20.320 Problem Set 1
September 10, 2009

#

This problem set consists of three problems designed to reinforce your knowledge of protein structure and energetics and to develop your skills at computationally analyzing protein sequences and structures:

Questions one and two relate to the structure of influenza hemagglutinin, which is the protein that allows the flu virus to enter human cells. (The swine flu is called H1N1 because it carries type 1 Hemagglutinin and type 1 Neuraminidase.) Question three examines the protein that causes the deadly genetic disease cystic fibrosis.

General Instructions:

1. You are expected to state all your assumptions and provide step-by-step solutions to the numerical problems. Unless indicated otherwise, the computational problems may be solved using Python/MATLAB or hand-solved showing all calculations. Both the results of any calculations and the corresponding code must be printed and attached to the solutions.

2. You will need to submit the solutions to each problem to a separate mail box, so please prepare your answers appropriately. Staples the pages for each question separately and make sure your name appears on each set of pages. (The problems will get sent to different graders, which should allow us to get the graded problem set back to you more quickly.)

3. Submit your completed problem set to the marked box mounted on the wall of the fourth floor hallway between buildings 8 and 16.

4. The problem sets are due at noon on Friday September 18th. There will be no extensions of deadlines for any problem sets in 20.320. Late submissions will not be accepted.

5. Please review the information about acceptable forms of collaboration, which was provided on the first day of class and follow the guidelines carefully.

20.320 Problem Set 1
Question 1

#

Hemagglutinins are a general class of factors that increase the affinity of red blood cells for each other, causing clumps to form (the clumping of red blood cells is referred to as hemagglutination). Although some hemagglutinins are expressed under normal conditions (for instance, blood group antigens and the Rh factor), many pathogens express hemagglutinins and hemagglutinin-like proteins to help them adhere to and invade host cells more effectively. For example, the influenza viruses express hemagglutinin glycoproteins on their surfaces that play a key role in the initial binding between virus and host cell.

Influenza hemagglutinin is particularly interesting because it exploits several features of the cell's endocytic pathway to protect the virus from degradation and to facilitate its release into the cytoplasm. Once the virus attaches to the exterior of the cell it is internalized in a membrane-bound compartment called an endosome, which fuses with a lysosome to begin digesting what the cell internalized. Key to this digestive process is the acidification of the endosome, since the enzymes involved in digestion are only active when the pH is substantially lower than in the cytoplasm. The influenza virus is able to exploit this acidification process using hemagglutinin. The hemagglutinins on the viral surface undergo a pH-dependent conformational change, exposing a hydrophobic pocket that can insert into the membrane of the endosome and fuse the endosomal and viral membranes together. This allows the virus to escape degradation and transit into the cytoplasm.

Structural data for both native HA (3EYJ.pdb) and HA at endosomal pH (1HTM.pdb) are posted on the Course website.

a) Write a Python program to parse the PDB files and extract the phi and psi angles for the HA_2 chain (chain 'B' in the PDB files) of Hemagglutinin in its native state and at endosomal pH. Use this to create a Ramachandran plot for both structures. (Note: Since chain 'B' in PDB file 1HTM only contains residues 40-153 of chain 'B' in PDB file 3EYJ, only consider those residues.) For this problem, use the Biopython package. Biopython is set of tools for biological computation written in Python and is free to download here: http://biopython.org/wiki/Download Source code for the PDB package can be found here: http://www.biopython.org/DIST/docs/api/Bio.PDB-module.html

   Use the following code segment as a model for parsing a PDB file:

```
for model in Bio.PDB.PDBParser().get_structure("HA_Native", "3EYJ.pdb") :
        polypeptides = Bio.PDB.PPBuilder().build_peptides(model["B"])
        for poly_index, poly in enumerate(polypeptides) :
```

   The following command is used to print the phi and psi angles of a polypeptide:

```
poly.get_phi_psi_list()
```

b) What do the Ramachandran plots tell you about the secondary structure of HA in these two conformations?

c) Plot phi angles vs. residue number for the two conformations on the same plot. (Each position on the x-axis should have two data points, representing the phi angle of that residue in the two structures). Make a similar plot for psi angles.

d) Based on the phi/psi angles, determine the number of residues that are alpha helical in one structure but not in the other. Define a helical residue as one where phi is between -57 and -71˚, and psi is between -34 and -48˚.

20.320 Problem Set 1
Question 2

#

Key to the function of influenza hemagglutinin is its pH-dependant conformational change in the endosome, fusing the viral membrane with the endosomal membrane and allowing release of the virus into the cytoplasm.

a) Of the principal forces responsible for maintaining the tertiary structure of a protein, which would be most strongly affected by the acidification of the surrounding environment?

Structural studies have shown that several histidine residues play a key role in mediating this pH-dependent conformational change of influenza hemagglutinin.

b) What property of histidine makes it especially suited to this role? Which other amino acid residues could potentially serve the same function? Be sure to justify your choices.

One way to determine which residues are vital to the structure and function of a protein is to align the sequences of many variants of the protein and look for conserved residues (those that are present in the same position in each protein variant). We can do this by comparing the hemagglutinins across various serotypes of human influenza A. On the Course website, you will find a document containing the amino acid sequences of several influenza hemagglutinins (H1, H2, H3, H5, H7, and H9).

c) Use CLUSTALW to find histidine residues that are conserved across all six sequences. Attach the CLUSTALW alignment, highlighting the residues you find. CLUSTALW is available on Athena clusters, or you can find a web client here: *http://www.ebi.ac.uk/Tools/clustalw2/index.html*

d) Assuming the pH of the acidified endosome is 4.5, which types of residues would you expect to see complexed with these key histidines? Based on your CLUSTALW analysis, identify the other residues that are likely involved with this pH-dependant transition.

e) Use Biopython to compute the distance between the alpha carbons of the conserved histidine residues you identified in Part (c) and the other conserved residues you identified in Part (d). Report the minimum distance you find for each conserved histidine. Does any pair of residues seem especially close together? Some hints:
   1. For this exercise, as with Question 1, only consider residues in the “B” chain of hemagglutinin at endosomal pH. This sequence is posted on Course website in FASTA format.
   2. It will help to repeat your CLUSTALW alignment from Part (c) with this new sequence – this will help you find the residues you are looking for.
   3. You can copy and paste the FASTA sequence directly into the list of hemagglutinin sequences you analyzed in Part (c).
   4. You should only be looking for residues that are conserved across all seven sequences in your new alignment.
   5. `residue["CA"].coord` returns the coordinates (*x, y, z*) of the alpha carbon of `residue`

20.320 Problem Set 1
Question 3

#

Cystic fibrosis (CF) is a genetic disorder caused by a mutation(s) in the cystic fibrosis transmembrane conductance regulator (CTFR) gene. CTFR, the protein product, is a traffic ATPase that transports chloride ions across epithelial cell membranes. Mutations lead to improper folding of CTFR and prevent proper chloride ion transport across these cell membranes. The ΔF508 mutation, aka the deletion of the phenylalanine (F) at position 508, is the most common mutation associated with cystic fibrosis.

a) Explain why the deletion of the phenylalanine (F) at position 508 might lead to misfolding (discuss the amino acid & its impact on structure).

The ΔF508 mutation along with several other known mutations that cause CF, occur in a region of the CTFR known as a nucleotide binding domain (NBD1). In an experiment, (Qu & Thomas JBC 271:13 1996 p. 7261-7264) NBD1 and NBD1 with the ΔF508 mutation (NBD1ΔF) were tested for folding yield at different temperatures.

b) Calculate the ΔG_{fold} (kcal/mol) at 37°C and 25°C of NBD1 and NBD1Δ using the following data: From the paper, we know that "at 2 μM final NBD1 concentration and 37°C, 63% of the wild type polypeptide folds into the soluble conformation, while only 38% of the ΔF508 assumes the folded conformation. At 18 μM final polypeptide concentration and 25 °C, 29% of the wild type domain reaches the native state in contrast to 19% of the ΔF508 mutant." Are these values reasonable? Explain.

This same group determined the free energy change of denaturation ΔG_D of wild-type NBD1 along with various other mutants from known CF cases at 37°C in a separate publication (Qu et al, JBC 272:25 1997 p 15739-15744), using somewhat different experimental conditions from the 1996 paper.

Protein	$\Delta G_{D,0}$ **(kJ/mol)**	$\Delta\Delta G_{D,0}$ **(kJ/mol)**
NBD1	15.5	
NBD1ΔF	14.4	-1.1
NBD1-R553M	16.6	1.1
NBD1ΔF-R553M	14.1	-1.4
NBD1-S549R	16.7	1.2
NBD1-G551D	16.6	1.1

c) Given the values of the ΔGs calculate the K_{fold} (ratio of folded to unfolded) of wild-type NBD1 and all of the mutants. ($\Delta G_{fold} = -\Delta G_D$)

d) Is the ΔG_{fold} for the wild type in Part (c) the same as your answer to Part (b)? Explain.

20.320 Analysis of Biomolecular and Cellular Systems
Fall 2012