8.592J–HST.452J: Statistical Physics in Biology

Assignment # **5**

Protein Folding

1. *Designed Random Energy Model (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 Ω_{str}, while the number of sequences is $\Omega_{seq} \gg \Omega_{str}$.

(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' τ. Calculate and plot the designed native energies $E_N(\tau)$ as a function of the design temperature τ.

*****

2. *Charged Random Energy Model:* Use the random energy model to investigate the freezing of a charged heteropolymer. Assume that there are g^N possible globular states of the polymer, whose energies are randomly selected from a Gaussian distribution of mean zero, and variance

$$\sigma^2 = u^2 N + c^2 \left(\frac{Q^2}{R}\right)^2.$$

The second term in the above formula is a rough estimate of the variations in Coulomb energy from different ways of distributing a charge Q over a volume of size R.

(a) Find the energy E_c at which the entropy vanishes, and the corresponding freezing temperature T_c.

(b) For compact globular states, how should Q^2 scale with N for the freezing temperature to be asymptotically independent of N?

*****

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

(a) Find a 20×20 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.

*****

4. *Analysis of protein structures:* Calculate ϕ and ψ torsion angles in `Rasmol` for a given protein (see the commands below). Make (ϕ, ψ) "Ramachandran" diagrams by plotting ϕ along the x and and ψ along the y axis; one (ϕ, ψ) point for each amino acid.

(a) Do amino acids that are part of different secondary structure elements (helices, sheets) land in the same or different islands on the (ϕ, ψ) diagram? You can find secondary structure elements in fields `HELIX` and `SHEET` of the protein structure file (aka PDB file). Explain your observations.

(b) Find amino acids that have unusual (ϕ, ψ) angles (i.e. deviate from the many clouds of points). What types of amino acids tend to have "unusial" (ϕ, ψ) conformation? Discuss.

(c) Visualize protein structure in `Rasmol`, following the sequence of commands below, and select those with "unusual" (ϕ, ψ) conformation. Do they tend to be close to the ligand?

Some sample proteins to explore (PDB files provided on the Assignment page):
Hemoglobin (alpha chain) 4HHB_A.PDB
Immunoglobulin domain 1TEN.PDB

You can use the following sequence of `Rasmol` commands to generate a good view of a protein, and the `fipsi.dat` file of (ϕ, ψ) angles

```
set background white
wireframe off
ribbons
color structure
select ligand
cpk
color green
select protein
write RDF fipsi.dat
```

To select a particular set of amino acids, (e.g. 128 and 156) you can do the following

```
select 128,156
cpk
color red
```

*****

8.592J / HST.452J Statistical Physics in Biology
Spring 2011