Synopsis

February 5, 2015

1 REM(Derrida, 1985) + Shaknovich and Gutin, 1989

First paper proved that energy of different compact folds can be treated as **IID variables**. Then the results of REM draft implies that:

$$< X >= \begin{cases} \frac{T}{T_c} & \text{if } T < T_c \\ 1 & \text{otherwise} \end{cases} \tag{1}$$

where T_c is the critical temperature and random variable **X** captures the idea of energy separation present within the confirmations of a random chain for a given sequence (energy matrix). It is defined as:

$$X = 1 - \sum_{i}^{M} p_i^2 \tag{2}$$

Here **M** is the total number of conformations for a given random chain of fixed length. p_i is the Boltzmann/stationary probability of each of these M conformations defined as:

$$P_i = \frac{exp(-E_i/k_BT)}{\sum\limits_{i}^{M} exp(-E_i/k_BT)} \tag{3}$$

E_i is the energy of the fold i defined as:

$$E_i = \sum_{i,j}^{N} B_{i,j}\delta(r_i^m - r_j^m) \tag{4}$$

Here, $\delta(r_i^m - r_j^m) = 1$ if i and j monomers are lattice neighbours. 0 otherwise. $B_{i,j}$ is the interaction/contact energy between monomers i and j drawn from a **Guassian Distribution** with a particular mean and variance.

Conclusion: If there is a unique ground state with large Boltzmann weight then $X \sim 0$. However, if many conformations have comparable occupation then $X \sim 1$. The former case implies that the separation is high enough and the ground state energy is pronounced.

2 Shaknovich and Gutin, 1990a

This paper derives an equation for the probability P_ϵ that for a random chain, ground fold m_0 will dominate, i.e., $p_{m_0} > 1 - \epsilon$ with probability defined as:

$$P_\epsilon = \frac{\sin(\pi X_0)}{\pi X_0} \epsilon^{X_0} \tag{5}$$

where $X_o =< X >$ defined in equation 1.

Conclusion: At high temperature $T > T_c$, $X_0 \sim 1$. Therefore $P_\epsilon \sim 0$ by equation 5 and thus there are no sequences that can fold. When the teperature decreases below T_c, the fraction of sequences that are able to fold grows drastically. For example, taking $\epsilon = 0.01$, which corresponds to 99% ground state dominance, we obtain $P_\epsilon \sim 0.1$ at $T = T_c/2$. This means that under this condition every tenth sequence will have one ground state fold with Boltzmann probability of 99%.

3 Kinetics of protein folding by Sali, Shaknovich and Karplus, 1994

A 3-D lattice model of protein is considered to investigate the properties required for its folding to a ground state(native state). Native state is the fully compact conformation having lowest energy among all the fully compact conformation. A total of 200 sequences with random interactions are generated and subjected to Monte Carlo Simulations to determine which chains find the ground state in a reasonable time. Comparison of folding and non-folding sequences are used to identify the features that are required for fast folding.

3.1 Important points

1. The necessary and sufficient condition for a sequence to fold rapidly to a ground state is thus stated as:

   - **Thermodynamic condition**
     The sequence must have folded conformation which is unique, thermodynamically stable and corresponds to the native structure. The native conformation is thermodynamically stable if it is a pronounced energy global minimum.
   - **Kinetic condition**
     The denatured chain can fold into this conformation under the appropriate solution conditions.

2. Energy surface of a protein is "rugged" (Bryngelson and Wolynes 1987, 1989, Shaknovich and Gutin 1990a). So folding requires that there exists a teperature high enough for the folding process to occur yet low enough so that the ground state is thermodynamically stable (equation 2).

3. Model:

   - Native configuration is the minimum among the fully compact conformation. This is achieved by introducing the overall hydrophic compactness condition. So the lowest energy conformation is known.
   - **Definition: Folding sequences** are those that can find the global minimum a number of times independant of the intial conformation. Non-folding are those that can't find in a reasonable time.

4. Methods:

   - 103346 compact structures
   - B_0 mean and σ_B standard deviation for the energy matrix. σ_B measures degree of heterogeneity.
   - Assumption: B_0 sufficiently negative so that global energy minimum is compact.
   - MCMC simulation:

- Starts from random conformation
- One application of Metropolis criterion = One monte carlo step

5. Choice of parameters:
   Definition: **Foldicity** is the fraction of monte carlo runs that start with a random configuration and end in native.
   Folding sequence, if, native conformation is structurally unique and foldicity is high under conditions where native structure is thermodynamically stable.

   (a) Maximum monte carlo steps set to 50×10^6. Simulation stopped when native reached first time.

   (b) Since relative values are important, σ_B set to 1.

   (c) To select B_0, variation of foldicity of a particular sequence with B_0 is noted. The most negative B_0 which still gives significant folding is selected to increase the probability that global min is compact. $B_0 = -2$ selected.

   (d) Conditions for **temperature**: (1) Thermodynamic stability of native state must be ensured (2) Foldicity should be as high as possible. $< X >$ increases with temperature. Since we want low $< X >$, first requirement puts an upper limit on temperature. At the temperature where $X^{csa} = 0.8$,(where X^{csa} is X defined for only the compact conformations), native conformation Boltzmann weight is more than 0.4 for significant fraction of the 200 sequences($\sim$ 40/200). Foldicity increases with temperature. Foldicity was high in the range $X^{csa} = 0.5$ to $X^{csa} = 0.9$ for one particular sequence. Considering these, T at which $X^{csa} = 0.8$ was selected.

3.2 Results

1. Database for analysis

   - 200 guassian distribution generated random sequences was subjected to MCMC simulation.
   - 30 sequences have $Foldicity > 0.4$ and 146 sequences did not fold at all.

2. Test of metastability of folded states

   - 146 unfolding sequences are tested to determine whether any of them end up in a single metastable state/local minima.
   - Each sequence was run for 10 independant folding simulations. Conformation with lowest energy was compared for all the 10 simulations for each of these 146 sequences. None of these sequences folded into same local energy minimum and thus the present model does not support the metastable native state model and is in accordance with results of Honeycutt and Thirumalai, 1992.

3. Comparison of lower part of energy spectrum of the ensemble with CSA chains

   - Only CSA chains are considered instead of the ensemble of self-avoiding chains.
   - Sample energy spectra of 10 random sequences showed that if teh discrete part of energy spectrum is sparse for CSE chain then it is likely to be sparse for all chain, i.e., X is strongly co-related to X^{csa}.

4. Relation between foldicity and energy spectrum

   - Folding is more associated with sequence having pronounced energy global minimum.
   - $< X^{csa}(T) >$ for non folding sequence is higher than that of folding sequences.
   - $(E_1 - E_0)$ and $T(X^{csa} = 0.8)$ are strongly correlated and determines whether or not a given sequence is a folding sequence.

5. Association between foldicity and conformation and native state

   - **Order of contact:** Absolute difference between the indices of two monomers in the chain
   - Contacts with less order are more likely to occur as compared to contacts with higher order. This resulted is in accordance with Wetlaufer, 1973.