

PHY 411-506 Computational Physics II
Chapter 12: Interdisciplinary Topics
Lecture 1

Monday April 7, 2008

Lecture Outline

Protein Folding	3
Protein Sequences	3
Folding and Structure	3
Monte Carlo simulation of a 2-D protein	5

Protein Folding

Protein Sequences

- Polymer chains of $\sim 50 - 10,000$ monomers
- Monomers are Amino Acids
- Each type of protein has a specific sequence of monomers
- The Central Dogma of Molecular Biology explains protein synthesis
 - ◇ DNA contains the Genetic Code for manufacturing proteins
 - ◇ Information in DNA is Transcribed into mRNA
 - ◇ Information in mRNA is Translated into a specific protein

Folding and Structure

- The functional properties of proteins depend on their 3-dimensional *folded* configuration: a protein must be folded exactly right or it will not work! How proteins fold is one of the great unsolved problems of molecular biology.
 - ◇ Wikipedia on Protein Structure

- Myoglobin
- Hemoglobin
- ◇ UB's Dr. Herbert Hauptman developed an efficient algorithm for determining crystal structure
- ◇ The Protein Data Bank is a repository of crystalline structures
- ◇ Database of Macromolecular Movements
- The number of possible folded configurations is exponentially large. For example, if each successive covalent bond between amino acids in the chain has 3 stable orientations, then the number of configurations of N amino acids is $\sim 3^N \simeq 10^{477}$ if $N = 1000$. It would take many times longer than the age of the universe to enumerate these configurations. How a protein finds its unique folded configuration is a mystery.
 - ◇ Wikipedia on Protein Folding
 - ◇ See Levinthal Paradox for more information
 - ◇ You can help understand Folding@home

Monte Carlo simulation of a 2-D protein

- Lattice Monte Carlo methods have been used to model protein folding. As a very simple model, consider a chain of molecules of fixed length N with the molecules constrained to the sites of a two dimensional square lattice.
 - ◇ The strong *covalent bonds* between amino acids are modeled by unbreakable bonds between nearest neighbor molecules in the chain. Under this constraint, a molecule can only move to a diagonally opposite nearest neighbor site, and it can move only if that site is empty and the other two sites in the square are occupied. An exception to this rule is that the two molecules at the ends of the protein can move to any of 3 unoccupied nearest neighbor sites if they are not occupied.
 - ◇ Weaker *hydrogen bonds* between amino acids that are not covalently bonded can be modeled by an interaction energy between molecules that happen to occupy nearest neighbor sites.
 - ◇ *Solvent interactions* are very important in realistic protein folding.

These can be modeled by assigning an energy to each molecule based on the number of *empty* nearest neighbor sites, which are assumed to be occupied by solvent molecules. A positive energy is *hydrophobic* and a negative energy *hydrophilic*.

- ◇ The dynamics of the protein can be implemented for example using the Metropolis algorithm. The system is assigned a temperature T . One Metropolis step consists in selecting a molecule at random and making a trial move. If the move is possible, it is accepted or rejected based on a test similar to that used in simulating the Ising model.
- Here is a simple Protein Folding Applet based on these ideas.
 - ◇ You can look at ProteinFolding.java to see how the algorithm can be coded.