Viewing Protein 3D Structures with Deep View

A quick guide to the visualisation components of Deep View

Version 1.3 (public)

Licence

Introduction

Whilst much useful information about a protein can be derived from looking at both DNA and protein sequence information the only way to get a complete understanding of its function is by looking at its 3D structure. Proteins operate in a 3D environment and by looking at them in this context it is possible to explain observations which could not be accounted for by other analyses.

There are now a huge number of protein structures made publicly available through the protein data bank (PDB). In this course you will be shown how to access these structures and manipulate them on your own computer.

Deep View

The basis for this course is a software package called Deep View (formerly called the Swiss PDB Viewer). This is a free package written by a group at the Swiss Institute for Bioinformatics. It is a pretty comprehensive molecular modelling package providing a variety of visualisation options as well as allowing homology modelling and energy minimisation. It also interacts with the POV-Ray raytracing package to produce publication quality molecular graphics.

This course will not cover the homology modelling aspects of Deep View, but will demonstrate how to use the visualisation components to analyse both single and multiple structures.

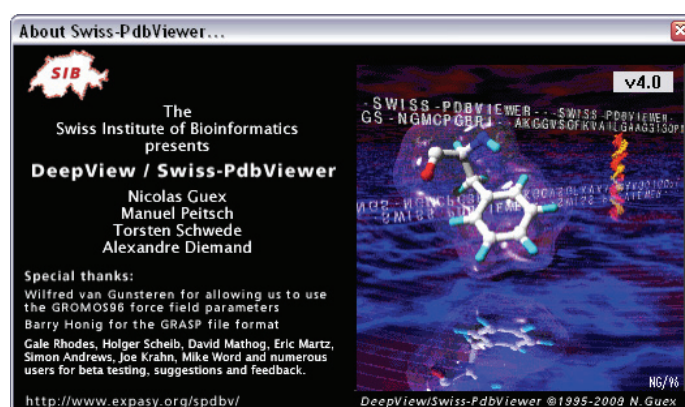You can download Deep View from http://spdbv.vital-it.ch/

Basic Deep View Controls

Starting Deep View

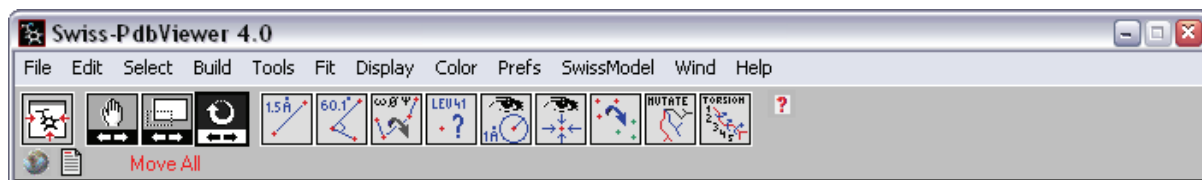Deep View is started by double-clicking the icon after installation.

You should see a splash screen appear.

Clicking on this will make it go away, and will leave you with the main program toolbar.

Opening Files

To open a PDB file into Deep View simply select File → Open from the File menu, and select the file. If the PDB file was generated by Deep View then your original colours and view will be preserved. If not, a default view will be presented.

Importing Files

If you want to bring in a structure directly from the PDB database, you can do this as long as you have the PDB accession code. This is a 4 letter/number code, and is usually quoted in any paper describing a structure (assuming the structure has been made public). To import a file you should select File → Import.

Simply enter the PDB accession code into the "Name" field of the form and press the button which says "PDB File" in the section "Grab from server". If you're sure that you've downloaded this file before then you can use the button in the "Grab from disk" section, which will save you having to transfer the file all the way from Switzerland again!

Searching for Structures

If you believe that a structure exists for a protein you're interested in, but you do no know its accession code, you can perform a text-based search from within Deep View. To do this, use File → Import again, but instead of an accession code, type a search string into the "Name" field. The search string is a series of words prefixed with either a "+" (if you want this to be included), or a "-" (if you want to exclude a term). For example, the string:

"+protease –serine –human"

Will find all non-human, non-serine proteases. The results you get back will look something like this:

If you want to download any of these structures simply click on the accession codes highlighted in red on the left. The first column will import a single protein chain from a structure, whereas the second column will import the whole structure.

Advanced Searching Tools

The search interface provided within Deep View is quite simple and may not be suitable if you have a more complex query you wish to perform. In these cases you can use one of the internet-based search systems to query the PDB database. One of the best interfaces is the Macromolecular Structure Database at the EBI, which can be found at:

http://www.ebi.ac.uk/msd/

In their "Services" section there are a number of tools which allow you to query the PDB. Probably the most useful of these is the MSDLite Search system which allows you to perform test based queries as well as adding functional filters based on the source or properties of the structures.

Basic Viewing Controls

When you first open a molecule in Deep View you should have the main toolbar open, and attached to this should be a viewing window containing your molecule. You may have other windows open as well, but we'll ignore those for the moment and come back to them later.

The first thing you'll want to do is to be able to move the molecule around.

Rotate, Translate and Zoom

The 3 basic controls for moving a 3D molecule are rotate, translate and zoom. These can all be accessed just using a mouse. To use these controls, position your mouse within the molecule window and use the controls shown below. For each control there is also a button placed on the main toolbar which can be used to achieve the same manipulation.

Movement	Mouse Control	Description	Toolbar Button
Rotate	Left mouse button	Up / Down rotates about the x-axis. Left / Right rotates about the y-axis.	
Translate	Right mouse button	Simply drag the molecule in the direction you wish to move it.	
Zoom	Both mouse buttons	Move up to zoom out, move down to zoom in.	

Although these controls give you a fairly flexible system of movement, they do not provide the fullest range possible, as they do not allow you to perform rotations or translations about the Z-axis (the one coming directly at you out of the screen). To achieve this extra movement there are a series of keyboard modifiers you can use in combination with the controls shown above.

F5 = Restrict movements to the **X-axis**
F6 = Restrict movements to the **Y-axis**
F7 = Restrict movements to the **Z-axis**
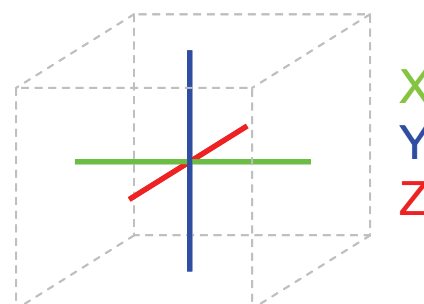
Centre View

A very useful tool is the furthest left button on the toolbar. This button will adjust the translation and zoom of your molecule so that it is all visible, and sits in the middle of the screen. It will also adjust the centre of rotation so that it falls within the middle of whatever is visible.

Slab

One problem with looking at 3D structures is that it's difficult to see what's going on in the middle of the structure as interior residues become obscured by those on the outside. To get around this you can use the Slab display mode. This shows you an interactive slice through your protein. Only where a residue falls inside this slice is it displayed.
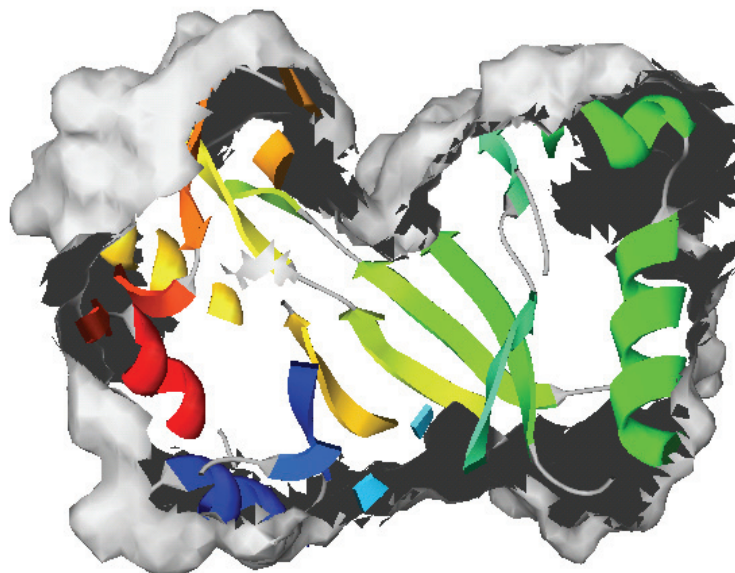
You turn the Slab display on and off by selecting Display → Slab from the main menu. Once you are in the Slab view all the movement controls work in the same way as before. One extra refinement is that if you want to drag your molecule through the slab (so you can see residues at the front or back) then you should use the Translate tool with F7 held (to restrict movement to the z-axis).
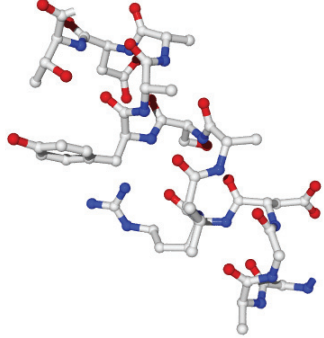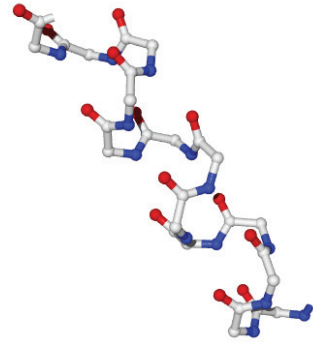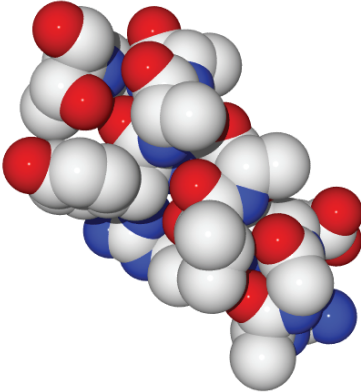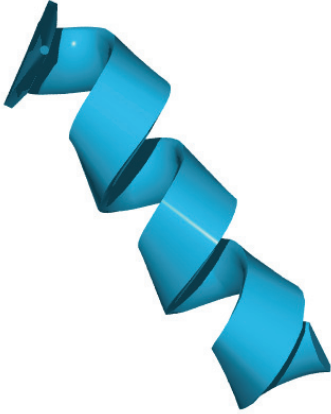
If you want to change the thickness of the slab in your display you can set this value under Preferences → Display.

Different Structural Representations

When you look at a molecule in a molecular modelling package you aren't looking at a realistic approximation of what that protein would really look like. Instead you get a cut-down stylised representation of the molecule, which although it is less accurate, conveys much more useful information about the molecule. There are a number of different ways you can represent a molecule in this kind of display, and Deep View allows you to mix and match them in any way you want. Some of the most common representations are shown below:

	This is a standard "ball and stick" representation of a stretch of protein. All the atoms in the structure are visible as points in space or small spheres, and the covalent bonds joining them are shown as lines or tubes.
	This is a similar representation to the ball and stick, except that it shows only the backbone residues. The atoms in the sidechains of the various amino acids are omitted. This produces a cleaner, less cluttered model in which larger structures are easier to visualise.
	This is a spacefilling representation. Covalent bonds are not shown, and atoms are represented by larger spheres which are sized to represent the size of the electron shell around the atom. This is a more realistic view, but is only useful when applied sparingly as it quickly obscures the view of other parts of the model.
	The final representation is the ribbon. This is a very simplified view where neither atoms nor bonds are shown. Instead a tube structure runs along the path traced out by the carbon backbone of the protein. Where the tube passes through a recognised region of secondary structure it is flattened to a ribbon shape. This representation is very useful for looking at larger structures within a protein.

The Control Panel

Most of the manipulation of your structure will take place in the Control Panel. If you can't see this window then select Window → Control Panel from the Window menu on the main toolbar.
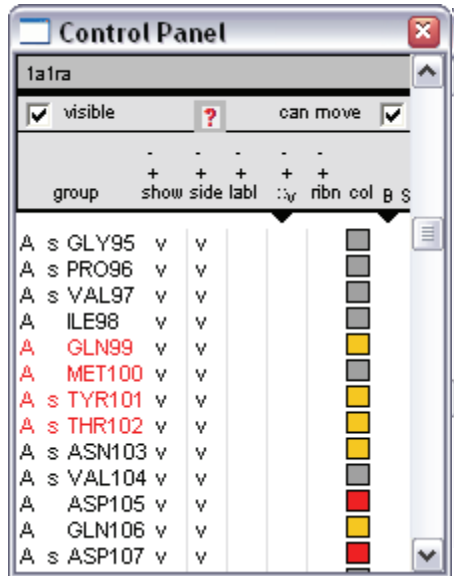
The top of the control panel gives you some basic information about the structure you are looking at. It tells you what the name of the current structure is, and whether it is visible and able to be moved (these options will be useful later when you're working with multiple structures).

Below this header information are a series of columns. These allow you to change various display options for some or all of the residues in your structure. A summary of what each of the different columns means is shown below.

Column Header	Description
Group	This text provides a description of a residue in your structure. It contains information about chain, secondary structure, amino acid and position. The format of the string is shown below
Show	This column denotes whether the backbone atoms of each reside are visible or not.
Side	This column indicates whether the residues side chain is visible. By default a side chain is never visible unless the backbone is also visible.
Labl	Turning this option on will place a small floating label next to the residue. By default this label will show the amino acid name and position.
: :	This interestingly named column (!) actually shows whether each residue in the structure has a surface associated with it or not.
Ribn	This says whether a ribbon representation of the structure is present for each residue
Col	This last column indicates what colour is being applied to each reside. Usually this is filled with a block of colour, but can also be a white box with a "-" sign in it, indicating that the objects default colour is being used.

The formatted string for the group identifier is laid out as below

Secondary Structure
S = beta-sheet
H = alpha-helix

Chain Identifier
This is optional
in structures
containing only
1 chain

Residue Number
Often numbered from the
full length protein, so
may not start at 1

A s GLY95

Residue Name
(standard 3 letter code
for amino acids)
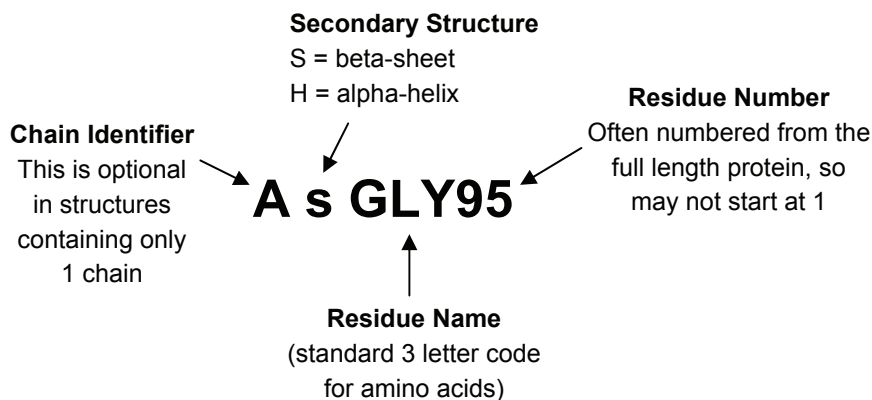
On two of the columns (Surface and Colour) there is a little black triangle ▼ below the column name). This indicates that the column can show more than one set of information. If you click on the triangle a list will drop down showing you the different options for that column. We will return to these options when discussing colours and surfaces.

Changing Attributes in the Control Panel
You will see that next to some residues there is a tick mark in some of the columns. Where a tick is present it means that that attribute is turned on for that residue. All of these attributes can be toggled by clicking on them with the left mouse button.

For instance, in this example you could add a label to residues 95 and 97 by clicking where the red crosses are shown.

If you want to change an attribute for all residues in a structure you can do so by clicking the right mouse button anywhere inside the appropriate column. In this example, clicking anywhere in the red area with the right mouse button would add a ribbon view for all residues in the structure.

Changing Colours

Deep View allows you to alter the colour of your molecules in pretty much any way you can think of. There are a number of inbuilt colour schemes, or you can define your own. Also, you can apply separate colour schemes to each different structural representation of your molecule.

Standard Colour Schemes
The standard colour schemes in Deep View are found under the "Color" menu on the main toolbar, in the first division at the top. They mostly colour your molecule according to some physical property of the residues in it. A full description of the most useful colour schemes is shown below:

Scheme Name	Colours
CPK	Coloured per atom by element type: Carbon — White Oxygen — Red Nitrogen — Dark Blue Sulphur — Yellow Hydrogen — Light Blue
Type	Coloured according to residue properties Acidic — Red Basic — Blue Polar — Yellow Hydrophobic — Grey
Secondary Structure	Coloured according to secondary structure Helices — Red Sheets — Yellow Loops — Grey
Secondary Structure Succession	Loops are coloured grey. All main secondary structure elements are coloured from cold (blue) to hot (red) running from the N to the C terminus.
Chain	Colours each chain in a structure a different colour. Useful if you are looking at a complex of two or more molecules.
Accessibility	Colours the structure according to the solvent accessibility of the residues. Cold colours indicate that residues are inaccessible, hot colours mean they are solvent exposed.
Other Colour	Brings up a colour selector and allows you to pick any colour you like to colour your molecule by.

Changing what you're apply colours to

To apply a colour you just select the appropriate entry from the Colour menu. The thing which catches people out is when they have a ribbon representation showing, they select a new colour scheme, and nothing happens. This is probably because they have just changed the colour of their backbone and sidechain (which they can't see).

Deep View allows you to apply a separate colour scheme to each of the following:

- Backbone and sidechain
- Backbone
- Sidechain
- Ribbons
- Surfaces
- Labels

act on Backbone + Sidechains
act on Backbone
act on Sidechains
act on Ribbon
act on Label
act on Molecular Surface

To change what you're applying your colours to you should go into the Color menu, and select the top entry which reads "Act on XXXXXX" (where XXXXXX is what you are currently applying colours to). You will then see a list of all the different features you can apply colours to.

Changing Colours in the Control Panel

In addition to the method outlined above, you can also use the Control Panel to change the colour scheme for your molecule.

| ✓ backbone + side |
| backbone |
| sidechain |
| ribbon |
| label |
| molecular surface |

To change what you're applying colours to you should click on the little ▼ symbol just to the right of the Col column. This will bring up a little popup menu from which you can choose what you're working on. The letters over this symbol indicate what you're working on at the moment (BS = backbone and sidechain, R=ribbon etc.).

You can change colours by using the coloured blocks in the Col column. These work in the same way as all the other Control Panel columns in that if you click in a box with the left mouse button you change the colour for one residue. If you click in a box with the right mouse button you change the colour for all residues.

Making things look Pretty

You will no doubt have noticed that whilst all the operations so far have been working fine, the image you have in the main display isn't exactly knocking you out with how beautiful it looks! This is a deliberate decision on the part of the Deep View developers. They give you the choice of presenting your molecule in a fast and simple wireframe representation (which is what you get by default), or as a slower but much prettier solid shaded representation (which is what we're coming to).

Deep View uses a system called OpenGL to make your molecules look solid. This is the same technology as is used in many 3D games. This means that if your computer is good for playing games, then it's also good for doing molecular modelling.

Levels of OpenGL

There are 3 different levels of graphical complexity within Deep View and more and more of your molecule will become solid as you step through them. The details of what is solid in each level is shown below.

	Atoms	Bonds	Surfaces	Ribbons
Wireframe	No	No	No	No
3D	No	No	Yes	Yes
Solid 3D	Yes	Yes	Yes	Yes

Wireframe	3D	Solid3D
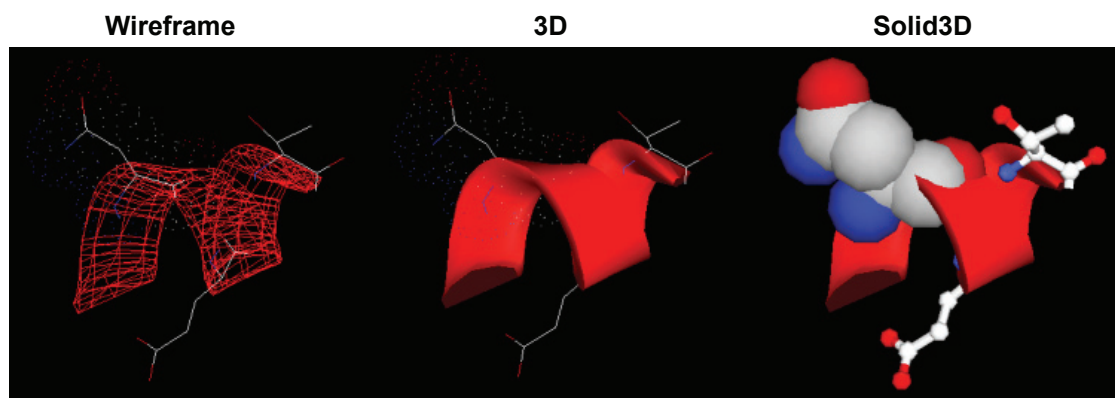
You can change these display settings under the "Display" menu, with the last two items at the bottom. You should note that the two 3D settings are additive, so the "Render in Solid3D" will also turn on "Render in 3D", but won't turn it off again when you turn solid 3D off.

Working with Selections

Often when looking at a molecule you would like to do something to a subset of the residues within it. In order to do this you need to be able to select some residues, either individually, or on the basis of some property calculation.

Using the Select Menu

Under the "Select" menu there are tools which allow you to make selections on the basis of various functional criteria. The most commonly used options within this menu are detailed below.

Name	Selection Criteria
All	Selects all residues in the current molecule
None	Deselects all residues in the current molecule
Inverse	Inverts the current selection
Visible Groups	Selects everything which is currently visible (either as a backbone, sidechain or ribbon). This selects everything which is theoretically visible, even if it isn't actually currently visible in the molecule window
Group Kind	Selects all of a particular amino acid, DNA base etc.
Group Property	Selects on the basis of the biophysical properties of your residues (acidic, basic, hydrophilic, hydrophobic etc.)
Secondary Structure	Selects residues based on the type of secondary structure they are part of.
Accessible aa	Selects residues whose solvent accessibility is greater than a user-supplied cutoff value. If you want to select residues whose accessibility is less than a certain value then you should use this function in combination with the Inverse selection function.
Neighbors of selected aa	Selects all the residues which fall within a sphere of user defined radius around all the currently selected residues.
Groups close to another chain	Selects all resides which fall on the region of contact between two chains in the same molecule. The cutoff for how close they must be can be specified by the user.

Making Selections with the Control Panel

In addition to the selection tools in the Select menu you can also make selections using the group column in the Control Panel.

In the Control Panel all selected residues (however they were selected), have their names highlighted in red.

Clicking on a chain identifier will select all the residues in that chain.

Clicking on a secondary structure identifier will select all of that piece of secondary structure. Note that this is different from the "Secondary Structure" in the Select menu. In the control panel you can, for example, select all the residues in **one** alpha helix, whereas in the Select menu you can only select the residues in **all** alpha helices.

Clicking on a residue name will select that residue.

Adding Selections Together

Often you will want to combine the results of more than one selection, for example you may want to select all tyrosines and tryptophans. To make another selection, whilst preserving the residues you already have selected you should hold down the **Control** key whilst making the second selection. This also applies in the Control Panel, where holding the **Control** key will allow you to select multiple residues in the group column.

In the Control Panel you also have an extra option, which is that you can use the **Shift** key to select all the residues between any two points. This would be useful if you wanted to select everything between residues 10 to 50.

Applying Attributes to Selections

Once you have made your selection you will want to do something with it. So far, all that has happened is that the names in the Control Panel have gone red. To do something useful with your selections you need to move back to the Control Panel.

We saw previously that clicking in a column turns an attribute on or off for one residue (if you clicked the left mouse button) or all residues (if you'd used the right button). The last part of this story is that **if you click on the column headers in the Control Panel you turn those attributes on for those residues which are selected, and off for everything else**.
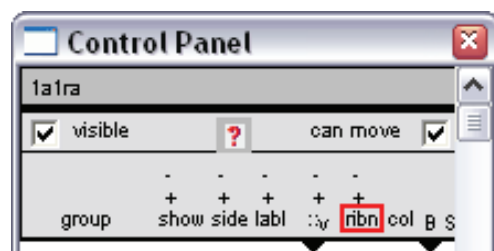
In this example, clicking anywhere in the red square would introduce a ribbon view only for those residues which had previously been selected. Any ribbons which had already been present would be removed.

This technique works for all of the column headers (except group) in the Control Panel.

If you already have some attributes turned on, and you just want to add in or remove some new residues which are selected you can do this by clicking on the + and – symbols above the column name. This will leave your existing attributes unaffected but will turn the attribute on or off just for the currently selected residues.

Adding Surfaces

Whenever you're working with structures it is easy to take the representation of the molecule which is displayed too literally. Most structural representations (eg: ball & stick, ribbon) make little attempt to present an accurate view of what a protein actually look like. Instead, they give you a schematic view which provides more information, but in a less realistic fashion.

The way to think of most molecular views is the same way as you might think of the schematic map of the London underground. It's less accurate than the real thing, but makes it a lot easier to get out the information you need.
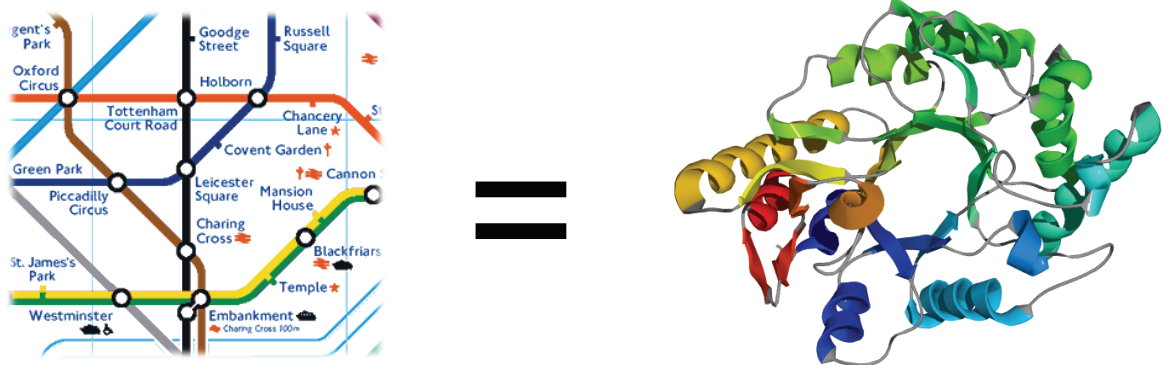
If you want a more realistic view of your molecule then you should consider adding a surface representation to all or part of it. It can be surprising the effect this can have. For example, in the molecule shown above, there isn't actually a hole in the middle of it, it's just the representation used which makes it look like there is.

Different Kinds of Surface

Once you decide that you want to add an extra dose of realism to your structure you still need to decide what kind of a surface to add. Deep View offers three different kinds of surface representation, and they're all good for different things.

Van der Waals Surface

This kind of surface is also sometimes referred to as a spacefilling view. Instead of all your atoms being represented by either points in space, or small spheres, in a Van der Waals representation they are expanded out to cover the full area which would be occupied by their electron shells. This then fills out most of the space surrounding each residue.

The image below shows the same 3 amino acid peptide sequence, only in the second image the middle amino acid has had a Van der Waals surface added to it.
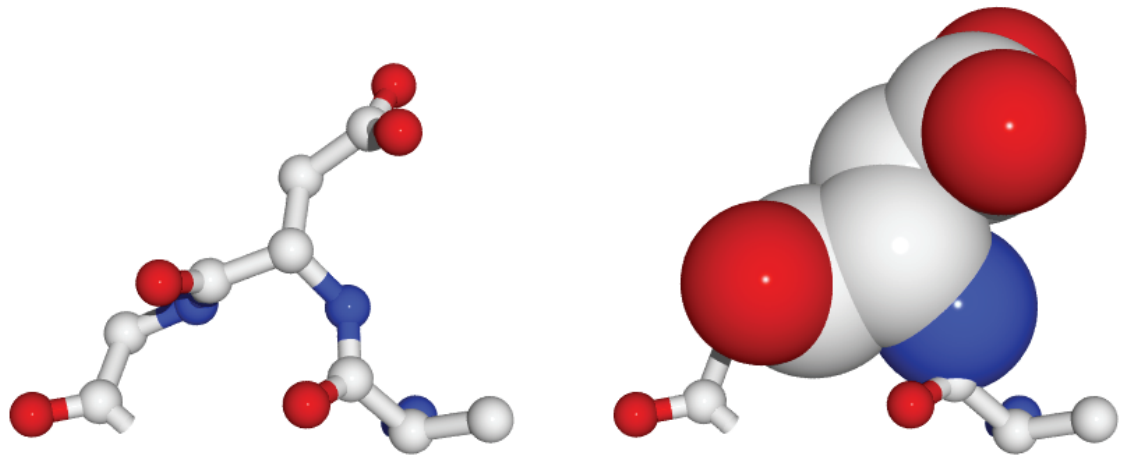
Van der Waals surfaces are particularly useful because you can add them to only a few residues. They are often used to highlight residues in an active site, or to show how a substrate would fill a binding pocket.

You can add Van der Waals surfaces to a molecule directly using the Control Panel. Just underneath the surfaces column (: :) there is a ▼ symbol. If you click on this you get a drop down menu showing the various kinds of surface available. Select VDW from this list and then use the surface column in the control panel to add a surface to whichever residues you like.

In order to see a Van der Waals surface, the underlying atoms must be visible. This means that the Backbone and Sidechain columns in the Control Panel must be ticked for the surface to have any effect.

You should note that unless you are using OpenGL + Solid 3D the Van der Waals surface will only show up as a series of dots around your atoms.

Molecular Surface
The other main type of surface you will find useful is the molecular surface. This type of surface is usually applied to whole molecules rather than individual residues. It adds an outer skin to the molecule, stretched around an imaginary Van der Waals surface underneath. This means that only the outside of the molecule is covered and that any other underlying molecular representations are still visible.
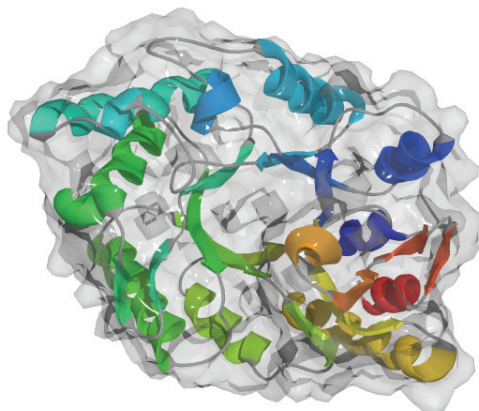
This is an example of a transparent molecular surface being applied to a molecule in such a way that you can still see the underlying ribbon representation beneath it.
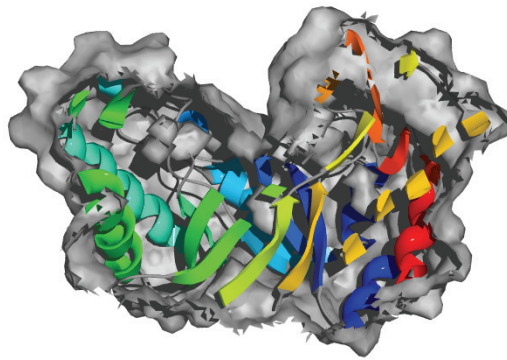
In this image a solid molecular surface has been applied to a molecule, then a cross section has been taken through it using the "Slab" option in the display menu. This allows you to see inside a molecule, but still retain an idea of where the external edge of the protein would be.

Another nice feature of molecular surfaces is that you can choose to exclude parts of your molecule when you are creating them. This is useful where you have a structure which is a complex of two molecules (eg an enzyme and its substrate). In these cases you wouldn't usually want to see the surface of the complex, but rather the surface of the main protein passing under the substrate.
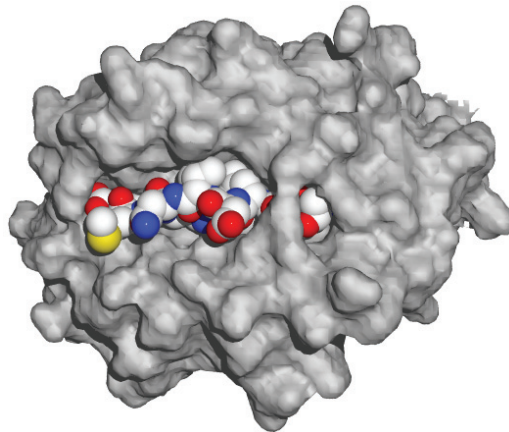
This image shows a Van der Waals representation of a peptide substrate sitting in the peptide binding groove of an MHC molecule which has had a molecular surface applied to it. This gives a good impression of how closely packed the two molecules are in the complex.

Molecular surfaces differ from Van der Waals surfaces in that they must be calculated before they can be displayed. Once you have calculated a molecular surface you can add and remove it using the Control Panel, the same way as for all other features.

Before you calculate a surface you should check your surface preferences in Preferences → Surfaces.
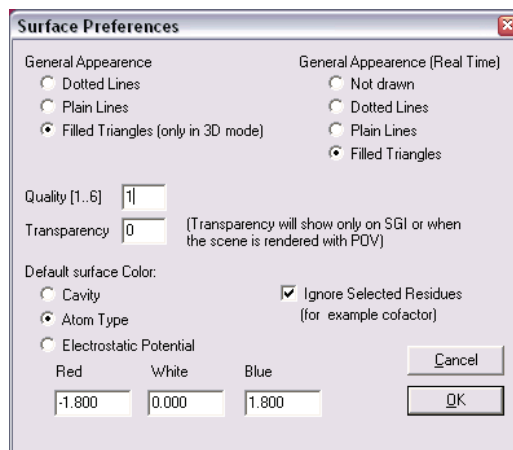
For the most part you can leave these settings alone, but there are a couple of things to beware of.

The main thing to check is that the surface quality is set to a sensible value. This determines how complex your surface will be. High quality surfaces take up a **lot** of memory and will slow things down considerably. For most purposes you will never need a surface quality higher than 3, and usually a quality 1 or 2 surface will suffice.

The other option to acknowledge is the one which says "Ignore Selected Residues". This means that whenever you calculate a molecular surface, any residues which are selected when you perform the calculation will be ignored and the surface will pass under them. Therefore if you want to calculate a surface for a whole protein you must use Select → None before starting the calculation. If your structure contains a substrate then you should select only the substrate before calculating the surface.

To calculate a molecular surface use Tools → "Compute Molecular Surface". Once the surface has been calculated you can manipulate it like any other feature using the Control Panel.

If you want to permanently delete a calculated surface you can do this by selecting "File → Discard → Surface".

Solvent Accessible Surface
The final kind of surface available within Deep View is the solvent accessible surface. The idea of this surface is to give an impression of what the molecule would look like to a water molecule. The solvent surface is therefore like a more diffuse molecular surface, only covering the exterior of the protein, but is like the Van der Waals surface in that it requires no calculation and can just be added via the Control Panel.

A solvent accessible surface is only a theoretical concept and doesn't represent any real surface. Therefore it is only ever displayed as a series of diffuse dots in Deep View, and never as a solid surface.

You probably won't make much use of this particular surface type.

Finding Regions of Interest

Once you are happy using the visualisation tools within Deep View you can start to use the program to look at interesting regions within your structure. The problem then becomes how to find the interesting regions. There are a couple of simple ways to identify the parts of the protein you should be looking at.

Letting other people do it!

By far the most important source of information about a specific protein will be the published analyses of the group who deposited the structure. It is always a good idea to retrieve the paper in which the structure was published, as well as any subsequent papers which cite it. This will nearly always provide you with the information you are looking for about your structure.

The PDB Header

When a structure is deposited, as well as the coordinates of the structure, the PDB file also has space for extra annotation. Most of this annotation will be facts and figures relating to how the structure was solved, and as such will only be of interest to other crystallographers, but there is often information in the header file which is of more general use. When looking at a new structure it is always worth having a look at the PDB header.

To access the PDB header of a file within Deep View you need to press the button which looks like a sheet of paper, on the top left of the main toolbar. This will then open a new text window which contains the header information.

The amount of information which can be found in the header will vary greatly depending on the structure, but at the very least it should always contain a reference to the paper in which the structure was described;

```
JRNL        AUTH   H.BRANDSTETTER,D.TURK,H.W.HOEFFKEN,D.GROSSE,         1ETR   8
JRNL        AUTH 2 J.STUERZEBECHER,P.D.MARTIN,B.F.P.EDWARDS,W.BODE      1ETR   9
JRNL        TITL   REFINED 2.3 ANGSTROMS X-RAY CRYSTAL STRUCTURE OF     1ETR  10
JRNL        TITL 2 BOVINE THROMBIN COMPLEXES FORMED WITH THE           1ETR  11
JRNL        TITL 3 BENZAMIDINE AND ARGININE-BASED THROMBIN             1ETR  12
JRNL        TITL 4 INHIBITORS /NAPAP$, 4-/TAPAP$ AND /MQPA$:           1ETR  13
JRNL        TITL 5 A STARTING POINT FOR IMPROVING ANTITHROMBOTICS      1ETR  14
JRNL        REF    J.MOL.BIOL.                   V. 226  1085 1992     1ETR  15
JRNL        REFN   ASTM JMOBAK  UK ISSN 0022-2836             070 1ETR  16
```
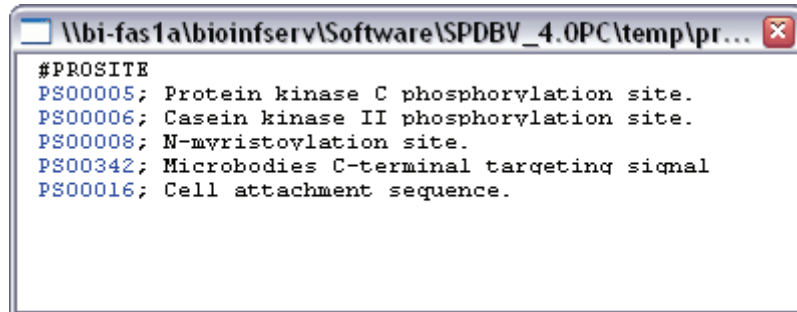
If you have a PDB file saved by Deep View which contains more than one structure then the headers will be concatenated so you may need to scroll down the page to find the headers for the structure you're interested in.

Finding Motifs

A simple way of finding sites of functional interest is to scan the protein sequence for any sequence-based motifs which have functional significance. These can then be mapped onto the structure. Deep View allows you to search through a motif database called Prosite, and to map the results of the search onto your structure.
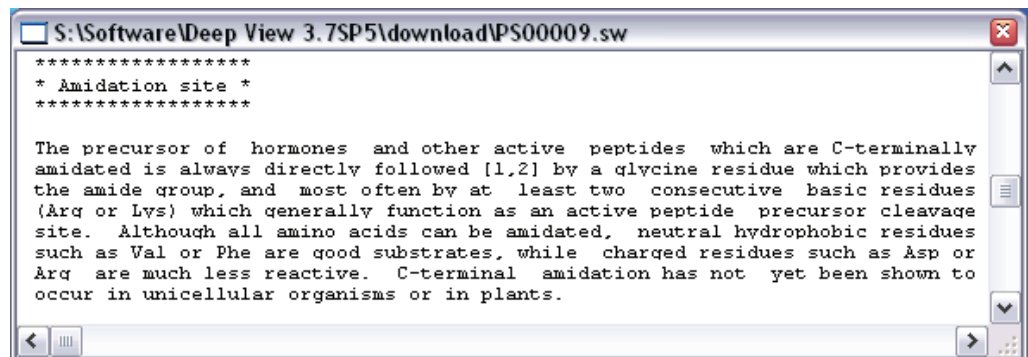
To run a prosite search on the current molecule simply select Edit → Search for Prosite Pattern. You should see a new window appear containing the list of patterns which were found in your sequence.

To see the position of the pattern within your structure you should click on the description of the pattern. To get a more detailed description of the pattern and what functional implications it has you should click on the prosite identification number.

You should be aware that the patterns in prosite fall into two distinct categories.

The most common category is short sites which are frequently found, and will turn up in most proteins. These generally represent sites for some sort of post-translational modification. The presence of one of these sites in your sequence doesn't mean that your protein has that modification, but if you know that your protein is modified in that way then this will give you the likely position of the modification. If you are not sure whether a particular site falls into this category then if you look at the page of information for the site they will all have:

CC   /SKIP-FLAG=TRUE;

At the top of the entry.

The other sites are generally longer, and will be annotated with a specific (usually catalytic) function. These sites represent the consensus sequences for the active sites of different classes of enzymes. Not every class of enzymes has a pattern which can be recognised in this way, but they are reasonably widespread. If you find one of these sites in your sequence then this is a strong indication that your protein has that functionality, and shows you where the active site is.

Looking around substrates

One of the easiest ways to find the residues which are likely to be of interest to you in a structure is to try to find a version of the structure which has a substrate in it. Usually when you perform a text search, the description of the structure will be something like XXXXXX complexed with YYYY, so it's fairly easy to spot if a substrate is present.

Once you have a structure with a substrate then you will want to be able to select those residues which are close to the substrate. The tools to do this have already been described in the "Working with Selections" section.

First you should identify the substrate. Substrates are usually listed in the Control Panel after the end of the normal protein chain, and before any water molecule. They usually have residue numbers which are a long way out of sequence from the rest of the protein. If your substrate is not a peptide then you can usually select it by using Select → Group Kind → Hetatom.
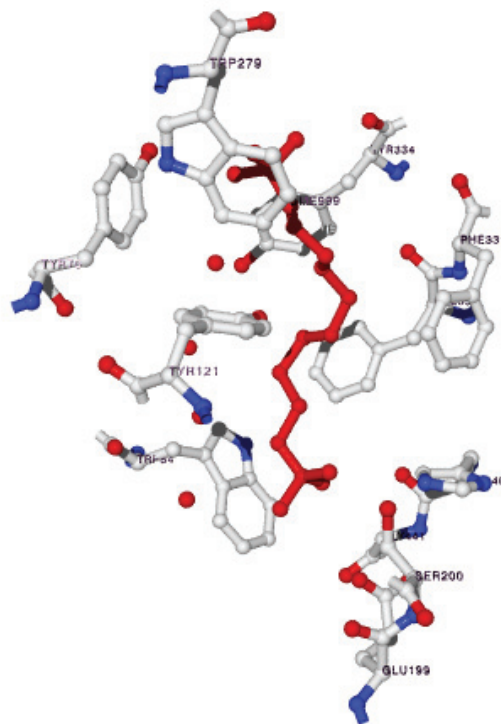
Once your substrate is selected you then want to also select any surrounding residues. You can do this by using Select → Neighbours of Selected Amino Acid, then choose "Select groups which are within", and choose a value of around 4-5Å.

Once your residues are selected then you can use the Control Panel to colour them or display them in a different way to the rest of the protein so they are highlighted.

This image shows the residues surrounding the red substrate.

Working with Multiple Structures

By now you should have enough information to allow you to manipulate and view a single structure in pretty much any way you like, but Deep View doesn't limit you to only viewing one structure at a time, in fact you can have up to 12 structures open at once.

Working with multiple related structures can provide you with a lot more information about your protein as it allows you to look at the differences between related structures and relate them to known functional differences.

Loading Multiple Structures

Once you have loaded in your first structure you can use exactly the same methods to open a second one. The two structures will appear simultaneously in the molecule view window, but will be unaligned when first opened. The two structures will not be merged together, but will be put into separate **layers**.

Choosing an Active Layer

A new concept you need to pick up when working with multiple structures is that of the active layer. At any time only one layer can be active. It is the active layer whose details appear in the control panel and on which all actions will operate (eg colour changes, surface calculations etc). When you have more than one layer open it is useful to look at the "Layers Info" window, which can be opened from the Window menu on the main toolbar.

Layers Info																	
?Layer	**taxon**	**sel**	**grp**	**vis**	**mov**	**axis**	**CA**	**O**	**H**	**Hbnd**	**Hdst**	**side**	**HOH**	**cyc**	**Aln**	**Wmdl**	**SelGrp**
ltax	Thermoascus aurantiacus	v	1	v	v			v	v	v			v	v	v	*	0
2exo	Cellulomonas fimi	v	2	v	v			v	v	v			v	v	v		0

This window shows you a quick summary of the status of the various layers you currently have available. The columns which you are likely to use are detailed below

Column	Description
Layer	The layer column gives the names of the structures you have open in each layer. One of the names will be highlighted in red and this is the currently active layer. You can make any layer active by clicking on its name in this column
Vis	This column shows whether this layer is currently visible. This option overrides any options you may have set in the columns of the Control Panel, and is the same thing as the "visible" tick box at the top of the control panel.
Mov	This column shows whether this layer will move when you apply any of the standard movement control. Having this attribute set to different values for different layers you can alter the relative alignment between two structures.
SelGrp	This is a counter to show how many residues are currently selected in each layer.
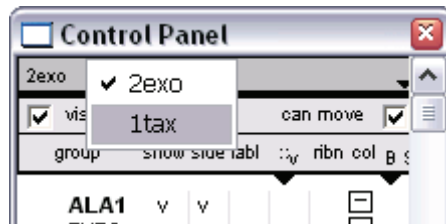
The Active Layer can also be set using the Control Panel. At the top of the Control Panel is a small piece of text showing the currently active layer. If you click on this text you should see a popup menu appear showing all the currently available layers. Simply select the one you want to be active.

Aligning Structures

Automated Alignment

When you first load in multiple structures they will not be aligned, but will be in the orientation they were originally saved in. In order to make meaningful comparisons between structures it is necessary to align them so that you can quickly see which parts of the structures correspond to each other.

You have already seen that it is possible to adjust the relative alignment of two structures by turning off the "Mov" attribute for one of them in the Layers Info window, and then moving the other one. **This is an extremely bad way of aligning structures and you will drive yourself mad if you try to do it this way!**

The best way to align two structures is to use the Tools available under the "Fit" menu. In particular, the "Iterative Magic Fit" tool is the one you should use for nearly all structural alignments. This tool will perform a sequence-level alignment of your structures to identify corresponding pairs of residues. It then adjusts the alignment of the structures so that as many of these pairs of residues as possible lie next to each other.

When you first select the Iterative Magic Fit tool you will be presented with some options for how you wish to perform the alignment.
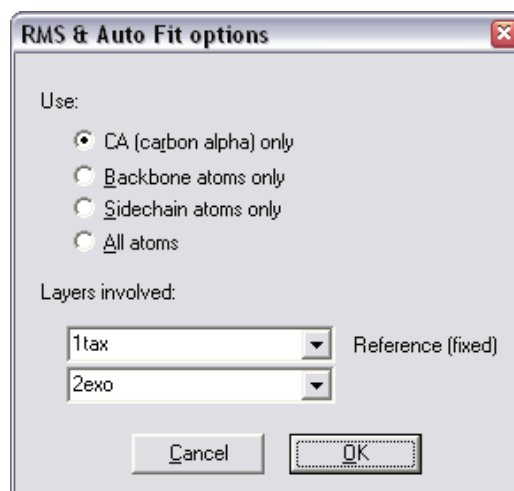
The first option allows you to select which residues will be used to align around. When you are aligning structures, the more atoms you try to align, the slower the alignment will be. For whole protein alignments, the structural alignment is basically just the alignment of the carbon backbones. Adding in sidechains isn't going to change anything and will just slow things down. Therefore:

- When aligning two normal proteins you only need to select "CA (carbon alpha)".

- For very small proteins you might be better off selecting "Backbone atoms only"

- For peptide fragments you might consider "All atoms"

The other thing you can change is the layers involved. When you only have two structures open this isn't much of a choice! The only thing to remember is that the layer in the top box will remain static and the layer in the second box will be moved to align with it.
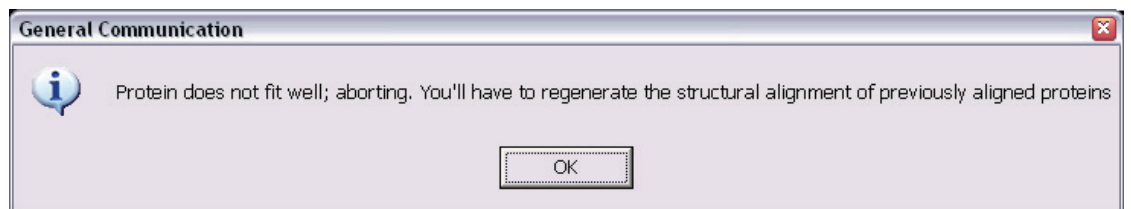
Once you're happy with the settings press OK to align your structures. Usually this works well, but occasionally you will see a message saying:

General Communication

Protein does not fit well; aborting. You'll have to regenerate the structural alignment of previously aligned proteins

OK

This usually means that the two structures aren't actually related and don't align. Occasionally this message can occur because the sequence divergence between the two sequences was too high. In this case you can attempt to align the structures using the manual alignment method described later.

Manual Alignment

Sometimes an automated alignment isn't appropriate. This is usually because:

1) The sequences in your structures are very diverse, and the automated alignment cannot figure things out on its own.

2) You want to bias your alignment around a particular set of resides (say the components of an active site) rather than align the whole protein.

In these cases you can perform a manual alignment. To do this you must select the corresponding residues in the two structures. Your residues must be selected in corresponding pairs, and they must be the same residue in each structure. You need at least 3 pairs of residues to be selected to perform this kind of alignment.

Probably the easiest way to make this kind of selection is with the alignment window. You can make this visible by selecting Window → Alignment from the main menu.

Alignment

1tax   ASAQSVDQLIKARGKVYFGVATDQNRLTTGKNAAIIQANFGQVTPENSMKWDATEPSQGNFNFAGAD
2exo   ATTLKEAADGAGR- DFGFALDPNRLSEAQYKAIADSEFNLVVAENAMKWDATEPSQNSFSFGAGD
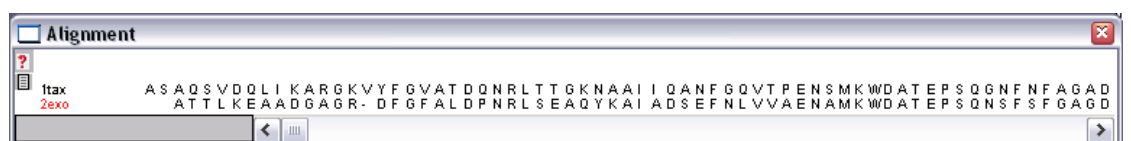
The alignment window shows you an alignment based on which residues are closest together in 3-dimetional space.

Selections in the alignment window work just like selections in the Control Panel. You can make discontinuous selections by holding down the Control Key. Being able to see the sequence of all layers makes it easy to select corresponding pairs of residues, as shown below:
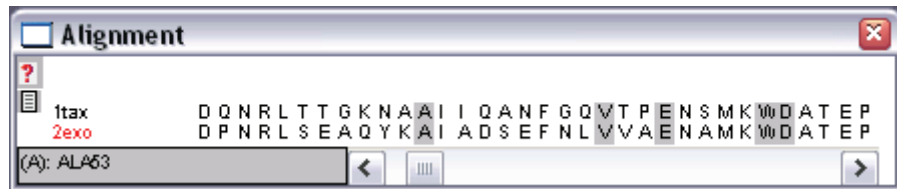
Once you have selected your residue pairs you can start the alignment by selecting Fit → Fit Molecule (From Selection) from the main menu. You then go through the same process as for an iterative Magic Fit to align your molecules. Because you are using only a small number of residues to calculate your alignment you may want to use the "All atoms" option for performing the alignment.

Aligning more than two structures

The above methods work fine for aligning two structures, but they provide no option for aligning more than two. As of yet there is no multiple alignment option for structures, but there is a simple way around this.

If you want to align three or more structures you must do so by deciding that one of them should be the reference structure, and then aligning all of the other structures to it. Because structural conservation within a protein family is so high this actually works really well for creating a family alignment.

You perform the individual alignments in the same way as described above, but you must make sure that when you select the layers involved your reference structure must always appear in the top box.
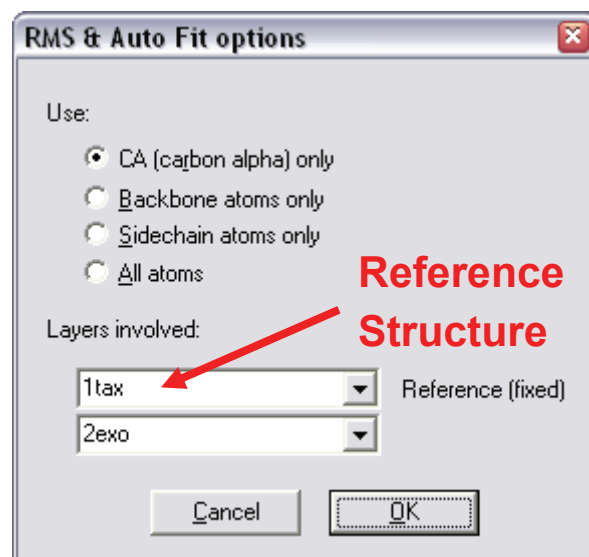
The last thing to mention about multiple alignments is that once you have aligned all your structures the sequence alignment in the Align window will probably be wrong. You can force it to be updated by selecting Fit → Generate Structural Alignment.

RMS Values

Something which is often used to represent the degree of relatedness of two structures is an RMS value. The RMS value is the Root Mean Squared value for the average distance, in Angstroms, between corresponding atoms. The lower the RMS value, the more closely related the two structures are. By definition, a structure compared to itself has an RMS value of 0. Structures which are related will usually have an RMS value between 0 – 2Å.

When you align two structures using either of the methods described above, Deep View will automatically calculate an RMS value for you. This is displayed in red text just below the buttons on the main toolbar.

When using the Iterative Magic Fit the program will use all of the corresponding atoms it can find. For the manual alignment it will use only the pairs of residues you specified.

Colour Schemes to use with alignments

Once you have your structures aligned there are a couple of extra colour schemes which become useful to you.

Layer

Colouring your molecules by layer simply assigns a different colour to each open layer. This is a quick and easy way of being able to tell which residues come from which structure when they're all overlaid. The other advantage to this (and all other colour schemes) is that the colouring applies not only to the view window, but also to the letters in the alignment window and the colour blocks in the Control Panel.
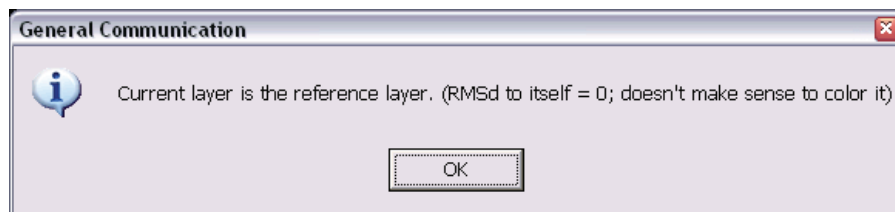
Alignment Diversity

This colour scheme colours your models by the degree of conservation in their underlying sequence alignment. As with many of the other colour schemes it works on a cold → hot colour series where cold colours are the most conserved positions and hot colours are the most divergent. The alignment diversity colour scheme will automatically apply itself to all the layers you currently have open.

RMS

The RMS colour scheme colours your model by the amount of separation in 3D space between corresponding pairs of residues in your structure. It doesn't matter which residues are present in your sequences, this is a measure of how divergent the actual structures are. You will usually find that whereas two structures are quite divergent at a sequence level, they will usually be pretty well conserved at a structural level.

The RMS value for each residue is calculated by reference to your original reference structure. This means that if you have two structures aligned together, only one of them can be coloured by RMS. If you try to apply an RMS colour scheme to the reference layer then you will see a warning like this:

General Communication

(i) Current layer is the reference layer. (RMSd to itself = 0; doesn't make sense to color it)

OK

Usually it's a good idea when using the RMS colour scheme to hide the reference layer (Turn off the vis attribute in the Layers window), and just display the other layer(s) coloured by RMS.

Saving Your Work

If you have spent time creating a large family alignment, or setting up a specific view of a protein then you don't want to have to throw it all away when you close Deep View down. Fortunately there are a couple of options you can use to save the work you have done.

Saving Structures / Views

Single Layers

If you have a model which contains only one layer then you can save this using:

File → Save → Current Layer

This will save the layer as a PDB file. The file created is a standard PDB file but with extra Deep View formatting codes in the header which tell the program how the view was configured when you saved. Deep View will read these codes and re-create the view when you next open the file.

If you want to use the PDB file in another program then it should work OK. You should be aware though that saving data from Deep View does not transform the underlying coordinates of the structure, so the view you have will not be preserved in anything other than Deep View. If you need to transform the raw data then speak to the Bioinformatics group who can advise you on how to do this.

Multiple Layers

If you have a structural alignment you wish to preserve then you can do this by using:

File → Save → Project (All layers)

This will create a single PDB file which contains all of the coordinates from all of the structures along with all of the view information to allow the scene to be recreated. You may well find that if you transfer this multiple PDB file to another structure display program it does not open correctly. If you need to do this, please contact the Bioinformatics group who may be able to help.

What is lost when saving?

When you save your work in Deep View not everything is preserved by default. The things you will lose unless you take some further action are:

- Any molecular surfaces you had calculated
- Any electrostatics potentials you had calculated

Surfaces and electrostatic potentials can be saved, but they require you to perform an extra step.

Saving and Loading Surfaces and Electrostatics

To save a surface from the currently active layer use:

File → Save → Surface

To save an electrostatic potential from the currently active layer use:

File → Save → Electrostatic Potential

Please note that these only save from the current layer. If you have an alignment of 10 proteins all with surfaces coloured by electrostatic potential then that's 21 separate saves you have to do. I know this seems a lot, but that's just how it is!

To load a surface back onto a structure you first need to reload the structure and make it the active layer. You can then reload the surface using:

File → Load Surface → SPDBV

Likewise for Electrostatic Potentials you would use:

File → Load Electrostatic Potential → SPDBV

Creating Publication Quality Graphics

Whilst the quality of the graphics in Deep View is quite good, especially in OpenGL mode, it is necessarily limited by the need to be able to update the image several times per second. There comes a point though where you would like to create a high quality image for use in a report or poster and for this Deep View interacts with a separate program called POV-Ray

POV-Ray

POV-Ray is a ray-tracing program, and doesn't specifically have anything to do with molecular modelling. It allows you to describe a 3-dimentional scene in terms of simple geometrical shapes and then add lights around it and define a view-point. POV-Ray will then simulate this scene and generate an image of what it would look like if you could see it. Because of the way the scene is described POV-Ray can create images of any resolution for a given scene, and can therefore produce extremely high quality output.

Creating POV-Ray files with Deep View

Deep View contains a function which converts its internal representation of your molecule into POV-Ray's scene language. To create a POV-Ray file from your current Deep View view simply select:

File → Save → POV-Ray Scene

It doesn't matter which display mode you are using when you are saving your POV scene (3D or wireframe) and the results will always be the same. However, the scene will reflect your current view exactly, including the level of zoom and the current view position. You should therefore ensure that you can see everything you need to see before going any further.

When you save a POV-Ray scene, two files will be created. The first ends with a .pov extension and contains the preferences settings for your scene (colours, textures, lights etc.) and the second ends with .inc and contains the actual shape descriptions Deep View used to describe your scene

A word of warning at this point! If you are creating POV-Ray files for a protein which has a molecular surface on it, then the .inc file can become very large. This isn't a problem as such as the images will still be generated as you expect, but don't be surprised if you get a file of several megabytes, and the POV-Ray files take a few seconds to be generated.
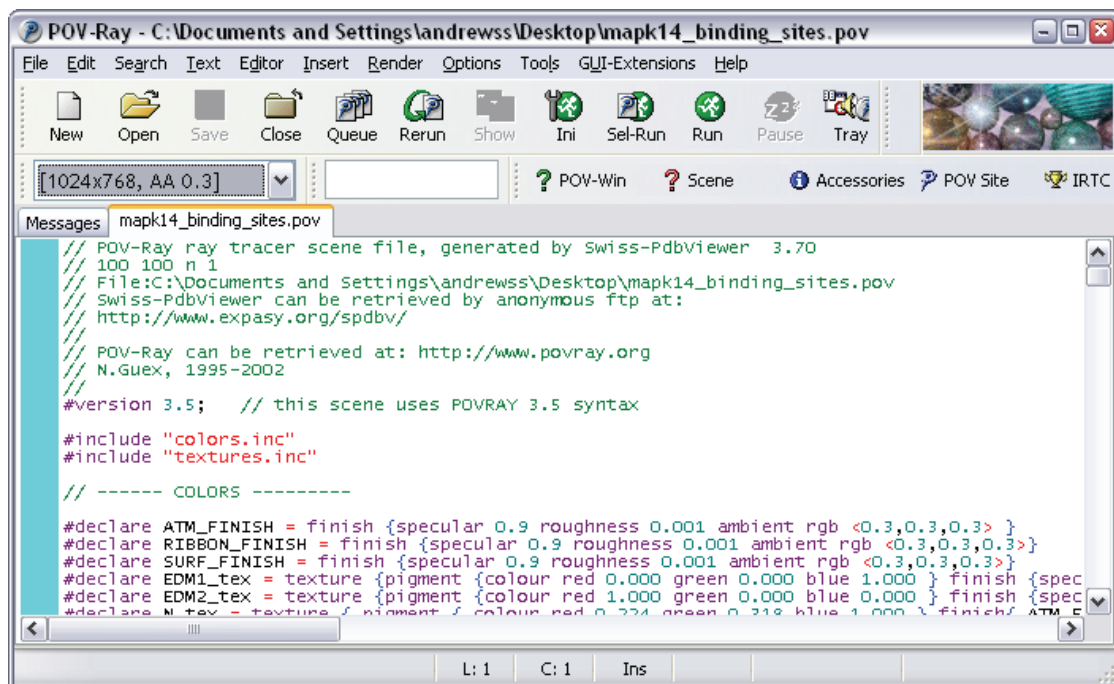
Creating an image from the POV-Ray files

Once you have your .pov and .inc files you want to use them to make an image. This requires the POV-Ray program to be installed somewhere for you to use.
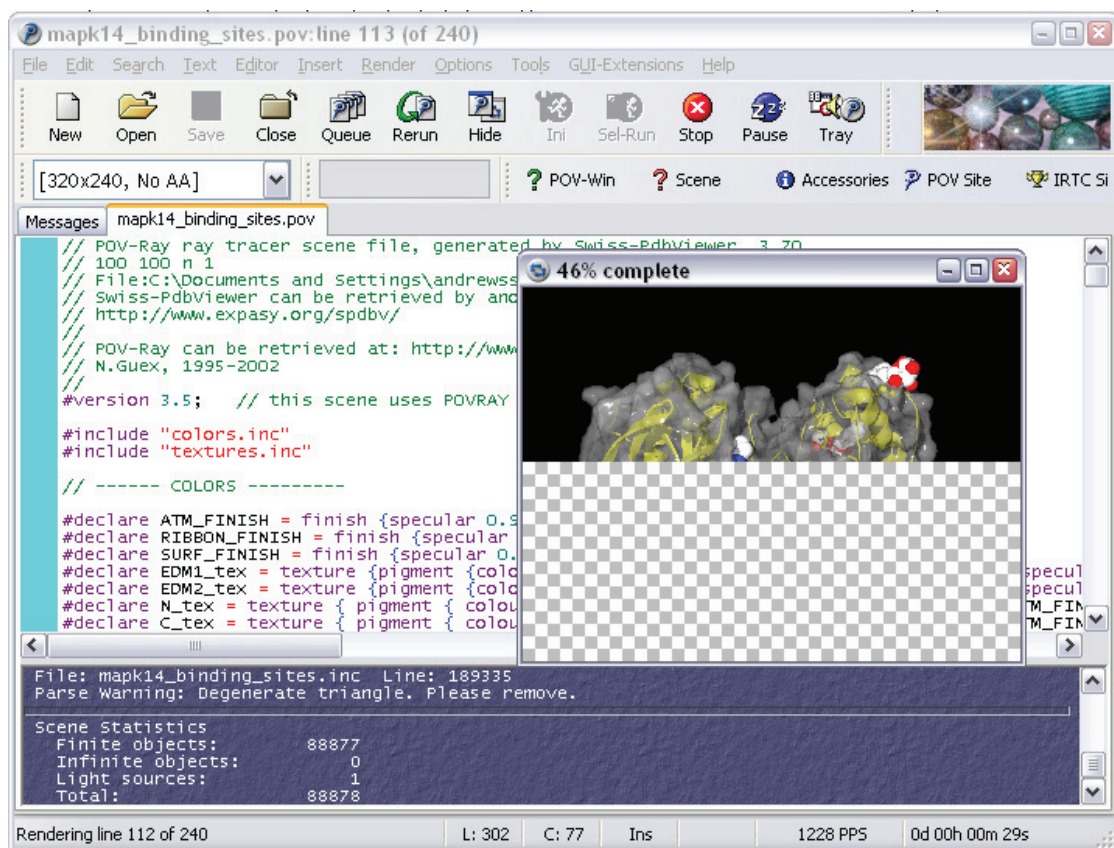
You can download and install povray from www.povray.org

Once you have saved your files you should be able to double click on the .pov file to load it into provray

You can use the drop down dialog box under the New / Open icons to select the size of image you want. Then press the green 'Run' button to start generating your image.

Your rendered image will be saved into the same folder as your original .pov file. There is no need to save the image once the render has completed.