Phys347 **Assignment #4** **Due Nov 6**

Reading: We have covered most of chapter 8. Read all of Ch 8 except 8.2.4. You can also begin to read Ch 10. We will cover up to 10.2.2 as well as 10.4.1 and 10.4.2.

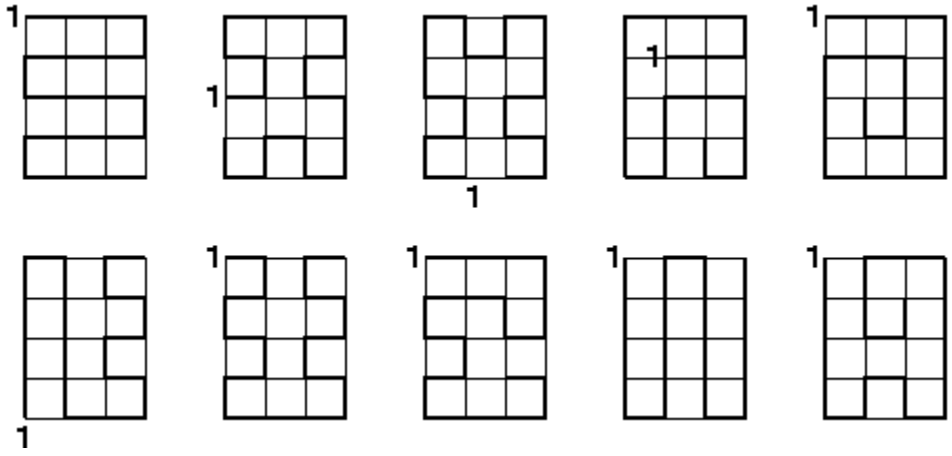
Problem 1: Back of the envelope protein folding

You are going to estimate how long a protein sequence needs to be in order for it to fold. Our model for the protein is going to be a linear chain of amino acids, each in the shape of a sphere of radius r. This chain of spheres then compacts into a final protein fold that has a spherical shape with radius *R*.

(a) Estimate how many amino acids make up this protein, N_{tot}?
(b) Approximately how many amino acids reside on the surface of the sphere, N_{surf}?
(c) Estimate how many amino acids reside in the core, completely away from the surface, N_{core}?
(d) The amino acids in the core are hydrophobic while those on the surface are hydrophilic. We will assume that a protein will only fold properly if the number of amino acids in the core is equal to or greater than the number of amino acids on the surface. Using your results from (a), (b) and (c) find the value of *R/r* that gives the same number of surface amino acids as core amino acids.
(e) Using the value of *R/r* found in (d), how many amino acids are necessary for this spherical protein to fold? Your result will be a slight overestimate, as the typical size of a protein sequence is 200-300 amino acids. Sadly, scientists do not yet have a good theory for how proteins fold.

Problem 2: Protein Folding:

Let us assume that 20mer HP sequences can only fold into the following set of 10 compact structures (in reality there are many, many, many more). The "1" indicates the starting amino acid of each structure.



Consider the following sequence: HPPHHHPPPHPHPHHHPHPH

a) For each structure translate the structure into a corresponding buried surface vector, *s*, where sites on the inner 6 locations are considered 'core' sites and have $s = 1$ whereas the surrounding 14 sites are considered 'surface' sites and have $s = 0$.

b) Evaluate the energy of the given sequence on each of the above structures using the solvation model: $E=\sum_i h_i s_i$ where $h_H=-1$ and $h_P=0$ and the structure is given by $s_C=1$ and $s_S=0$ where C represents a buried core site and S represents a surface site.

c) What is the ground state energy? Is the ground state degenerate (i.e. there is another structure with the same energy)?

d) Your at room temperature. Given your answer in (c) what is the most probably energy, E? What is that probability as a number? Does it correspond to the lowest energy? (We're ignoring all the unfolded configurations, but this calculation still proves a point about how degeneracy affects the likelihood).

Problem 3: Radius of Gyration: (8.2 from book)

You're going to prove what seems to be a simple result for the radius of gyration of a freely jointed chain. It turns out to be a lot more involved. But it is one of those nice results that many use without knowing why it is that way. Now you'll be able to say that you know how to derive it.

Prove: $\left\langle R_G^2 \right\rangle = \frac{L\xi}{3}$

The radius of gyration is defined to be:

$R_G^2 = \frac{1}{N}\sum_{i=1}^{N} (\vec{R}_i - \vec{R_{CM}})^2$ and the center of mass is $\vec{R_{CM}} = \frac{1}{N}\sum_{j=1}^{N} \vec{R}_j$

a) Put the formula for R_{CM} into R_G and after some work show that this becomes

$$R_G^2 = \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} (\vec{R}_i - \vec{R}_j)^2$$

b) Note that every i-j pair occurs twice in the above sum, so that we can rewrite this as:

$$R_G^2 = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=i}^{N} (\vec{R}_i - \vec{R}_j)^2$$

For large N these two sums go over to integrals and by taking the average over all configurations we arrive at:

$$\left\langle R_G^2 \right\rangle = \frac{1}{N^2}\int_0^N du \int_u^N dv \left\langle \left| \vec{R}(u) - \vec{R}(v) \right|^2 \right\rangle \quad .$$

Now there are $(u\text{-}v)$ links of size, b, between the positions u and v. This is just a random polymer of this length. So we can use the result that the average squared end-to-end distance for a polymer of this length is $\left\langle \left| \vec{R}(u) - \vec{R}(v) \right|^2 \right\rangle = (u-v)b^2$ where $b = 2\xi$. Put this result into the above integral and carry out the integration to finish the proof.

c) We usually measure the length of a genome in terms of the number of basepairs, N_{bp}, not L. Rewrite

the equation for the radius of gyration in terms of N_{bp} instead of L, using the fact that one basepair has a length of 0.34 nm ~ 1/3 nm.

d) Using the result in (c) calculate the radius of gyration of bare human DNA in solution that has $N_{bp} \sim 3 \times 10^9$. How does this compare to the size of a typical eukaryotic nucleus that has a radius, R ~ 3 μm?

Problem 4: Entropy of 1D random polymer (8.1c from book):

a) Calculate the change in entropy in extending the end-to-end distance from R=0 to R=r of a 1D random polymer. To do this realize that the number of states for a given R is proportional to the probability we derived in class P(R,N). Use the large N limit where this probability becomes a Gaussian distribution.

b) From your result in (a) what is the free energy difference due to extension?

c) Given your result in (b) what is the force? And hence what is the effective spring constant of the random polymer?