Computational Studies of Protein Folding

ADDISON WISTHOFF AND ANDREW CLELAND; ADVISOR: JOELLE MURRAY
LINFIELD COLLEGE, MCMINNVILLE, OREGON

Abstract

Proteins are known to fold into tertiary structures that determine their functionality in living organisms. The goal of our research is to better understand the protein folding process. Using MATLAB, we created an algorithm that models the folding process via a Monte Carlo time step approach. Specifically, amino acids in the chain at each time step are allowed to fold to certain locations according to a set of rules. These rules are based on two main criteria: folds must maintain bond length and should be thermally and energetically favorable. One central goal of our research is to examine whether the folding process can be viewed through the lens of self-organized criticality. In particular we are interested in whether there are features of the folding process that are independent of the size of the protein.

What is Protein Folding?

Protein folding is the process in which a string of amino acids under goes folding to find an equilibrium state where system energy is minimized. The particular sequence of amino acid types in the protein determine the unique tertiary structure formed at the end of the folding process. It is these structures that determine the protein's biological function. A number of diseases are believed to be caused when proteins fails to fold to the proper structure. So a better understanding of the folding process may help shed light on how these diseases could be prevented or treated.

What is Self Organized Criticality?

Self Organized Criticality (SOC) is a theory of complex dynamic processes that posits that complex systems evolve on their own into critical states where a small disturbance, by a domino effect, can result in larger events called avalanches. Avalanches come in all sizes, ranging from insignificant to catastrophic.
The first and simplest example used to illustrate self-organized criticality is the sand pile model. Imagine dropping grains of sand, one at a time, on a flat surface. The grains will form small piles. Eventually the slope of one of more these piles will become steep enough that a grain slips and tumbles down the hill. This causes other grains to tumble as well, thus creating an avalanche. Per Bak, a computation/theoretical physicist who has pioneered the theory of SOC, has shown that (insert book reference here).

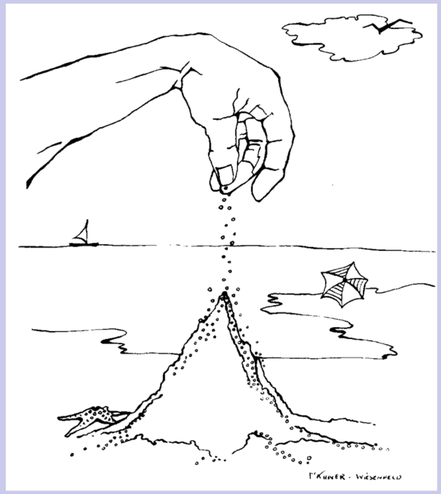

Figure (1): Illustration by Elaine Wiesenfeld. Per Bak uses a sand pile as a illustration of a self-organized critical system.

Our Algorithm

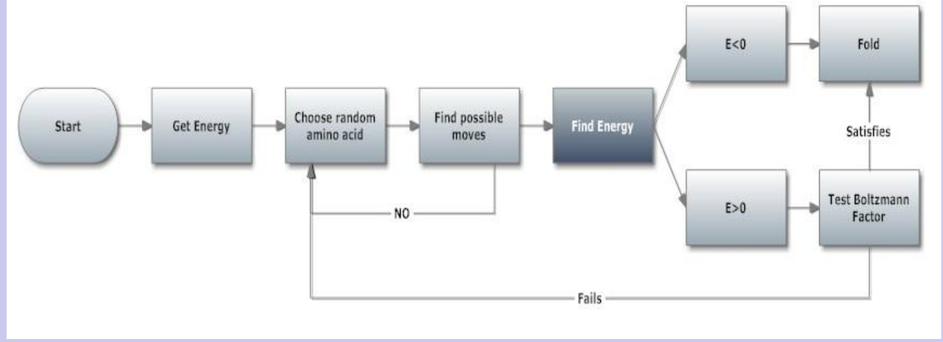

Figure (2): This is a flow chart of our algorithm.

Our model works by a simple algorithm illustrated in figure (2) above.

1. This algorithm starts by finding the current energy level of the system and temporarily stores it for comparison later.
2. From there an amino acid is selected at random for an attempted fold. The code looks for the possible locations it can move the amino acid to without breaking the following set of rules:
   a. Bonds cannot stretch or break
   b. Amino acids cannot exist in the same location
3. From there a random legal move is selected and we calculate the energy of the system if the amino acid were to move there. If this energy value is less than the previously stored value, the move is allowed.
4. If the move raises system energy, there is still a random chance the fold will occur, with the probability of the fold depending on the system's thermal energy. This is accomplished using the Boltzmann factor, which is a function of the change in the system's energy and the temperature of the system. The larger the increase in energy, the less likely a fold will occur at given temperature. The higher temperature, the more likely the fold will be allowed.
5. For each fold where energy increases, a random number is generated between 0 and 1 and is compared to the Boltzmann factor. The fold is allowed only if the Boltzmann factor is greater than the random number.
6. This process is repeated for a specified number of iterations.

With our model we can look into the folding process and also see if our model has scaling features. We can also see if this process follows self-organized critical behavior.

Results of an example run with 200,000 iterations

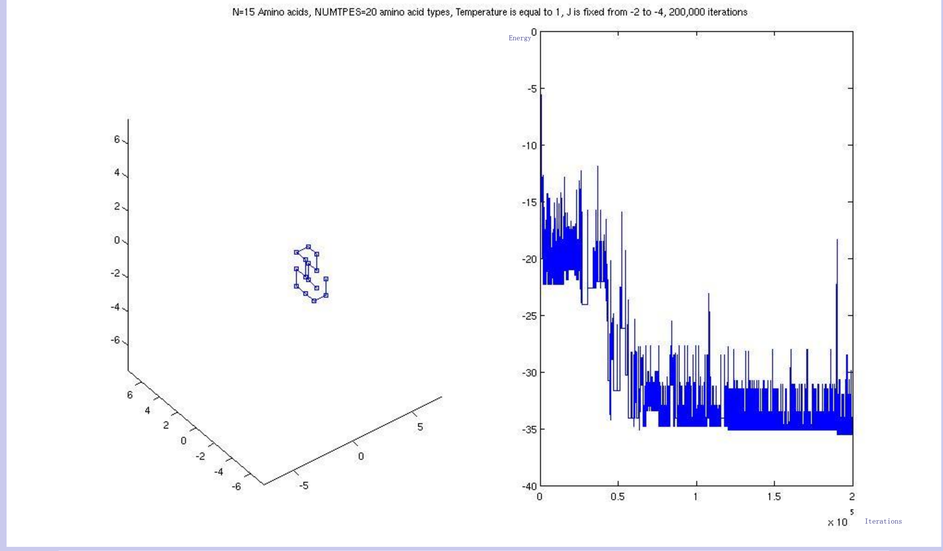

Figure (3): Left: an example of a 3D protein with 15 amino acids after 200,000 iterations of folding. Right: energy of the protein by iteration.

Results

- Power laws are a common occurrence when dealing with SOC.
- The power law of interest can be expressed as $f(s) = As^{\tau}$ where f is the frequency of avalanche of size s, A is a constant, and τ is the slope when you take the logarithm of both sides.
- In SOC systems, the tau $[\tau]$ parameter does not depend on system size (how big the sand pile is, or in our case, how many amino acids are in the protein).
- In our model, we measure avalanche size s by counting how many folds occur consecutively that lower energy. Frequency f is the number of avalanches of a particular size that are observed per run.
- We estimate the parameter tau using an ordinary least-squares fit.
- The figures at right show a comparison of tau estimates for two different sized proteins (N=15 and N=60, where N=number of amino acids in the chain). Each frequency represents an average of 30 runs.
- Our results are so far inconclusive whether tau is invariant with protein size. This is an area that calls for further research.

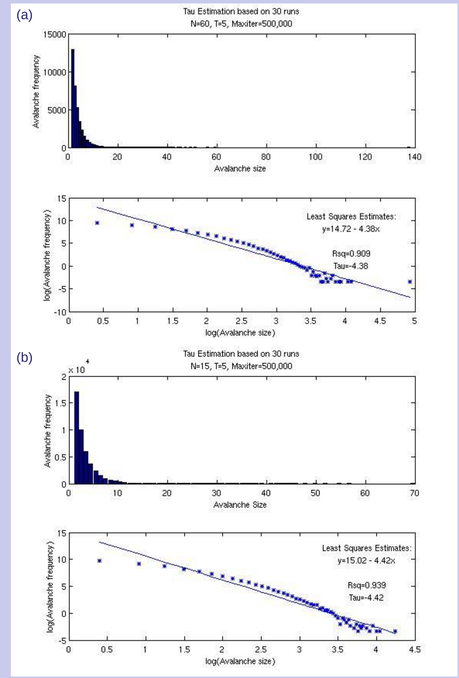

Figure (4): Examples of estimating tau for (a) N=60 amino acids and (b) N=15 amino acids. Shows both the histogram of avalanche frequency by avalanche size and a log-log plot where tau is estimated by least squares.

What's Next?

- We would like to estimate tau with more statistics . So far due to long computational times, we have only been able to base our estimates on 30 runs (each with 500,000 iterations).
- To make our model more realistic, we are considering expanding it to incorporate different bond lengths between every amino acid pair.

Acknowledgements and References

Acknowledgements:
Linfield College Physics Department

References:
P. Bak *How Nature Works: The Science of Self-Organized Criticality*. Copernicus, New York, NY (1996)
"Computational Physics", Nicholas J. Giordano, Hisao Nakanishi, Pearson/Prentice Hall (2006)

Linfield College