Protein Structure and Visualisation

Introduction to PDB and PyMOL
Feedback Persons

http://www.bio-evaluering.dk/
<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.00-8.45</td>
<td>Introduction to PDB &amp; PyMOL</td>
</tr>
<tr>
<td>8.45-9.00</td>
<td>Break</td>
</tr>
<tr>
<td>9.00-11.00</td>
<td>PyMOL tutorial</td>
</tr>
<tr>
<td>11.00-11.15</td>
<td>Break</td>
</tr>
<tr>
<td>11.15-11.30</td>
<td>Summary</td>
</tr>
<tr>
<td>11.30-12.00</td>
<td>Quiz</td>
</tr>
</tbody>
</table>
Protein Data Bank

- [http://www.rcsb.org/](http://www.rcsb.org/)
- Contents
- File structure
  - Types of structures
- Structure reports & summaries
- Quality check
- Searching
- Molecule of the Month
PDB Growth 1971-2013
PDB File Header

HEADER IMMUNE SYSTEM 16-AUG-04 1X7Q
TITLE CRYSTAL STRUCTURE OF HLA-A*1101 WITH SARS NUCLEOCAPSID
TITLE 2 PEPTIDE
COMPND MOL_ID: 1;
COMPND 2 MOLECULE: HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, A-11

REMARK 3 DATA USED IN REFINEMENT.
REMARK 3 RESOLUTION RANGE HIGH (ANGSTROMS) : 1.45
REMARK 3 RESOLUTION RANGE LOW (ANGSTROMS) : 30.00

SEQRES 1 C 9 LYS THR PHE PRO PRO THR GLU PRO LYS
HET SO4 B4001 5

HETNAM SO4 SULFATE ION

HELIX 1 1 ALA A 49 GLU A 53 5

CRYS1 58.465 80.623 56.570 90.00 116.79 90.00 P 1 21 1 2
### PDB File Atom Coordinates

<table>
<thead>
<tr>
<th>ATOM</th>
<th>1</th>
<th>N</th>
<th>THR A 1</th>
<th>25.200 26.068 37.670</th>
<th>1.00</th>
<th>25.43</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATOM</td>
<td>2</td>
<td>CA</td>
<td>THR A 1</td>
<td>26.443 26.547 37.135</td>
<td>1.00</td>
<td>16.70</td>
<td>C</td>
</tr>
<tr>
<td>ATOM</td>
<td>3</td>
<td>C</td>
<td>THR A 1</td>
<td>27.568 25.589 37.431</td>
<td>1.00</td>
<td>13.12</td>
<td>C</td>
</tr>
<tr>
<td>ATOM</td>
<td>4</td>
<td>O</td>
<td>THR A 1</td>
<td>27.577 25.073 38.554</td>
<td>1.00</td>
<td>15.92</td>
<td>O</td>
</tr>
<tr>
<td>ATOM</td>
<td>5</td>
<td>CB</td>
<td>THR A 1</td>
<td>26.745 27.891 37.843</td>
<td>1.00</td>
<td>20.41</td>
<td>C</td>
</tr>
<tr>
<td>ATOM</td>
<td>6</td>
<td>OG1</td>
<td>THR A 1</td>
<td>25.564 28.674 37.550</td>
<td>1.00</td>
<td>26.40</td>
<td>O</td>
</tr>
<tr>
<td>ATOM</td>
<td>7</td>
<td>CG2</td>
<td>THR A 1</td>
<td>27.995 28.594 37.359</td>
<td>1.00</td>
<td>22.25</td>
<td>C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HETATM</th>
<th>3284</th>
<th>O3</th>
<th>GOL B5007</th>
<th>23.356 27.459 10.492</th>
<th>1.00</th>
<th>29.71</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>HETATM</td>
<td>3285</td>
<td>O</td>
<td>HOH A5007</td>
<td>19.520 54.389 40.841</td>
<td>1.00</td>
<td>10.97</td>
<td>O</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ATOM</th>
<th>1</th>
<th>N</th>
<th>GLY A 1</th>
<th>9.360 23.828 42.512</th>
<th>1.00</th>
<th>45.34</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANISOU</td>
<td>1</td>
<td>N</td>
<td>GLY A 1</td>
<td>6582 3670 6975 2086 1854 1681</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATOM</td>
<td>2</td>
<td>CA</td>
<td>GLY A 1</td>
<td>8.104 24.285 41.984</td>
<td>1.00</td>
<td>33.31</td>
<td>C</td>
</tr>
<tr>
<td>ANISOU</td>
<td>2</td>
<td>CA</td>
<td>GLY A 1</td>
<td>5450 1898 5309 -137 1446 394</td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATOM</td>
<td>3</td>
<td>C</td>
<td>GLY A 1</td>
<td>7.740 25.713 42.380</td>
<td>1.00</td>
<td>28.54</td>
<td>C</td>
</tr>
<tr>
<td>ANISOU</td>
<td>3</td>
<td>C</td>
<td>GLY A 1</td>
<td>4079 1849 4915 -6 728 561</td>
<td>C</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### PDB File Fields

<table>
<thead>
<tr>
<th>COLUMNS</th>
<th>DATA TYPE</th>
<th>FIELD</th>
<th>DEFINITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – 6</td>
<td>Record name</td>
<td>&quot;ATOM&quot;</td>
<td></td>
</tr>
<tr>
<td>7 – 11</td>
<td>Integer</td>
<td>serial</td>
<td>Atom serial number.</td>
</tr>
<tr>
<td>13 – 16</td>
<td>Atom</td>
<td>name</td>
<td>Atom name.</td>
</tr>
<tr>
<td>17</td>
<td>Character</td>
<td>altLoc</td>
<td>Alternate location indicator.</td>
</tr>
<tr>
<td>18 – 20</td>
<td>Residue name</td>
<td>resName</td>
<td>Residue name.</td>
</tr>
<tr>
<td>22</td>
<td>Character</td>
<td>chainid</td>
<td>Chain identifier.</td>
</tr>
<tr>
<td>23 – 26</td>
<td>Integer</td>
<td>resSeq</td>
<td>Residue sequence number.</td>
</tr>
<tr>
<td>27</td>
<td>AChar</td>
<td>iCode</td>
<td>Code for insertion of residues.</td>
</tr>
<tr>
<td>31 – 38</td>
<td>Real(8.3)</td>
<td>x</td>
<td>Orthogonal coordinates for X in Angstroms</td>
</tr>
<tr>
<td>39 – 46</td>
<td>Real(8.3)</td>
<td>y</td>
<td>Orthogonal coordinates for Y in Angstroms</td>
</tr>
<tr>
<td>47 – 54</td>
<td>Real(8.3)</td>
<td>z</td>
<td>Orthogonal coordinates for Z in Angstroms</td>
</tr>
<tr>
<td>55 – 60</td>
<td>Real(6.2)</td>
<td>occupancy</td>
<td>Occupancy.</td>
</tr>
<tr>
<td>61 – 66</td>
<td>Real(6.2)</td>
<td>tempFactor</td>
<td>Temperature factor.</td>
</tr>
<tr>
<td>77 – 78</td>
<td>LString(2)</td>
<td>element</td>
<td>Element symbol, right-justified.</td>
</tr>
<tr>
<td>79 – 80</td>
<td>LString(2)</td>
<td>charge</td>
<td>Charge on the atom.</td>
</tr>
</tbody>
</table>
Worldwide Structural Genomics

- “Fold space coverage”

- Complete genomes
  - Disease-causing organisms
  - Model organisms

- Membrane proteins

- Protein-ligand interactions

Hou et al., PNAS 2003, 100: 2386-2390
Protein Folds in PDB

No new folds!
Protein Structure Visualisation

PyMOL
What is PyMOL?

- Open-source molecular viewing program

http://www.pymol.org
Benefits

• It’s free!
  – For academia…
  – …not for industry.
  – Version 0.99rc6
  – Users can always compile the latest version.

• But you should contribute!

• Pay to get support, manual, latest version etc.
Potential Weaknesses

• Few!

• Not a fully integrated modelling environment.

• Not fully developed for experimental structure determination/fitting.

• Mostly for qualitative analyses.

• *No undo function*...
PyMOL

- Representations
  - Lines, sticks, ribbon, spheres, cartoon(s)
- Surfaces
  - Transparency, quality
- Ray-tracing (rendering)
  - Modes
• Every molecule (pdb file) is an object.
• Selections refer to objects
  – Make smaller or composite objects
• Changes in representation can affect objects or selections.
Links

• PDB (protein structure database)
  – www.pdb.org/

• PyMOL home:
  – http://pymol.sourceforge.net/

• PyMOL manual:
  – http://pymol.sourceforge.net/newman/user/toc.html

• PyMOL Wiki:
  – http://www.pymolwiki.org/index.php/Main_Page

• PyMOL settings (documented):
  – http://cluster.earlham.edu/detail/bazaar/software/pymol/modules/pymol/setting.py
Installation

• Windows

• Mac OS X
  – Special behaviour
  – Rename to get plugin menu!
    • MacPyMOL → MacPyMOLX11Hybrid

• Linux
PyMOL scripting
Scripts

• PyMOL is based on the programming language Python.

• PyMOL will read python commands.

• PyMOL also has a set of native commands.

• See PyMOLWiki and elsewhere for examples.
  – http://www.rubor.de/bioinf/tips_python.html
  – http://pldserver1.biochem.queensu.ca/~rlc/work/pymol/
Automating Tasks

• Scripts are useful for automating tasks
  – Repeated analyses
  – Reading in large numbers of structures
  – Making illustrations!
    • get_view
    • set_view
    • viewport

• Saves space!
  – Relative to PyMOL session files (.pse)
A Useful Example

Using PyMOL commands:

```python
list=[]
iterate (name ca),list.append((resi,resn))
print list

[('ASP', '1'), ('CYS', '2'), ('ALA', '3'), ('TRP', '4'), ('HIS', '5'), ('LEU', '6'), ('GLY', '7'), ('GLU', '8'), ('LEU', '9'), ('VAL', '10'), ('TRP', '11'), ('CYS', '12'), ('THR', '13')]
```

or using a Python script (in PyMOL):

```python
from pymol import cmd,stored
stored.list=[]
cmd.iterate("(name ca)","stored.list.append((resi,resn))")
print stored.list

[('1', 'ASP'), ('2', 'CYS'), ('3', 'ALA'), ('4', 'TRP'), ('5', 'HIS'), ('6', 'LEU'), ('7', 'GLY'), ('8', 'GLU'), ('9', 'LEU'), ('10', 'VAL'), ('11', 'TRP'), ('12', 'CYS'), ('13', 'THR')]
```
color_b.py

• To colour proteins according to the values of the B-factor column (can be replaced with any desired value).

Syntax: (see script header)

color_b (c;a or c;b), mode=ramp, gradient=bwr, nbins=30, sat=0.5, value=1
to color chains A and B with the Blue-White-Red gradient in 30 colors of equal numbers of atoms in each color.

http://www.mcglnmr.ca/ProtSkin/
PyMOL Log Function

• Records all actions as commands in a PyMOL session
  – GUI clicks
  – Written commands

• Needs to be opened at start of session

• Is only written when closed (File → Close log)
Example 1

Large subunit:
Catalytic mechanism

Small subunit:
Specificity

Protein
mRNA
rRNA
Example 2

René Magritte

Ceci n’est pas une pipe.
Aesthetics of Molecular Images

• Personal taste

• General guidelines:
  – Focus on relevant parts
  – Strip away unnecessary information
  – Find good viewing angle!
  – Choice of representation
  – No need for excessive graphics

• Good figures are (almost) self-explanatory.
Molecular Images

• Molecular graphics is an example of selective reduction of complexity.
Molecular Images

- Molecular graphics is an example of selective reduction of complexity.
Molecular Images

- Molecular graphics is an example of selective reduction of complexity.
Molecular Images

- Molecular graphics is an example of selective reduction of complexity.
Molecular Images

- Molecular graphics is an example of selective reduction of complexity.
Molecular Images

- Molecular graphics is an example of selective reduction of complexity.
Molecular Images

- Molecular graphics is an example of selective reduction of complexity.
Molecular Images

• Molecular graphics is an example of selective reduction of complexity.
Molecular Images

- Molecular graphics is an example of selective reduction of complexity.
Figures in PyMOL
Representations

ray_trace_mode 0
Representations

ray_trace_mode 1
Representations

ray_trace_mode 2
Representations

ray_trace_mode 3
Representations
Representations
Focus on the Relevant Parts

• Homologous peptides in different proteins
• Intuitive colouring
Residue Conservation

• http://www.mcgnmr.ca/ProtSkin/
T-Cell Receptor Interactions
Cross-Sections

• Proteasome
Electrostatic Surface Potentials
Composite Images

- MHC-I binding groove
- Images combined in Adobe Illustrator.
Simplified Composite Images
Cross-Sections
Comparing Active Sites
Compiled Graphics Objects
PyMOL Links

• PyMOL home:
  – http://www.pymol.org/

• PyMOL manual:
  – http://pymol.sourceforge.net/newman/user/toc.html

• PyMOL Wiki:
  – http://www.pymolwiki.org/index.php/Main_Page

• PyMOL settings (documented):
  – http://www.pymolwiki.org/index.php/Settings
<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.00-8.45</td>
<td>Introduction to PDB &amp; PyMOL</td>
</tr>
<tr>
<td>8.45-9.00</td>
<td>Break</td>
</tr>
<tr>
<td>9.00-11.00</td>
<td>PyMOL tutorial</td>
</tr>
<tr>
<td>11.00-11.15</td>
<td>Break</td>
</tr>
<tr>
<td>11.15-11.30</td>
<td>Summary</td>
</tr>
<tr>
<td>11.30-12.00</td>
<td>Quiz</td>
</tr>
<tr>
<td>Time</td>
<td>Activity</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>8.00-8.45</td>
<td>Introduction to PDB &amp; PyMOL</td>
</tr>
<tr>
<td>8.45-9.00</td>
<td>Break</td>
</tr>
<tr>
<td>9.00-11.00</td>
<td>PyMOL tutorial</td>
</tr>
<tr>
<td>11.00-11.15</td>
<td>Break</td>
</tr>
<tr>
<td>11.15-11.30</td>
<td>Summary</td>
</tr>
<tr>
<td>11.30-12.00</td>
<td>Quiz</td>
</tr>
</tbody>
</table>
## Programme

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.00-8.45</td>
<td>Introduction to PDB &amp; PyMOL</td>
</tr>
<tr>
<td>8.45-9.00</td>
<td>Break</td>
</tr>
<tr>
<td>9.00-11.00</td>
<td>PyMOL tutorial</td>
</tr>
<tr>
<td>11.00-11.15</td>
<td>Break</td>
</tr>
<tr>
<td>11.15-11.30</td>
<td>Summary</td>
</tr>
<tr>
<td>11.30-12.00</td>
<td>Quiz</td>
</tr>
</tbody>
</table>