Seminar: Cell Visualization

Protein Data Bank (PDB)

by

Thomas Waltemate

Overview

- **What is the Protein Data Bank?**
- **How to use?**
- **File format**
- **Viewers**

What is the Protein Data Bank?

- **(ww)PDB = (worldwide)Protein Data Bank**
  - Subbranches: PDBe, PDBj, BMRB...

- **“The Protein Data Bank (PDB) archive is the single worldwide repository of information about the 3D structures of large biological molecules, including proteins and nucleic acids”**
  - ➔ Data bank for structures of bio molecules

What is the Protein Data Bank?

- **Established 1971 (with 7 structures)**
- **Now: More than 80.000 structures**
- **Techniques: Electron Microscopy, X-ray crystallography, NMR (nuclear magnetic resonance)**

How to use?

- **URL: www.pdb.org**
- **Search: PDB ID, molecule name, author, keyword...**
  - PDB ID is a unique four character code consisting of letters and numbers (example: 1N9U)
- **Browse: Category, organism, sequence length... and more**
- **Molecule of the Month by David S. Goodsell**

How to use?

- **Download via “Download Files” and select the desired file format**
  - FASTA Sequence
  - **PDB**
  - mmCIF
  - **PDBML/XML**
  - …

- **Statistics: http://www.ebi.ac.uk/pdbsum/**

File format

- **PDB file (*.pdb)**
  - One datum per line
  - Each line starts with a different keyword
    - MODEL, ENDMDL, ATOM, HETATM, HELIX, MATRX...
  - For each column is defined which information it holds
  - Restriction: Only 80 digits per line

- http://www.wwpdb.org/docs.html for more information

File format

```
MODEL        1
ATOM      1  O5'  DT A   1       35.410  47.335  16.900  1.00 20.00           O
ATOM      2  C5'  DT A   1       34.452  47.177  15.853  1.00 20.00           C
ATOM      3  C4'  DT A   1       33.480  46.058  16.177  1.00 20.00           C

...


HETATM  224  K      K A  13       22.093  37.317  10.277  0.50 20.00           K
HETATM  225  K      K A  14       23.197  39.697   9.729  0.50 20.00           K
HETATM  226  CAA R14 A  15       18.156  36.833  21.916  0.10 20.00           C

...

ENDMDL
MODEL        2
```

File format

- **PDBML/XML (*.xml)**
  - Typical XML structure
  - Easy to parse
  - Easy to extend
  - Human readable

  - Big files

```
<PDBx:atom_siteCategory>
   <PDBx:atom_site id="1">
      <PDBx:B_iso_or_equiv>62.14</PDBx:B_iso_or_equiv>
      <PDBx:Cartn_x>38.672</PDBx:Cartn_x>
      <PDBx:Cartn_y>32.138</PDBx:Cartn_y>
      <PDBx:Cartn_z>2.778</PDBx:Cartn_z>
      <PDBx:auth_asym_id>A</PDBx:auth_asym_id>
      <PDBx:auth_atom_id>N</PDBx:auth_atom_id>
      <PDBx:auth_comp_id>ASN</PDBx:auth_comp_id>
      <PDBx:auth_seq_id>431</PDBx:auth_seq_id>
      <PDBx:group_PDB>ATOM</PDBx:group_PDB>
      <PDBx:label_alt_id xsi:nil="true" />
      <PDBx:label_asym_id>A</PDBx:label_asym_id>
      <PDBx:label_atom_id>N</PDBx:label_atom_id>
      <PDBx:label_comp_id>ASN</PDBx:label_comp_id>
      <PDBx:label_entity_id>1</PDBx:label_entity_id>
      <PDBx:label_seq_id>4</PDBx:label_seq_id>
      <PDBx:occupancy>1.00</PDBx:occupancy>
      <PDBx:pdbx_PDB_model_num>1</PDBx:pdbx_PDB_model_num>
      <PDBx:type_symbol>N</PDBx:type_symbol>
   </PDBx:atom_site>
   <PDBx:atom_site id="2">
  ...
   </PDBx:atom_site>
```

Viewer

- **Simple Viewer**
- **Jmol**
- **quteMol**
- **My Viewer**
- **Many more...**

Viewer

- **Simple Viewer**
  - A VERY simple viewer provided by PDB
  - Java Web Start
  - Only one type of Visualization (secondary structure)

Viewer

- **Jmol**
  - Java based (also as Java applet)
  - Cross platform
  - Multiple visualization types
  - Stereoscopic rendering

Viewer

- **quteMol**
  - Cross platform (precompiled+source for Windows; source for Linux)
  - Multiple Visualization types
  - Advanced rendering techniques
    - Ambient occlusion, shadow mapping...

Viewer

- **My Viewer (bachelor thesis)**
  - Cross platform (Linux and Windows)
  - Multiple Visualization types
  - Advanced rendering techniques
    - Shadow mapping, ray casting of quadratic surfaces
  - Stereoscopic rendering
  - Future work
    - Screen space ambient occlusion
    - Cell Visualization + Membrane mapping

Thank you for your attention!