Protein Folding Simulation and Self-Consistent Potential Functions

Fumi Takazawa [1]
takazawa@kuicr.kyoto-u.ac.jp

Masaki Sasai [2]
sasai@info.human.nagoya-u.ac.jp

Minoru Kanehisa [1]
kanehisa@kuicr.kyoto-u.ac.jp

[1] Institute for Chemical Research Kyoto University, Uji, Kyoto 611-0011, Japan
[2] Graduate School of Human Informatics, Nagoya University, Nagoya 464-8601, Japan

1 Introduction

Protein folding is one of the most fundamental problems in biophysics. Molecular dynamics (MD) simulations aim to provide the most detailed description of the behavior of atomic models of proteins during their folding processes. However, current MD simulations, even with the most powerful computational techniques, can barely explore more than a few nanoseconds for an all-atom model of a protein and its solvent environment. On the other hand, the folding experiment takes place not in the range of hundreds of microseconds and but usually in milliseconds, seconds or minutes. In order to simulate the events in longer time ranges, potential functions are required to be simplified to represent molecular structures in less detail. Statistical potentials extracted from a dataset of known structures in native states are useful for the prediction, such as by 3D-1D (threading) methods. In 3D-1D prediction, the correct structure is recognized as a most stable state among template structures. However, such statistical potentials cannot always simulate successful protein folding. One of the reasons is that the statistical potentials are extracted from the various protein structures in native states, lacking the information of those in unfolded or misfolded states. Potential functions for folding simulation should have the ability to evaluate energy of each state that appears during folding simulation. In this paper, we present self-consistently adjusted potentials that are obtained by the following procedure. We start from the statistical potentials obtained from known structures in native states, generate a series of structures by MD simulations of protein folding, and re-adjust the potentials to better represent both native and non-native structures.

2 Methods

2.1 Pairwise potential

We express the energy of the structure as a sum of pairwise potentials. When the residues p and q are found at positions i and $i+k$ in structure μ of the library, a Gaussian function whose center is at $r = r^{pq,\mu}_{i,i+k}$ is summed into the potential $V^{pq}_{k(l)}$:

$$V^{pq}_{k(l)} = -\frac{1}{N_k\sqrt{2\pi c_k}}\sum_{\mu}\sum_{i}\exp\left(-\frac{(r-r^{pq,\mu}_{i,i+k})^2}{2c_k}\right)$$

Here p and q represent 2 of the 20amino acids, N_k is the number of pairs that appear in the library, c_k is chosen to be $i-j$, and $r^{pq,\mu}_{i,i+k}$ is the spatial distance between C_β atoms. For glycine, C_α is used instead of C_β. The potentials are constructed from a library of 150 known protein structures. Two proteins do not have more than 25% sequence identity in the library.

2.2 Surface exposure

A contribution from surface exposure and packing is defined as follows. The number of other C_β atoms surrounding the C_β atom of a given amino acid residue is considered as a cost function of surface exposure of the residue.

2.3 Molecular dynamics simulation

We assume that the chain obeys the following Langevin equation of motion [1]:

$$\frac{dp}{dt} = -\frac{\partial V}{\partial r} - \gamma p + \xi(t)$$

$$\frac{dr}{dt} = \frac{p}{m}$$

Here p is the momentum. Mass of all atoms are considered to be equal; $m = 1$. $\xi(t)$ is a random force, which is assumed to be Gaussian white noise and its variance is related to temperature T by the fluctuation-dissipation relation;

$$< \xi_i(t_1)\xi_j(t_2) >= m\gamma T \delta_{ij}\delta(t_1 - t_2)$$

γ can be estimated from Stokes relation, $\delta = 6\pi\eta a$, assuming η to be the viscosity of water at room temperature.

2.4 Series of structures appearing in folding funnel

In order to adjust the potential, not only the structures in the native state but also a series of structures in unfolded or misfolded states are needed. The series of structures of a given protein in the folding is created by folding simulation using a modified potential, which is constructed from the training data set that additionally contain one hundred native-state structures of the protein.

3 Results and Discussion

The modified potential made it possible to recognize the native state with a higher probability in actual MD folding simulation.

Therefore it can be considered to contain funnel-like folding energy landscape. During the folding simulation, the potential should be able to help the structure to escape from local minima and enhance the probability to reach the global minimum of the native state. We added the surface-exposure-cost function (see 2.2) to the pairwise potential (see 2.1), which did improve the accessibility to the global minimum in our MD simulations (see 2.3).

We have applied the above procedure to several proteins with different folds and adjusted the parameters to better represent the folding process. We will use the adjusted potentials for folding simulations, where we plan to use MD simulation together with the Monte Carlo simulation.

Acknowledgements

This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas, 'Genome Science', from the Ministry of Education, Science, Sports and Culture of Japan.

References

[1] Sasai, M., Conformation, energy, and folding ability of selected amino acid sequences, *Proc. Natl. Acad. Sci. USA*, 92:8438, 1995.