Using Protein Data Bank and Astex Viewer to Study Protein Structure

Teppo Valtonen
ASN 50768A
Contents

1. Introduction .................................................................................................................. 3
2. Protein Data Bank ......................................................................................................... 4
   2.1 PDB Lite .................................................................................................................. 5
3. Astex Viewer ................................................................................................................. 6
   3.1 Using the Viewer ..................................................................................................... 6
4. Conclusion ..................................................................................................................... 8
1. Introduction

Finding a way to predict the 3D structure of proteins based on their amino acid sequence has long been a wet dream among bio scientists. Even today, the era of super computers, ever increasing calculation capacity and information networks the dream has not turned into reality.

The alternative is to explicitly measure the structures of all interesting molecules and list the results in an open database using standardized language. That is what the Protein Data Bank is all about. Researchers across the world study and measure different molecules and then deposit their research findings to the database maintained by a non-profit organization.

Storing the information however is not enough. Being able to exploit them is crucial in order to benefit from the storage. Thus versatile and easy to use tools are important for the information to be truly open.

Astex Viewer is a lightweight viewer designed to display coordinate and sequence information form molecules. 1 It runs as a Java Applet, which enables platform independent operation. It has been developed by Astex Technology Limited in Cambridge Science Park and further modified by the Macromolecular Structure Database Group of the European Bioinformatics Institute, under license from Astex Technology Limited.

In the following first the Protein Data Bank, PDB Lite and a pdb-format is introduced. Then the Astex Viewer and access to it are presented.

---

1 Astex Viewer Documentation, http://bip.weizmann.ac.il/astviewd/ViewerDocumentation.html
2. **Protein Data Bank**

By definition the Protein Data Bank (PDB) “is a single worldwide repository for the processing and distribution of 3D structure data of large molecules of proteins and nucleic acids”\(^2\). The organization maintaining the database is the Research Collaboratory for Structural Bioinformatics. The current holding is almost 30000 structure files of different macromolecules. The amount is said to increase with 50 to 100 entries weekly.

The structures of the molecules are stored in pdb-format. The pdb-file is a pure text file that describes the molecule atom by atom. For example in the coordinate section of the file each line presents the atomic coordinates for standard residues:

<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>73 - 76</td>
<td>LString(4)</td>
<td>segID</td>
<td>Segment identifier, left-justified.</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>

As the files are being saved in the database each is given a unique PDB ID. The ID is a four character alphanumeric identifier and it is assigned upon the deposition of the structure. Users can search the files for example with the ID, the authors of the files or using full text search. There are different search tools available for different purposes. One, Search Lite, is described more in detail below.

---

The archive can be accessed freely worldwide either via the main PDB Web site at http://www.pdb.org/, or one of its mirror sites. The mirror sites are located at different universities and research centers all over the globe. The links are listed at the main page of the main web site.

2.1 PDB Lite

The PDB Lite is a simple search interface based upon the Search Lite tool. It is maintained at different mirror sites, for example European Bioinformatics Institute (EBI) in the United Kingdom. The PDB Lite uses the OCA Browser database that allows access to the main PDB database. The user interface allows simple text search and returns a list of hits.

The difference between using the PDB Lite mirrors and the main PDB database is that the mirrors often allow direct access to the software developed at institute maintaining the mirror. For example Astex Viewer is a tool for viewing the structure molecules. EBI participates in developing the program and thus supports viewing the found molecule with it.
3. Astex Viewer

Astex Viewer is a tool designed for viewing coordinate and sequence information from molecules saved in pdb-format. The program has originally been developed by Astex Technology Limited but Macromolecular Structure Database Group of the European Bioinformatics Institute is modifying the program under a license from Astex Technology.

The viewer runs as a Java Applet. This removes platform dependencies. All one needs is a Java capable browser. The applet can be accessed for example via EBI PDB Lite mirror.

3.1 Using the Viewer

Since the viewer operates as a Java applet, getting it running is simple and straightforward task:

- Open PDB Lite EBI mirror site http://oca.ebi.ac.uk/oca-bin/pdble for example via link at the main PDB site.
- Search for protein, e.g. ‘human growth hormone’
- Retrieve the data matching the query
- Pick desired molecule from the list
- Select “View the molecule in 3D on AstexViewer”

![The graphical user interface of Astex Viewer.](image)

**Figure 1.** The graphical user interface of Astex Viewer.
The user interface is clear with three separate panels (see Figure 1.): a palette with function buttons, a structure panel with 3D view of the molecule structure and sequence panel that displays the sequence information of the molecule as a string.

With the buttons in the palette the user can select for example different ways to draw the molecule. The available styles are a bond-stick diagram (shown in Figure 1.), a ribbon model (Figure 2. a) and a net surface drawing (Figure 2. b).

![A ribbon (a) and a surface (b) model of the molecule.](image)

**Figure 2.** A ribbon (a) and a surface (b) model of the molecule.

Also different colorings can be selected to highlight for example ligands. In addition graphs can be drawn to open in additional windows. Separate atoms can be selected with the mouse and the selection is pointed out in all views and graphs (see Figure 3.).

![Protein tyrosine 111 zoomed and highlighted (a) and the corresponding Ramachandran plot (b, selection in purple).](image)

**Figure 3.** Protein tyrosine 111 zoomed and highlighted (a) and the corresponding Ramachandran plot (b, selection in purple).
4. **Conclusion**

Protein Data Bank is obviously a well kept, active and abreast of times research database. It offers free world wide access to up-to-date molecule structure information. Until predicting the 3D structure of a protein based on the amino acid sequence becomes a reality, this kind of storage is a necessity in studying the interaction between molecules.

Astex Viewer provides easy and quick way for visualizing the data. As a Java Applet it is platform independent but still fully functional program. Together the Protein Data Bank and the Astex Viewer are an effective tool to start studying the structures of and interactions between proteins.