Monte Carlo Simulations of Protein Folding using Lattice Models
June 16, 2006

Ryan Cheng, Carnegie Mellon University
Mentor: Dr. Kenneth Jordan, University of Pittsburgh

Introduction:

Understanding the conformations of protein molecules is of great importance due to the direct relationship between their structure and biological function. Termed the so-called native state, the biologically active form is generally adopted at the global free energy minimum located among an enormous number of local conformational minima. As a result, locating the global minimum through computational methods has attracted considerable attention. These computational approaches to optimization problems include heuristic Monte Carlo-based algorithms [**1, 2**]. However, it is understood that an examination of all possible conformations is computationally impossible due to the number of degrees of freedom in a protein [**3**]. Thus, simplifications are often employed such as two-dimensional and three-dimensional lattice models where each successive monomer occupies an adjacent lattice point [**4**]. One common lattice model is the Hydrophobic-Polar (HP) model, in which the hydrophobic effect is assumed to be the driving force in protein folding [**5**]. This greatly simplifies the interaction energies between non-bonded monomers by characterizing each amino acid in a peptide sequence as being either hydrophobic or polar [**5**]. Thus, this model allows for the feasible simulation of protein folding by greatly reducing the degrees of freedom and simplifying the energy landscape.

One heuristic technique used in optimization is simulated annealing (SA) [**2**], which is based on the idea of heating a material to form a melt followed by slow cooling

to form a perfect crystal. As performed by Kirkpatrick et al., this technique often utilizes the Metropolis Monte Carlo (MMC) algorithm [**1**]. Generally for MMC, all moves that lower the energy will be accepted while all moves that increase the energy will be accepted based on a Boltzmann-like criterion [**1**]. Additionally, the annealing schedule in SA allows for the lattice chain to escape local energy minima while searching for the global minimum. Thus, SA provides a method of locating low-energy conformations.

Methodology:

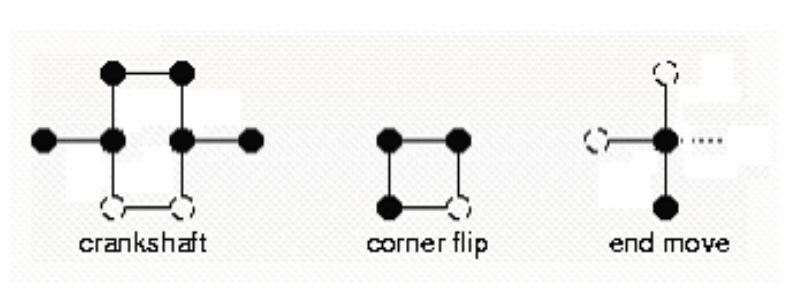
The monomer chains will be generated on a two-dimensional square lattice with periodic boundary conditions [**4**]. These chains will be grown to a desired length, *n*, using the "random walk" technique [**4**]. Additionally, the use of a "self-avoidance" criterion will take into account the effects of excluded volume by forbidding any lattice point from being occupied by more than one monomer [**4**]. Lattice simulation of a known peptide chain will then be considered using the HP model [**6**].

Following generation of the lattice chain, the energy of the conformation will be given by $H = \sum_{i<j} E_{ij}\delta$, where *i* and *j* denote monomers on the lattice and E_{ij} represents the topological contact energy between *i* and *j* [**5**]. This contact energy is only nonzero between adjacent non-bonded monomers and as a result, the Kronecker delta (δ) is 1 if *i* and *j* satisfy the topological contact criterion and 0 otherwise [**5**]. Additionally, based on the HP model, the possible non-bonding contact energies between hydrophobic and polar monomers are E_{HH}, E_{HP} and E_{PP} [**6**].

Subsequently, simulated annealing (SA) will be used to locate the low energy conformations of the generated chains. Annealing will be initially carried out from an

effectively high temperature (T) and slowly cooled to a sufficiently low temperature. Re-arrangement of the configuration during Metropolis steps will be carried out using a form of the Verdier-Stockmayer algorithm, which uses "crankshaft", "corner flip", and "end" moves [**4**]. These moves on a two-dimensional lattice can be seen in **figure 1** on a square lattice where the possible new positions are shown in white.

Figure 1:



Expected Results:

Results from the simulation on a two-dimensional lattice should provide insight on the conformations of the varying monomer chains under investigation. The conformations observed in the simulation should compare well with known literature conformations. Additionally, adoption of low-energy folds through simulated annealing should be characterized by a melting transition on a plot of heat capacity vs. temperature. Subsequently, the information obtained from this simplified model can be used to carry out complicated models that provide a more realistic simulation of protein folding. Such models include use of a three-dimensional lattice and more realistic interaction energies.

References:

[1] Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *Journal of Chemical Physics* (1953). 21: 1087-1092

[2] Kirkpatrick, S; Gelatt, C. D.; Vecchi, M. P. *Optimization by Simulated Annealing*. Science (1983), Volume 220: 671-679.

[3] Levinthal, C. *Mössbauer Spectroscopy in Biological Systems.* Edited by Debrunner, Tsibris, and Münck. University of Illinois Press (1968).

[4] Leach, A. Molecular Modeling: Principles and Applications, 2nd Edition. Prentice Hall (2001): 423-431.

[5] Dil, K. *Theory for the folding and stability of globular- proteins.* Biochemistry (1985), 24: 1501-1509.

[6] Klimov, D. K.; Thirumalai, D. *Criterion that Determines Foldability of Proteins.* Physical Review Letters (1996), Volume 76: 4070-4073.

[7] Bahar, I.; Atilgan, A. R.; Jernigan, R. L.; Erman, Burak. *Understanding the Recognition of Protein Structural Classes by Amino Acid Composition.* Proteins: Structure, Function, and Genetics (1997), 29: 172-185.