Physics 4251 / Fall 2014
Problem Set 3, Sept. 5
Due: Monday, Sept. 15 before class

Problem 1: Force of an entropic spring

A 1D freely-jointed chain composed of N segments each of length a responds to an applied force, F, by generating a restoring force, F_{restore}. In static equilibrium $F_{\text{restore}} = -F$; the restoring force contributes energy to the object/device applying the tension. The free energy of the **system** can thus be written as $G = -FL - TS$ (recall that the freely-jointed chain cannot store internal energy).

(a) What is the expected length, L, in the case of no applied force (purely entropic)? What about at $T = 0$ (purely enthalpic)?

(b) In class we determined the relationship between force and end-to-end length in the limit that $L << Na$, where Na is the fully extended length. Now, find the length that minimizes the free energy, i.e., $\partial G/\partial L = 0$, without taking this limit. Express your final result in the most concise form possible.

Problem 2: The cleavage mechanism of the autocatalytic transporter EspP

For this problem, we'll examine the structure of the protein EspP. EspP is an autotransporter expressed in the outer membrane of Gram-negative bacteria that secretes its own N-terminal domain to the outside of the cell. After secretion, this domain, a virulence factor, is auto-cleaved by the transporter and released.

(a) Download and examine the PDB for EspP (PDB ID 3SLT) in VMD. The site of catalysis is residue Asn1023, which can undergo "asparagine cyclization" to cleave the peptide bond between it and residue 1024. However, to crystallize the pre-cleavage state, Asn1023 had to be mutated to another amino acid. Which amino acid was used in this structure?

(b) Make a figure of the residues in the active site around residue 1023. Use a cartoon representation for the protein, but do not include the barrel. Then use licorice for all the residues within 8 Å of residue 1023. Color them by restype. Make a separate representation for 1023 such that it stands out from the others. Render and print this figure. Does anything stand out about the residues in the vicinity of 1023?

(c) Using the tkconsole of VMD, make atom selections to answer the following questions: how many negatively charged amino acids are present in the EspP structure? How many positively charged ones? (Hint: you need to include something in your selection that identifies each residue only once.)

(d) Now we will do a pKa calculation to determine the actual charge states of ionizable residues. Go to the PropKa website at `propka.ki.ku.dk` and input the PDB ID and click submit. Although a lot of data are output, focus on the part near the bottom entitled "summary of this prediction". Pick three residues from the list, one that is typically negatively charged, one typically positively charged, and one histidine. Using the Henderson-Hasselbalch equation, calculate the probability that each is protonated at pH 7.0.

(e) Are any of the residues predicted to be in a non-standard protonation state at pH 7.0? Where are these residues located in relation to the active site? Which residue is closest? Does it make sense to alter the protonation state of this residue? Why or why not?

Material for this problem derives from the paper "Molecular basis for activation of a catalytic asparagine residue in a self-cleaving bacterial autotransporter." T. J. Barnard, J. Gumbart, et al. *J*MB **415**:128-142, 2012. It is available on the course website.

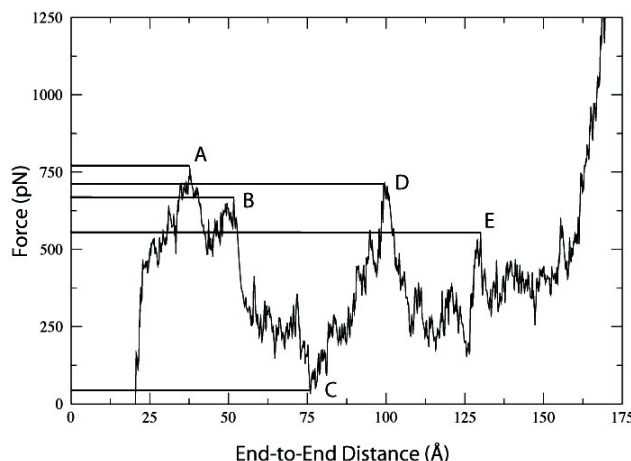**Problem 3: Worm-like chain model applied to protein extension**

Figure 1: This plot compares the force and the separation during a constant velocity pull of .05 Å/ps of a ubiquitin monomer between the N terminus and Lys48. Points A, B, C, D, and E represent the breaking of different secondary structures. This is analogous to the stretching of a rubber band; the force is at a maximum immediately before it breaks.

The worm-like chain model has proven to be a useful polymer model for both DNA and proteins. It assumes a continuously flexible, uniform rod, but in its simplest form does not account for higher order structure. Nonetheless, it has been used successfully for interpreting experimental AFM data. In this model, the force required to extend the polymer can be expressed as

$$\frac{Fl_p}{k_BT} = \frac{z}{L_C} + \frac{1}{4(1 - z/L_C)^2} - \frac{1}{4} \tag{1}$$

where z is the extension, L_C is the contour length, and l_p is the persistence length. We will now use this model to examine data provided from a simulation of unfolding ubiquitin from the N terminus to K48.

(a) Determine the behavior of z in the high and low force limits ($F \to \infty$, $F \to 0$). Keep in mind physical limitations on z. (If you cannot do this using exact methods, you may also try numerically solving the equation for large and small values of F to determine the dependence of z).
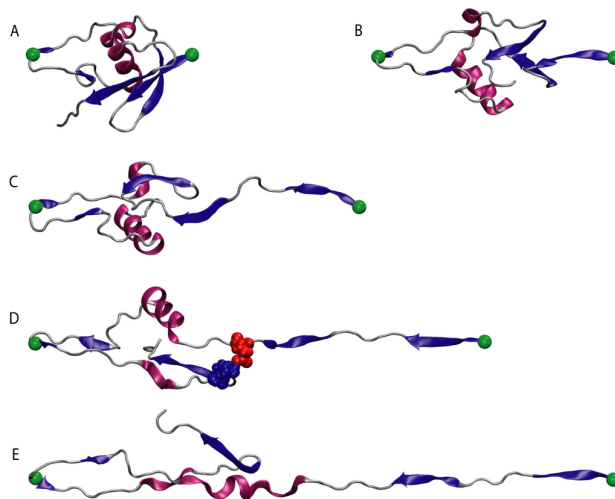
Figure 2: An illustration of points A-E in the plot shown in Figure 1.

(b) The simulation was performed at 300 K. In units of pN*Å, what is the thermal energy, k_BT?

(c) Open the file 1UBQ.pdb in VMD. In the Graphical Representations window, select "Trace" for the protein; it should resemble a jointed chain. Measure the length of a handful of segments, each one representing one peptide (using the "Labels → bonds" option and the Labels window). Use an average value to calculate L_C, the extended length, between the N terminus and K48.

(d) Since our model does not take into account the detailed interactions involved in unfolding a protein, we will only examine the last part of the extension, from 132 Å to 171 Å (see Figure 1). The data from this portion of the simulation is provided in the file SIMDATA.dat. The first column is extension in units of Å and the second column is force in units of pN. Fit this data to the model above using our computed values of k_BT and L_C. What is the fit value of the persistence length, l_p? (Note: while you may have to manipulate the form of the data or the equation, this problem does not require advanced programs to solve).

(e) Plot your equation now along with the data to visually examine your fit. A widely accepted value of the persistence length is 4 Å. How does your computed value compare to this?

Material for this problem comes from the Ubiquitin Case Study available at
www.ks.uiuc.edu/Training/CaseStudies/.