

Protein Bioinformatics

260.655

Tuesday 13 April 2010

Sean Prigge and Gundula Bosch

Goals:

[1] Use the PDB website and Jmol

[2] Use the NCBI website and Cn3D

[3] Use the ExPASy website and Swiss-Pdb DeepView

Approx. Time

50 minutes

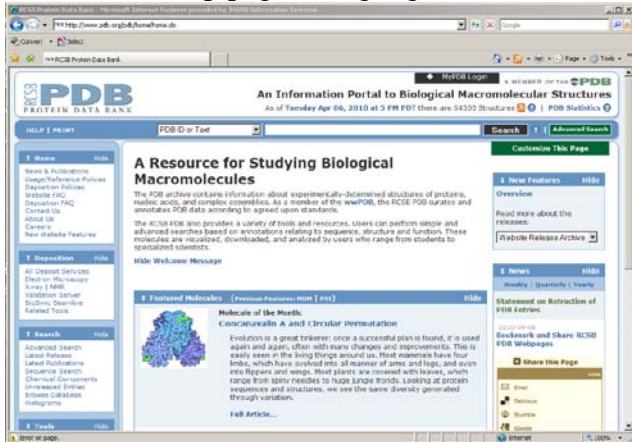
20 minutes

10 minutes

[1] Use PDB website and Jmol.

Go to the PDB top page (use google to find the site)

<http://www.pdb.org/pdb/home/home.do>



[1A] Find human myoglobin structure 2MM1

Type '2MM1' in the search window

Observe the information in the 5 boxes under the 'summary' tab

Primary Citation

Molecular Description

Source

Ligand Chemical Component

Derived Data

Now click on the Jmol viewer on the right

Rotate the structure with your left mouse button

Q How many helices does this structure have?

9

Find the amino acid which is bound to the heme iron

To do this, we need to see all of the atoms:

Right click [Macs: Ctrl-click] to open the Jmol menu

Select 'Style' then 'Scheme' then 'Ball and Stick'

Zoom in on the structure using the 'Zoom' tool on the menu, or your mouse wheel (if it has one)

[Macs: use the "alt" key + mousepad]

You can point your cursor to any atom and an atom label box will appear after a second.

Q. Which amino acid makes a bond with the heme iron?

H93

How can you view this amino acid more clearly?

Center on it: Menu 'Set picking' then 'Center'. Then click on an atom.

Now Zoom in using the zoom tool.

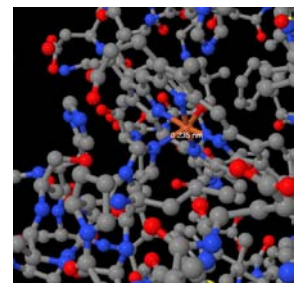
Q. What is the length of the bond between the iron and H93?

'Measurements' then 'Click for distance measurement'

Carefully click once on each atom of the bond

.23nm

Save a jpg of this view: 'Save' then 'JPG Image'



[1B] Malaria ferredoxin

Go back to the PDB front page

Type 'ferredoxin' in the search window

Q. How many hits are there?

339
1IUE

Go back and type 'malaria ferredoxin'

Click on the PDB ID code '1IUE'

Q. How many amino acids are reported in this structure?

98

Q. What ligands are bound to this protein?

FeS,NA

Now explore the structure in the Jmol viewer

Q. How many beta-strands and beta-sheets does this structure have?

8,2

Show the hydrogen bonding pattern in this structure:

'Style', then 'Hydrogen Bonds', then 'Calculate'

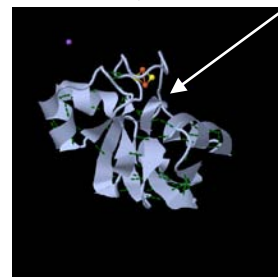
Make these hydrogen bonds more visible:

'Style', then 'Hydrogen Bonds', then '0.15 Å'

'Color', then 'Hydrogen Bonds', then 'Green'

Examine both of the beta-sheets and look at the hydrogen bonding patterns

Note that one of the strands does not make hydrogen bonds with the others



Check the Phi, Psi angles of an amino acid in this 'strand'

We need to see the main chain (backbone) atoms

'Select', 'Protein', 'Backbone'

(note: now only the 393 backbone atoms are listed next to the word 'Select')

'Style', 'Scheme', 'Ball and Stick'

Now center and zoom on an atom from this strand

'Set picking' then 'Center'. Then click on an atom.

'Zoom'

Look specifically at C77

'Measurements', 'Click for torsion (dihedral) measurement'.

Then click on the 4 atoms that define phi

Leu76 C, Cys77 N, Cys 77 CA, Cys 77 C

(phi for C77 is -69)

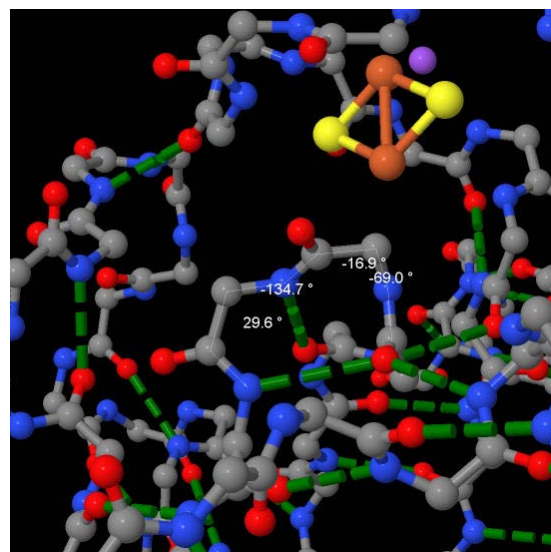
Now determine psi:

Cys77 N, Cys 77 CA, Cys 77 C, Thr 78 N

(psi for C77 is -17)

Measure phi and psi for T78 also:

(phi= -135, psi= 30)



Q. C77 and T78 are part of what kind of secondary structure?

Beta Turn

(Note the hydrogen bond between L76 and C79)

Alternative method:

Display only the amino acids from L76 to C79:

'Select', 'None'

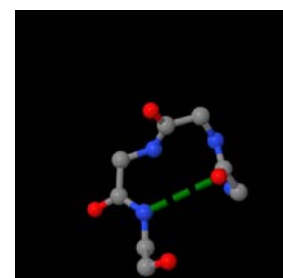
Now 'Select' should be followed by '(0)'

'Set picking', 'Select group'

Now click on any atom from the four amino acids from L76 to C79

Now 'Select' should be followed by '(27)'

Go into the 'Select' menu and check the 'Display Selected Only' box



Now examine the iron atoms that are bound to this protein

First, you have to change the display to show all atoms

'Select', 'All'

'Style', 'Scheme', 'Ball and Stick'

Play with the select tool to make the iron binding site easier to visualize.

Make sure you uncheck the 'Select' menu 'Display Selected Only' box

Now you can 'Select', 'Protein', 'Backbone'

Followed by 'Style', 'Scheme', 'Cartoon'

You can also remove atoms, such as the waters:

'Select', 'Hetero', 'All water'

'Style', 'Atoms', 'Off'

You will probably want to recenter and zoom on one of the iron atoms as well

Q. What are the 4 amino acids which bind to the iron atoms?

Q. What is the distance between the two iron atoms?

C39,44,47,77
.27nm

[1C] *E. coli* GlpG

A member of class is interested in an *E. coli* rhomboid protease called GlpG
Go to the top page of the PDB and search for GlpG

Q. How many hits are there?

7

These proteins are not all from *E. coli*. Use the following strategy to find *E. coli* GlpG:

On the PDB top page, click on the green 'Advanced Search' button

In the drop down menu choose query type 'Text Search'

Type 'GlpG' in the text box, and press the 'Result Count' button

This should indicate that there are 7 GlpG structures

Now click on the words 'Add Search Criteria' to open a second search box

In the drop down menu select query type 'Source organism browser' (in the **Biology** submenu)

Now type in '*Escherichia coli*' and click on the 'Find in tree' button

Click on the parent group of *Escherichia coli*

Now click on the 'Result Count' button for this search

Now click on the 'Submit Query' button to get only structures of *E. coli* GlpG

There are still several structures. Let's look at 2IC8.

Click on 2IC8 and look at the information provided.

Read the abstract and note the molecular description and ligands

Now explore the structure in Jmol

What are the ligands bound to this protein and why are they there?

The abstract describes this as a membrane-bound protein. Where does the membrane bind?

Let's identify the hydrophobic residues

You can try your own approach or try the one below:

'Select', 'Protein', 'All'

'Style', 'Scheme', 'Ball and Stick'

Now color the protein

'Select', 'Protein', 'Polar Residues'

'Color Atoms', 'Violet'

'Select', 'Protein', 'Nonpolar Residues'

'Color Atoms', 'White'

Now let's expand the size of the atoms to their full van der Waals radius:

'Select', 'Protein', 'All'

'Style', 'Atoms', '100% van der Waals'

There is a white band around the center of the protein

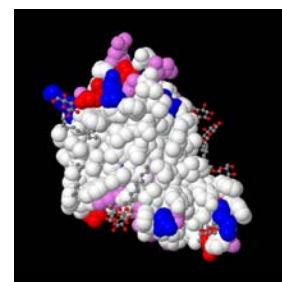
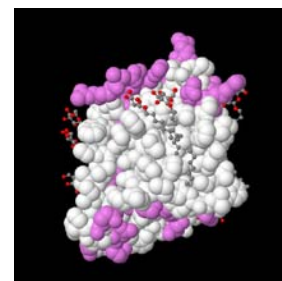
Notice where the hydrophobic parts of the detergent molecules bind

You can additionally highlight the charged residues

Use the selection tools to color:

basic residues (+) → blue

acidic residues (-) → red



Now change the van der Waals radius to a small value

'Select', 'Protein', 'All'

'Style', 'Atoms', '15% van der Waals'

Note the charged (red or blue) amino acids buried in the core of the protein

In several cases, a basic (blue) amino acid is found in the protein core

Q Which amino acid is this?

H

Q Think about the side chain pKa of this amino acid. Is it charged?

No

*Advanced

Find the catalytic serine of this serine protease.

[Hint: In serine proteases, the serine is hydrogen bonding distance from a catalytic histidine]

[1D] Glycoprotein D from Herpes Simplex Virus-2 (HSV-2)

Find this protein in the PDB:

```
mgrltsgvgt aallvvavgl rvvcakyala dpslkmadpn rfrgknlpvl dqltdppgvk  
rvyhiqpsle dpfqppsipi tvyyavlera crsvllhaps eapqivrgas dearkhtynl  
tiawyrmgdn caipitvmey tecpynkslg vcpirtqprw syysdfsavs ednlgflmha  
pafetagtly rlvkindwte itqfilehra rasckyalpl rippaacfts kayqqgvtd  
sigmlprfip enqrvalys lkiagwhgpk ppytstllpp elsdttnatq pelvpdpdp  
salledpagt vssqipnwh ipsiqdvaph hapaapsnpg liigalagst lavlviggia  
fwvrrraqma pkrlrlphir dddappshqp lfy
```

Use a FASTA sequence search

Click '**Advanced Search**' on the PDB top page

Choose '**Sequence (Blast/Fasta)**'

Paste in the above sequence and click on '**Result Count**'

Now click on '**Submit Query**' and look at the list of results

Note: the results are now sorted by BLAST expect values (E Values)

Also note the sequence alignment window for each structure

Scroll down the list of results and note the E Values, the sequence alignment and the titles of the structures

Q. How many of the structures in the list are relevant?

4

Now return to the '**Advanced Search**' page and redo the sequence search

This time, change the '**E Cut Off**' from 10.0 to .001

The irrelevant results are now filtered out

Click on the '**Submit Query**' button and look at these four structures

Q. Should we use the highest E value structure (1JMA) to explore glycoprotein D?

Let's use 2C36 since it is more complete (285aa) and has a low resolution (2.11Å)

Click on 2C36, look over the information about this structure and then open Jmol

As you rotate this structure you will see that the amino acid chain is broken

Q. What residues are missing from this structure?

257-267

Examine the ligands bound to this protein

There are two green spheres as well as a violet sphere.

Q. What atoms do these spheres represent?

Cl, Zn

Find the amino acids which bind the zinc atom

Approach:

Center and zoom on the zinc atom

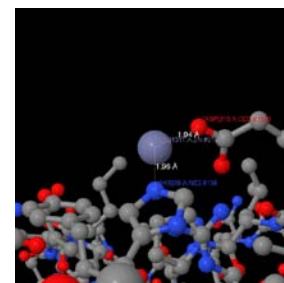
Select all protein atoms

Display all protein atoms in **Ball and Stick** style scheme

Measure distances from amino acid atoms to the zinc atom

Q. What amino acids make bonds with the zinc?

H39, D215



If you have time, find the amino acids responsible for binding the chlorine ions

Q. What is the distance between the chlorine ions and the amino acids?

Q. Do these amino acids make covalent bonds with the chlorine ions?

No

Now examine the trisaccharide ligand

Q. How is this ligand bound to the protein?

N-linked glycosylation to N94

In class we talked about the covalent disulfide bonds that cysteine amino acids can make

This protein has several disulfide bonds

Q. How many disulfide bonds are there?

Q. Which cysteine residues form disulfides?

Q. Are there cysteine residues not in disulfide bonds?

Approach:

Select protein backbone atoms

Display backbone atoms as style scheme wireframe

Select protein side chains

Remove side chain atoms by displaying as style atoms off

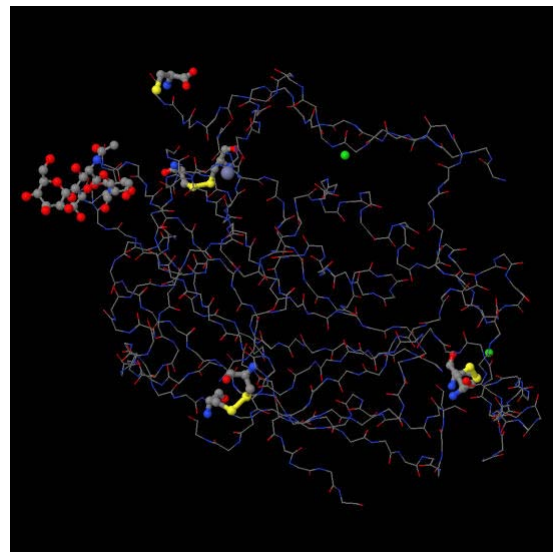
Remove side chain bonds by displaying as style bonds off

Now display the cysteine atoms/bonds in Ball and Stick mode

'Select', 'Protein', 'By Residue Name', 'Cys'

Display the cysteines with Ball and Stick style scheme

For easier viewing, you can increase the bond and atom sizes



[1E] Tuberculosis two component proteins

Here are two sequences of interest from one of your classmates.

Sensor X3:

```
TTGCAATGCATTGTACACGCTGATGCCGCTGGGCGCCGCGGTGGTGGACACCCATCGCGAC
GTTGTCTACCTCAACGAACGGGCCAAA
GAGCTAGGTCTGGTGCAGCAGCCGAGCTCGATGATCAGGCCTGGCGGGCCGCCGCGCAG
GCGCTGGGTGGTGAAGACGTCGAGTTCGACCTGTGCGCGCGCAAGCGGTGCGCCACGGGT
CGATCCGGGCTATCAGTGCATGGGCATGCCGGTTGCTGAGCGAGGAAGACCGCCGGTTC
GCCGTGGTGTTCGTGCACGACCACTCGGATTATGCGCGGATGGAGGCGGCTAGGCGTGAC
TTCGTGGCCAACGTCAGTCACGAGCTCAAGACGCCCGTCGGTGCCATGGCTCTACTCGCC
GAGGCGCTGCTGGCGTCGGCCGACGACTCCGAAACCGTTCGGCGGTTGCGCGAGAAGGTG
CTCATTGAGGCCAACC GGCTCGGTGACATGGTCGCCGAGTTGATCGAGCTATCCCGGCTA
CAGGGCGCCGAGCGGCTACCCAATATGACCGACGTCGACGTCGATACGATTGTGTGCGAA
GCGATTTACGCCATAAGGTGGCGGCCGACAACGCCGACATCGAAGTCCGCACCGACGCG
CCCAGCAATCTGCGGGTGTGCGGCGACCAACTCTGCTGGTTACCGCACTGGCAAACCTG
GTTTCCAATGCGATTGCCTATTGCGCGCGGGTCTGCTGGTGTGATCAGCCGTGCGCGT
CGCGGTGCCAACATCGAGATCGCCGTACCGACCGGGGCATCGGCATCGCGCCGGAAGAC
CAGGAGCGGGTCTTCAACGGTCTTCCGGGGGACAAGGCGCGCTCGCGTGCCACCGGA
GGCAGCGGACTCGGGTTGGCCATCGTCAAACACGTGCGGGCTAATCACGACGGCACCATC
CGCGTGTGGAGCAAACCGGGAACCGGTCAACGTTACCTTGGCTCTTCCGGCGTTGATC
GAGGCCTATCACGACGACGAGCGACCCGAGCAGGCGCGAGAGCCGAACTGCGGTCAAAC
AGGTCACAACGAGAGGAAGAGCTGAGCCGATGA
```

Regulator X3:

```
GTGGCGCTGATGACCAAGTGTGTTGATTGTGGAGGACGAGGAGTCGCTGGCCGATCCGCTG
GCGTTTCTGCTGCGCAAGGAGGGCTTTGAGGCCACGGTGGTGACCGATGGTCCGGCAGCT
CTCGCCGAGTTCGACCGGGCCGGCGCCGACATCGTCCTGCTCGATCTGATGCTGCCTGGG
ATGTGCGGTACCGATGTATGCAAGCAGTTGCGCGCTCGGTCCAGCGTTCCGGTGATCATG
GTGACCGCCCGGATAGCGAGATCGACAAGGTGGTGGCCTGGAGCTGGGCGCTGACGAC
TACGTGACCAAGCCCTATTCCGGCACGCGAGTTGATCGCACGCATCCGCGCGGTGCTGCGC
CGTGGCGGGCAGCAGGACTCGGAGATGAGCGATGGCGTGCTGGAGTCCGGGCCGGTTCGC
ATGGATGTGGAGCGCCATGTCTCTCGTGAACGGTGACACCATCACGCTGCCGCTCAAG
GAGTTCGACCTGCTGGAATACCTGATGCGCAACAGCGGGCGGGTGTGACTCGCGGACAA
CTGATCGACCGGTCTGGGGTGGGACTACGTGGGCGACACCAAGACGCTCGACGTCCAT
GTCAAGCGGCTGCGCTCCAAGATCGAAGCCGACCCGGCTAACCCGGTTCCTTGGTGACG
GTGCGCGGGCTGGGCTACAACTCGAGGGCTAG
```


Let's start with the first sequence (Sensor X3).

Go to the NCBI top page (use google to find the site)

<http://www.ncbi.nlm.nih.gov/>

From the NCBI top page, click on the 'Popular Resources' link on the right and select 'BLAST'

Click on 'nucleotide blast'

Paste the DNA sequence above into the sequence window

Use the pull down box in 'Choose Search Set' and select 'Nucleotide collection (nr/nt)'

We can limit our search to *Mycobacterium tuberculosis* by typing this in the 'organism' box

Then click on the 'Blast' button (bottom left)

The top hits are all identical sequences from different strains of *M. tuberculosis*

Scroll down to the first hit and click on '[two component system sensor histidine kinase senX3](#)'

Get the PROTEIN sequence:

```
MTVFSALLLAGVLSALALAVGGAVGMRLTSRVVEQRQRVATEWS
GITVSQMLQCIVTLMPLGAAVVDTHRDVVYLNERAKELGLVRDRQLDDQAWRAARQAL
GGEDVEFDLSPRKRSATGRSGLSVHGHARLLSEEDRRFAVVFVHDQSDYARMEAARRD
FVANVSHELKTPVGAMALLAEALLASADDSETVRRFAEKVLI EANRLGDMVAELIELS
RLQGAERLPNMTDVDVDTIVSEAI SRHKVAADNADIEVRTDAPSNLRVLGDQTLTVTA
LANLVSNAIAYS PRGSLVSI SRRRRGANIEIAVTDRGIGIAPEDQERVFERFFRGDKA
RSRATGGSGGLGLAIVKHVAANH DGTIRVWSKPGTGSTFTLALPALIEAYHDDERPEQA
REPELRSNRSQREEELSR
```

Return to the PDB and use the Blast/Fasta sequence search tool to find similar proteins

Q. How many hits do you get with an E Value of 10?

46

Now try an E Value of (0.00001) (E-5)

Q. Look over the 12 results of this search. Are they all relevant structures?

Yes

Q. The top two hits come from what organism?

T. maritima

Explore structure 2C2A in Jmol

Q. How many proteins are shown?

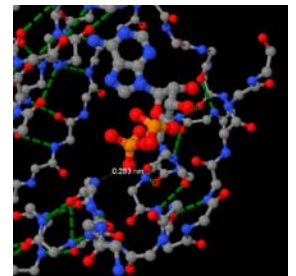
2

Q. How would you describe this multimer?

homodimer

Zoom in on the ADP ligand bound to one of the monomers

Display the main chain in ball and stick style scheme



Observe the interactions between the first helix in the structure and the ADP

Q. Why does the N-terminus of the helix form part of the binding site?

The helix dipole helps to bind the negatively charged phosphates

Now look at the dimer interface

Q. What secondary structure elements are involved in dimerization?

2 helices

Think of a way to better visualize the dimer interface

Approach:

Select protein backbone

Display as cartoon style scheme

Select protein side chains

Display as ball and stick style scheme

Now calculate a van der Waals surface for monomer 1:

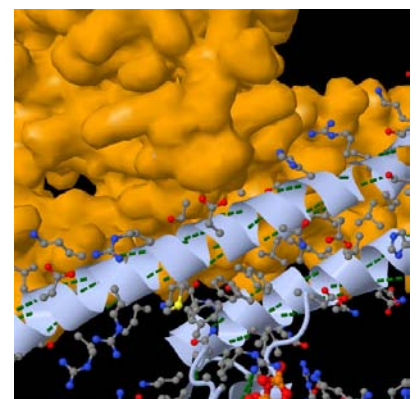
In the main menu select the box under 'All 2 models', '1.1:1'

Select all protein atoms

Select 'Surfaces', 'van der Waals Surface'

Select 'Surfaces', 'Make Opaque'

Now display both monomers:



In the main menu select the box under ‘**Model 1/2**’, ‘**All**’

Explore structure 3DGE in Jmol

Q. How many proteins are shown?

Q. How would you describe this multimer?

Q. What protein is bound to the sensor kinase?

4

Homodimer of heterodimers
the regulator

Now examine the interface between the sensor and regulator

Approach:

This structure was entered as one ‘model’,

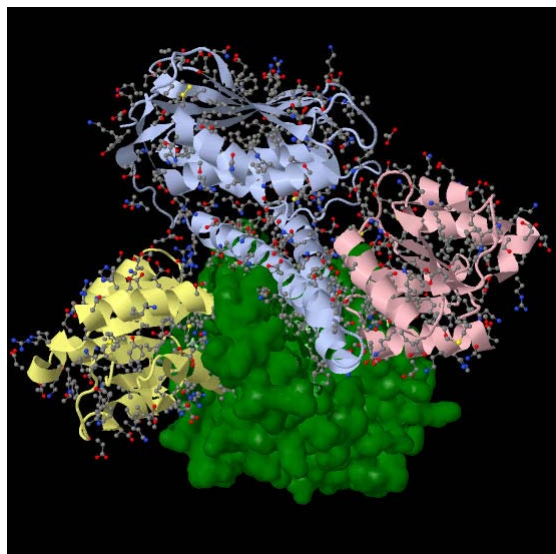
If we want to make a surface around one protein molecule, we can use the ‘**Set picking**’ tool

‘**Select**’, ‘**None**’

‘**Set picking**’, ‘**Select molecule**’

Now click on one atom of a molecule and calculate a surface for this molecule

You can repeat this procedure to calculate surfaces for other monomers



[2] Use NCBI website and Cn3D.

[2A] Find human myoglobin structure 2MM1

APPROACH 1:

Go to the NCBI top page (use google to find the site)

<http://www.ncbi.nlm.nih.gov/>



Type 2MM1 in the search window

Q. How many Proteins were found?

1

Q How many Structures were found?

1

Q How many publications reference structure 2MM1?

7

Click on 'Structure' and then click on the thumbnail picture of the structure

Click on this if you are lost →



There are links here to several other databases (including the PDB)

Now click on the 'Structure View in Cn3D' button

Rotate and explore this structure with your mouse

Measure the distance between the iron atom and the distal histidine

One method – go to the upper menu and 'style', 'rendering shortcuts' 'ball and stick'

Find the distal histidine ('View' 'Zoom In' may be helpful)

(H93)

Double click on any atom of the distal histidine

-Note that it turns yellow in the molecule and in the sequence window

Experiment with selecting residues in the sequence window and the structure

APPROACH 2:

What if we don't know the PDB code 2MM1?

Type 'myoglobin' in the search window

How many structure hits are there?

312

Go back and type 'human myoglobin'

Can you find 2MM1?

How many structure hits are there?

9

How many have been determined by X-ray crystallography?

9

APPROACH 3:

What if we only have the protein sequence?

GLSDGEWQLVLNVWGKVEADIPGHGQEVLRIRLFKGHPEPTLEKFDRFKHLKSEDEMKASEDLKKH
GATVLTALGGILKKKGHAEIKPLAQSHATKHKIPVKYLEFISEAIIQVLQSKHPGDFGADAQGA
MNKALELFRKDMASNYKELGFQG

From the NCBI top page, click on the 'Popular Resources' link on the right and select 'BLAST'

Click on 'protein blast'

Paste the protein sequence above into the sequence window and click on the 'Blast' button (bottom left)

Q. Can you find 2MM1 in the list of hits?

Q Which hits are from structures?

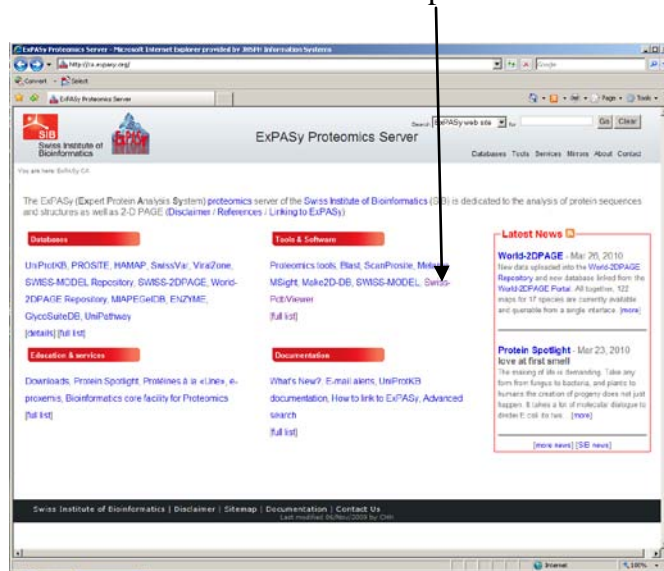
(Note that there is a red box with a white 'S' to link back to Cn3D and the PDB)

Redo the BLAST search, but this time type 'Homo sapiens' in the 'Organism' box.

[3] Use the ExPASy site and Swiss-Pdb 'DeepView'

Install the Swiss-Pdb Viewer 'Deep View' from the ExPASy web site:

<http://ca.expasy.org/>



To use an external viewing program like Deep View (or PyMol) you should download a file from the PDB

Go back to the top page of the PDB and type in the PDB code '2mm1'

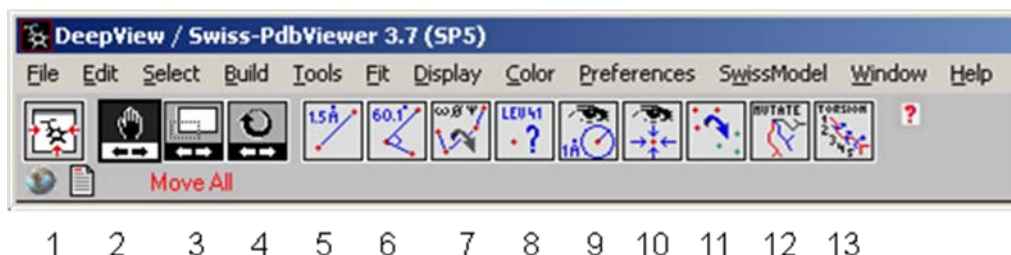
You can type in one of the other PDB codes we used in part 1 if you want

Click on the 'Download Files' icon (On the upper right)

Choose 'PDB file (text)' and save the file to your computer (2MM1.pdb)

Open the DeepView on your computer

You should see this tool bar:



Use the 'file' tab to open the PDB file that you saved to your computer

Select 'Open PDB File'

Experiment with the 13 buttons on the tool bar:

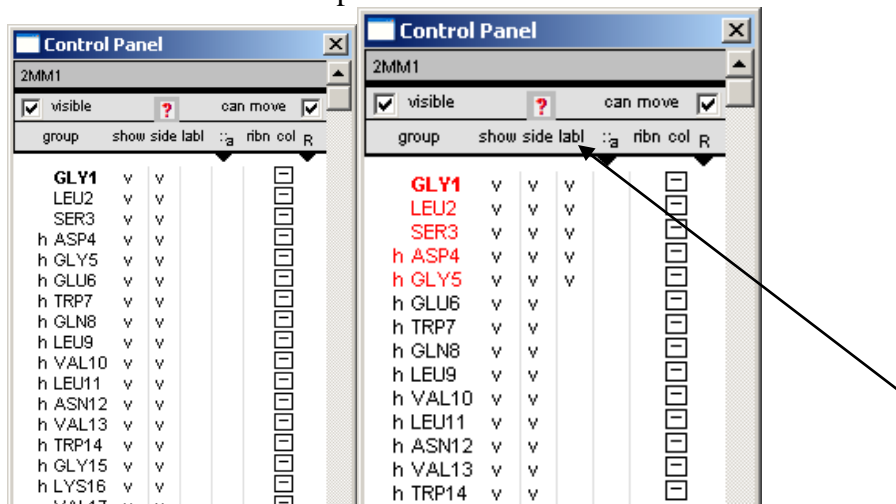
- 1 center
- 2 move (translate)
- 3 zoom in/out
- 4 rotate
- 5 click, select two atoms, determine distance in angstroms
- 6 measure bond angles (pick center atom, then two more atoms)
- 7 measure dihedral angles (w, f, j)(omega, phi, psi) from a selected atom.
- 8 identify an atom (and the group to which it belongs).
Type: CA, CB, O
Group: LYS116, etc. x,y,z atom coordinates
- 9 display groups within a particular distance (e.g. 10A) from a selected atom. Note the selection on the control and graphics panels.
- 10 center display on a selected atom
- 11-13 advanced tools

Use the pull-down menu options such as the following:

Select → Group property → non-polar (and visualize this)

Select → Secondary structure → helices

Go to window → control panel.



Shift/click to select the first five amino acid residues of myoglobin.
They should appear red.

Click “labl” (i.e. label)(see arrow, above right).

Those five residues now have a “v”.

Inspect the display panel; those five residues are labeled.

