Homework 7

PHZ 3151

Due Monday, April 27

Please submit your code and plots wherever requested. Results can be handed in either as a hardcopy, or as an electronic document (e.g. tex, latex, MS word, or even a .pdf) sent via email.

In this project, we will make a simple model of protein folding. This is of enormous importance to biological sciences, because the manner in which a chain of amino acids (a.k.a. a protein) folds determines its functionality. For example, the ability of a protein to bind with other molecules in solution depends on the structure of the protein. When a protein loses its structure, for example due to a mutation in a gene or even due to high temperatures, a living organism is likely to get sick or die.

Amazingly, proteins fold into their useful structure in a repeatable fashion, in spite of the fact that the energy landscape (or more accurately, free-energy landscape) contains a vast number of local minima with energies that differ very little. It is still not very well understood how proteins successfully fold so repeatably. Computational approaches are widely applied, but the problem is still extremely challenging and many questions are unanswered.

We have seen that a polymer can be modeled as a self-avoiding random walk (SAW). In the case of a protein, interactions between the amino acids, and also differences in their affinity to solution (water), tend to make the amino acid chain fold in a way that is not completely random. As we know, real proteins tend to produce the same structure during folding. While a real protein is too difficult, we can make a simple model that highlights some of the main features.

1. Write a code to implement a Metropolis Monte-Carlo algorithm to study a simple model of protein folding. Take the energy of the protein to be given by,

$$H = \sum_{<m,n>} J_{A(m),A(n)}$$

where the sum is over nearest neighbors that are *not* connected by a covalent chemical bond. The $A(m)$ labels the particular amino acids of the chain. As in nature, consider 20 possible amino acids, and choose the interaction coefficients $J_{A(m),A(n)}$ randomly from the interval from -1 to $+1$. Since we are considering pair interactions, with 20 possible amino acids, there are $\frac{20\times 19}{2} + 20 = 210$ independent interaction coefficients $J_{A(m),A(n)}$. Your code should let the structure evolve in two dimensions.

Make a random chain of $N = 50$ and study at temperatures $T = 1, 2, 4, 6, 8, 10$. Plot the average end-to-end length and energy as a function of T. Start at $T = 10$ and work your way down using simulated annealing as we discussed. Run at each temperature for a large number of Monte-Carlo steps (i.e. at least 10^4) before you start to take statistics on the energy and end-to-end chain length.

Repeat for another random realization of the coefficients $J_{A(m),A(n)}$ again randomly chosen over the interval from -1 to $+1$. Do you results differ in a noticeable way that you can determine above the statistical fluctuations? This might tell you something about the dependence of the tertiary structure on the primary structure (e.g. via the interaction coefficients).

We showed in class that the heat capacity could be written as

$$C = \frac{\langle H^2 \rangle - \langle H \rangle^2}{k_B T^2}$$

Likewise, the susceptibility χ, which measures the change in the average length resulting from an external force, can be shown to be,

$$\chi = \frac{\langle |\vec{r}_N - \vec{r}_1|^2 \rangle - \langle |\vec{r}_N - \vec{r}_1| \rangle^2}{k_B T}$$

where $\vec{r}_N$ and $\vec{r}_1$ are the coordinates of the ends of the protein chain. Using these expressions, compute the heat capacity and susceptibility as a function of temperature.

In each case where we are taking averages, make sure to allow some MC steps to occur first to allow equilibration to the new temperature. Generally, equilibration can be assumed to occur when quantities are fluctuating but not drifting with MC steps.