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

Assignment # **3** Due: **10/13/25**

Polymers & Protein folding

1. *Freely–jointed rods:* The flexibility of a long polymer arises from fluctuations of segments much longer than its persistence length ξ_p. The important parameter is thus not the degree of polymerization (number of monomers) $\mathcal{N}$, but the number of unconstrained degrees of freedom, or the *Kuhn* length $N \approx \mathcal{N}/\xi_p$. As a model of the resulting elasticity, consider N rods of fixed lengths $|\vec{c}_i| = c \approx \xi_p$ that are connected end to end. Two points separated by n such rods are displaced by a vector

$$\vec{r}_n = \vec{c}_1 + \vec{c}_2 + \cdots + \vec{c}_n \, .$$

For freely jointed rods, each of the vectors $\vec{c}_i$ is independently chosen to point in an arbitrary direction.

(a) Use the central limit theorem to estimate the probability distribution function $\mathcal{P}_n(\vec{r}_n)$, for large n. Note that quite generally, for sums of independent random variables, the characteristic functions are multiplied, i.e.

$$\tilde{\mathcal{P}}_n(\vec{k}) \equiv \left\langle \exp\left[i\vec{k}\cdot\vec{r}_n\right]\right\rangle = \prod_{i=1}^{n} \left\langle \exp\left[i\vec{k}\cdot\vec{c}_i\right]\right\rangle \, .$$

(b) Calculate the characteristic function $\tilde{\mathcal{P}}_1(\vec{k}) \equiv \left\langle \exp\left[i\vec{k}\cdot\vec{r}_1\right]\right\rangle$ for a single rod.

(c) Noting the independence of segment, what is the characteristic function $\tilde{\mathcal{P}}_n(\vec{k}) \equiv \left\langle \exp\left[i\vec{k}\cdot\vec{r}_n\right]\right\rangle$ for the n-rod segment?

(d) Write the expression for $\mathcal{P}_n(\vec{r}_n)$ as an integral, and evaluate it by the saddle point method for large n.

*****

2. *Bending a charged polymer:* As indicated in the figure below, bending a polymer reduces the distances between its monomers.

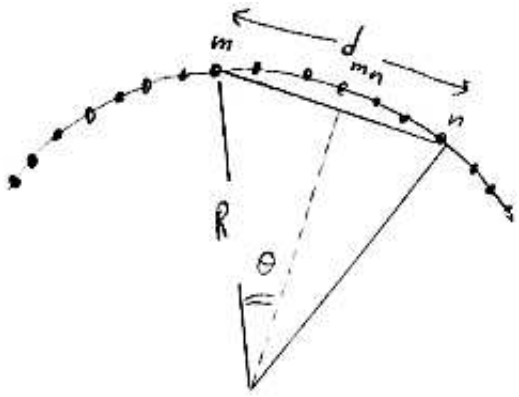
(a) For two monomers which are at a distance $d_{mn} = b|m-n|$ when the polymer is straight, compute the change $\delta d_{mn}(R)$, when the polymer is bend into a circle of radius $R \gg d_{mn}$.

(b) Assume that unit charges at each monomer interact through the Debye potential

$$V(d) = k_B T \ell_B \frac{e^{-d/\lambda}}{d},$$

where ℓ_B and λ and the Bjerrum and screening lengths, respectively. Compute the change in this pairwise energy if the distance is changed from d to $d + \delta d$.

(c) Show that bending increases the energy of the charged polymer by an amount proportional to L/R^2, where $L = Nb$ is the total length of the polymer.



(d) The overall bending modulus for a charged polymer can be written as $\kappa = \kappa_b + \kappa_e$, where κ_b is the cost of deforming the backbone, while κ_e comes from the additional electrostatic energy. Compute κ_e using the model explored in the previous parts.

(e) Is electrostatic energy a significant component of the cost of bending a double stranded DNA? (For DNA in water, $b \approx .17$nm, $\ell_B \approx .7$nm, $\lambda \approx 1$nm, and $\ell_P \approx 50$nm. What about single-stranded DNA or RNA?)

*****

3. *Packaging DNA in a phage:* After an infected bacterium has duplicated the DNA and coat of an infecting phage, a new phage is assembled with the aid of protein motors. In the case of bacteriophage $\phi 29$, a 20,000 base pair dsDNA has to be packaged in a capsid, which is a cylinder of radius $r = 42$nm and height $h = 47$nm. Inside the capsid the DNA is arranged like a spool, first winding in a helical shell next to the wall, and then forming successively tighter shells moving inwards. A typical separation between strands in this structure is 2.3nm. Single molecule experiments have shown that the work required to pack the DNA in the capsid is approximately $10^5 k_B T$ at room temperature ($T = 300^o$K). In the following, *use order of magnitude estimates* to determine what sets this energy scale.

(a) Estimate the entropy of the DNA in solution, using a persistence length of $\ell_P \approx 50$nm. Can the loss of this entropic free energy account for the work of packaging?

(b) Estimate the energy cost of bending DNA into the helical form found in the capsid. (Express the rigidity parameter κ in terms of the persistence length ℓ_P.) Is bending energy a significant fraction of the overall work of packaging?

(c) Estimate the electrostatic energy of DNA in the capsid: Assume unit charges along the DNA at a spacing of $b \approx 0.17$nm, which interact through a Debye–Hückel potential of screening length $\lambda \approx 1$nm, with charges on nearby strands (separations of around 2nm). Can electrostatic energies account for the work of packaging?

Here are a two articles on the packaging of DNA in a phage: *P. K. Purohit, M. M. Inamdar, P. D. Grayson, T. M. Squires, J. Kondev, and R. Phillips,* Forces during Bacteriophage

DNA Packaging and Ejection- *Biophys. J., February 1, 2005; 88(2): 851-866;* and S. Tzlil, J. T. Kindt, W. M. Gelbart, and A. Ben-Shaul, Forces and Pressures in DNA Packaging and Release from Viral Capsids- *Biophys. J., March 1, 2003; 84(3): 1616-1627.*

*****

4. *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.

*****

5. (Optional) *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.

(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. (PDB files for a number of proteins are available on the webpage.)

*****

6. (Optional) *Kinetics of protein folding:* [Adapted from Gutin *et al.*, J. Chem. Phys. **108**, 6466 (1998).] Assume protein folding proceeds through a folding nucleus which has the free energy $F^\ddagger = E^\ddagger - k_B T \log M^\ddagger$. The folding nucleus that serves as a transition state for the folding reaction has $M^\ddagger$ possible states. The typical folding time needed to climb over this free energy barrier is

$$t = \tau_0 \exp\left(\frac{F^\ddagger - F}{k_B T}\right),$$

where T is the temperature, and τ_0 is an elementary time step.

(a) Use a random energy model of $M = g'^N$ states to calculate F as a function of temperature T, and calculate the folding time $t(T)$ for two regimes $T > T_c$ and $T < T_c$. Plot $\ln t(T)$ as a function of $1/T$.

(b) Consider a limit of $T \to \infty$ and express the folding time as a function of the total number of conformations $M = g'^N$ and the number of states in the folding nucleus $M^{\ddagger}$. Interpret your result.

(c) Find a temperature T_{opt}, which provides the fastest folding, compare it to T_c. Compare the optimal folding time with the folding time from "non-designed" REM at T_c. Make conclusions about folding kinetics for random sequences (REM) and designed sequences (designed REM).

*****