

Molecular visualization using Rasmol

Robert Jorissen, Ludwig Institute for Cancer Research, Melbourne branch

The purpose of this practical session is to introduce you to examining the structures of proteins and associated biomolecules. We hope that you will begin to start thinking about the structure of your protein of interest as part of thinking about its behaviour.

In this session, you will download a number of files which each contain the coordinates of one or more proteins. Then, you can render the protein(s) in different ways, including rendering selected regions of the protein(s), to investigate different aspects of structure. The structure can be analysed with respect to how it relates to function. The molecular structure viewer Rasmol will be used in this practical session to view various biomolecules.

This practical session has eight components:

1. Downloading a protein coordinate file from the Protein DataBank.
2. Initial viewing of the protein coordinate file using Rasmol.
3. Selective display of items in Rasmol
4. Using Rasmol to look at the structure of G-CSF and its receptor binding
5. Using Rasmol to analyse enzyme-substrate interactions
6. Visualization of sequence information using a homologous protein
7. Visualization of a protein - DNA complex
8. Using Rasmol for images and presentations

The first three sections cover finding, retrieving and viewing the structure of a protein. The next four sections cover using Rasmol to look at systems of biological interest. The last section discusses the construction of images of the molecular system using Rasmol and using Rasmol from a PowerPoint presentation.

1. Downloading a protein coordinate file from the Protein DataBank

The RCSB Protein DataBank (PDB) is the major repository for the three-dimensional structures of biomolecules such as proteins and nucleic acids. The PDB is accessible over the Internet. In this part, you will use the PDB site to search for a structure of the insulin receptor kinase and then download it.

Get to the main PDB site:

- Start a web browser program such as Netscape Navigator. The application is accessible from the Network Applications folder at the bottom of the screen.
- At the address line, type in <http://www.rcsb.org/pdb/> (An easier-to-remember method is to go to the Google search engine at <http://www.google.com> and search for "pdb". The top hit should be the RCSB PDB site.)
- Scroll down the page until you find the "Search the Archive" section.
- Type in some key words about the protein you are looking for. Note that you may need to use the word "and" between keywords. Typing in "protein kinase" will return entries which contain these two words as a phrase whereas typing "protein and kinase" will return entries which contain the words "protein" and "kinase" but not necessarily near each other. (Similarly, you can use the words "or" and "not" in your query to aid your searching.)
- Press the button marked "Find a structure".

You should now see a list of entries. Each entry begins with a four character code, beginning with a number and then three other characters which can be letters or numbers. There will also be a brief description of the molecular system whose structure was determined.

- Find the entry called 1ir3 and click the hyperlinked word "EXPLORE" for this entry. This entry is the form of the insulin receptor kinase which is phosphorylated and has substrates bound.
- If you did not see the entry with the code 1ir3, go back to the search page and try another set of keywords, or at least put "and" between some keywords.

You should now be at the main summary page for the PDB entry 1ir3. The main section of the page is a summary about the protein and the techniques and associated data about solving the structure. On the left hand side of the page are a number of links including:

- Summary information - this is the link for the current page
- View Structure - allows you to quickly view the structure if your browser is suitably configured
- Download/Display File - used for viewing and/or downloading the text file that contains the coordinates of the structure plus other associated information
- Sequence Details - the protein/nucleic acid sequences whose structures were determined

- Click on the link “View Structure” on the left hand side of the web page.

There are a number of applications that can be used to view the structure. You should have Java activated for your browser. If so, then you should see a button marked “QuickPDB” below the links to the different options.

- Click on the button “QuickPDB” If the button is not present, change the preferences of your browser to activate Java and then reload the page.

A second window should appear containing a black window showing the C_α trace of the protein in red. The lines connect the C_α atom positions from successive amino acids. You can rotate the molecule by starting with the cursor in this window, pressing the mouse button (left mouse button if you are using a PC) and then moving the mouse. The α-helices should be apparent. The largest β-sheet can be made out with a bit of effort.

Having glanced at the structure of your protein, you can now download the file which contains the coordinates.

- Close the QuickPDB window
- In the browser, click on the link “Download/Display File”
- Scroll to the bottom of the page to section titled “Download the Structure File”.
- Click on the letter X corresponding to no file compression and PDB format. If you see a warning prompt displaying several options, click on the button marked “Save File As...”.
- For the purposes of this practical session, save the file as “1ir3.pdb” (change to lower case for later convenience if it is not already) in the Bioinformatics subdirectory on the desktop. This directory was set up for you to use for the practical sessions in the Bioinformatics workshop.

You should also view the so-called header of the PDB format file which contains information that is relevant to the solved structure.

- Go the PDB browser window that you used to download the PDB file of the insulin receptor kinase and scroll up to the section entitled “Download File”. If you do not have this window, go to the PDB site, find the PDB entry with PDB code 1ir3 and then select the link marked “Download/Display File”.
- Click on the link corresponding to PDB format, HTML format and “header only”. This corresponds to the so-called header region of the PDB-format file.

There are a number of REMARKS in the header of the PDB file that include: the identities of the main components of the biomolecular system (COMPND); how the biomolecules were generated (SOURCE); the journal article that describes the structure (JRNL); the method of structure determination (EXPDTA); missing residues (REMARK 6); and missing atoms (REMARK 470). In some crystal structures, atoms of flexible sidechains may not be able to be resolved and residues from loops might also be missing.

2. Initial viewing of the protein coordinate file using Rasmol.

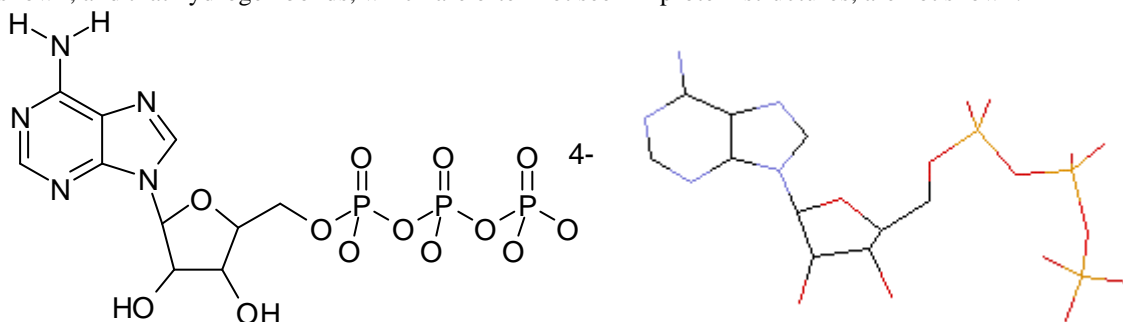
Rasmol was one of the first molecular visualization programs that could be used to view protein structures on a PC or Macintosh. Although there are more recent programs available that can be used (and some of there are Rasmol derivatives), Rasmol remains a useful program for the visualization and analysis of biomolecules. In this section, you will take a first look at the structure with PDB code 1ir3 and determine which molecules are present in the coordinate file.

- Begin the Rasmol program by selecting “RasMac v2.6” from the Applications folder at the bottom of the screen. (Note that the PC Windows version of Rasmol is sometimes called RasWin.)
- In the “File” menu, select the “Open ...” option. Then, go to the Bioinformatics directory, where you saved the file 1ir3.pdb, and open the file 1ir3.pdb.

The lines you see on the screen are the bonds between each atom in the macromolecular system. Each vertex (end of line and where lines intersect) represents an atom and each atom should be colored according to its atom type. Some of these colors are:

light grey	Carbon
red	Oxygen
white	Hydrogen
light blue	Nitrogen
yellow	Sulphur
orange	Phosphorous

For example, the molecule ATP which you will have seen in textbooks looking something like the diagram on the left will be rendered in many molecular visualization programs resembling the diagram on the right. Note that bond orders are not shown, and that hydrogen bonds, which are often not seen in protein structures, are not shown.



Unless you are familiar with the molecule(s) you are looking at, the interpretation of what you are currently seeing on the screen is usually difficult. In this section, you will use some alternative representations to get a better feel for the structure(s) you are looking at.

The table below indicates which mouse buttons (and sometimes keys) are required to rotate, translate etc the molecule in Rasmol. Moving the mouse while holding the required mouse button/key(s) will perform the operation as listed below.

	Macintosh	PC/Linux/UNIX
Rotate molecule	Mouse button	Left mouse button
Translate molecule	Apple key*	Right mouse button
Zoom (move mouse up/down)	Shift key/ mouse button	Shift key/ left button
Rotate molecule about the plane of the screen	Shift key/Apple key*	Shift key/right button

* The Option key has the same effect as the Apple key in the Macintosh version of Rasmol.

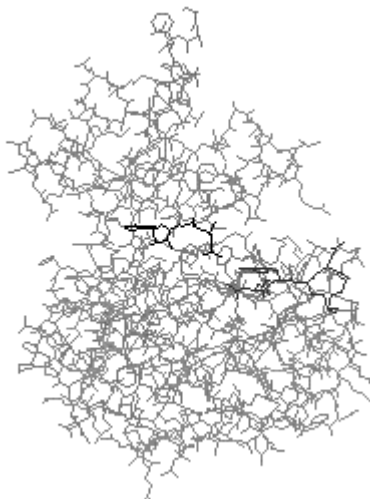
Yes, go and use these buttons to move around the molecule. You will be using the rotate, translate and zoom features frequently as you look at various aspects of its structure.

The menu items have several options for changing how the molecule is rendered. These are handy for obtaining an initial idea about the molecular system whose coordinate file you downloaded.

To gain an initial idea about what molecules are present, go to the **Colours** menu and select the **Chain** option. For this section, this sort of instruction will be written as:

- Colours: Chain

It may be useful to rotate your molecule to that shown in the figure below. This is an orientation which clearly shows the two substrates and their proximity to one another.



If there is only one molecule, then colouring “by chain” will result in the molecule being colored white. Otherwise, the molecules will have different colors. Protein chains are colored blue, green, yellow etc. Non-peptide and non-nucleic acids are usually colored red. Importantly, some molecules with one heavy atom such as water or metal ions may not be seen with this representation.

To view the system where each of the atoms are rendered as spheres, perform the menu selection:

- Display: Spacefill

You will see the molecules you saw previously rendered as spheres. Also, you will notice additional red spheres scattered around the protein. These are the oxygen atoms of water molecules which were resolved in the crystal structure of the insulin receptor kinase. The hydrogen atoms were not observed, as is the case for many crystal structures of macromolecules.

When you start Rasmol, two windows are generated. The first window is the graphics window which contains the image of the molecular system. The second window is the command window where you can type commands and see the results of clicking on atoms (identity of an atom, a bond distance etc). Find this window and move it so that the other Rasmol window does not obscure it. (If you are using a computer using a Microsoft operating system, you may have to move the cursor to the bottom of the screen to bring up the so-called Start Bar and then click on the icon to bring up the RasMol Command Line window.)

Click on some atoms of the molecular system. In the command line, you will see the name of the atom and the name of the residue (as a three letter code) of the atom you clicked on. If you click on a protein or nucleic acid atom, you may also see a letter that identifies which chain the atom is part of.

Q: How many molecules are present? (Count water molecules collectively as one group) What are the three letter codes for the non-protein molecules? How are the different polypeptide chains distinguished?

The menu items can also be used to look at the fold of any proteins present. Try the following menu commands:

- Display: Cartoons

This draws the molecule as a cartoon that highlights the helices and β -strands of any polypeptide regions. (Rasmol also draws an analogous diagram for any nucleic acids but they are less impressive to look at.) Molecules other than peptides and nucleic acids are not drawn (with few exceptions such as heme group). To colour by secondary structure element:

- Colours: Structure

Q: The insulin receptor kinase can be viewed as two segments between which one of the substrates is bound. What is the most frequently occurring secondary structure element in each of the top and bottom segments?

Another method for viewing the fold of the protein is to use the menu option:

- Display: Backbone

Using this representation, all of the C_{α} atoms of the polypeptides are joined together by thick lines. (For nucleic acids, the phosphorus atoms are joined up.) Clicking on the atoms leads to the identity of the atom being shown in the command window. Confirm that the atoms joined together with this backbone representation are all C_{α} atoms.

3. Selective display of items in Rasmol

In this section, you will begin to use the command line window in Rasmol. While use of the menu items is a quick way to change the display of the molecular system you are looking at, much of the power of Rasmol derives from the command line commands that allow you to select particular parts of the molecular system and alter they are displayed.

You should still have the insulin receptor kinase structure in Rasmol. Bring up the command line window. At the prompt, type

zap

Your molecule will now have disappeared. To get it back, type

load HD:Desktop Folder:Bioinformatics:lir3.pdb

where “*HD*” is replaced with the name of the hard disk e.g. “Mac 614”. This command loads the molecule with the name *lir3.pdb* which is in the subdirectory “*HD:Desktop Folder:Bioinformatics*”. Note that the nomenclature used to specify the directory and file name is specific to the Macintosh operating system. For the PC running a Microsoft operating system, you would type in a command like:

load c:\ Bioinformatics \lir3.pdb

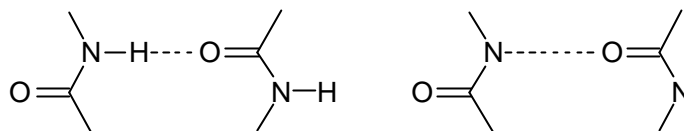
Your molecule should now be back on the screen with wireframe rendering and colored according to atom type.

Changing how the molecular system is displayed consists of two steps: selecting the part of the molecular system whose appearance you wish to alter and typing a command to render the selected part of the system in a particular way. The examples in this section will not only demonstrate the power of selections but illustrate some aspects of protein structure.

select all
wireframe off
select mainchain
wireframe 10

The first two commands select all of the atoms in the system and turns off the wireframe style of rendering. The third command selects only the protein mainchain atoms (N, C_{α} , C and O). The fourth command displays these selected atoms in wireframe representation. The number “10” refers to the radius of the bonds. You can change this to a larger number such as 50.

- Use the zoom facility of Rasmol to get a close up of a part of the protein. Identify the mainchain atoms N (amide nitrogen), CA (C_{α} carbon), C (carbonyl carbon) and O (carbonyl oxygen) by clicking on these atoms.
- Type **hbonds** in the command window. Identify some hydrogen bonds and convince yourself that hydrogen bonds are being made between the carbonyl oxygen atom and the unseen amide hydrogen atom. Hydrogen atoms are present in only a few crystal structures but the hydrogen bonds can generally be deduced from the heavy (non-hydrogen) atom positions.



The amide nitrogen atom and the carbonyl oxygen and carbon atoms should be crudely co-linear; their angle should be between greater than 120° in magnitude and the N-O distance should be less than 3.5 \AA . (N.b. these are approximate values)

- Type **hbonds off** in the command window. Use the zoom feature to view the entire protein again.

At the command line, type:

restrict mainchain and sheet

The command “restrict” performs a selection in the same manner as the “select” command. However, the “restrict” command also removes from the screen any part of the protein that is not part of the selection. This command is a quick way of removing non-selected parts of the molecular system from the screen.

The selection contains two elements which are linked by the word “and”. The two components of the selection are “mainchain”, which has been described, and “sheet”, which refers to any part of a protein which are assigned as present in a β -sheet. The “and” statement requires that each atom selected fulfils the criteria of being a mainchain atom and being part of a residue which is part of a β -sheet.

- Bring the largest β -sheet to the center of the graphics window of Rasmol. At the command window, type **cartoon**

Note that the cartoon representation has only been rendered for the selected β -strands. This rendering is similar to that which occurs upon selecting the “Cartoons” options from the “Display” menu.

Q: Are the b-strands in the largest b-sheet parallel or antiparallel?

Similarly, we can use Rasmol to look at the helices of the insulin receptor kinase.

cartoon off

hbonds off

wireframe off

select helix and mainchain

wireframe 10

- Determine the start and end residue number of one of the helices by clicking on atoms of these residues. Type the command “restrict” followed by the range of amino acids of the helix e.g. “restrict 1-20”.

Look at the hydrogen bonds of this helix by typing at the command line:

hbonds

Rasmol can perform measurements on the structures displayed. This is done by change the “picking” mode. The default picking mode is “ident” which identifies the atom which is clicked on. To measure distances, type:

set picking distance

Measure a number of hydrogen bond distances (heavy atoms) in the helix.

Q: What is the range of hydrogen bond distances in the helices (where the heavy atoms were used)?

To measure angles, type:
set picking angle

Q: What is the range of C-O...N angles in the hydrogen bonds?

To measure some torsion or dihedral angles, type:
set picking torsion

- Measure some amide bond torsion angles by clicking on successive C_α, N, C, C_α atoms of the helix.

Distances can also be displayed on the screen. This is useful for displaying hydrogen bonds which are not detected by Rasmol's "hbonds" command. This facility can be accessed by typing:
set picking monitor

- Click on the nitrogen and oxygen atoms of a hydrogen bond.

To remove a distance from the window displaying the molecule, click on the two atoms whose distance was measured. To turn off all distance monitors, type:
monitors off

To restore the appearance of the protein to that which we started with:
select all
wireframe
hbonds off
set picking ident

Another important type of interaction in proteins is that between the hydrophobic parts of the protein, most notably the packing of hydrophobic sidechains. To visualize the packing of hydrophobic residues around a given residue, we can use the "select within" statement to view all of the atoms within a specified distance of a selection. Here, you will look at the interactions of hydrophobic residues about the sidechain of Leu 1199 of the insulin receptor kinase.

wireframe off
select within (7.5,1199)

This command selects all atoms within 7.5 Å of any atom within the Leu 1199, including the atoms of Leu 1199 itself. The selection "1199" selects any atom with residue number 1199. More complicated selections can be used as the second argument in the brackets.

wireframe 20
select 1199
wireframe 50

The residue Leu 1199 should now be visible, rendered as thick sticks and surrounded by other residues (some of which are only shown in part) as thinner sticks.

- Remembering that the nitrogen atom is coloured blue and oxygen atom is red, find the mainchain atoms of Leu 1199. Now find the sidechain of Leu 1199. You should see that the sidechain consists only of carbon atoms since the hydrogen atoms are not visible. To highlight the sidechain of Leu 1199, type:

select 1199 and sidechain
color green

We now wish to highlight the interaction of the Leu 1199 sidechain with other hydrophobic sidechains such as that of Phe 1195. (This is the only aromatic residue near Leu 1199 and you should be able to identify it from its planar six-membered ring.) You can demonstrate the closeness of the sidechains of Leu 1199 and Phe 1195 in at least two ways.

select 1195 or 1199

spacefill

The “or” statement requires that each selected atom must, in this case, be part of a residue numbered 1195 or part of a residue numbered 1199. An alternative way to specify this sort of selection is by typing “select 1195,1199”. This approach makes it easier to select a long list of residues e.g. “select 1190-1200, 1205, 1210-1212”.

You should now see Phe 1195 and Leu 1199 shown as spheres, with the sidechains of both residues in close proximity, consistent with the two residues interacting. For an alternative way of showing the interaction between the two residues:

spacefill off
dots 200

This displays a series of dots which represents a van der Waals surface for the atoms i.e. a surface at the van der Waals radius of each atom. The density of dots is indicated by the value of 200 used in the “dots” statement. Note that the “dots” rendering is different to other sorts of rendering since the van der Waals surface is only displayed for the selection that was valid for when the “dots” statement was invoked i.e. if you want to have a van der Waals surface for a lot of atoms, they all have to be included in the last “select” statement.

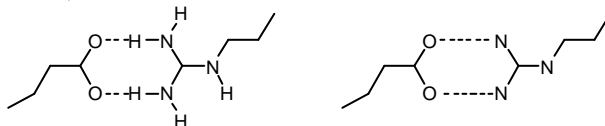
Q: What are the other residues which make hydrophobic interactions with the sidechain of Leu 1199?

There is one other sort of interaction in proteins worth looking at which is the salt bridge. One definition of a salt bridge is a pair of oppositely charged residues (e.g. Glu and Arg) which form at least one hydrogen bond between them. We will look at the salt bridge formed by Glu 1179 and Arg 1253.

select all
dots off
color cpk
wireframe
select 1179, 1253
wireframe 50

Note that the atoms which the ‘color cpk’ command acts on are part of the current selection (from the ‘select helix’ command) although this is not obvious from this example). ‘cpk’ is the specification for colouring according to atom type. (The ‘cpk’ refers to Corey, Pauling and Koltun who developed the original plastic models.)

As you have probably worked out, the last two commands select for the two residues of interest and then display them using wireframe rendering and with thicker bonds than the rest of the molecule. The hydrogen bonds between the two residues are illustrated by the left hand side of the next figure. However, since hydrogen atoms are not present in this solved structure of the insulin receptor kinase, you will see a figure resembling that on the right (but without the hydrogen bonds shown).



Hide the parts of the molecule not close to the two residues which make up the salt bridge by typing at the command line:

restrict within (7.5,(1179,1253))

Note that the selection “1179,1253” was enclosed in brackets to distinguish it as defining the set of atoms which the selected atoms must be within 7.5 Å of.

Now type
hbonds

Notice that the hydrogen bonds shown are only between the mainchain atoms of the insulin receptor kinase and do not appear for the sidechains of our salt bridge. Since the hydrogen bonds can be identified by eye, distance monitors can be used to indicate them.

set picking monitor

- Measure the distances between the heavy atoms which are involved in the hydrogen bonds.

Q: The arginine sidechain makes an additional hydrogen bond to a mainchain carbonyl group. Which residue is this carbonyl group a part of? (Note that there are two nearby carbonyl groups but one should be excluded because the geometry is inconsistent with the formation of a hydrogen bond.)

(The salt bridge between Glu 1179 and Arg 1253 is sandwiched between the sidechains of two proline residues. For extra brownie points, determine which residues they are and highlight them by showing their van der Waals surface.)

Charged residues occur in the interior of globular proteins infrequently and their charged ends generally interact with other polar and/or charged groups. Another example of this is Asp 1191 whose charged carboxyl group makes two hydrogen bonds with mainchain amide groups.

The activation of the insulin receptor kinase requires phosphorylation of three tyrosine residues. In the PDB file of the insulin receptor kinase, these residues have been renamed from TYR to PTR, and they can be found using the last two commands from the following sequence of commands:

```
select all  
wireframe  
hbonds off  
monitors off  
set picking ident  
select not protein  
wireframe 50
```

The “select protein” command would select atoms belonging to amino acid residues (Ala, Asn, ..., Trp, Val). The “not” keyword selects for all atoms which do not meet the requirement of meeting the “protein” selection. The phosphotyrosine residues and the ATP analogue are shown as thick wireframe rendering. Clicking on either the ATP analogue or the three phosphotyrosine residues reveals that the atoms in these groups are considered hetero-atoms. These atoms can also be selected by “selected hetero”.

It is worth exploring the rendering of the insulin receptor kinase fold.

```
restrict not protein  
wireframe 50  
select protein
```

At this stage, only the ATP analogue and the phosphotyrosine residues are displayed but the protein atoms are the most recently selected. Thus, you can choose several renderings of the fold such as “backbone”, “trace” and “cartoon”.

- Explore the renderings listed above with and without a numerical value after rendering command e.g. “backbone 75”. Remember to turn off the display of the protein (e.g. “backbone off”) prior to choosing a new way of representing the fold.

Secondary structure elements could be further highlighted further by colouring them. For example:

```
select sheet  
color yellow  
select helix  
color cyan
```

Similarly, different protein chains can be colored differently. The peptide substrate in this structure of the insulin receptor kinase is denoted as chain B (the kinase is chain A). Therefore this peptide can be selected and then colored by:

```
select *:b  
color green
```

There are two important features of this selection which are of note. The first is the asterisk. This is the so-called wildcard which selects everything. Its power comes from its use in conjunction with other selections such as residue name. The “:b” part of the command refers to chain B. The colon followed by the letter of the chain can be placed after a selection to restrict that selection to the chosen chain. Thus, “select 9-11:b” is equivalent to “select 9-11 and *:b”. In some cases, the colon is redundant but it is safer to use the colon when referring to a given chain as part of a selection.

Selection of atoms uses a similar syntax to the selection of chains but with the full stop instead of the colon. For example, “select ala.ca” selects all of the C_α atoms in alanine residues.

Up to now, you have typed in many commands into Rasmol using the command line window. An alternative is to create a script file which is a text file consisting of a set of Rasmol commands.

On the Macintosh:

- Start the bbsedit text editor. This should be found in the Applications folder at the bottom of the desktop.
- Type the following commands into the bbsedit editor:

```
zap  
load HD:Desktop Folder:Bioinformatics:lir3.pdb  
select all  
wireframe off  
cartoons
```

This set of commands closes the previous structure in Rasmol (“zap”), loads the insulin receptor kinase structure afresh, turns off the default wireframe rendering and displaying the fold of the structure using the cartoon rendering.

- Save the file in the “Bioinformatics” directory. Call the file “ir3_script”.
- In Rasmol, type at the command line:

```
script HD:Desktop Folder:Bioinformatics:ir3_script
```

This commands executes the Rasmol commands in the file “ir3_script”.

Scripts are useful for situations where lots of commands must be put together. The sequence of commands can be edited without having to retype all of the commands.

A summary of some Rasmol selections and commands

There are different ways to select items which can be acted on:

select *.c	select all carbonyl carbon atoms (i.e. all atoms with the name C)
select ala	select atoms in all alanine residues
select ala.c	select all carbonyl carbon atoms in alanine residues
select as?	Select atoms in either asparagine (asn) or aspartate (asp)
select *.n?	select all nitrogen atoms
select [k25]	select atoms in molecule(s) with the residue name “K25” (the square brackets are needed as the residue name has numeric digits in it)
select *120	select atoms at residue 120 of all chains
select *:b	select all atoms in chain B (also, “select *b”)
select within (7.5, *:b)	select all atoms within 7.5 Å of any atom in chain B.
select 94,96,98-120	select atoms in residues 94 or 96 or in the residues 98-120.

Other items can be selected as a group:

select protein	select all protein atoms
select nucleic	select all nucleic acid (DNA, RNA) atoms
select hetero	select all non-protein, non-nucleic atoms which have been specified as hetero-atoms in the PDB file e.g. ligand atoms
select helix	select all atoms in a helix
select sheet	select all atoms in a β -sheet
select hydrophobic	select ala, gly, ile, leu, met, phe, pro, trp, val
select polar	select asn, asp, arg, cys, gln, glu, his, lys, ser, thr, tyr
select charged	select asp, arg, glu, his, lys

There are many actions that can be applied to a selection:

wireframe 30	render selected atoms as wireframe (thickness 30)
backbone 30	protein C α atoms and nucleic P atoms joined by lines (thickness 30)
cartoons	show proteins and nucleic acids in selection as cartoon
spacefill	Display selected atoms as VDW spheres
spacefill 30	Display selected atoms as spheres of size 30
dots 200	show dots (density 200) around selected atoms with radius equal to preset VDW radius; previously rendered dots are removed

Picking modes:

set picking ident/set picking	identify picked atom
set picking distance	measure distance between two picked atoms
set picking angle	measure angle of three sequentially picked atoms
set picking torsion	measure torsion of dihedral angle of three sequentially picked atoms
set picking monitor	measure distance between two picked atoms and display in the molecular graphics window

4. Using Rasmol to look at the structure of G-CSF and its receptor binding

The skills and Rasmol commands you learnt in the first three sections will be applied to looking at protein G-CSF (Granulocyte Colony Stimulating Factor). G-CSF stimulates proliferation and differentiation of neutrophils from precursor cells and is required for normal neutrophil production *in vivo*. The initial steps of G-CSF-induced signalling are the binding of G-CSF to its receptor followed by dimerization of the receptor.

A large number of alanine mutants of G-CSF were constructed and tested to determine regions of the ligand involved in receptor binding. In the first part of this section, you will look at the structure of G-CSF and locate the residues whose mutation to alanine reduced activation of the G-CSF receptor.

- Download the PDB file with code 1rhg from the PDB Internet site. This structure is of human G-CSF. (Bull and dog G-CSF structures are also in the PDB.) While the structure of some of the residues of the human G-CSF structure are missing, their absence does not impact on this exercise.
- Use Rasmol to view the human G-CSF structure. Show that there are three copies of G-CSF present.
- Change the rendering of G-CSF and determine the fold of the protein. You may wish to restrict the viewing to the A chain G-CSF molecule.

Q: What are the predominant secondary structure elements of G-CSF?

Fifteen mutants of G-CSF noticeably reduced the *in vitro* response of the G-CSF expressing cells relative to wild-type G-CSF. These mutants are shown on the next page:

L15A	Q25A
E19A	L31A
K34A	L47A
K40A	L54A
V48A	S142A
L49A	R146A
D122A	
L124A	
F144A	

The residues in the right hand column were considered to alter the structure of G-CSF and hence lead to reduced binding due to a change in the structure.

Use the following commands to highlight intramolecular interactions between Gln 25 and other residues in the G-CSF molecule.

```
select all  
wireframe  
restrict within (7.5,25a) and *:a  
select 25a  
wireframe 50
```

The third command restricts the display of the G-CSF those residues in chain A (“*:a”) which are within 7.5 Å of Gln 25 in chain A (“within (7.5,25a)”).

Change the mode of picking (what occurs when you press the mouse button when the cursor is over an atom) from the “ident” mode to “monitor” mode by typing at the command line:

```
set picking monitor
```

- Measure distances between the polar atoms of the Gln 25 sidechain and those of other G-CSF residues. Satisfy yourself that some of these interactions are hydrogen bonds. Note that, unlike the insulin receptor kinase structure, some hydrogen atoms are present in this structure.
- Apply a similar procedure used for looking at Gln 25 to the residues Ser 142 and Arg 146 of the G-CSF A chain. Identify hydrogen bonding and/or charged interactions that suggest that the roles of these residues are in maintenance of the G-CSF structure. (Note that repeating the process for Arg 146 would yield different results for the B and C chains.)

The residues Leu 31, Leu 47 and Leu 54 are buried in the G-CSF structure and so the reduction of G-CSF-induced activity upon their mutation was interpreted in terms of alteration of the protein’s structure rather than interference with binding to its receptor.

- Using a similar procedure that that applied to residue Leu 1199 of the insulin receptor kinase in the previous section, show that the sidechains of Leu 31, Leu 47 and Leu 54 of G-CSF chain A are buried.

The nine residues listed in the left column of the above table were considered to be potentially part of G-CSF binding sites for its receptor.

- Use Rasmol to render the backbone trace of the G-CSF.
- Using spacefilling or wireframe representation, show the positions of the nine residues of G-CSF considered to interact with the G-CSF receptors.

Q: How many binding sites does G-CSF appear to have? Given that the stoichiometry of the G-CSF - G-CSF receptor complex on cells is thought to be 2:2, propose a schematic model for this complex.

Several years after this mutational study, the crystal structure of G-CSF bound to a fragment of the extracellular region of G-CSF receptor was published. The authors solved the structure for two ligand molecules and two receptor fragments. However, the receptor fragment was missing one domain which had been previously shown to be involved in the formation of the 2:2 ligand-receptor complex. We can highlight the mutated G-CSF residues on this crystal structure and assess how likely the crystal structure represents the physiologically relevant G-CSF - G-CSF receptor complex. However, we will first explore some of the features of this crystal structure.

- From the PDB Web site, download the structure of the G-CSF - G-CSF receptor complex with PDB code 1cd9. Save the coordinate file into the “Bioinformatics” directory in which you saved the insulin receptor kinase structure.
- Display the contents of the file 1cd9 using Rasmol. Change the rendering and colouring of the molecules to answer the next five questions.

Q: Other than the protein molecules, how many other types of molecules are present.

- Select the cysteine residues in and display them using wireframe rendering (**wireframe 50**). Then, type the command **ssbond 50** to show the disulfide bonds.

Q: How many disulfide bonds are there in G-CSF and G-CSF receptor? How many unpaired cysteines are there in each of these two proteins?

- Find one example of a non-water, non-protein molecule present. Colour it using atom-based colouring (select the molecule and then colour using the “cpk” specification).

Q: What do you think this molecule is?

- Find the amino acid closest to this molecule. Then, change the “picking: mode to one that allows you to measure distances. Find the shortest distance between the “hetero” molecule and the amino acid residue?

Q: What is the residue type you have found? What is the the shortest distance between the hetero group and the amino acid residue? Is this distance indicative of a chemical bond or a non-bonded interaction? What do you think you are seeing?

- Display the backbone of one of the receptor molecules.

Q: A domain is a portion of the polypeptide chain which can fold as an independent unit. The fragment of the G-CSF receptor in the crystal structure contains two domains. Where, approximately, is the boundary between the two domains?

OK, now you can look at the complex and assess it with respect to the mutational studies of G-CSF.

- Display the protein (ligand and receptor) chains using backbone rendering.
- Using spacefilling rendering, display the ligand residues which are thought to be potentially in ligand-receptor interfaces (left column from table of G-CSF mutants). Please note that the G-CSF numbering in this structure differs from that of the G-CSF structure. [Add 1 to the G-CSF numbering of the mutants to see them on the structure.](#)

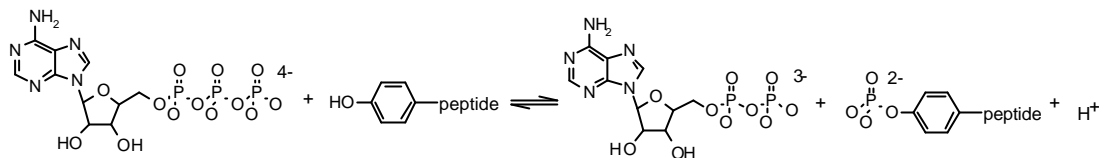
Q: Which of the residues are present in a protein-protein interface? Which of the protein-protein interfaces in the structure are consistent with the G-CSF mutation data?

- Binding of a particular monoclonal antibody suggests that the N-terminus of G-CSF is exposed when G-CSF is bound to its receptor in the active, signalling complex. Locate the position of the N-terminus of G-CSF in the crystal structure.

Q: Is the N-terminus of G-CSF buried or exposed in the crystal structure of the complex? Is this observation consistent with the receptor dimerization interface of the structure?

5. Using Rasmol to analyse enzyme-substrate interactions

In this section, you will look at the interactions between the tyrosine kinase domain of the insulin receptor with a bound substrate and a substrate mimic. Tyrosine kinases are enzymes which are generally found in cell signalling pathways. They catalyse the transfer of the (terminal) gamma-phosphate of bound ATP to the tyrosine residue of a bound peptide. A schematic diagram of this reaction is shown below.



- Use Rasmol to visualize the structure of the molecules in the PDB file with code 1ir3. Color the system by “Chain” using this option from the “Colours” menu. This identifies three components of the system: the protein kinase domain (chain A), the peptide substrate (chain B) and the ATP analogue (residue ANP). Note the proximity of the end of the phosphate chain of the ATP analogue and the tyrosine residue of the peptide substrate. Clearly these regions must in close proximity since this is the region when the transfer of the phosphate chain occurs. Color the system using atom-based coloring using the “CPK” option in the “Colours” menu.

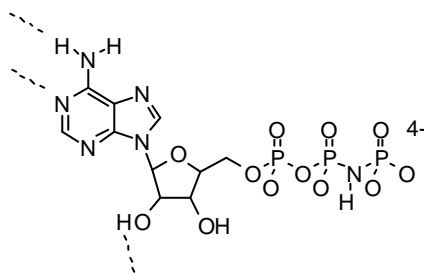
Here, you will begin by looking at the interactions between the protein kinase domain and the non-hydrolysable mimic of ATP called AMP-PNP. In the PDB file, this residue has the three letter code ANP.

- Using the Rasmol command line, display only the residue ANP.

The nitrogen atom (colored blue) in the triphosphate chain prevents the gamma phosphate from being transferred to the bound tyrosine. Thus, the enzyme reaction does not occur and the crystal structure of the stable, non-reacting complex is able to be solved.

- Using the “select within” and “wireframe” commands, display the atoms 6 Å within the residue ANP. You may wish to select ANP and display it using thicker wireframes than the surrounding atoms.

In tyrosine kinases, ATP (as well as its non-hydrolysable analogues) is held in place by a number of hydrogen bonds, salt bridges and hydrophobic interactions. The first step is to identify the hydrogen bonds which interact with the adenine and ribose moieties. There are at least three hydrogen bonds made between AMP-PNP and the kinase. The AMP-PNP atoms which are involved in hydrogen bonds are shown in the figure below.



- Noting that only the heavy atoms are present in the complex, find the hydrogen bonds which are present between AMP-PNP (residue ANP) and the tyrosine kinase domain of insulin. Use the “set picking monitor” command and then pick the atoms which are part of hydrogen bonds. Note that you will have to imagine the hydrogen atoms on the amino groups (amide bond and amino group of the adenine moiety of AMP-PNP) and on the hydroxyl groups of the ribose moiety of AMP-PNP

Q: What are the pairs of heavy atoms which are involved in the three hydrogen bonds?

- There are several hydrophobic residues which interact with the adenine moiety above and below the plane of the adenine moiety. Using similar commands to those used to look at hydrophobic packing in Section 3, identify these hydrophobic residues.

Q: What are the residues in the insulin receptor kinase which make hydrophobic interactions with the adenine moiety above the plane of the ring? Below the plane of the ring?

The triphosphate chain of the ATP analogue is held in place by a number of polar interactions, mostly with charged species.

- The residue Lys 1030 interacts with one of the oxygen atoms of the alpha phosphate group (phosphate group closest to the ribose moiety). Measure the distance between the sidechain nitrogen atom of Lys 1030 and the nearest oxygen atom of AMP-PNP.

Q: The residue Lys 1030 is highly conserved in tyrosine kinases. Mutation of this residue in a number of tyrosine kinases does not greatly alter ATP binding but abolishes tyrosine kinase enzymatic activity. What do you think the role of Lys 1030 in the enzyme-catalysed reaction is?

Type the commands

**select within (7.5,anp) and not protein and not anp
spacefill 100**

A number of spheres will appear around AMP-PNP. If the molecules are colored using CPK coloring (element-based coloring), most of the spheres will be red - these are the oxygen atoms of water molecules. The two green spheres are (positively charged) magnesium ions. These ions help to correctly orient the phosphate chain for transfer of the gamma phosphate to the nearby tyrosine residue.

- For convenience, remove the water molecules from the display window.
- Measure the distances ($< 3 \text{ \AA}$) between the magnesium ions and the phosphate oxygen atoms of AMP-PNP.
- Measure the distances ($< 3 \text{ \AA}$) between the magnesium ions and the protein kinase.
- Find hydrogen bonds between the protein kinase and the phosphate oxygen atoms. (You will have already detected that between Lys 1030 and an alpha phosphate oxygen.)

Q: What do you expect will be effect of mutating Asn 1137 to alanine? Reduced ATP binding? Reduced tyrosine kinase activity?

The substrate other than the Mg_2 AMP-PNP is the tyrosine-containing peptide substrate.

- Display the peptide (chain B) on its own in the Rasmol molecular display window. Sufficient electron density was visible only to solve for the structure of six of the eighteen peptide residues. Of these, four of these residues were of interest: Asp 9; Tyr 10, Met 11 and Met 13. Identify these residues in the peptide substrate.

Some of the interactions between the tyrosine kinase domain and the bound peptide substrate are mainchain-to-mainchain hydrogen bonds.

- Display the protein kinase residues near the peptide substrate using the "select within" command. Then, identify the hydrogen bonds between the mainchains of the kinase and the peptide substrate. Hint: while the "hbonds" command does not locate inter-chain hydrogen bonds, the interacting strands have been assigned as β -strands.

Q: What are the kinase and peptide residues which make these hydrogen bonds.

Using phage display techniques, the preferred residues for the amino acid position immediately N-terminal to the site of tyrosine phosphorylation (P-1 position) have been shown to be those with a negative charge i.e. glutamate and aspartate.

- Identify the basic residues (Arg, Lys and His) in the insulin receptor kinase which are close to Asp 9 of the peptide substrate.

Q: Do the basic residues near Asp 9 make hydrogen bonds to Asp 9 or are the only interactions charge-charge interactions?

The same phage display experiment indicated that hydrophobic residues were preferred for the P+1 and P+3 residues (which correspond to Met 11 and Met 13 of the bound peptide).

- Using similar techniques of visualization as for looking at the hydrophobic interactions in Section 3, identify the residues which interact with the sidechains of Met 11 and Met 13 of the peptide substrate.

Q: Do the residues surrounding Met 11 and Met 13 support the observation for the preference of hydrophobic residues for the P+1 and P+3 positions in the peptide substrate.

The tyrosine residue of the peptide substrate is positioned to accept the gamma phosphate of bound ATP.

Q: Which is the hydrophobic residue in the tyrosine kinase which the tyrosine sidechain sits atop?

Q: Which are the two charged residues of the tyrosine kinase which appear to make hydrogen bonds with the phenol group of the peptide tyrosine residue? The closeness of these residues to the phosphate accepting tyrosine residue suggests that these residues are involved in the catalysis mechanism.

6. Visualization of sequence information using a homologous protein

There is much interest in the insulin receptor and its ligand binding properties due to the problems of insulin signalling in some diabetics. To date, the structure of the ligand binding region of the insulin receptor has not been reported. However, the structure of the first three domains of the related insulin-like growth factor 1 (IGF-1) receptor has been published and is present in the PDB. A number of alanine mutants of the N-terminal regions of the insulin receptor have been measured for changes in binding affinity compared to the wild-type receptor. In this section, you will map the insulin receptor mutations onto the IGF-1 receptor structure to determine whether part of the receptor shows a potential binding site.

The binding data for the insulin receptor mutations is shown below. Nb NS is not secreted.

Mutant	K _d (mutant)	K _d (mutant)/ K _d (wt)	Mutant	K _d (mutant)	K _d (mutant)/ K _d (wt)
E6A	1.44	1.02	F39A	35.7	25.32
D12A	9.12	6.47	F46A	2.24	1.59
R14A	-	>40	Y60A	NS	-
R19A	1.46	1.04	Y64A	-	>40
E24A	0.93	0.66	Y67A	3.22	2.28
E30A	1.08	0.77	F88A	2.31	1.64
K40A	1.6	1.13	F89A	5.17	3.67
R42A	1.62	1.15	Y91A	3.85	2.73
E44A	4.84	3.43	F96A	NS	-
D45A	0.74	0.52			
D59A	0.91	0.65	M11A	1.66	1.18
R65A	1.07	0.76	I13A	16.31	11.57
E70A	1.54	1.09	N15A	352.5	250.00
K73A	0.75	0.53	L17A	0.49	0.35
D74A	1.41	1.00	H32A	2.06	1.46
R83A	0.83	0.59	L33A	NS	-
E97A	NS	-	Q34A	18.1	12.84
R114A	0.48	0.34	I35A	NS	-
			L36A	14.1	10.00
L87A		6.67	L37A	NS	-
			M38A	4.43	3.14
			T41A	1.05	0.74
			L61A	1.68	1.19
			L62A	NS	-
			L63A	NS	-
			V66A	1.09	0.77
			L69A	0.89	0.63

			N90A	8.83	6.26
			L93A	1.93	1.37

You will extract the sequence of the human insulin receptor N-terminal region and then perform a BLAST search using this sequence against the sequences from the PDB.

- Using an Internet browser, go to a site which contains a protein sequence database (e.g. ANGIS, NCBI, SRS). Find the entry with the accession number P06213 (human insulin receptor).
- Start a text editor bbsedit (or a new window in this text editor). Copy the human insulin receptor sequence from the sequence ²⁸LYPGE...DNEEC¹⁸². (This is the sequence of the first domain of the insulin receptor. The first 27 amino acids are the signal peptides which are cleaved to generate the mature protein.) Remove the numbers so you are left only with letters and spaces.
- Start a new Internet browser window and go to a site which allows a BLAST search to be performed e.g. NCBI BLAST site. Paste the edited insulin receptor sequence into the sequence window. Select the PDB database to perform the search against.

The top scoring sequence alignment is the alignment of interest. You will need this sequence alignment for the rest of this section.

- Download the structure with PDB code 1igr from the PDB. View it in Rasmol. A recommended rendering of the residues is spacefill for residue 1-183 and backbone for residues 184-467
- Color the residues as according to their values of $K_d(\text{mutant})/K_d(\text{wt})$ (ratio of mutant to wild-type binding of insulin):
 Red: $K_d(\text{mutant})/K_d(\text{wt}) > 40$ Orange: $K_d(\text{mutant})/K_d(\text{wt}) 10-40$
 Yellow: $K_d(\text{mutant})/K_d(\text{wt}) 2.4-10$ Green: $K_d(\text{mutant})/K_d(\text{wild type}) < 2.5$
 White: Alanine mutant not secreted Blue: untested
- From the coloring of the residues, locate the region in the first domain of the IGF-1 receptor which, in the similar insulin receptor, is likely to bind ligand.

7. Visualization of a protein - DNA complex

The protein NFAT (nuclear factor of activated T cells) and the protein dimer made up of Fos and Jun (called AP-1) cooperatively bind to DNA. In this section you will look at the DNA binding regions of these proteins.

- Download the file with PDB code 1a02 from the PDB site to the "Bioinformatics" directory as you have downloaded other coordinate files from the PDB site.

The skills you have learnt from previous sections will enable you to answer the questions listed below. You may ask one of the demonstrators for assistance if necessary.

Q: How many chains are present in the system? What are the one letter identifiers for each of the protein/nucleic acid chains present? E.g. NFAT is chain N. This can be found in the summary section for the PDB entry 1a02.

- Only display the nucleic acid chains. (Hint: use the restrict command in Rasmol.) Find the ribose groups, phosphate groups and the bases. This is to remind you structure of DNA.

Q: With the use of the Rasmol "hbonds" command, determine how many hydrogen bonds are present between adenine and thymine and how many hydrogen bonds are between cytosine and guanine. (Yes I know this can be found from books, but looking at the structure and being able to rotate and translate a DNA structure might be more meaningful.) Why do you think that DNA fragments with high GC content have higher melting points than those with lower GC content? (It's a gimme question - humour me...)

Q: What are the secondary structures of the three proteins? How many domains do you think NFAT has?

- There are two grooves in DNA: the major groove and the minor groove. Use the spacefilling representation and the backbone representation to identify the two grooves. (Rotate the molecule to provide a better indication of the three-dimensional nature of the structure.)

Q: Where do most of the interactions between the NFAT-Fos-Jun complex and DNA occur: in the major groove or the minor groove? Are there any interactions between the protein complex and the other groove?

Q: Due to the phosphate component of its backbone, DNA is a highly charged molecule. Which charged residues of the protein complex are interacting with the DNA fragment. Don't list all of them, just list which residue types and whether there are a lot, a few or none for each of the charged residues. Hint: use the "select within" command.

8. Using Rasmol for images and presentations

In this section, you will learn how to create static images for use in presentations, and how to execute a Rasmol script from within a PowerPoint presentation. These features may be useful for displaying molecular systems in your lab meetings.

There are two methods which can be used to generate static images of your molecular system. The first consists of generating a file of an image (e.g. PICT, .BMP, GIF formats) from Rasmol.

- Use Rasmol to change the appearance of your molecular system for your image. Generate the image of your molecular system in Rasmol as desired. If you want an image with a white background, type in the command "set background white". Change the size of the Rasmol molecular display window to the size of the image you wish to generate.
- In the "Export" menu, choose the format of the image file you wish to create. If you are using a Macintosh, you are recommended to choose PICT format if you are inserting the image into a Macintosh format file e.g. a word processing document on the Mac. In the dialog which appears, choose a name for the file and save the image in the "Bioinformatics" directory.
- Open up the "Bioinformatics" directory. Double click on the file you saved in the previous step. You should see your image in an image editing program. Close the program.

The image you have just generated could be inserted into a document. An alternative method for generating an image that can be inserted into a document is outlined below. Do not carry out these steps in the practical session as they are only included for your future reference.

- Use Rasmol to change the appearance of your molecular system for your image.
- In the "Edit" menu, select the option "Copy". This places a copy of the image displayed in Rasmol into the Clipboard.
- In another program such as a word processor, paste the contents of the Clipboard (often achieved by choosing the "Paste" option in the "Edit" menu) into the document.

Executing a Rasmol script from a PowerPoint presentation allows you to display your molecular system as you desire and to rotate etc the molecular system to give an impression of the three-dimensional nature of the molecular system. It is even possible to animate molecules. To run a Rasmol script from PowerPoint on the Macintosh operating system is a convoluted process that requires several sets of steps.

On the Macintosh, you need to generate a Rasmol script which, when you double click, will start Rasmol and run the script. (Note that a quirk of the Macintosh of version Rasmol is that only one copy of the program can be run at a time.)

- Render the molecular system in Rasmol.
- At the command line, type
save script HD:Desktop Folder:Bioinformatics:script

You may need to edit the Rasmol script that was generated by the program. You need to choose a text editor that can read the script (SimpleText cannot do this) and one that does change the file type (which Microsoft Word does).

- In the Applications directory, open up the text editor program “bbsedit”.
- Open the script that you have previously saved in the “Bioinformatics” directory.
- Make an alteration to the script.
- Save the script (“Save” option in the “File” menu).

You now need to create an Apple script that starts the Rasmol script.

- Open up the Script Editor for Apple Script. A copy of the Script Editor has been placed in the “Bioinformatics” directory on the desktop for you to use. (This program is usually found in the directory “*HD*: Applications:Apple Extras:Apple Script:Script Editor”.)
- Click on the “Record” icon in Script Editor.
- In the “Bioinformatics” directory, double click the script file.
- Click on the “Stop” icon in Script Editor.
- Save the Apple Script file in the “Bioinformatics” directory. Call the file “AppScript”. Choose the format as “Application” or “Classic Applet” depending on the version of the Script Editor.

Now you are ready to use PowerPoint.

- In the “Applications” directory at the bottom directory, open the PowerPoint application.
- Create a PowerPoint slide. To insert a button, select the “Insert” menu and then the “Picture” and then “AutoShapes” options. This may vary according to the version of PowerPoint you are running.
- If you have inserted an “AutoShapes” picture, a small window should appear titled “Action Settings”. If you do not have this, or you wish to link some text with executing the Rasmol script, click on the relevant object with the Ctrl key held down (right mouse button on the PC) and select the “Action Settings...” option. (At some stage, a prompt box may have appeared suggesting that you save your PowerPoint presentation before proceeding. Save the file in the “Bioinformatics” directory.)
- Click on the button next to “Run Program:” and then click on “Browse...”. Find the “AppScript” Apple Script file in the “Bioinformatics” directory and click the two “OK” buttons..
- Test the presentation by choosing the “View” menu and the “Slide Show” option. The slide should appear as part of a presentation. Press the icon/text that you have linked to the running of Rasmol. The Rasmol program should appear but with nothing in the window because you have not told PowerPoint which file to open. Quit the Rasmol program and press the “Esc” key to return to editing your PowerPoint presentation.

If you are using the PC version of PowerPoint, you need to:

- Create a Rasmol script either by creating a text file or by modifying a script generated by the Rasmol program.
- Follow the steps above which relate the PowerPoint file. When you select the application that the icon/text links to, search for the Rasmol application. You then need to modify the line which appears for the Program by adding the location of the script file. The resultant line should read something like:
C:\Programs\Rasmol2.6\RASWIN.EXE -script C:\structures\ir3_script.txt
where the “-script” keyword indicates that script containing the Rasmol commands to be executed. If you wish to read in a PDB-format file, replace “-script” with “-pdb” (or nothing since “-pdb” is the default) followed by the full name of the PDB-format file.

References for Rasmol on the Internet

<http://www.openrasmol.org/>

This site is arguably the *de facto* home page for Rasmol. The Rasmol manual is on this site at:

<http://www.openrasmol.org/doc/rasmol.html>.

<http://www.umass.edu/microbio/rasmol/index2.html>

This is a site from Eric Martz who was once an evangelist for Rasmol. The site contains a wealth of material including tutorials. Some of the useful pages in this site are:

<http://www.umass.edu/microbio/rasmol/makescrp.htm> for making movie scripts using Rasmol and

<http://www.umass.edu/microbio/rasmol/otherof.htm> which lists other molecular viewing programs. You may find

that some of these are friendlier than using Rasmol (no typing in commands). Protein Explorer, a Rasmol derivative which is Eric Martz's current obsession, can be accessed from <http://www.umass.edu/microbio/rasmol/>

<http://news.bmn.com/hmsbeagle/109/reviews/sreview>

This web page on the BioMedNet site contains an independent review of the latest version of Rasmol.

References to the biological systems visualized in this practical exercise:

Identification of residues critical to the activity of human granulocyte colony-stimulating factor. Reidhaar-Olson JF et al., *Biochemistry*. 1996. 35: 9034-9041.

The structure of granulocyte-colony-stimulating factor and its relationship to other growth factors. Hill CP et al., *Proc Natl Acad Sci U S A*. 1993. 90: 5167-5171.

Atomic structure of the GCSF-receptor complex showing a new cytokine-receptor recognition. Aritomi M et al., *Nature*. 1999. 40: 713-717.

Identification of ligand-binding site III on the immunoglobulin-like domain of the granulocyte colony-stimulating factor receptor. Layton JE et al., *J Biol Chem*. Accepted.

Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. Hubbard SR. *EMBO J*. 1997. 16: 5572-5581.

Catalytic specificity of protein-tyrosine kinases is critical for selective signalling. Zhou S et al., *Nature*. 1995. 373: 536-539.

Crystal structure of the first three domains of the type-1 insulin-like growth factor receptor. Garrett TPJ et al., *Nature*. 1998. 394: 395-399.

Mapping of an NH₂-terminal ligand binding site of the insulin receptor by alanine scanning mutagenesis. Williams PF et al., *J Biol Chem*. 1995. 270: 3012-3016.

Replacements of leucine 87 in human insulin receptor alter affinity for insulin. Nakae J et al., *J Biol Chem*. 1995. 270: 22017-22022.

Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. Chen L et al., *Nature*. 1998 392: 42-48.

Gene expression elicited by NFAT in the presence or absence of cooperative recruitment of Fos and Jun. Macian F et al., *EMBO J*. 2000. 19: 4783-4795.

Acknowledgements

Paul Gooley, Nathan Hall, Sue Legge, Robert Flegg, Tony Kyne.