

Rasmol

(Learning how to do simple structure visualization)

November 20, 2002

Sequence Analysis & Consulting Service

Seminar Series

Susan Johns

(johns@cgl.ucsf.edu)

Launch a telnet session on socrates by clicking on the socrates icon in the long window on the left hand side of the screen and responding with your account name and password.

Then, copy over to your account some data to work with by entering the following commands. [The first one does the actual copying. And the second one moves you into this area with the data.

```
ras_setup<rtm>  
cd images<rtm>
```

As a result of these commands, you are now in a new sub-directory called **images** in your account where the necessary files are located.

Rasmol is a structural visualization program that runs on a number of platforms. Originally written by Roger Sayle, it has now been taken over by Herbert Bernstein. UC Berkeley even has a Rasmol enhancement project for the original version of the package.

Rasmol can be used as a simple viewer or as a development tool to display structural concepts through scripts. To use the software in more than the most simplistic mode requires the user to be familiar with the fundamentals of structure data files (PDB).

PDB (Protein Data Bank) is the single worldwide archive of structural data of biological macromolecules. Originally run by the Brookhaven National Laboratories, it is now managed by Research Collaboratory for Structural Bioinformatics (RCSB).

When PDB was started in 1971 as the archive for biological macromolecular crystal structures, there were only 7 structures. In the 1980's the number of deposited structures began to

increase dramatically due to changes in crystallographic technology, NMR structures inclusion and changes in the community views on data sharing.

Access to the data has changed over the years (web vs. magnetic tape). The sources of the structures now includes x-ray structures, NMR, cryoelectron microscopy and theoretical modeling. The user base has changed from a select set of experts to a very diverse group of researchers in biology, chemistry and computer science, educators and students at all levels.

In October 1998, the management of PDB was taken over by Research Collaboratory for Structural Bioinformatics (RCSB). Their goal is to create a resource based on the most modern technology to facilitate the use and analysis of structural data, this providing an enabling resource for biological research.

Due to the age of the resource, its data format may seem archaic to current users. Developed at the beginning of the computer age, the database has a rigid ascii format for its various types of data lines.

The information is presented in a specific order in lines that need to start with a specific term. In general, the compound type, its name, source, author and journal information precedes a series of remark lines with comments. The SEQRES lines contain the reported structure's residues given in three letter code. Data of the type of residues, formulas and crystal information is given next. Followed by the actual coordinate data, given in a single line for each solved atom of the structure.

While changes have been made over the years to meet the needs of new data sources, the format is a space intensive means of storing data. PDB does have a file format FAQ page to assist users with questions.

While Rasmol can be run in stand alone mode on the computer in front of you, we will be using socrates as the source of program and the data to use. To do this requires that we have an x-emulator program running on the machine. Go to the SACS folder in the Macintosh HD window. Double click on the MacX Application folder to open it. Double click on the MacX icon (a triangle formed by three connected balls) to start the program. A window will come up showing the loading of the software and then disappear.

In your telnet window, do a directory listing to find out how many files are there with the pdb file extension. These are the files that will be explored initially.

Use the more command to look at these files, one at a time. Notice that the second and third pdb files are the example files looked at previously. The first file is for the ORNITHINE AMINOTRANSFERASE structure.

Start up rasmol by entering the term rasmol at the socrates prompt. Two windows come up. The white contains information on the program and the RasMol> prompt. The black one is for displaying the structure. Move the black window on the terminal screen so that you have as little overlap as possible with the white window. Have the browser window sticking out behind the black one.

On socrates, data to be used with the program is best kept in the same directory as where the program was fired up. Otherwise, the complete directory path needs to be given for each data file used. Since there are three files here to use, they can be loaded into the program by entering a command at the RasMol> prompt similar to this one, `load file<rtm>`.

Load in the data files, one at a time, starting with the smallest one (9lyz.pdb). Notice that the information lines given in the command window seem to indicate that there is no data in this file and yet there is a structure in the data window. (You may need to click on the black window to make the structure visible.) This is due to the fact that the program expects to be using protein data and this file contains structural information for a substrate which uses "HETATM" lines instead of "ATOM" lines to store its coordinates.

The structure is given in as a wireframe image, the default display mode. The bonds are very thin. To increase the size of these bonds enter `wireframe 0.2` at the prompt. Rotate the structure around using the scroll bar controls at the bottom and tight side of the display window.

Explore the display options by pulling down the "Display" menu on the image window. Wireframe is what is shown on the screen now. The "Backbone", "Ribbons", "Strands", and "Cartoons" display modes are for use with protein structures and will cause this structure to disappear from the screen. The default colors shown are the standard for the various atoms, white for carbon, blue for nitrogen and

red oxygen.

When you are finished exploring the structure, enter `zap` at the RasMol> prompt to clear the image window.

Next, load in the middle sized data file (2mlt.pdb). This one is of a small protein, melittin. It is comprised of two chains, A and B and has two sulfate ions (the yellow and red structures).

Since this is a protein structure, the protein only display options will work on the protein parts of the structure. The sulfate ions will disappear however from the image. From the "Display" menu, select "Cartoons". The small sticks have been replaced by large structures denoting the secondary structural elements contained in the PDB file. In this case, four helices and two small random sections. Then go to the "Colours" menu and select "Structure". The image is now colored by the default coloring scheme of the program. Random sections are white, helices are hot pink, strands if there are any are yellow and turns blue.

Explore how these changes impact the other display options of the program. Since there are two chains in this structure (A, B), try to select the members of one of these chains and color it green. You will need to enter the following commands at the prompt to do this.

```
select (*chain_name)
color green
```

You may need to click in the display window to have this change take place. When you are finished, enter `zap` to clear the screen.

Load in the largest data file (2can.pdb). This one is of a medium sized protein, ORNITHINE AMINOTRANSFERASE COMPLEXED WITH L-CANALINE. It is comprised of three chains, A, B and C, and has the substrate canaline associated with each chain.

Color the three chains red, yellow and green using the instructions given before.

```
select (*chain_name)
color desired_color
```

You may need to click in the display window to have the change take place.

The three canaline molecules are given in the structure at atoms 3820-3838, 7683-7691 and

11536-11544. To have these atoms displayed, requires command lines as given below. The final image will be a CPK model of the canaline. Color all three of the groups this way.

```
select (atomno >= 3830) and (atomno <= 3838)
colour white
cpk
```

When you are finished, enter **zap** to clear the screen.

Rasmol can also be used to create scripts to display structural concepts. In this next section, you will use three established scripts to get an idea of what can be done with Rasmol in this context.

The three scripts are called, bases2.script, myo.script and tryp.script. These scripts put text information in the command window that explains what is being shown in the display window. The scripts have pauses built into them so that a viewer can rotate the structures to better understand them prior to moving on to the next image.

At the command window prompt enter the following command and follow the instructions in the command window to work your way through each script. Do all three scripts.

```
script script_name
```

When finished with the scripts, enter **zap** at the prompt to remove the structure.

Rasmol has a whole series of commands that allow you to work with the image in the black window. To explore these commands, load in the file called **title.coords** into the program. This file contains data that spells out the word, trypsin.

Notice that the term trypsin is upside down on the screen. To get the data in its proper orientation enter rotate x 180 at the RasMol> prompt.

Enter the following series of commands at the RasMol prompt to color the word yellow and create spheres of size 300 on the screen.

```
select all
colour yellow
spacefill 300
```

All the extra lines are gone and just the yellow spheres that appear to be connected into the

letters of the word. Now enter zoom 300 and see what happens.

Move to the beginning of the word by entering **translate x 100** at the prompt. Use the cursor to select first one end of the cross bar on the T and then the other end. Information is shown in the command window showing the atom name, its number in the structure and group and chain data.

Enter the following list of commands at the prompt to determine the distance between the centers of the two selected atoms on the T. Replace #1 and #2 in the example with the actual atom numbers.

```
set monitor on
set fontsize 15
colour labels blue
monitor #1 #2
```

Distance is given in blue on the screen. To turn off the monitor, enter set monitor off.

Return to the original size of the word by entering **reset**. Rotate it again to get the correct orientation. Zoom in again, zoom 300. You can print the image on the screen using the write command at the prompt, but that would use lots of ink. Change the color of the background by entering, background white. Now export the image on the screen by entering the following.

```
write ps myfile.ps
```

Exit from the RasMol program by entering quit at the prompt.

Print off you image by using the following command in the telnet window.

```
lpr -Pclass myfile.ps
```

It is now time to run Rasmol on the local machines. To do this requires some data to work with. Go to the PDB site and enter the term 6lyz into the "Enter a PDB ID or keyword" box. Click on the "Find Structure" button to make the search. From the results page, click on the "Download/Display File" link. In the "Download the Structure File:" section of the page click on the link that represents a PDB formatted file without compression. Respond to the warning window by clicking on "Save file" and put that file into the Temp location on your machine.

On the local machine, Rasmol is located in the RasMol v2.6 folder. Once the folder is open,

double click on the RasMac v2.6 icon to start the program. The two windows come up. Separate them so there is little overlap and then enlarge the size of the display window. From the "File" menu at the top of the screen, select "Open" and then go to the Temp folder to select the 6lyz file.

Lysozyme is a mixed protein with helices, sheets and turns recorded in its structure. Use the "Cartoons" option from the "Display" menu and the "Structure" option from the "Colours" menu to display lysozyme in its determined secondary structural elements.

Explore this structure comparing the options that you used in the earlier socrates session with what is available in this version of the software.

When finished select the "Quit" option of the "File" menu to exit the program.

Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov and Philip E. Bourne, " The Protein Data Bank", Nucleic Acids Research, 2000, Vol. 28, No. 1 235-242

URLs used in this seminar:

Roger Sayle's latest version of RasMol (getting and installing)
<http://www.umass.edu/microbio/rasmol/getras.htm>

Herbert Bernstein's latest version of RasMol (getting and installing)
<http://www.bernstein-plus-sons.com/software/rasmol/>

UC Berkeley's enhanced version of the original RasMol
<http://mc2.CChem.Berkeley.EDU/Rasmol/>

PDB format description
http://www.rcsb.org/pdb/docs/format/pdbguide2.2/guide2.2_frame.html

PDB FAQ (file format)
<http://pdb.rutgers.edu/format-faq-v1.html>

Citing the PDB
<http://www.rcsb.org/pdb/citing.html>

Latest paper on the PDB
<http://nar.oupjournals.org/cgi/reprint/28/1/235.pdf>

RasMol online documentation
<http://www.sacs.ucsf.edu/Training/rasmol/rasmol.html>

RasMol postscript documentation
<http://www.sacs.ucsf.edu/Training/rasmol/manualUS.ps>

RasMol refcard
<http://www.sacs.ucsf.edu/Training/rasmol/refcardUS.ps>