

# Assignment 1

J. T. Sage

Physics G731

Fall 2007

Due Friday, Sept. 14.

The goal of this assignment is for you to develop the ability to access for yourselves the huge amount of structural information that is available on biological molecules.

Go to the protein data bank (<http://www.rcsb.org/pdb/>) and download the atomic coordinates for one protein that interests you. If you are not sure what to pick, the “Molecule of the Month” feature provides basic information on several interesting proteins, together with the 4-character accession code that will allow you to obtain the coordinates. “Protein Spotlight” at <http://www.expasy.org/spotlight/> is a similar feature.

1. Name the protein and list any molecular groups other than amino acids in the structure. These should be identified as “hetero” groups in the coordinate file. Identify the number of polypeptide chains in the structure (there may be one or more, depending on the protein that you choose) and the number of amino acids in each chain.
2. Locate and download a molecular viewer. I recommend RasMol, which is available free at <http://www.openrasmol.org/> or <http://www.umass.edu/microbio/rasmol/index2.htm>. A manual is also available. (Other popular viewers include DeepView and VMD.) Load your structure into the viewer. Restrict the view to a single copy of the molecule, if more than one is shown. Use the ‘set background white’ command to save printer ink, and print three views of the molecule to turn in:

- (a) A spacefilling view, which shows all atoms as spheres and gives an idea of the actual space occupied. To give an idea of the charge distribution on the surface of the protein, show acidic amino acids in red, basic amino acids in blue, and other amino acids in white.
  - (b) A backbone view, showing only a tube connecting the alpha carbons. This reveals the underlying structure. Display all the atoms of any “hetero” groups that you identified in the last part in a “ball-and-stick” format.
  - (c) A ball-and-stick model of a single amino acid. On your printout, label (by hand) the alpha carbon, the amide nitrogen, the carbonyl oxygen, and the side chain.
3. Find the amino acid sequence for the protein. Estimate the following basic physical properties of the polypeptide (that is, omit any prosthetic groups): the mass (in Da), the volume (in nm<sup>3</sup>), and the net charge (in electronic charges) at pH 7. (For full credit, use the complete sequence given in the file header—you may find that some amino acids are omitted from the coordinate listing because their positions have not been determined.)
4. List a sequence of DNA bases (A, C, T, or G) that could code for the first five amino acids in the protein, including a start code, and for the last five amino acids in the protein, including a stop code. (Because of the redundancy in the code, there may be more than one possible answer.) If your protein contains more than one chain, do this for chain A only.