DEPARTMENT OF PHYSICS, BROWN UNIVERSITY

Monte Carlo Simulation of Protein Folding

Yangrui Hu

March 8, 2018

Proteins are one of the most important components of cells and organisms. Proteins perform a large number of functions in biochemical processes such as energy storage, transfer, intercellular communication, and work as catalysts in biochemical reactions. As a kind of macromolecules, a protein is composed of many chains of amino acids of which there are twenty kinds known to exist in nature. Different combinations of amino acids form thousands of kinds of protein molecules. The primary structures of proteins are their linear amino-acid sequences, and their specific structures determine their functions and properties.[2] When a protein works in a biological activity, its structure is the high-order structure, generally speaking, as known as the folding state. The first step to study how a protein works or to design a protein with a new kind of function is to figure out its folding structure. According to Anfinsen's research[1], heating proteins or putting them in some chemical environments can cause protein denaturation and unfolding. However, if we put them back into the original environment, the polypeptide chain will fold back into its native state in just a second. Anfinsen concluded that the resulting three-dimensional structure (native state) is determined by the amino acid sequence or primary structure[1]. This conclusion is the essence of protein folding problem. In this project, I will simulate the process of the protein folding with the Monte Carlo method and present some properties of its native state.

To build a computable model, forces or energies which influence the folding structure need to be considered. There are four key factors taken into account[3]: Van der Waals force between non-neighboring amino acids; the hydrogen bonding; interactions between amino acids and water molecules or other chemical molecules in the solution; and the thermal fluctuation. The Van der Waals force can be described by Lennard-Jones potential: $U(r) = 4\varepsilon[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^{6}]$, which is attractive when the distance r is not too large. The hydrogen bonding is also attractive and leads to the folding structure. As for interactions between amino acids and water

molecules, some amino acids are hydrophilic or water-loving, and others are hydrophobic or water-fearing. Hydrophilic amino acids tend to unfold to increase the probability of touching water molecules. On the other hand, hydrophobic amino acids tend to fold. In general, corresponding energies of forces mentioned above are in the order of $k_B T$. So, the thermal fluctuation leading to the increase of the disorder of the system has to be considered.

Then, suppose an amino acid chain with N amino acids is located on a 2D grid, and each amino acid is on a grid point, forming a self-avoiding walks(SAW) pattern. The amino acid sequence can be represented by an $N \times 1$ array $A[N]$, where $A[i] = 1, 2, 3, ..., 20$, corresponding to twenty kinds of amino acids. The interaction energy between non-neighboring amino acids m and n is $J_{A(m),A(n)}$. So the energy of the protein is $E = \sum_{<m,n>} J_{A(m),A(n)}$, where $< m, n >$ means a pair of non-neighboring amino acids. In order to simplify the problem, set J_{ij} as a random value across the allowed range first. Given the initial condition, we can use Markov Chain Monte Carlo(MCMC) method to do the simulation and the algorithm is summarized as following[4]:

- Choose a node of the chain randomly and suppose its coordinate is (x_0, y_0).
- Randomly choose one of its neighbors (x_n, y_n).
- If moving the node to the new position (x_n, y_n) doesn't change the length of the chain, calculate the change in energy ΔE after the movement.
- Apply Metropolis algorithm to decide whether accept this movement: if $\Delta E < 0$, accept; else the probability of the acception is $e^{-\beta \Delta E}$, where $\beta = \frac{1}{k_B T}$, k_B is Boltzmann constant, and T is the temperature of the system.
- Repeat this process a large number of times

After a long time, when the system approaches to the thermal equilibrium, we can find the final state, study statistical properties of this system, and discuss effects of some parameters on the final state, such as the interaction energy J_{ij}, temperature, and the length of the chain N.

REFERENCES

[1] Christian B. Anfinsen. The formation and stabilization of protein structure. *Biochemical Journal.*, 128(4):737–749, 1972.

[2] Christian B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(4096):223–230, 1973.

[3] C N Pace, B A Shirley, M McNutt, and K Gajiwala. Forces contributing to the conformational stability of proteins. *The FASEB Journal*, 10(1):75–83, 1996. PMID: 8566551.

[4] Jay Wang. *Computational Modeling and Visualization of Physical Systems with Python*. John Wiley & Sons.