Protein Structure and Visualization

Applied Bioinformatics lecture 6
David L. Tabb
Overview

• Introduction to Worldwide Protein Data Bank

• Introduction to molecular visualization tools
Where do structures come from?

As of January 25, 2011
X-ray crystallography: 57,513 proteins
Nuclear magnetic resonance: 7,632
Electron microscopy: 236

X-ray structures typically have higher resolution.
In solution, proteins explore a range of motion.

Images courtesy of Bensaccount and MartinSaunders.
Worldwide Protein Data Bank

- Maintains a single Protein Data Bank Archive of macromolecular structural data
- Spans Rutgers, UCSD, EMBL, EBI, Osaka Univ., Japan Science and Technology Agency, and U. Wisconsin at Madison
- Stores 3D structures in structured text files of PDB or mmCIF format
- Began in 1971 at Brookhaven with 7 structures
New structures accumulate slowly

ftp://snapshots.wwpdb.org/README
Molecule of the Month: Structural View of Biology

- Protein Synthesis
- Enzymes
- Health and Disease
- Biological Energy
- Infrastructure and Communication
- Biotechnology and Nanotechnology

List View of Archive By: Title | Date | Category

http://www.rcsb.org/pdb/motm.do
Pinning down the right structure

http://www.rcsb.org (searched for BGAL_ and opted to “Refine Query”)

Which species?

Older may not mean lower-res

What tech?
Beta-galactosidase in 3D

http://www.rcsb.org/pdb/explore/explore.do?structureId=1JZ7, visualized in Jmol
What is in a PDB file?

PDB files store the information that appears in 3D structures. They are text, with lines of many different types. Here, the “ATOM” lines define the coordinates in molecular space occupied by the atoms of an Arg of 1JZ7, a structure for beta-galactosidase.

<table>
<thead>
<tr>
<th>Serial</th>
<th>Name</th>
<th>Location</th>
<th>Chain</th>
<th>Residue</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>Occupancy</th>
<th>Temp</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATOM</td>
<td>1</td>
<td>N</td>
<td>ARG</td>
<td>A</td>
<td>-4.432</td>
<td>-58.076</td>
<td>-5.586</td>
<td>1</td>
<td>44.13</td>
<td>N</td>
</tr>
<tr>
<td>ATOM</td>
<td>2</td>
<td>CA</td>
<td>ARG</td>
<td>A</td>
<td>-4.158</td>
<td>-56.638</td>
<td>-5.649</td>
<td>1</td>
<td>25.95</td>
<td>C</td>
</tr>
<tr>
<td>ATOM</td>
<td>3</td>
<td>C</td>
<td>ARG</td>
<td>A</td>
<td>-4.343</td>
<td>-56.101</td>
<td>-7.091</td>
<td>1</td>
<td>21.64</td>
<td>C</td>
</tr>
<tr>
<td>ATOM</td>
<td>4</td>
<td>O</td>
<td>ARG</td>
<td>A</td>
<td>-3.638</td>
<td>-56.468</td>
<td>-8.024</td>
<td>1</td>
<td>100.00</td>
<td>O</td>
</tr>
<tr>
<td>ATOM</td>
<td>5</td>
<td>CB</td>
<td>ARG</td>
<td>A</td>
<td>-3.091</td>
<td>-55.899</td>
<td>-4.796</td>
<td>1</td>
<td>36.95</td>
<td>C</td>
</tr>
<tr>
<td>ATOM</td>
<td>6</td>
<td>CG</td>
<td>ARG</td>
<td>A</td>
<td>-3.314</td>
<td>-55.984</td>
<td>-3.273</td>
<td>1</td>
<td>28.52</td>
<td>C</td>
</tr>
<tr>
<td>ATOM</td>
<td>7</td>
<td>CD</td>
<td>ARG</td>
<td>A</td>
<td>-2.867</td>
<td>-54.788</td>
<td>-2.436</td>
<td>1</td>
<td>20.48</td>
<td>C</td>
</tr>
<tr>
<td>ATOM</td>
<td>8</td>
<td>NE</td>
<td>ARG</td>
<td>A</td>
<td>-2.587</td>
<td>-55.235</td>
<td>-1.088</td>
<td>1</td>
<td>30.16</td>
<td>N</td>
</tr>
<tr>
<td>ATOM</td>
<td>9</td>
<td>CZ</td>
<td>ARG</td>
<td>A</td>
<td>-1.766</td>
<td>-54.647</td>
<td>-0.227</td>
<td>1</td>
<td>21.64</td>
<td>C</td>
</tr>
<tr>
<td>ATOM</td>
<td>10</td>
<td>NH1</td>
<td>ARG</td>
<td>A</td>
<td>-1.14</td>
<td>-53.525</td>
<td>-0.481</td>
<td>1</td>
<td>18.62</td>
<td>N</td>
</tr>
<tr>
<td>ATOM</td>
<td>11</td>
<td>NH2</td>
<td>ARG</td>
<td>A</td>
<td>-1.623</td>
<td>-55.23</td>
<td>0.949</td>
<td>1</td>
<td>29.27</td>
<td>N</td>
</tr>
</tbody>
</table>

http://www.wwpdb.org/docs.html
RasMol: early molecular viewer

- Roger Sayle, a senior at Imperial College, worked with biophysicists to create a 3D molecular viewer in 1989.
- During his Ph.D. (1993), the tool had matured sufficiently for wide distribution.
- Scripting, shadows, and new views attracted thousands of users.

http://www.umass.edu/microbio/rasmol/pershist.txt
Jmol: workbench for visualization

• Chime (1996) had brought structures to the web, but it had lost its commercial support.
• Jmol evolved as an open-source toolbox for molecular structures at three levels:
  – As a library to be integrated in other software
  – As a web tool for communicating structures
  – As software to support tutorial projects

Essential Jmol commands

• Select: chooses a set of atoms that correspond to a description. Examples:
  – Select carbon  #picks all carbons in model
  – Select [HIS]  #picks all histidines
  – Select [HIS]92  #picks only His at position 92

• Color: changes the color of the selected set
  – Color white  #changes color to white
Summary

- Public databases of protein structures are growing, but not as quickly as sequence DBs.
- Tools for examining structures have taken advantage of the 3D graphics in PCs.
- Understanding scripting commands can enable more powerful use of viewers.
Challenges

• What is the highest resolution structure for human cyclin-dependent kinase 2?

• Where do potential acetylysine sites of human histone H3.3 appear in the structure?