

Algorithms in Bioinformatics I, WS2002/3  
**Assignment sheet # 12**

Daniel Huson

January 14, 2003

**1 Protein structure: PDB files (2 points)**

Three dimensional protein structures are often described using the PDB file format. Go to the PDB database (<http://www.pdb.mdc-berlin.de/pdb/>) and download the PDB files corresponding to the following identifiers: 2BOP, 1RNB and 1CD8, plus two others that you find interesting. What information can you obtain about these proteins? Please give a short summary.

**2 Parsing the ATOM lines in PDB files (3 points)**

Design and implement a Java class PDB that scans a PDB file and parses all ATOM lines in the file. Print statistics counting how many residues of each type occur in the sequence. (Hint: you may first want to search the web to find a description of the PDB format).

**3 Drawing a protein structure (5 points)**

Add a method to your PDB class that pops open a window and draws a simple picture of the protein, using small discs to represent the atoms and joining consecutive atoms by lines. Color atoms contained in  $\alpha$ -helices red, atoms contained in  $\beta$ -sheets blue, and others black. Run your “viewer” on all five PDB files.

**Due by 10am, Monday, January 20, 2003.**