1. What is the Protein Data Bank?

The (worldwide) Protein Data Bank (short: (ww)PDB) is a data bank for 3D structures of large biological molecules or to cite the definition from www.pdb.org: "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." It was established in 1971 with only seven structures. Nowadays it offers more than 80.000 structures of molecules. The main techniques that are used to record the different structures are electron microscopy, x-ray crystallography and nuclear magnetic resonance (NMR).

2. How to use the Data Bank?

The data bank offers multiple ways to get to the desired structure.
The first and easiest way to find what you need is to use the search engine. It is possible to search a molecule by molecule name, author, other keyword or by it's PDB ID. The PDB ID is a unique four character code that can be used to identify one specific molecule. This makes it easy to find a structure (and information about it) again, if only the PDB ID is known.
The second way to find what you looking for is to browse the data bank by categories, for example: organism, sequence length, cell component, molecular function and many more.
Also there is the category "Molecule of the Month" by David S. Goodsell. At this place a molecule and its function is presented every month. So if someone is not looking for something specific and is just interested in the topic of molecule structures maybe this could be a good start.
If one has found a structure of interest it is possible to download it for free. The only thing to do is, click the Download button and pick the desired format. There are a few different formats (FASTA Sequence, PDB, mmCIF, PDBML/XML...). The next chapter introduces two of these formats.

3. File formats

The most common format is the standard pdb-format, but it has a few disadvantages, so there is already a newer format, the pdbml-format, to replace it.

The pdb-format is pretty simple. It has a fixed number of columns (80) per line and it is exactly specified which columns hold which information. This makes it

easy to parse this type of file, but makes it also inflexible. That restricts the number of atoms of one molecule theoretically to 99,999 atoms per structure, for example. Also it is determined by a keyword at the beginning of the line what information this line offers.

Because the standard pdb-format has these disadvantages, a new format was developed, which follows a standard xml scheme. That format is, because of the xml structure, also easy to parse, but produces also very big files.

Conclusion: The standard pdb-format is still widely used, although it has it's disadvantages. The main reason for this could be the large files, that the pdbml-format produces.
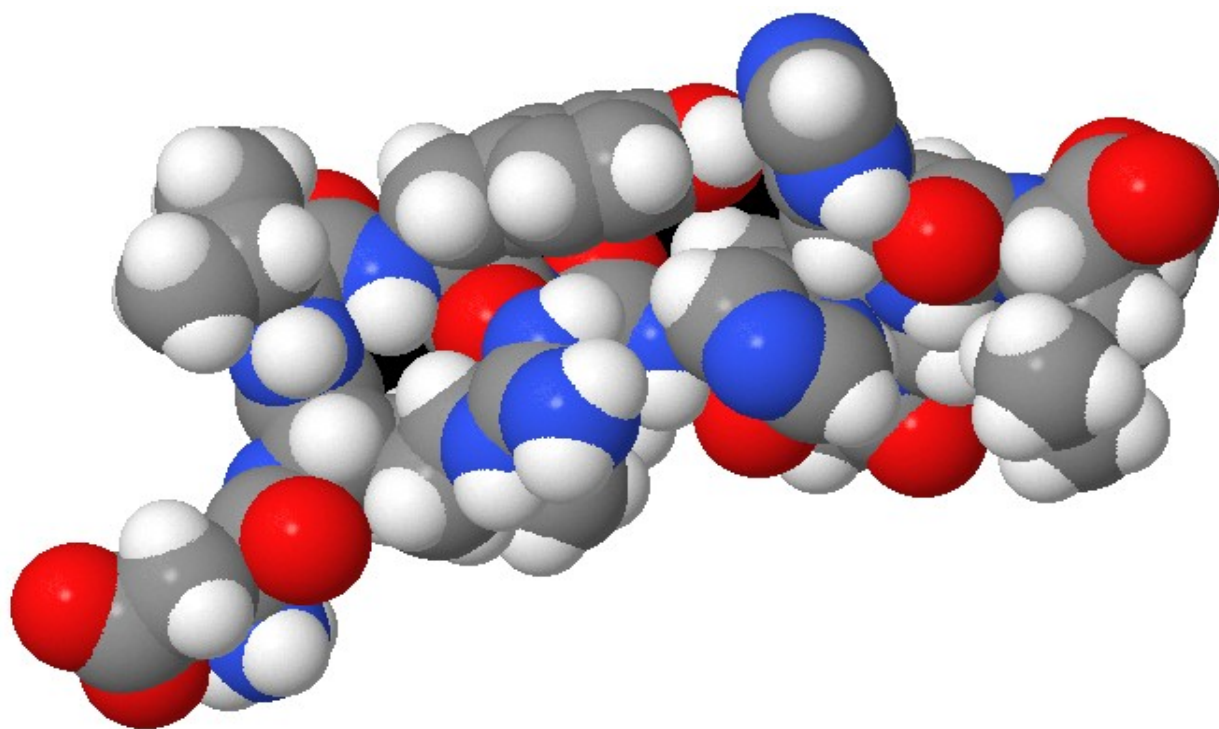

3. Viewers

In this chapter a few (popular) viewers for pdb files are presented.
First of all there is the simple viewer, which is provided directly on the pdb site to get a quick and 'simple' look at pdb structures. It uses Java Web Start and can only visualize the molecule in it's secondary structure (see Picture 3.1).
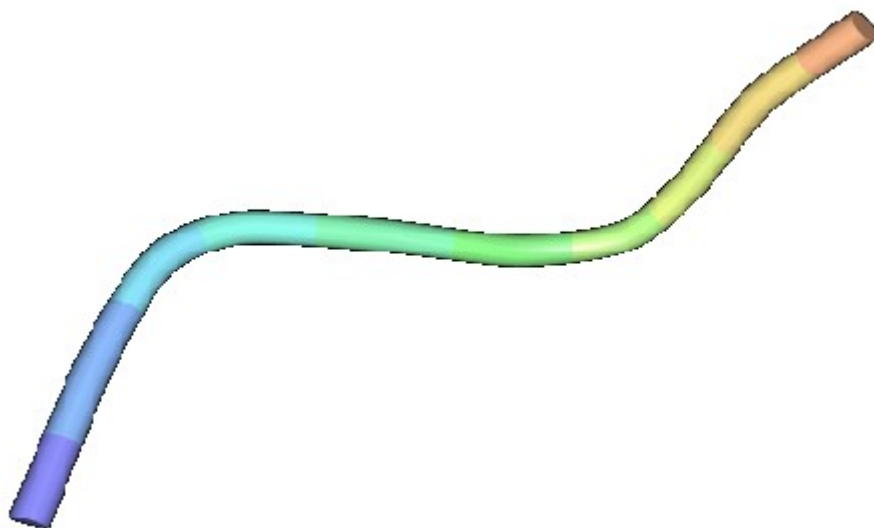

Picture 3.1: Visualization provided by "Simpleviewer." (PDB ID: 1N9U)

A very popular viewer is Jmol. It is a Java based viewer which has many functions, multiple visualization types (including stereoscopic rendering) and is fast, cross-platform running and also available as a Java applet. The main disadvantage of this viewer is it's not so good looking visualization (see Picture 3.2)

Picture 3.2: The molecule 1N9U with Van der Waals radii
visualized with Jmol.

Another well-known viewer is quteMol, which biggest strong point is its good looking visualizations and the variety of them. It uses shadow mapping, ambient occlusion and some kind of phong shading to provide a very good perception of depth and the shape of a molecule (see Picture 3.3). It also provides multiple visualization types (Van der Waals, balls and sticks).
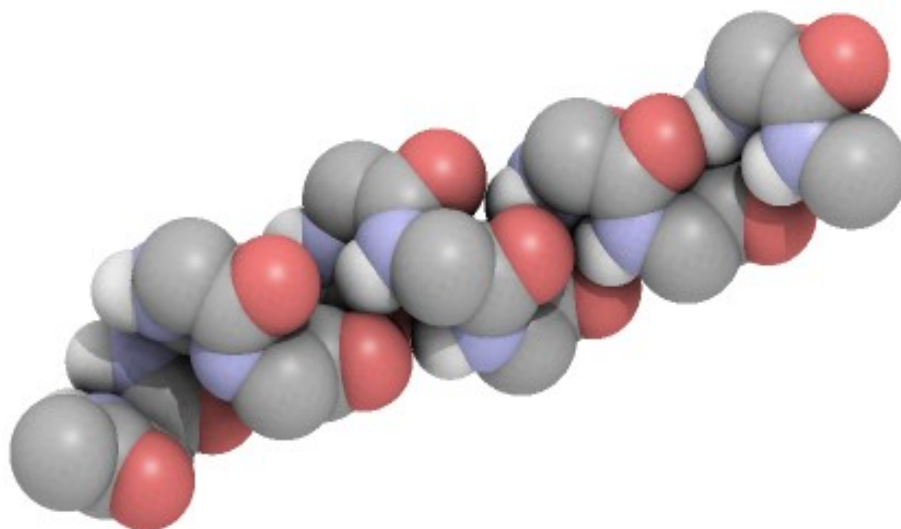

Picture 3.3: Visualization of 1PJD in quteMol.

A not so well-known viewer is one, that was part of a bachelor thesis and now is further developed in a master thesis. It uses screen space ambient occlusion (SSAO), shadow mapping and ray casting of quadratic surfaces to provide a good visualization while still have enough performance to render even large structures in real-time(also in stereoscopic 3D). Picture 3.4 shows an example with SSAO and shadow mapping enabled.
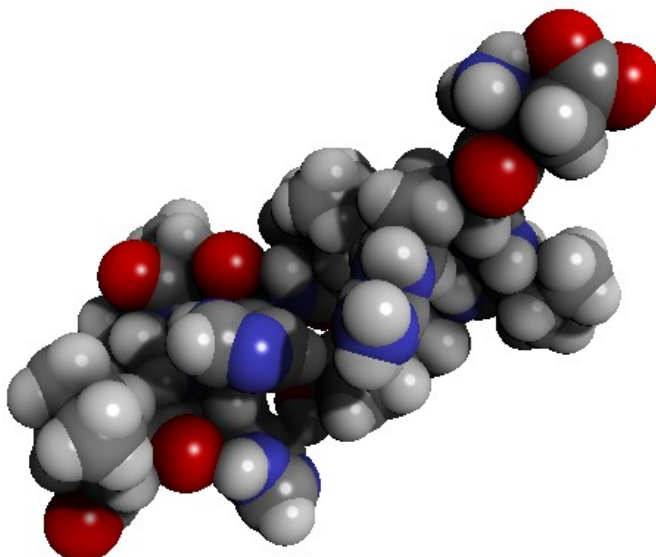
Picture 3.4: Visualization from viewer of bachelor/master thesis with SSAO and soft shadows (1N9U).
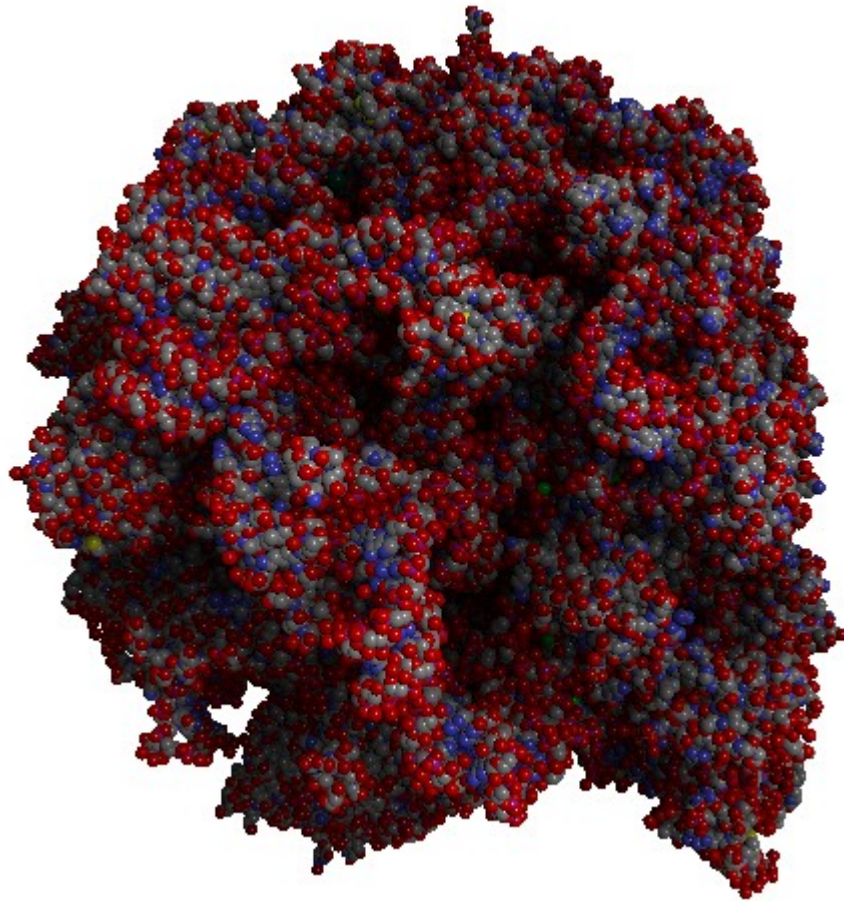
Picture 3.5: Large structure (99319 atoms) visualized
with viewer from bachelor/master thesis
(SSAO + soft shadows enabled; PDB ID: 3I55).

4. Conclusion

This elaboration only gives a small overview on the variety of viewers and the pdb file formats. Maybe it gives interested people a hint of what is possible and how to make use of the PDB.

Sources:

- www.pdb.org
- qutemol.sourceforge.net
- jmol.sourceforge.net
- www.wwpdb.org/docs.html
- Bachelor Thesis: GPU-Based Molecule Rendering