

## Z. Practical: Protein structure visualisation

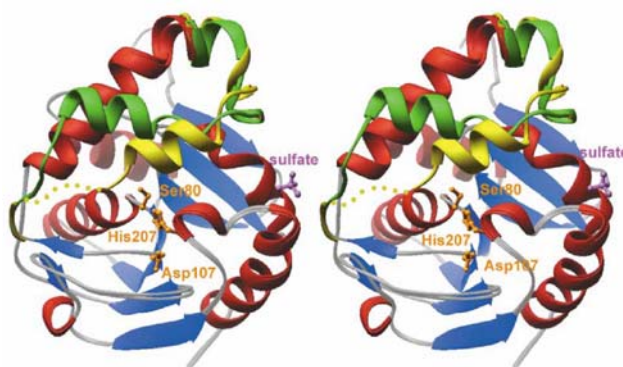
**Verantwortlich:** Milton T. Stubbs  
**Raum:** Rechnerraum im Institut für Biotechnologie

**For possible changes / more info:** see <http://www.biochemtech.uni-halle.de/xray/Praktika>

### Z.1. Introduction: Structural biology

Open any textbook or journal today, and more often than not, you will be confronted with colourful diagrams depicting the three-dimensional structure of biomolecules. In some cases (advertisements), these are designed to be eye-catching, depicting the ‘modernity’ of the company selling the product. Alternatively, the structures are presented to indicate details of a fundamental biological process. If you want to adapt your protein to perform novel functions, then the X-ray structure is enormously valuable.

Yet there is a fundamental problem in representing three-dimensional information on a two-dimensional page – it just cannot be done for complex structures. And the structures that are being solved are getting more and more complex.... How can one get around this? One way is to use stereo pictures (like the ‘magic eye’ books):



**Figure Z.1.** Can you see this in stereo? The figure shows two differing conformations of the active site lid in the thioesterase domain of surfactin synthetase (Bruner *et al.*, Structure **10**, 301-310, 2002), which is responsible for peptide cyclisation

Yet this only allows us to see what the authors of the paper wanted us to see. On the other hand, all published structures are (*supposed to be*) deposited in the Protein Data Bank (**PDB**). All the information is there for us to take a look for ourselves. This is not to say that the authors are trying to pull the wool over our eyes – but the structure contains so much information, they cannot describe **everything** that might be interesting to **us**.

The goal of this practical will be to demonstrate how to extract a three-dimensional structure from the PDB, how to visualise it using **PyMol**, and how to make pretty (maybe) and informative (perhaps) representations of your structure. **PyMol** is freely available on the web (<http://pymol.sourceforge.net/>), and is platform independent – so you can install it on your own computer and spend hours of fun looking at your favourite macromolecule....

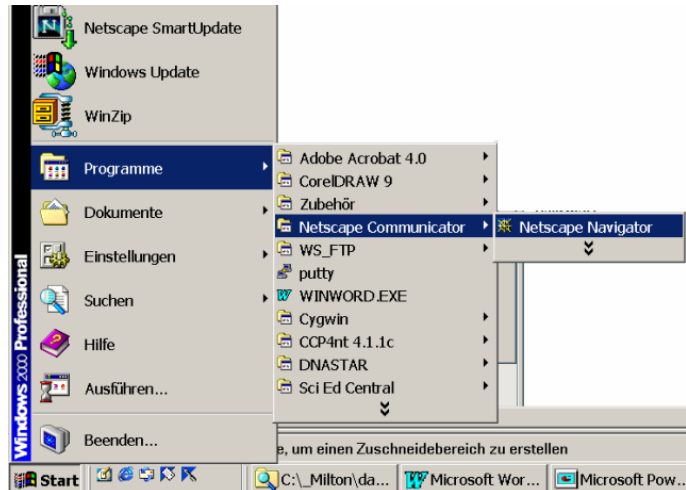
## Z.2 Protocol

The assistant will give you a paper beforehand describing a macromolecular structure. It is your task to search the PDB for the structure(s) described therein, to download these, and view it/them interactively. This will give you the opportunity to study the aspects described in the paper at first hand.

1. Download structure from PDB

Open Netscape (or Explorer):

- (1) hit **Start** with the left mouse button
- (2) move the mouse to Programme, then Netscape Communicator and finally Netscape Navigator; click with the left mouse button
- (3) type in the window **http://www.rcsb.org/pdb/**



which will bring you to the RCSB Protein Data Bank (PDB):



Now you have to select your structure. You can either give in the pdb code given in your paper, a keyword (e.g. **trypsin**), or type in the authors:

**PDB Query Result - Netscape**

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Bookmarks Location: http://www.pdb.mdc-berlin.de/pdb/cgi/resultBrowser.cgi

Instant Message Instant Message Internet Lookup New&Cool Happy New Year

**PDB**  
PROTEIN DATA BANK

**Query Result Browser**

Help PDB Home Contact us

Your query found 688 structures in the current PDB release and you have selected 0 structures so far. (There are currently 11 structures being processed or "on hold" matching your query!) You can select specific structures by clicking on the checkbox next to their id. If you do not select any structures, certain options will default to all structures. To examine an individual structure select the Explore link!

Pull down to select option: New Search Go

1-20

<input type="checkbox"/>	<b>1A0J</b>	Deposited: 01-Dec-1997 Exp. Method: X-ray Diffraction Resolution: 1.70 Å	(EXPLORE)
<i>Title</i> Crystal Structure Of A Non-Psychrophilic Trypsin From A Cold-Adapted Fish Species.			
<i>Classification</i> Serine Protease			
<i>Compound</i> Mol. Id: 1; Molecule: Trypsin; Chain: A, B, C, D; Ec: 3.4.21.4; Other Details: Benzamidine Inhibitor In The Active Site			
<input type="checkbox"/>	<b>1A0L</b>	Deposited: 03-Dec-1997 Exp. Method: X-ray Diffraction Resolution: 3.00 Å	(EXPLORE)
<i>Title</i> Human $\beta$ -Tryptase: A Ring-Like Tetramer With Active Sites Facing A Central Pore			
<i>Classification</i> Serine Proteinase			
<i>Compound</i> Mol. Id: 1; Molecule: $\beta$ -Tryptase; Chain: A, B, C, D; Ec: 3.4.21.59; Biological_Unit: Tetramer; Other_Details: In Complex With Amidino Phenyl Pyruvic Acid (Appa), A Synthetic Inhibitor			
<input type="checkbox"/>	<b>1A1Q</b>	Deposited: 12-Dec-1997 Exp. Method: X-ray Diffraction Resolution: 2.40 Å	(EXPLORE)
<i>Title</i> Hepatitis C Virus Ns3 Proteinase			
<i>Classification</i> Hydrolase			
<i>Compound</i> Mol. Id: 1; Molecule: Ns3 Proteinase; Chain: A, B, C; Engineered: Yes; Biological_Unit: Monomer			
<input type="checkbox"/>	<b>1A2C</b>	Deposited: 26-Dec-1997 Exp. Method: X-ray Diffraction Resolution: 2.10 Å	(EXPLORE)
<i>Title</i> Structure Of Thrombin Inhibited By Aeruginosin298-A From A Blue-Green Alga			
<i>Classification</i> Complex (Serine Protease/Inhibitor)			
<i>Compound</i> Mol. Id: 1; Molecule: $\alpha$ -Thrombin; Chain: L, H; Ec: 3.4.21.5 Mol. Id: 2; Molecule: Hirugen; Chain: I; Engineered: Yes Mol. Id: 3; Molecule: Aeruginosin298-A; Chain: R			
<input type="checkbox"/>	<b>1A5H</b>	Deposited: 17-Feb-1998 Exp. Method: X-ray Diffraction Resolution: 2.90 Å	(EXPLORE)
<i>Title</i> Catalytic Domain Of Human Two-Chain Tissue Plasminogen Activator Complex OFA Bis-Benzamidine			
<i>Classification</i> Hydrolase			
<i>Compound</i> Mol. Id: 1; Molecule: Tissue Plasminogen Activator; Chain: C, D; Fragment: Heavv Chain Fragment, Catalytic Domain; Ec: 3.4.21.68;			

Document: Done

In this case, 688 structures are far too many to find what we want – so we have to narrow down our search. We can combine our query with other keywords; e.g. **trypsin AND bovine** will pull out all pdb entries containing BOTH keywords (if we had used **trypsin OR bovine**, then we would get all entries containing EITHER keyword – i.e. a lot more than we want.... Other combinations include e.g. (**trypsin AND bovine**) NOT stubbs....

**PDB Query Result - Netscape**

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Bookmarks Location: http://www.pdb.mdc-berlin.de/pdb/cgi/resultBrowser.cgi

Instant Message Instant Message Internet Lookup New&Cool Happy New Year

**PDB**  
PROTEIN DATA BANK

**Query Result Browser**

Help PDB Home Contact us

Your query found 289 structures in the current PDB release and you have selected 0 structures so far. You can select specific structures by clicking on the checkbox next to their id. If you do not select any structures, certain options will default to all structures. To examine an individual structure select the Explore link!

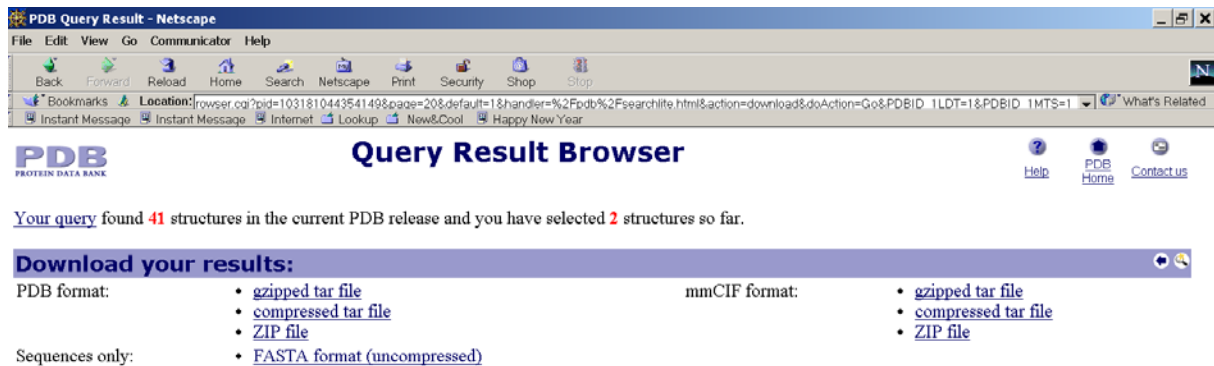
Pull down to select option: New Search Go

1-20

<input type="checkbox"/>	<b>1AAL</b>	Deposited: 09-Apr-1992 Exp. Method: X-ray Diffraction Resolution: 1.60 Å	(EXPLORE)
<i>Title</i> Structural effects induced by mutagenesis affected by crystal packing factors: the structure of a 30-51 disulfide mutant of basic pancreatic trypsin inhibitor.			
<i>Classification</i> Serine Protease Inhibitor			
<i>Compound</i> Bovine Pancreatic Trypsin Inhibitor (Bpti, Basic) Mutant With Cys 30 Replaced By Val and Cys 51 Replaced By Ala (C30V, C51A)			
<input type="checkbox"/>	<b>1AB9</b>	Deposited: 05-Feb-1997 Exp. Method: X-ray Diffraction Resolution: 1.60 Å	(EXPLORE)
<i>Title</i> Crystal Structure Of Bovine $\gamma$ -Chymotrypsin			
<i>Classification</i> Complex (Serine Protease/Peptide)			
<i>Compound</i> Mol. Id: 1; Molecule: $\gamma$ -Chymotrypsin; Chain: A, B, C; Ec: 3.4.21.1 Mol. Id: 2; Molecule: Pentapeptide (Tpgvy); Chain: D; Engineered: Yes			
<input type="checkbox"/>	<b>1ACB</b>	Deposited: 08-Nov-1991 Exp. Method: X-ray Diffraction Resolution: 2.00 Å	(EXPLORE)
<i>Title</i> Crystal and molecular structure of the bovine $\alpha$ -chymotrypsin-eglin c complex at 2.0 Å resolution.			
<i>Classification</i> Hydrolase(Serine Protease)			
<i>Compound</i> $\alpha$ -Chymotrypsin (E.C. 3.4.21.1) Complex With Eglin C			
<input type="checkbox"/>	<b>1ADZ</b>	Deposited: 19-Feb-1997 Exp. Method: NMR, 30 Structures	(EXPLORE)
<i>Title</i> The Solution Structure Of The Second Kunitz Domain Of Tissue Factor Pathway Inhibitor, NMR, 30 Structures			
<i>Classification</i> Hydrolase			
<i>Compound</i> Mol. Id: 1; Molecule: Tissue Factor Pathway Inhibitor; Chain: Null; Fragment: Factor Xa-Binding Domain (Domain II); Synonym: Tf $\beta$ i, Epi, Laci; Engineered: Yes; Mutation: N-Terminal Ins(Dykdddkk); Biological_Unit: Monomer			
<input type="checkbox"/>	<b>1AFQ</b>	Deposited: 12-Mar-1997 Exp. Method: X-ray Diffraction Resolution: 1.80 Å	(EXPLORE)
<i>Title</i> Crystal Structure Of Bovine $\gamma$ -Chymotrypsin Complexed With A Synthetic Inhibitor			
<i>Classification</i> Complex (Serine Protease/Inhibitor)			
<i>Compound</i> Mol. Id: 1; Molecule: Bovine $\gamma$ -Chymotrypsin; Chain: A, B, C; Ec: 3.4.21.1			

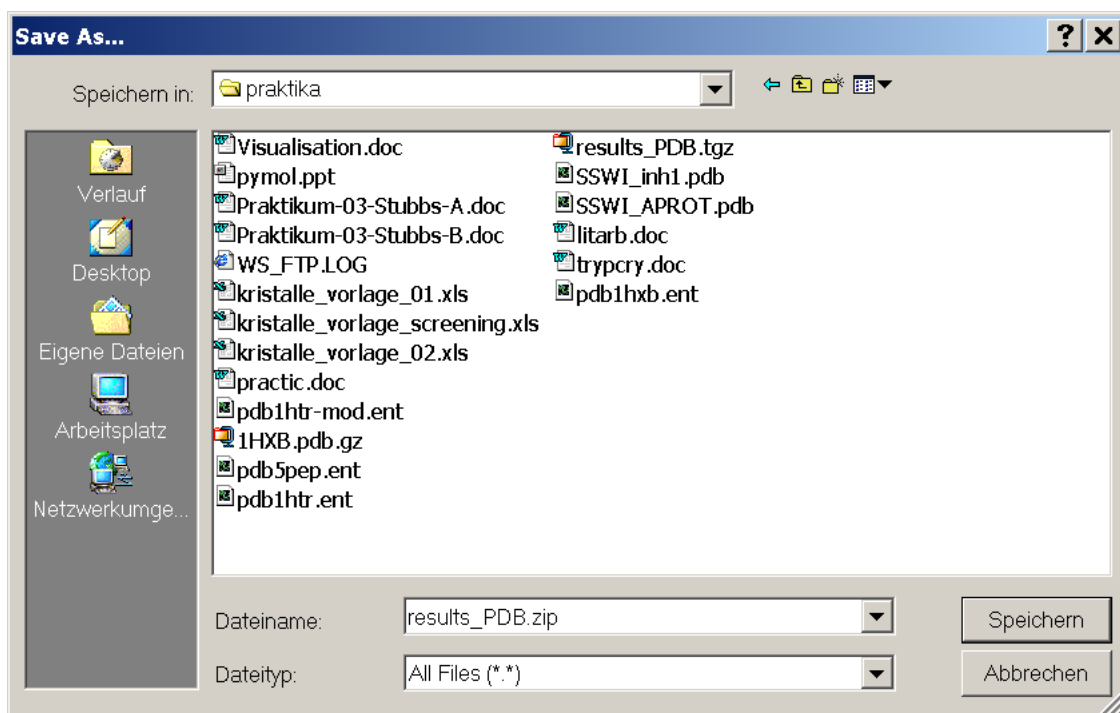
Document: Done

Now you can choose the structures you are interested in by clicking in the boxes; the next step is to go to the **Pull down to select option:**, and select **Download Structures or Sequences** followed by **Go**



Select the **ZIP file**, this is compressed and so downloads quicker....

The following window will appear – don't forget to note where you've saved your results(!) [c:\Praktikum\yourname].



Now you can exit Netscape and find the directory you saved the file in – you have to start the Windows Explorer (right mouse button on **Start**, then left mouse button on **Explorer**).

Double click on the file you downloaded; this will open WinZip, then you can **Extract** the pdb entries that you selected:

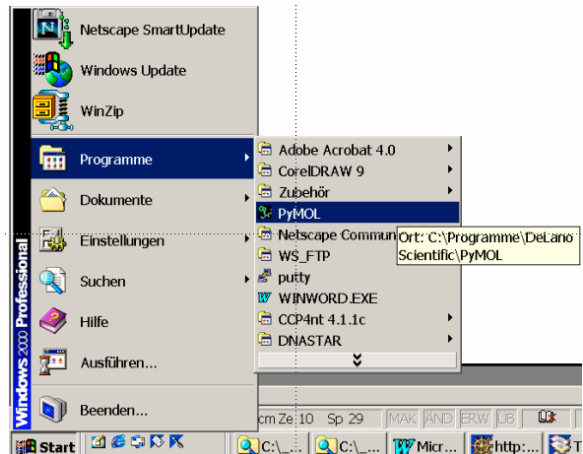


Now you need to start PyMol:

In PyMol, you have to say where your PDB files are. You do this using the command **cd** (= change directory):

**cd c:/praktika/yourname**

In the following, most of the necessary commands and concepts will be illustrated using a worked example.



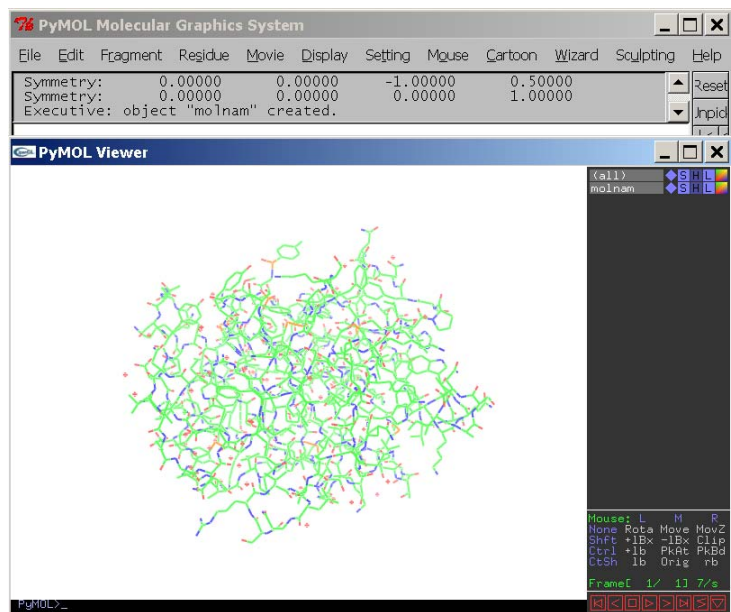
You have to load your PDB file into the program; to make it simpler later, we give this a **molecule name**:

**load pdbXXXX.ent, molnam**

There's your molecule! You can **rotate** it by holding down the **left** mouse button, **zoom** in using the **right** button, and **move** it around using the **middle** button.

With all those atoms, it's difficult to see; type in

**show cartoon, molnam**



Type also

**set cartoon\_smooth\_loops=0**

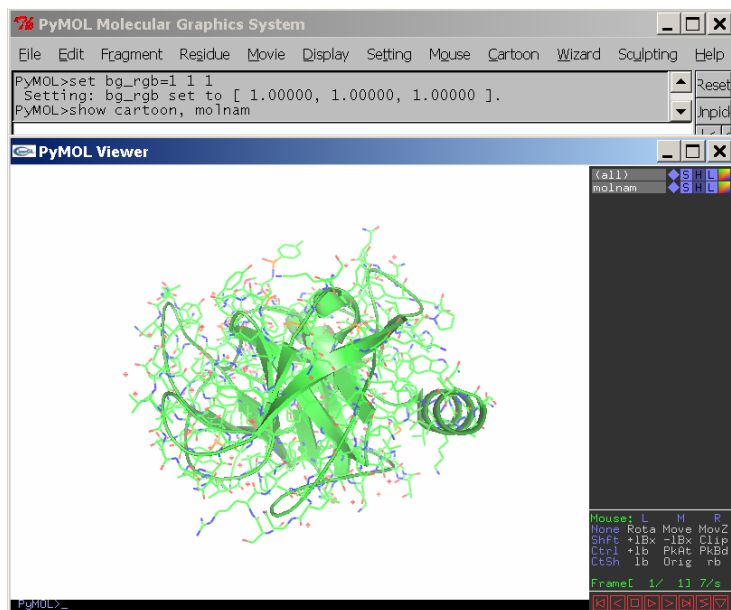
What happens?

It's still too complicated – try the following:

**hide lines, molnam**

**hide nonbonded, molnam**

Now all our atoms are gone (nonbonded are the solvent molecules); we're left with a simplified representation of the secondary structure ( $\alpha$ -helices and  $\beta$ -sheets).



Now we want to look at some specific atoms in the structure – in this case, the active site. This must be determined from the paper – ask the demonstrator for help. In this example, the active site serine is residue 195; in PyMol, atoms are denoted as

**/model/segment/chain/residue/atom**

Thus, to select the active centre, we type

**select active, /molnam///195**

We could also select more atoms:

**select active, /molnam///195 or /molnam///57 or /molnam///189**

Now we want to show them:

**show sticks, active**

Zoom in on the active site:

**zoom active**

You can change the colour of the active site residues by clicking on the little rainbow next to the (active).

What's in the neighbourhood of the active site? Type

**select close, active around 5.**

and then

**show sticks, close**

and display them in a different colour.

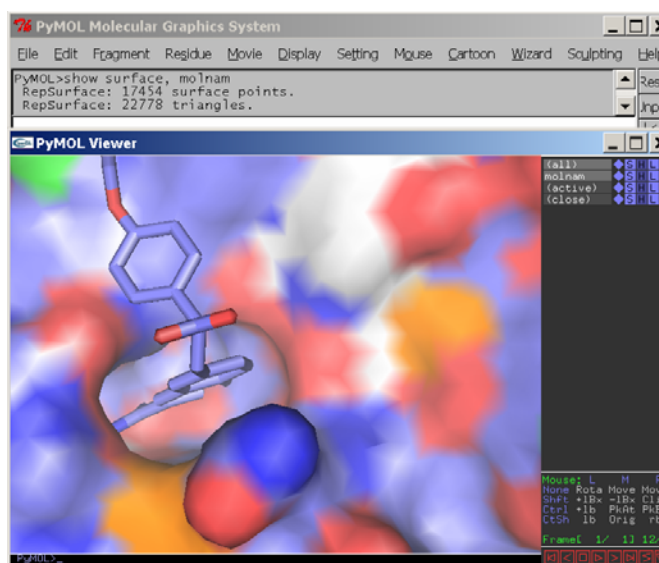
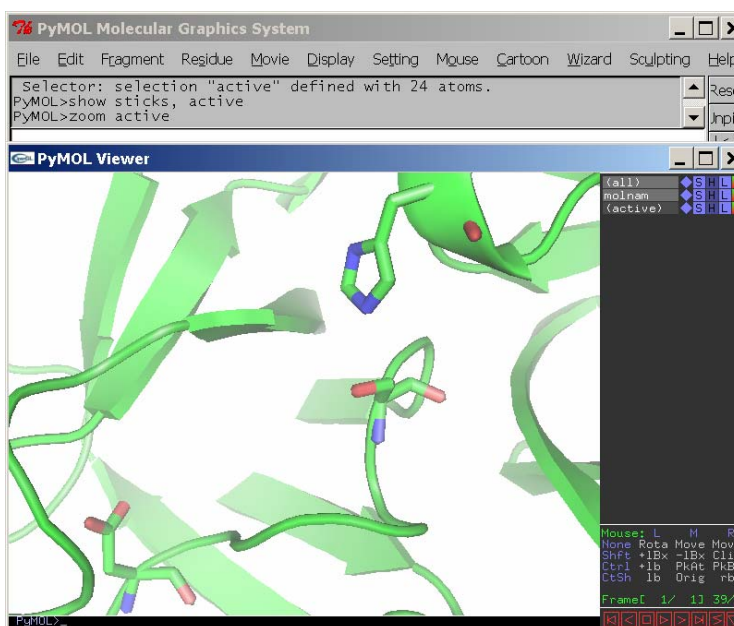
We can also show the protein surface:

**show surface, molnam**

In this example, we see an inhibitor in the active site. Type

**hide sticks, close**

**show lines, close**



and then click on the **Wizard**; a pop-up menu appears, and you can click on **Label**.

To pick an atom, press both the <Strg> key and the middle mouse button at the same time. This will give you the full ID of that atom: in this example below, it is /molnam///999.

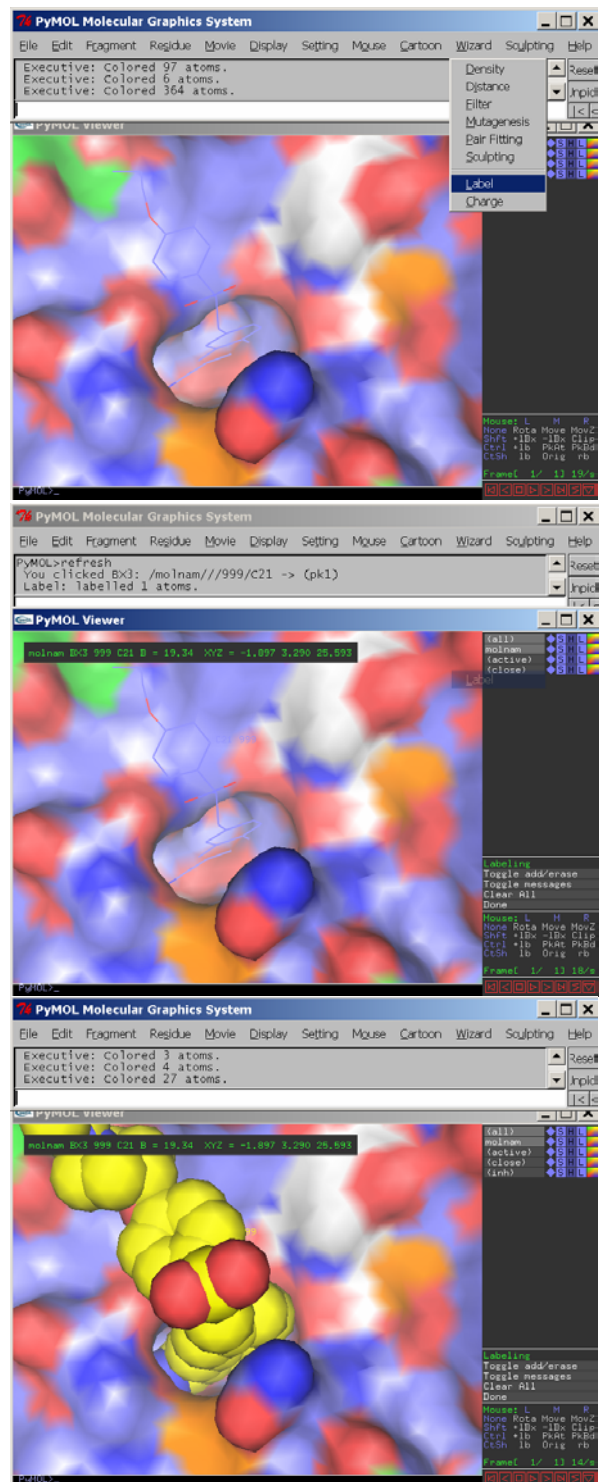
You can now select this to form a new object:

```
sele inh, /molnam///999  
show spheres, inh
```

and re-colour it.

To display the interactions, we can make the surface transparent:

```
set transparency = 0.3
```



What if we want to compare two structures? Let's load in a second one:

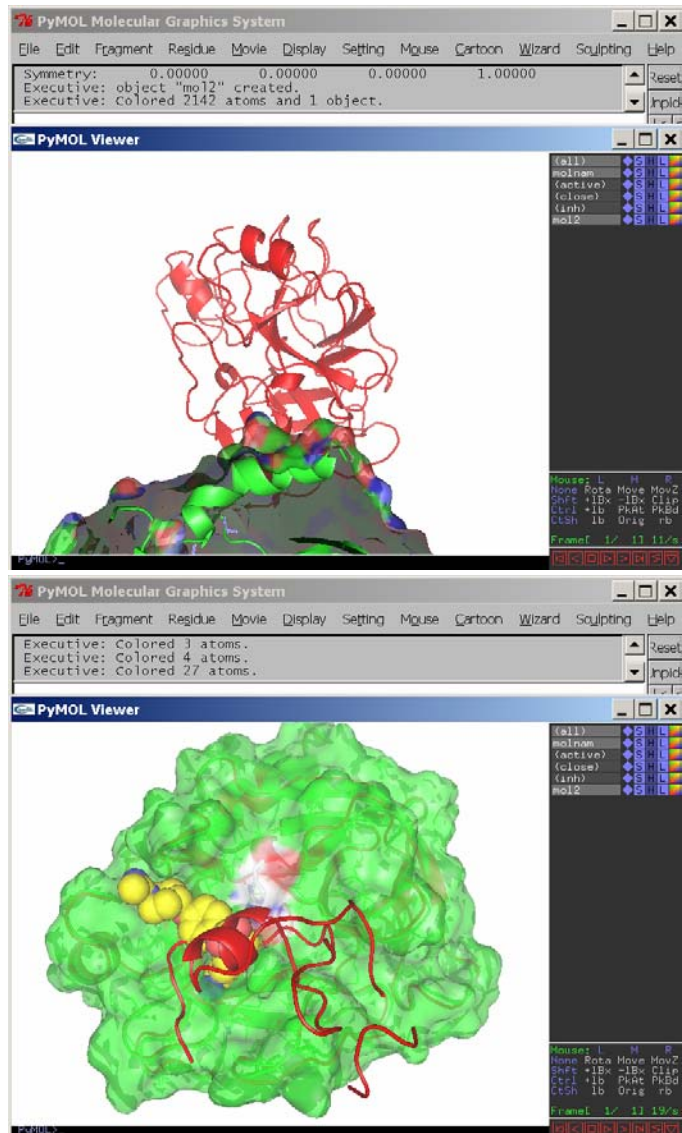
### load pdbYYY.ent, mol2

What's happened to our molecule? It has moved, as it comes from a different crystal. Colour **mol2** differently to distinguish it from **molnam**. Hide the lines and nonbonded atoms, and make a cartoon from **mol2**.

Now let's superimpose the two molecules on one another:

### align molnam, mol2

and rotate the molecules to look towards the active site cleft.



It is possible to run these commands from a **file** or **script**; you will be able to find more of these at

<http://www.biochemtech.uni-halle.de/xray/Praktika>

Should time permit, there will also be a demonstration showing how we can work with three dimensional structure information.....