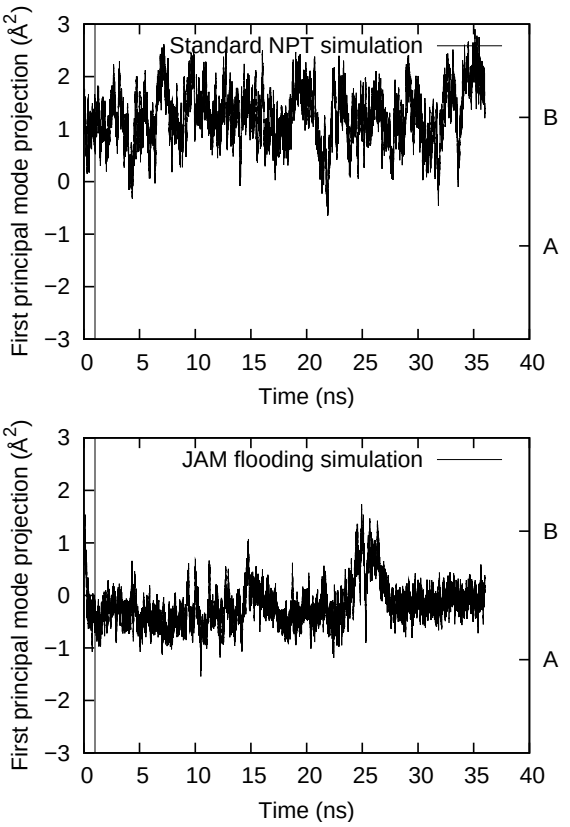

Figure 7: Typical trajectory of (top) canonical sampling simulation and (bottom) JAM flooding sampling simulation. Trajectory was taken for 36 ns, and first 1 ns was dropped from analysis as an equilibration time.

Figure 7 shows typical time course of trajectory projected on PC axis. Canonical simulation fluctuate around current known minima, while JAM flooding sweep the region between two minima. Although the latter even visits both two states, structural switch to another state is not sufficiently frequent in one simulation. This would be severe problem when we investigate the free energy difference. One possible reason is that the Gaussian flooding potential does not fit well in this case, or inappropriate E_m^{eff} value is used. For the detailed analysis of protein energy landscape, further refinement is necessary.

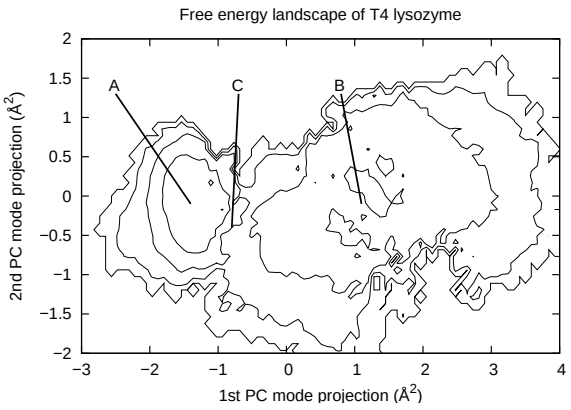

Figure 8: Free energy map obtained from T4 lysozyme simulation with JAM flooding. State A and B represents known stable structures, while state C denotes presumed transition state structure.

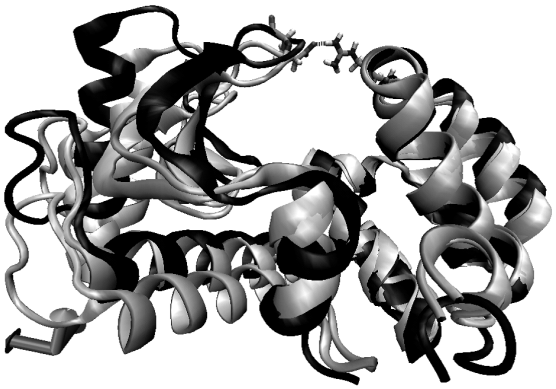

Figure 9: Transition state structure obtained on T4 lysozyme. (black) Presumed transition structure, corresponding state C in figure 8. (light gray) Open and closed reference structure.

Reweighting and free energy analysis Figure 8 illustrates a free energy map of T4 lysozyme obtained by reweighting calculation [12]. From free energy map we can see two known stable states, along with presumed transition state structure. Reconstructed transition state structure is represented as in figure 9. It has the hydrogen bonding between Glu-22 to Arg-137, which connects the two "heads" of T4 lysozyme, but does not have tight hydrogen-bonding at Arg-52, Thr-59, Glu-62, which locks the hinge angle. As both hydrogen bonds exist in "closed" structure and does not exist in "open" structure, it is reasonable to conclude that this presumed structure represents proper intermediate structure between these two, therefore enforcing the validity of this method. I propose that in the transition from "close" to "open" structure, first the three hydrogen bonds are cleaved, then hinge angle flucuates, and finally hydrogen bond that locks two heads of lysozyme will be cleaved. I note here, however, that this molecule need an extra caution on analysis, as it is liable to be artifact from the reasons mentioned above.

4 Conclusion

I modified multi-minima chemical flooding to adapt JAM model, yielding JAM flooding method, which gives probability density of protein structure with reasonably good approximation. The method could be straightforwardly applied to small molecules, showing equivalent convergence compared to the previously reported methods. The method was also capable of drawing energy landscape and finding transition structure, in an application to macromolecules. Further refinement of the process and the validation is now on the progress.

5 Acknowledgment

Part of the computations were performed using Research Center for Computational Science, Okazaki, Japan.

Appendix

A Derivation of force

In the present research, the umbrella potential is defined as,

$$V_{\mathrm{umb}}(\mathbf{\Gamma}) = \sum_{m}^{\mathrm{minima}} E_m^{\mathrm{eff}} \exp(-\Delta\boldsymbol{\sigma}_m^T K_m^{-1} \Delta\boldsymbol{\sigma}_m) \quad (16)$$

$$\equiv \sum_{m}^{\mathrm{minima}} w\left(\Delta\boldsymbol{\sigma}_m(\mathbf{\Gamma})\right). \quad (17)$$

Where $w(\Delta\boldsymbol{\sigma}_m)$, and $\Delta\boldsymbol{\sigma}_m(\mathbf{\Gamma})$ is,

$$w(\Delta\boldsymbol{\sigma}_m) = E_m^{\mathrm{eff}} \exp(-\Delta\boldsymbol{\sigma}_m^T K_m^{-1} \Delta\boldsymbol{\sigma}_m), \quad (18)$$

$$\Delta\boldsymbol{\sigma}_m(\mathbf{\Gamma}) = \boldsymbol{X}_m^T \boldsymbol{Q} - \bar{\boldsymbol{\sigma}}_m = \boldsymbol{X}_m^T M^{1/2} \mathbf{\Gamma}_{\mathrm{bestfit}} - \bar{\boldsymbol{\sigma}}_m. \quad (19)$$

$\mathbf{\Gamma}_{\mathrm{bestfit}}$ is best-fit structure of $\mathbf{\Gamma}$ to reference one, and M is a diagonal matrix that represents the mass of atoms. Force is calculated from this relation. Deriving potential energy by coordinate, we get

$$F_i = -\frac{\partial V_{\mathrm{umb}}}{\partial \Gamma_i} = -\sum_{m}^{\mathrm{minima}} \sum_{j} \frac{\partial w_m}{\partial \Delta\sigma_{mj}} \frac{\partial \Delta\sigma_{mj}}{\partial \Gamma_i} \quad (20)$$

$$= 2 \sum_{m}^{\mathrm{minima}} \sum_{j} E_m^{\mathrm{eff}} k_j^m \Delta\sigma_{mj} \exp(-\Delta\sigma_{mj}^2 k_j^m) \frac{\partial \Delta\sigma_{mj}}{\partial \Gamma_i}. \quad (21)$$

Last derivative term vanishes for all j except when i and j point to the coordinate of the same atom.

$$\frac{\partial \Delta\sigma_{mj}}{\partial \Gamma_i} = \begin{cases} 0 & \text{same atom} \\ X_{mj} m_i^{1/2} \frac{\partial \mathbf{\Gamma}_{\mathrm{bestfit}}}{\partial \Gamma_i} & \text{otherwise.} \end{cases} \quad (22)$$

Here $\frac{\partial \mathbf{\Gamma}_{\mathrm{bestfit}}}{\partial \Gamma_i}$ is best-fit rotated structure differentiated by coordinates. The differentiated fraction is also the function of $\mathbf{\Gamma}$. As it is difficult to obtain this differentiation analytically, I solved this equation numerically.

$$\frac{\partial \mathbf{\Gamma}_{\mathrm{bestfit}}}{\partial \Gamma_i} = \lim_{\Delta\Gamma_i \to 0} \frac{1}{2\Delta\Gamma_i} (\mathbf{\Gamma}_{\mathrm{bestfit}}(\mathbf{\Gamma} + \Delta\Gamma_i \boldsymbol{e}_i) - \mathbf{\Gamma}_{\mathrm{bestfit}}(\mathbf{\Gamma} - \Delta\Gamma_i \boldsymbol{e}_i)). \quad (23)$$

At each step, we have to fit each molecule to reference structure, and also have to calculate perturbed value of best-fit rotation quaternion for each coordinate. Both requires the calculation of eigenvectors [18], thus fast calculation was necessary for speeding this method up. I used inverse power method to calculate eigenvectors quickly, reusing the vector at best-fit calculation. Conjugate gradient method was also used to solve linear equation that arises on the inverse power calculations. For the details of these methods, please see [19, 20].

B Molecular Dynamics Simulation

Molecular dynamics inspects the behaviour of the mechanical system through numerical calculation. If we have Hamiltonian $\mathcal{H}$, initial coordinates and momenta $\boldsymbol{q}(0), \boldsymbol{p}(0)$, we have the following,

$$\dot{\boldsymbol{q}} = \frac{\partial \mathcal{H}}{\partial \boldsymbol{p}}, \tag{24}$$

$$\dot{\boldsymbol{p}} = -\frac{\partial \mathcal{H}}{\partial \boldsymbol{q}}. \tag{25}$$

Solving these differential equations numerically, we can get the discrete time series of coordinates and momemta. We denote this as generalized coordinate on phase space, $\boldsymbol{\Gamma}$ for simplicity.

$$\boldsymbol{\Gamma}(t) = \begin{pmatrix} \boldsymbol{q}(t) \\ \boldsymbol{p}(t) \end{pmatrix}. \tag{26}$$

Probability density $\rho(\boldsymbol{\Gamma})$ are defined as the probability that system is found in the value $\boldsymbol{\Gamma}$.

$$\rho(\boldsymbol{\Gamma}, t) = \int_0^t \mathrm{d}t' \delta(\boldsymbol{\Gamma}(t') - \boldsymbol{\Gamma}). \tag{27}$$

If we can assume the ergodicity on the system, this value converges to one, time-independent probabilty density ρ,

$$\lim_{t\to\infty} \rho(\boldsymbol{\Gamma}, t) = \rho(\boldsymbol{\Gamma}). \tag{28}$$

Depending on the type of system, this ρ converges to different value. This 'type' of the system is called as "ensemble", and represents much of the characteristics of the system. *Canonical distribution* is a kind of ensembles, which can be observed on the constant temperature system with no particle exchange. In canonical ensemble, probability density ρ will converges to:

$$\rho_{\mathrm{canonical}}(\boldsymbol{\Gamma}) = \frac{\exp -\beta V(\boldsymbol{\Gamma})}{\int \mathrm{d}\boldsymbol{\Gamma}' \exp -\beta V(\boldsymbol{\Gamma}')}, \tag{29}$$

where V is the potential energy of the system.

In the simulation, we can never simulate for infinite time nor simulate for enough time to observe all possible configuration of molecules. Thus, we need an efficient algorithm to generate modified density probability distribution, to sample appropreate configurational spaces.

References

[1] U. H. E. Hansmann, Y. Okamoto, and F. Eisenmenger. Molecular dynamics, langevin and hybrid monte carlo simulations in a multicanonical ensemble. *Chem. Phys. Lett.*, 259(3):321–330, 1996.

[2] Christian Bartels and Martin Karplus. Multidimensional adaptive umbrella sampling: Applications to main chain and side chain peptide conformations. *J. Comp. Chem.*, 18:1450, 1997.

[3] Christian Bartels, Michael Schaefer, and Martin Karplus. Determination of equilibrium properties of biomolecular systems using multidimensional adaptive umbrella sampling. *J. Chem. Phys.*, 111(17):8048, 1999.

[4] G. M. Torrie and J. P. Valleau. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chemical Physics Letters*, 28:578–581, October 1974.

[5] D L Beveridge and F M DiCapua. Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.*, 18:431–492, 1989.

[6] Yuji Sugita and Yuko Okamoto. Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. *Chem. Phys. Lett.*, 329:261, 2000.

[7] E. Matthias Müller, A. de Meijere, and H. Grubmüller. Predicting unimolecular chemical reactions: Chemical flooding. *J. Chem. Phys.*, 116:897–905, January 2002.

[8] Akio Kitao, Steven Hayward, and Nobuhiro Go. Energy landscape of a native protein: Jumping-among-minima model. *PROTEINS*, 33:496–517, 1998.

[9] Oliver F. Lange, Lars V. Schafer, and Grübmuller Helmut. Flooding in gromacs: accelerated barrier crossings in molecular dynamics. *J. Comp. Chem.*, 27(14):1693–1702, 2006.

[10] Akio Kitao, Fumio Hirata, and Nobuhiro Go. The effects of solvent on the conformation and the collective motions of protein :normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. *Chem. Phys.*, 158:447–472, 1991.

[11] Angel E. García. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.*, 68(17):2696–2699, Apr 1992.

[12] Shankar Kumar, Djamal Bouzida, Robert H. Swendsen, Peter A. Kollman, and John M. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *J. Comp. Chem.*, 13(8):1011–1021, 1992.

[13] Satoru G. Itoh and Yuko Okamoto. Multi-overlap molecular dynamics methods for biomolecular systems. *Chem. Phys. Lett.*, 400:308–313, 2004.

[14] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[15] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with namd. *J. Comp. Chem.*, 26:1781–1802, 2005.

[16] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102(18):3586–3616, 1998.

[17] M.A. Marti-Renom, A. Stuart, A. Fiser, R. Sanchez, F. Melo, and A. Sali. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.*, 29:291–325, 2000.

[18] Berthold K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *J. Opt. Soc. Am. A*, 4(4):629–642, 1987.

[19] Anthony Ralston and Philip Rabinowitz. *A first course in numerical analysis*, chapter 9 and 10. Dover, New York, 2001.

[20] Richard Barrett et al. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, chapter 2. SIAM, Philadelphia, USA, 1994.