Do lattice protein simulations exhibit self-organized criticality?

ADDISON WISTHOFF ; 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 my research is to better understand the protein folding process through a lattice Monte-Carlo simulation. Specifically, amino acids in the chain at each time step are allowed to fold to certain locations according to two main criteria: folds must maintain bond length and should be thermally and energetically favorable. This simulation will then be used to examine whether the folding process can be viewed through the lens of self-organized criticality (SOC). In particular I am interested in whether there are features of the folding process that are independent of the size of the protein.

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.

Our Folded Protein

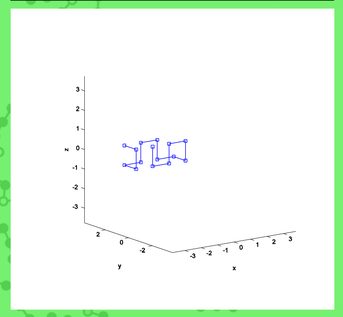


Figure (1): An example of a 3D cubic latice protein with 15 amino acids after 200,000 iterations of folding.

Sand Pile Model

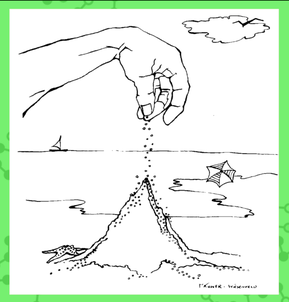


Figure (2): Illustration by Elaine Wiesenfeld. Per Bak uses a sand pile as a illustration of a self-organized critical system.

Power Law Behavior

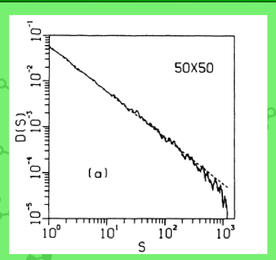


Figure (3): Demonstrating power law behavior in the sand pile model. [1]

Finite-size Scaling

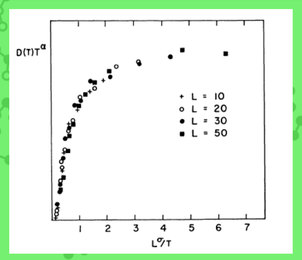


Figure (4): Demonstrating scaling behavior in the sand pile model. [1]

Simulating Protein Folding

- Protein folding is the process in which a string of amino acids folds
- The sequence of amino acid types determine the unique tertiary structure
- To aid in the folding process an annealing function has been inserted
  - Annealing is the process of lowering the temperature as the protein folds
- The annealing function follows a Boltzmann distribution, $T=T_0 e^{CE}$
  - T is temperature, T_0 is the initial temperature, C is a constant, and E is the energy of the protein; note E ≤ 0 in the simulation
- A HP (Hydrophobic and Hydrophilic) model is used
  - Two amino acid types

Our Algorithm

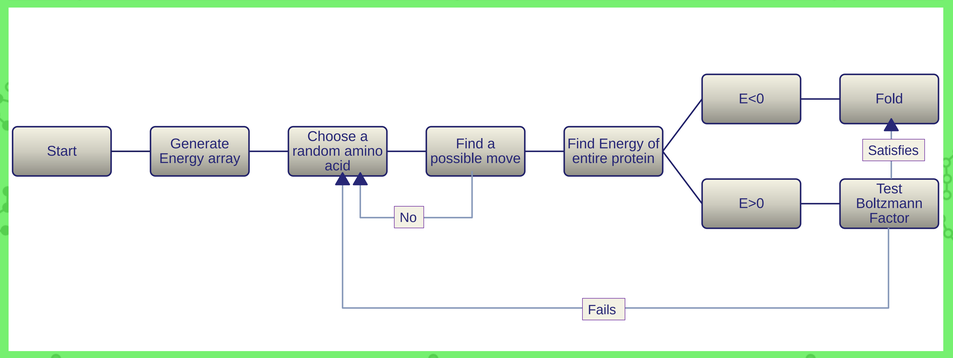


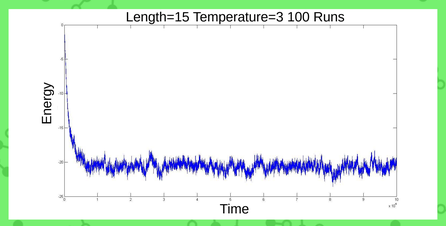
Figure (5): This is a flow chart of our algorithm.

Our model works by a simple algorithm illustrated in figure (3) above.

1. This algorithm starts by finding the current energy level of the system and temporarily stores it for comparison later.
2. The annealing function finds the temperature of the protein based on its current energy.
3. 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
4. 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.
5. 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.
6. 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.
7. 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.

The Annealing function



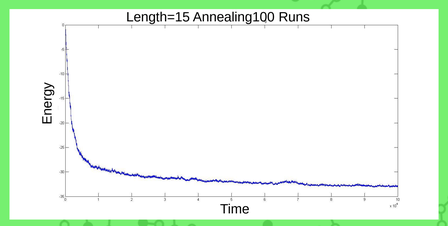
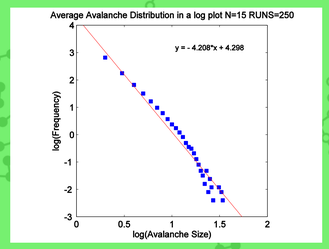


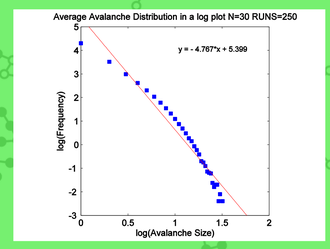
Figure (6): Left: The average energy verses time graph with a constant temperature. Right: The average energy verses time graph with annealing.

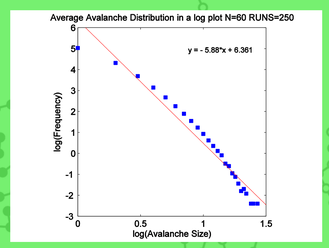
Results

- 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 [τ] 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.
- We use the Kolmogorov–Smirnov statistical test to confirm whether the data is in the distribution.[2]

Power Law Behavior







	Random J_{ij} model KS test (α=0.05)
N=15	pass
N=30	pass
N=60	pass

Figure (8): The results from the Kolmogorov–Smirnov statistical test. By passing the statistical test the above graphs in Figure(7) are indeed in the basis of the best fit line, confirming they follow a power law behavior.

Figure (7): Results from consecutive folds (Avalanches) that are graphed on a log-log plot to demonstrate their power law behavior.

What's Next?

- We would like to move away from the HP model and test whether a less generalized amino acid definition can prove similar results.
- To make our current model more realistic, we are considering expanding it to incorporate different bond lengths between every amino acid pair, as well as the stretching of bonds.
- Trying to find limitations to SOC by seeing what causes the folding process to not recover after a disrupting event (Preventing certain folds)

Acknowledgements and References

Acknowledgements:
Linfield College Physics Department
Andrew Cleland

References:
"Computational Physics", Nicholas J. Giordano, Hisao Nakanishi, Pearson/Prentice Hall (2006)
[1]Bak P, Tang C, Wiesenfeld K (1988) Self-Organized Criticality Phys Rev A Volume 38, p364
[2]Touboul J, Destexhe A (2010) Can Power-Law Scaling and Neuronal Avalanches Arise from Stochastic Dynamics? PLoS ONE 5(2): e8982

Linfield College