Random Energy Model (REM)

1. Designed REM: Consider a protein model in which for a given sequence and structure, the energy is randomly taken from the Gaussian probability density

\[ p(E) = \frac{1}{\sqrt{2\pi\Sigma^2}} \exp\left(-\frac{E^2}{2\Sigma^2}\right). \]

The total number of structures is \( \Omega_{\text{str}} \), while the number of sequences is \( \Omega_{\text{seq}} \gg \Omega_{\text{str}} \).

(a) A particular sequence has a (unique) native structure of energy \( E_N \). Calculate and plot the energy \( E(T) \) of this sequence as a function of temperature \( T \).

(b) For a particular structure, we attempt to design a good sequence by Monte Carlo sampling of representative sequences at a ‘temperature’ \( \tau \). Calculate and plot the designed native energies \( E_N(\tau) \) as a function of the design temperature \( \tau \).

********

2. Folding time: [Adapted from Gutin et al., J. Chem. Phys. 108, 6466 (1998).] Assume that to change from one compact structure to any other, the protein has to unfold to an intermediate flexible state of (higher) energy \( E^* \). If the starting configuration is at an energy \( E \), the typical (activation) time to overcome this energy barrier behaves as

\[ t_0(E) = \tau \exp\left(\frac{E^* - E}{k_B T}\right), \]

where \( T \) is the temperature, and \( \tau \) is an elementary time step. The folding time is then related to the number of accessible states (hence entropy) to be explored, by

\[ t_F(E) = t_0(E)n(E) = t_0(E)\exp\left(\frac{S(E)}{k_B}\right). \]

(a) Use a random energy model to calculate \( E \) and \( S \) as a function of temperature \( T \).

(b) Calculate the folding time \( t_F(T) \), and plot \( \ln t_F(T) \) as a function of \( 1/T \).

********

3. Amino-acid interactions: What can we learn by combining the REM with commonly used interaction potentials between amino acids?

(a) Find a \( 20 \times 20 \) matrix of interactions \( U(a,a') \) amongst amino acids, and calculate the mean \( \langle U \rangle \) and variance \( \langle U^2 \rangle_c \) of its elements. The commonly used Miyazawa–Jernigan
(MJ) interaction matrix can be found in S. Miyazawa and R.L. Jernigen, J. Mol. Biol. 256, 623 (1996). (Table 3 of this publication is available on the web-page for assignments.)

(b) Model the possible configurations of a protein by the ensemble of compact self-avoiding walks on a cubic lattice. (All lattice sites are visited by compact walks.) Calculate the number $n$ of non-polymeric nearest neighbor interactions for such configurations on an $N = L \times L \times L$ lattice, and deduce the ratio $n/N$ for large $N$.

(c) The number of compact walks on a cubic lattice asymptotically grows as $g^N$, with $g \approx 1.85$. Use this in conjunction with the results from parts (a) and (b) to estimate the folding temperature $T_c$ of a random sequence of amino-acids, and the corresponding energy $E_c$.

(Optional) (d) Select a protein, find its amino-acid sequence and construct a contact matrix corresponding to its structure. Use the interaction matrix from part (a) to estimate the energy of the native structure, and calculate the ratio $E_N/E_c$.

******