Assignment 4: Creating Ramachandran Plots from Known Protein Structures and the NDB

1. **Ramachandran Plots**. Our goal is to generate Ramachandran plots for a particular amino acid residue or a group of residues. We have assembled a database of 50 proteins based on the article: M.A. Williams, J.M. Goodfellow, and J.M. Thornton, "Buried Water and Internal Cavities", *Protein Science* **3**, 1224 (1994). These files can be found in the PDB directory of Insight II prepared for our course.

   Each of you must generate two Ramachandran plots. Check the chart below for your particular assignment (on the basis of your last name).

First letter of your last name	Subgroup 1	Subgroup 2
A–N	**Ala**, **Val**, **Leu**, **Ile**	**Gly**
O–R	**Asp**, **Asn**, **Glu**, **Gln**	**Pro**
S–V	**Lys**, **Arg**, **His**	**Ser**, **Thr**
W–Z	**Trp**, **Tyr**, **Phe**	**Cys**, **Met**

   Each plot should have the data points for the (ϕ, ψ) dihedral angles, corresponding to the assigned group of residues from all the proteins in the database. To find the values of the ϕ and ψ dihedral angles in a protein you can use $\boxed{\text{Protein}}$/List command from the **Biopolymer** module. Record these angles to a file. You can use the Fortran code posted on the website (aa_select.f) for searching the $\boxed{\text{Protein}}$/List output files (called PDA files) for the dihedral angles of selected residues. Alternatively, you can write a suitable program in a different language.

   If you use the code from the website, you will need to edit it to replace all occurrences of the names ALA, VAL, etc. by the abbreviations of the residues you are searching for. These abbreviations must be capitalized. Compile the code and execute it with each PDA file as input. The output file should contain only two numbers per line, the ϕ and ψ angles for the specified residues. Check the numbers for correctness by comparing a few lines from this output with the numbers in the PDA file.

   Note: For plotting, you must use Insight II so that all plots are uniform in size. We will overlay them in class! Follow the instructions below.

   Prepare a file with your data points collected from all the proteins for each of the assigned groups of residues. These files should have the extension tbl (filename.tbl). In addition, you must format these files for Insight II by inserting the 12 lines as indicated:

```
#
TITLE: Phi (deg)
MEASUREMENT TYPE:  Ang
UNITS OF MEASUREMENT: Deg
FUNCTION: dihedral
#
TITLE: Psi (deg)
MEASUREMENT TYPE:  Ang
UNITS OF MEASUREMENT: Deg
FUNCTION: dihedral
#
#
−34.5      144.9                    This is the first line of your numeric data.
:
:
```

Note that at the top of the file there is space for your own comments and you can use as many lines as you need. Insight II will start reading the data from the line without the character # on the first column. That first line should be as indicated above.

If you have done everything correctly, you are ready for plotting. Press the $\boxed{\text{Graph}}$ button on the left of the screen and select Get. Specify the data file name, dihedral1 as X_Function and dihedral2 as Y_Function. Keep the New_Graph on and execute. The zigzag appearance of the plot now requires fixing. Move the graph box near the bottom-left corner of the screen. First, connect to the object (your plot) with the command $\boxed{\text{Transform}}$/Connect. Second, move the plot by clicking on it and dragging to the desired position. The scatter plot will be produced in the next step. Select Point (only!) from the $\boxed{\text{Graph}}$/Modify_Display dialog box. Choose Star as the Point Symbol, scale it ten times and execute. Use $\boxed{\text{Graph}}$/Threshold to change the minimum value to −180.0 and the maximum value to 180.0 for both, X Graph Axis and Y Graph Axis. Scale each axis 4 times with $\boxed{\text{Graph}}$/Scale_Axis command. Change the color of any white elements in your plot with $\boxed{\text{Graph}}$/Color command. Change the background color to white. Print your plot (see instructions below) and repeat the procedure for the second data file.

Prepare a transparency for each of the two plots and bring to class. (Hand in the printed version of the plots with your homework.) Also e-mail the $\{\phi, \psi\}$ files you generated, sending each file separately and specifying your name and the group of residues in the Subject line.

2. **The** NDB. The next part of the homework will acquaint you to the Nucleic Acid Database, NDB, on which we will have a guest lecturer. The web address for NDB is: ndbserver.rutgers.edu/

First, explore the database to discover what is available, and look through the newsletter archives for current update information. Then, describe the structures available in the database and the numbers in each class (e.g., B-DNA, ribozymes). Explore the different features of NDB. There are many exciting structures and utilities.

3. **Sugar Conformations for Canonical A, B, and Z-DNA**. To appreciate the different features of canonical A, B, and Z-DNA forms, choose through the NDB entries one unmodified form in each DNA class with the largest number of residues possible. You can view the structure within NDB, or by porting the PDB file to Insight II and using the Molecule /Get from the **Viewer** module (though the latter may be more difficult).

Learn how to use the Report facility in the NDB Query Interface to generate a list of all the sugar pseudorotation angles (P) for each deoxyribose in each of the three structures you have chosen for your study. (Remember that there are two sugars per base pair). Print a table for each structure.

Now, on one figure for all the three structures, plot P versus the residue number. Exclude the two terminal base pairs of each structure for plotting purposes. You may need to connect the points corresponding to each structure for clarity.

Label the pattern clearly to indicate the A, B, and Z-DNA data. Hand in this plot, with a description of the patterns you had identified for the sugar conformation for the three canonical DNA forms, indicating the specific structures you have chosen.

Summary of Items to Hand In:

(a) Ramachandran plots for the two subgroups of amino acid residues assigned to you.
(b) Data on the structure class types and the amount of structures in each class available in NDB.
(c) The figure with P versus residue number for the A, B, and Z-DNA forms, with complete discussion.

Background Reading from Coursepack

- B. Cipra, "Computer Science Discovers DNA", in *What's Happening in the Mathematical Sciences*, pp. 26–37 (P. Zorn, Ed.), American Mathematical Society, Colonial Printing, Cranston, RI (1996).

- M. Gerstein and M. Levitt, "Simulating Water and the Molecules of Life", *Sci. Amer.* **279**, 101–105 (1998).

Printing Instructions

To save ink, change the background color from black to white at each printing of a color or black/white image. Go to Session /Environment, press the Background button and change the color to white.

Proceed to File , choose Export_Plot, postscript, Gray_Scale and optionally Ball_and_Stick. Use Save_Device_File to save a postscript file (the file will have the ".ps" extension). This file can be printed on any postscript printer you can access. Hand in your printout as part of the homework.

Fortran **program for selecting Ala, Ile, Val, and Leu data**
(see website)