# Self-Organized Criticality in Protein Folding with AMBER Parameters

Yura Sim, Joelle Murray
Department of Physics, Linfield College, McMinnville, Oregon 97128

## Introduction

Self-organized critical systems (SOC) are characterized by power law behavior and scale invariance.[1] Developed for a sand-pile model, SOC is recognized in many systems, from avalanches to forest fires.[2,3] The identification of protein folding as an SOC system is of interest in the solid phase synthesis of proteins as SOC mechanics can be used to better control how proteins behave especially with frequency shifting.

To show that protein folding is a self-organizing system, we show that protein folding has power law behavior and scale invariance.

A power law function is defined as $f(x)=ax^{k}$ where frequency $f$ is a function of size $x$.

The logarithm of the equation gives a linear function, $\log(x) = \log a + k \log x$

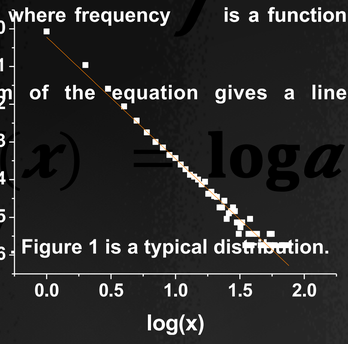



Figure 1 is a typical distribution.

**Figure 1: (Above)**
Typical power law distribution data.

For protein folding, $k$ is negative meaning that the frequency of large events is smaller than the frequency of large events. Thus for a system to be considered an SOC system some size must be linked to a frequency such that the logarithm of the two provides a decreasing linear function.

The scale invariance of an SOC system can be linked to the scalar $a$. A change in $a$ should not change $k$, meaning they are independent of one another.

After identifying protein folding as an SOC system, we apply frequency shifting.

Frequency shifting occurs when one controls a system with small events to replace large events. Many ski resorts or forestry managers force small events to occur, for ski resorts

## Methods

The energy of a protein is calculated with MATLAB by the equation below, from AMBER.[4]

$$E = \sum_{\text{Bonds}} k_b (b - b_e)^2 + \sum_{\text{Angles}} k_a (a - a_e)^2 \ldots$$

$$\ldots + \sum_{\text{Dihedral}} \frac{V_n}{2}\left[1 + \cos(n\varphi - \gamma)\right]^2 \ldots$$

$$\ldots + \sum_{j=1}^{N-1}\sum_{i=j+1}^{N}\left\{\acute{U}_{ij}\left[\left(\frac{r_{0ij}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{0ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\acute{U}_0 r_{ij}}\right\}$$

The first term refers to bond lengths. Each bond is assigned an equilibrium length $b_e$ and a spring stiffness $k_b$. The second term refers to angle strain in similar manner. The third term, refers to dihedral angles, angles between the planes of the bonds and the fourth term refers to van der Waals forces (Lennard-Jones) with electrostatic influences (Coulomb's Law).

After an initial energy is calculated an atom is moved around adjacent particles. If the new recalculated energy is below the energy of the prior iteration or within a tolerance limit, then the iteration is accepted. If the iteration is not accepted the movement is ignored.

Size is defined as both the number of consecutive folds and the magnitude of the energy change of a fold. Frequency is defined how often that size occurs relative to other sizes.

Structure conservation was calculated through the element standard deviation (ESD) of pairwise distance matrixes.

| | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0 | 1.5 | 4.8 |
| 2 | 1.5 | 0 | 2 |
| 3 | 4.8 | 2 | 0 |

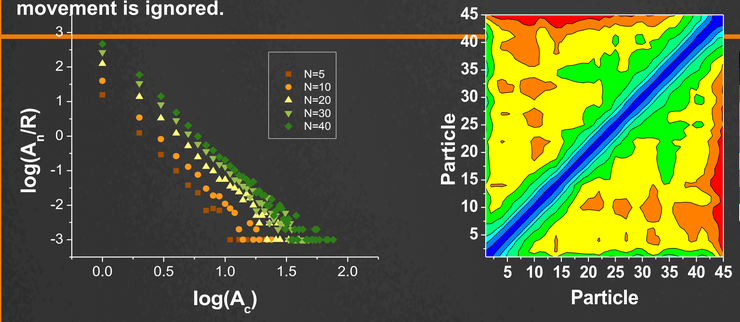
| | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0 | 1.6 | 4.5 |
| 2 | 1.6 | 0 | 2.1 |
| 3 | 4.5 | 2.1 | 0 |

| | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0 | 1.7 | 5.1 |
| 2 | 1.7 | 0 | 1.5 |
| 3 | 5.1 | 1.5 | 0 |

Particle 1 and 3 are 4.8 units apart from each other in run 1, which changes to 4.5 and 5.1 in other runs. The ESD at each position of the matrixes should then be the following,

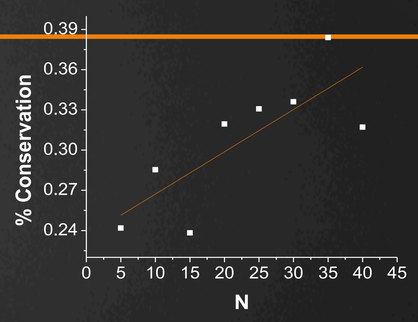
| | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0 | 0.1 | 0.3 |
| 2 | 0.1 | 0 | 0.32 |
| 3 | 0.3 | 0.32 | 0 |

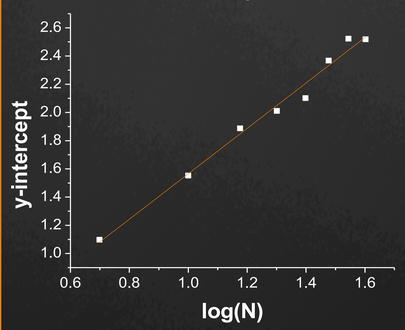
This means that the distance between particle 2 and particle 3 are more variable than the distance between 2 and 1. If one computes the average value of this matrix, we arrive at the value of 0.16. This gives us a score as to how conserved a model is. High scores infer low conservation, low scores infer high conservation.

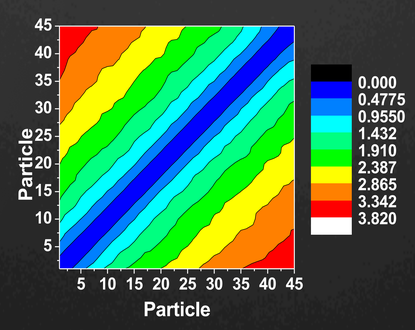


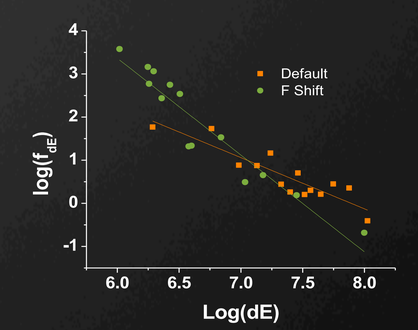







## Discussion

Figure 2 shows the probability distribution as size changes. All fits pass the two sample Kolmogorov–Smirnov test. This establishes that protein folding exhibits power law behavior, which passes the first test of SOC systems.

The slopes of fits in Figure 2 are similar (-3.36 ± 4%), which means that the variable $k$ is constant. However the intercept, $\log a$, changes as a function of size as seen in Figure 3. This means that the slope $k$ is independent of scalar $a$, which is related to the size $N$ and passes the second test of SOC systems. Thus protein folding can be defined as an SOC system.

Figure 3 displays an intuitive mechanic. The larger one makes a mountain the larger the avalanches become. Figure 3 shows that the same thing applies with protein folding.

The AESD's were calculated from pairwise distance matrixes (Figures 4 & 5). This reveals that conservation increases as a function of size (Figure 6). This intuitively makes sense, the longer a proteins gets the more likely there are to be familiar structures (like helices, or barrels, or sheets) that are recognized with the pairwise distance matrix. One is unlikely to find these structures consistently in small proteins.

In Figure 7 we increased the frequency of small energy folds by changing the tolerance limit of our programs to force small folds through. In doing so, there is a large increase of small folds but there is a reduction of large folds. We can apply this knowledge to protein folding with structure induction (SI).

SI[7] works by forcefully inducing small helixes throughout the protein structure.[7] These helices are rigid, much like double bonds in fatty acids, and reduce the tangling that occurs. Tangling is what prevents solid phase syntheses (SPS) of proteins from being successful. Whereas most proteins are anywhere from 100-300 residues, SPS methods have issues at 30 due to tangling. To better control the protein SI can be used to create helices, small folds, to reduce the frequency of tangling, large folds. This type of methodology is the type of behavior we predict in our model, the so called frequency shifting.

Future work on this project will monitor how SI influences real systems and how it influence our understanding of SOC.

## References & Acknowledgements

[1] Abaimov, Sergey G. (2008). "Self-organized Criticality in Complex Systems[...]"
[2] Bak, P. (1987). "Self-organized criticality: an explanation of 1/f noise".
[3] Turcotte, D. (1999). "Seismicity and self-organized criticality."
[4] Cornell W. D. (1995). "A Second Generation Force Field for the Simulation of Proteins[…]"