

Problem Set 2: Structural Biology

Due Tuesday, April 14 at noon in the Bi 1 closet

HOMEWORK INSTRUCTIONS

- 1) Turn in your homework stapled to this cover page.
- 2) Use separate sheets of paper for your answers.
- 3) Write or type your answers neatly.
- 4) Put your name on each page of your answers.
- 5) Box your answers, please, so that the grader can find them.

Points may be deducted if you don't follow these instructions!

ANSWER KEY

Name: _____

Section #: _____

Mail Code: _____

TA Names: _____

Date and Time turned in: _____

Number of pages including this one: _____

AFTER YOU FINISH:

Go to the Bi1 moodle site at <http://www.courses.caltech.edu/> and take the homework survey

There are **3** questions. The number of parts to each question is listed at the beginning of each; be sure to answer all the parts!

Grade:

Problem 1 _____

Problem 2 _____

Problem 3 _____

TOTAL: _____

Figure Submittal:

A color printer is not required for this set. If you don't have a color printer, simply print out the required images in B&W. You may also submit your figures electronically in accordance with the guidelines at: <http://www.its.caltech.edu/~bi1/assignments.html>

Required reading:

Pymol Tutorial (posted with your problem set)

Recommended reading:

Biological Science 3rd edition, pp. 43-79, B-8 to B-9

Problem 1 – Amino Acid Hydrophobicity (20 Points – 5 parts)

One experimental method used to quantify the hydrophobicity of amino acids is to measure the octanol/water partition (P). As it turns out, octanol serves as a good model environment for the hydrophobic core of a protein or the interior of a lipid bilayer in a membrane. Therefore, comparing the relative concentrations of various amino acids in the aqueous and organic phases of the mixture can provide insight into three-dimensional protein structure. The partition is defined as:

$$\log P = \log(\text{concentration in octanol}/\text{concentration in water})$$

A. (4 Points) Would you expect His to have a higher partition at pH 5 or at pH 8? Why? Explain in 2-3 sentences.

His⁺ is a Lewis acid, so it should be neutralized to His under basic conditions. Since His⁺ is an ion, it will be more soluble in water and have a lower partition. Thus, the partition is higher at pH 8.

B. (4 Points) Another approach used to quantify hydrophobicity is to measure the change in free energy (ΔG) for the transfer of each amino acid from water to octanol. Data for each amino acid are shown in Figure 1. According to this scale, would a 20-residue polyalanine sequence be sufficiently hydrophobic to partition into the hydrophobic core? What is the value of ΔG for this transfer?

Since ΔG for a single alanine is about .5 units, the value for a 20-alanine sequence is about 10. Since $10 > 0$, the sequence will not enter the hydrophobic core.

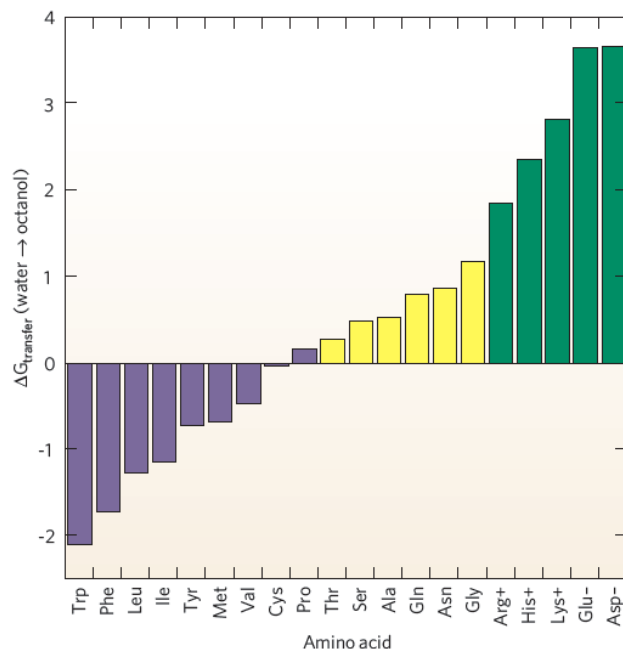


Figure 1. ΔG values for the transfer of amino acids from water to octanol

C. (4 Points) You now replace half of the alanine residues with leucine residues so that your peptide contains 10 alanines and 10 leucines. **What is the resultant value for ΔG ? Would this be expected to partition into the membrane?**

The value for leucine is about -1.2, for a total ΔG of -7. The mixed sequence will partition into the core.

D. (4 Points) Which amino acid best represents a polypeptide backbone with no sidechain? Where would a polypeptide of this amino acid partition according to Figure 1? What does that imply about the role of side chain hydrophobicity in driving the equilibrium in favor of amino acid insertion into the protein core? Explain in 2-3 sentences.

Glycine has no sidechain. Glycine has a large ΔG value and therefore would likely be outside the membrane. Large hydrophobic side chains are necessary for a protein to pass through the membrane.

E. (4 Points). Recent experimental studies of insertion into membranes indicate that predictions based upon transfer table are fairly accurate. However, octanol is obviously not a perfect model of the hydrophobic core, as it is a homogenous solvent. **What is another reason why the transfer of single amino acids into octanol is not a perfect model for the transfer of proteins into a membrane bilayer? Explain in 2-3 sentences.**

Transfer of individual amino acids into octanol is not a perfect model because the amino and carboxy groups of individual acids are charged, whereas they are not charged when incorporated into a polypeptide chain.

Problem 2 – Bacteriorhodopsin (35 points – 8 parts)

Biological membranes include a diversity of lipids, proteins and carbohydrates. The basic structure of a membrane is a bilayer made of phospholipids. Eukaryotic cells use membranes to physically isolate different organelles, including mitochondria, chloroplasts, trafficking vesicles, endoplasmic reticulum, the nucleus, etc. The composition of the membrane is different in each type of organelle.

Integral membrane proteins are proteins that are inserted into the phospholipid bilayer. They function in many processes, including transport of molecules across membranes and ATP synthesis. Membrane proteins constitute approximately 25% of all proteins and 50% of current pharmaceutical drug targets, so understanding their structures is an important goal in structural biology.

We will focus our attention on one membrane protein, bacteriorhodopsin, which is found in an archaea called *Halobacterium salinarum*. Bacteriorhodopsin is a light-driven protein pump involved in the process that generates energy from sunlight. To summarize briefly, bacteriorhodopsin uses energy from sunlight to pump protons from the cytoplasm to the external

medium, generating an electrochemical gradient along the membrane, which is used to produce ATP.

Bacteriorhodopsin was one of the first membrane proteins for which 3-D structural information was available. The protein contains several transmembrane domains and one molecule of retinal, a chromophore involved in converting light to chemical energy. In this exercise you will examine the structure of this membrane protein. You will first need to download the PyMol session file “Bacteriorhodopsin.pse” and the FASTA file “BR.fasta” from the Bi1 website:

<http://www.its.caltech.edu/~bi1/assignments.html>

A. (5 points) The topology of a membrane protein can be predicted computationally using methods that examine the hydrophathy, or degree of hydrophobicity, of the amino acids that comprise it. One hydrophathy-based prediction tool is TMHMM, located at:

<http://www.cbs.dtu.dk/services/TMHMM/>

You have been provided with the sequence of bacteriorhodopsin in FASTA format in the file BR.fasta. **Perform a hydrophathy analysis on bacteriorhodopsin by uploading this file into TMHMM. Hand in the computed hydrophathy graph as part of your homework.** If you wish to submit your figures for this set electronically, you must follow the guidelines for electronic submittals mentioned above.

How many transmembrane domains does TMHMM predict and at what residue does each predicted helix begin and end?

Figure. Seven transmembrane regions: 6-25, 38-60, 75-97, 104-126, 130-152, 164-183, 193-215. Note: Hydrophathy prediction methods are not normally this successful.

B. (5 points) Load the file Bacteriorhodopsin.pse, which contains an embedded PDB of a structure of bacteriorhodopsin solved to 1.5 Å resolution by X-ray crystallography, and also a PDB for the structure of sperm whale myoglobin solved to 1.4 Å resolution. The file is initially configured to display only the bacteriorhodopsin protein in cartoon form. Note that a membrane protein must be extracted from the membrane in detergent prior to crystallization. Because the structure does not include a lipid bilayer, we must analyze/interpret the structure in order to determine which parts are likely to be embedded in the lipid bilayer. **How many helices do you see in the crystal structure? Choose one helix. Specify the beginning and ending residues and measure the length (in Ångstroms, see p. 6 of PyMol tutorial for help with distance measurements). How does this compare to the result produced by the TMHMM prediction?** *For this question and the next, please note that if you inadvertently alter the preset selections or run into other difficulties with the display configurations, it is often easier to simply reload the session file and start over.*

Seven helices: 9-31, 36-63, 80-101, 105-128, 130-155, 164-192, 200-230. The lengths are 35.50, 36.05, 32.15, 39.90, 32.56, 43.30, 41.15.

C. (4 points) Color the “BRHydrophobic” selection yellow. (Selection: BRHydrophobic, Menu: C, Yellows → Yellow). Display the stick representations for the “BRHydrophobic” selection in addition to the cartoon (S Menu: Show→Stick). This selection has been constructed to highlight

the tryptophan and phenylalanine residues for you. **Do most of the hydrophobic side chains face the interior of the protein, the exterior of the protein, or are they roughly equally distributed? Does this distribution make sense for a membrane protein? Explain in 2-3 sentences.**

Some point to the outside, some to the inside: in a membrane, a hydrophobic side chain has no energetic reason to favor one position over the other.

D. (4 points) The main chain atoms in a polypeptide chain are polar, thus it is typically energetically unfavorable for a polypeptide chain, even one with hydrophobic side chains, to be transferred from a polar environment to a non-polar environment. However, the polarity of main chain atoms can be neutralized the hydrogen bonding patterns of secondary structures. **With this in mind, which protein structures could be accommodated in a cellular membrane: a single α -helix, a single β -strand, a small β -sheet (2-3 β -strands), or a β -barrel? (Note: GFP is an example of a protein with a β -barrel structure.) Why? Explain in 2-3 sentences.**

An alpha helix or beta barrel can be accommodated in a membrane because all polar groups on the backbone can be neutralized by hydrogen bonding.

E. (4 points) Change the background color by selecting “Display→Background→White”. Create an image file by selecting “File→Save Image...”. Print out the image, with the lipid bilayer drawn in directly on your printout, and submit it with your set. Be sure the protein is correctly oriented within the lipid bilayer.

Figure

F. (4 points) Hide the “Bactrhod” selection (Selection: Bactrhod, Menu: H, Everything) and display the “Myoglobin” selection as a cartoon. Color the “MYGHydrophobic” selection yellow. (Selection: MYGHydrophobic, Menu: C, Yellows → Yellow). Display the stick representations for the “MYGHydrophobic” selection in addition to the cartoon (S Menu: Show→Stick). This selection has been constructed to highlight the tryptophan and phenylalanine residues for you. **Do most of the hydrophobic side chains face the interior of the protein, the exterior of the protein, or are they roughly equally distributed? Does this distribution make sense for a soluble (i.e., not membrane-bound) protein? How does this compare to your answer for part C?**

Most (but not quite all) of the hydrophobic side chains point in, because they tend to avoid contact with polar water.

G. (5 points) Hide the Myoglobin structure. Display the bacteriorhodopsin structure as lines (Selection: Bactrhod, S Menu: Show→As→Lines) and color the entire structure blue (Selection: Bactrhod, Menu: C, Blues→Blue). Display the “Lys216” selection as sticks (S Menu: Show→As→Sticks). **What function does Lys216 serve in the bacteriorhodopsin molecule?** Please limit your answer to 2-3 sentences. (Hint: Display the labels of any groups interacting with the lysine side chain.)

The N of lysine binds to retinal.

H. (4 points) At which position in an α -helix would a proline residue be accommodated without disruption of the hydrogen bonding pattern: the N-terminus, the C-terminus, or the middle? Explain in 2-3 sentences. Can you find an example of a proline in this position in the bacteriorhodopsin structure? Hint, use the sequence above the structure to find prolines (single letter code is “P”), click on each P and see where it is in the structure.

In an alpha helix, the hydrogen on an N closer to the C terminus hydrogen bonds to the oxygen of the carbonyl on a residue closer to the N terminus. Proline has no hydrogen on its nitrogen available for hydrogen bonding. If there is a residue closer to the N terminus than proline, then this bond will not form, disrupting the structure of the helix. But if proline is at the N terminus, there is no residue for it to bond with, and the structure is not disturbed.

Problem 3 – DNA/Protein Interactions (45 Points – 9 parts)

Restriction endonucleases are enzymes that recognize short, specific sequences of DNA and cleave DNA at these sites. In this problem, you will use PyMol to examine the structure of a restriction enzyme bound to a short fragment of DNA. You will first need to download and open the PyMol session file “DNAProtein.pse” from the Bi1 website:

<http://www.its.caltech.edu/~bi1/assignments.html>

This file contains an embedded PDB file with some preset selections to assist you with this problem. The restriction enzyme (yellow; shown as sticks in an all-atom representation) is bound to a 16-mer of DNA containing its target sequence (blue; shown as sticks in an all-atom representation) along with two divalent cations (green). Typically, restriction enzymes use magnesium ions to catalyze the hydrolysis of the sugar-phosphate backbone of DNA. This structure was crystallized in the absence of magnesium to prevent the DNA from being hydrolyzed during crystallization. The crystals were then soaked in a calcium solution to repopulate the active sites with a divalent cation.

In this problem, you will attempt to identify the unknown restriction enzyme through a structural analysis of the protein/DNA complex.

A. (5 points) The backbones of the DNA strands create two grooves in the helical structure of DNA. The larger groove is referred to as the “major groove” and the smaller groove is referred to as the “minor groove”. Hide everything (Selection: All, Menu: H, Everything). Show the “StrandX” and “StrandY” as sticks (Menu: S, Show→As→Sticks) and color them blue (Menu: C, Blues→Blue). **Change the background by selecting “Display→Background→White”.** **Create an image file by selecting “File→Save Image...”.** **Submit the image, with the major and minor grooves clearly labeled directly on your figure, and submit it with your set.** **Which groove is more likely to host sequence-specific protein/DNA interactions? Why? Explain in 2-3 sentences.**

Figure. Most protein interactions occur in the major groove because its dimensions make the bases more accessible.

B. (5 points) Change your background back to black. Hide the “StrandY” selection (Menu: H, Everything). Display the CalciumX selection as a sphere (Menu: S, Show→Spheres). Color the StrandX selection by element (Menu: C, By Element→CHNOS [7th color scheme from the top]) and display the residue labels (Menu: L, residues). Show the CSiteX selection of residues as sticks, display the residue labels, and color the selection by element as before. **Identify one residue that interacts with the calcium ion. Measure the distance between the calcium ion and an oxygen atom in the sidechain of the interacting residue.**

The calcium ions interact with the Asp84 (2.3 Å, 3.5 Å) and Val94 (2.4 Å) residues of each monomer.

C. (5 points) Hide the “CSiteX” selection (Menu: H, Everything). **What is the distance from the calcium ion (green sphere) to the nearest phosphorus atom on the DNA backbone?** The phosphodiester linkage (-PO₄⁻) nearest the calcium ion is referred to as the “scissile phosphate”. This is where the restriction endonuclease cleaves the sugar-phosphate backbone of the DNA.

Hint: You may need to hide the calcium ion after completing your measurement to see the value displayed on your screen.

This distance is 2.96 angstroms.

D. (5 points) Determine the identity of the bases (A, C, T, or G) on the 5' and 3' side of the scissile phosphate. The next four bases are ATCT. Together, these six bases form the recognition sequence of the enzyme.

The scissile phosphate links the two nucleotides: 5'-AG-3'

E. (5 points) List the recognition sequence from part D alongside its complement. Indicate where the sequence is cleaved on each strand. Will this cleavage pattern result in blunt ends, sticky ends with 5' overhangs or sticky ends with 3' overhangs? The sequence bears a special relationship to its complement. **What is that relationship?**

5' -A | GATCT -3'
3' -TCTAG | A -5'

This cleavage pattern results in sticky ends with 5' overhangs. The recognition site is palindromic.

F. (5 points) How many hydrogen bonds does the base at the 5' end of the recognition site form with its partner on StrandY? Identify the heavy (non-hydrogen) atoms in each hydrogen bond and measure the length of each hydrogen bond (distance between the heavy atoms). Note that hydrogen atoms are not usually resolved in structures derived by X-ray crystallography, so hydrogen bonds must be inferred by the locations of the heavier atoms.

Two hydrogen bonds: oxygen (in T) to N-H (in A), 2.87 angstroms; N-H in T to N in A, 2.88 angstroms

G. (5 points) There are several online servers that you can use to determine which restriction enzymes will cleave a given sequence. Some restriction enzymes can cleave more than one sequence, so their cognate sequences may be listed with variables in place of bases, where **R** stands for puRines and **Y** stands for pYrimidines. (R = A or G, Y = C or T). For example, a restriction enzyme that can cleave the sequences TGCA or CGCG would be listed as YGCR. For this part, submit your answer to part **E** to the NEBCutter from New England Biolabs, a commercial vendor of restriction enzymes:

<http://tools.neb.com/NEBcutter2/index.php>

What are the potential candidates for the restriction enzyme present in this crystal structure?

The NEB Biocutter identifies Sau3AI, MboI, DpnII, BfuCI, BstYI, and BglIII as candidates. Only the last two have six base recognition sequences, so full credit should only be given for answers limited to these two candidates.

H. (5 points) In order to determine which restriction enzyme is present, we will look at the sequence-specific interactions between the protein and DNA at the 5' end of the recognition site. Hide everything (Selection: All, Menu: H, Everything). Show the "SeqSpec" selection as sticks (Menu: S, Show→As→Sticks) and color it by element (Menu: C, By Element→CHNOS [7th color scheme from the top]). **List all interactions between the protein and the first (5') base pair of the recognition site, including distances of any potential hydrogen bonds. What other base pairs could participate in all of the interactions that you have identified?** Note: In this context, a GC base pair is not the same as a CG base pair.

Bond between N of A and O of enzyme, 3.00 angstroms; bond between O of T and N of enzyme, 2.97 angstroms.

I. (5 points) Based upon what you learned in part **H**, you should be able to identify the restriction enzyme correctly. **Which restriction enzyme is present in the crystal structure?**

The interaction depends on the specific nucleotide in this position, so the enzyme must be BglIII.