Introduction to Jmol

Gaining an understanding of protein and nucleic acid structure is an important part of learning biochemistry. Your textbook has excellent illustrations, but they are limited by being two-dimensional representations of three-dimensional structures. A number of computer programs are available which display molecular structures that you can interact with by moving, rotating, zooming in and out, and displaying in different ways. Using these programs makes it much easier to gain an appreciation of three- dimensional structure.

Using a variety of programs, both freely available software and commercial programs, anyone can easily download the coordinates of a macromolecule and display it on a personal computer. Biochemists use such programs to explore protein and nucleic acid structure, to plan experiments, and to interpret experimental findings. The use of these programs has become one of the indispensable tools of every biochemist.

One of these programs, Jmol, it is used widely by biochemists. It requires the use of typed instructions in addition to menus, but most people find it easier to use than other completely menu-operated programs. It is among the easiest to learn of the programs that are available.

Downloading Jmol

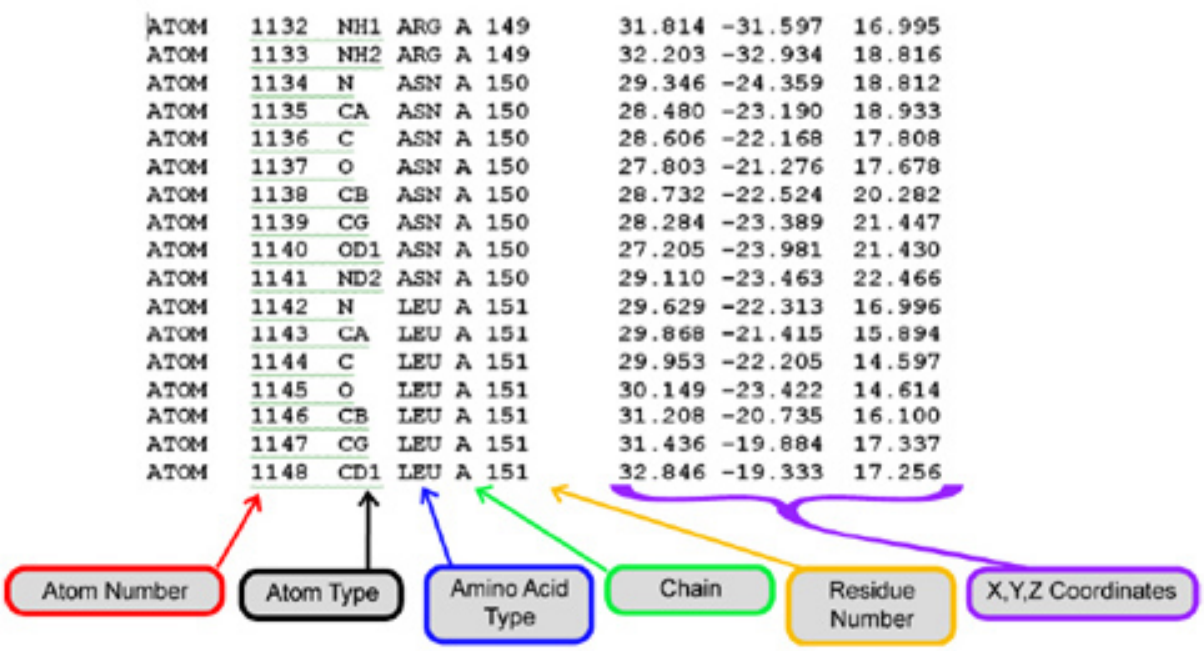
Jmol Can be Used in Two Ways:

1. **As an independent program on a desktop -** Jmol can be downloaded to run on your desktop like any other program. It uses a Java platform and therefore functions equally well in a PC or Mac environment.
2. **As a web application -** Jmol has a web-based version (oftern refered to as "JSmol") that runs on a JavaScript platform and therefore functions equally well on all HTML5 compatible browsers such as Firefox, Internet Explorer, Safari and Chrome. This web-embedded version of Jmol is what you currently see to on the right side of the screen.

The RCSB Protein Databank

Biochemists determine the three-dimensional structure of proteins and nucleic acids by X-ray crystallography, by NMR (nuclear magnetic resonance) spectroscopy, or by cryo-electron microscopy. **The most common molecular structure files we will be working with in this Jmol Training Guide is** RCSB Protein Databank (.pdb) files . A list of coordinates along with other information is deposited as a structure file in the RCSB Protein Databank (***http://www.pdb.org***) which is maintained on a server at Rutgers University in New Jersey. Molecular visualization software, such as Jmol, can use the coordinates stored in these file to create an interactive 3-dimensional visualization of a molecular structure. The result is a map of the structure in which it is possible to specify the location of every atom by its x, y, and z coordinates.

Once a structure has been determined, each atom in the structure is assigned an (X, Y, Z) coordinate to mark its location in 3-dimensional space. These coordinates are then stored in a file. The image below shows a short bit of code from inside of a structure file.



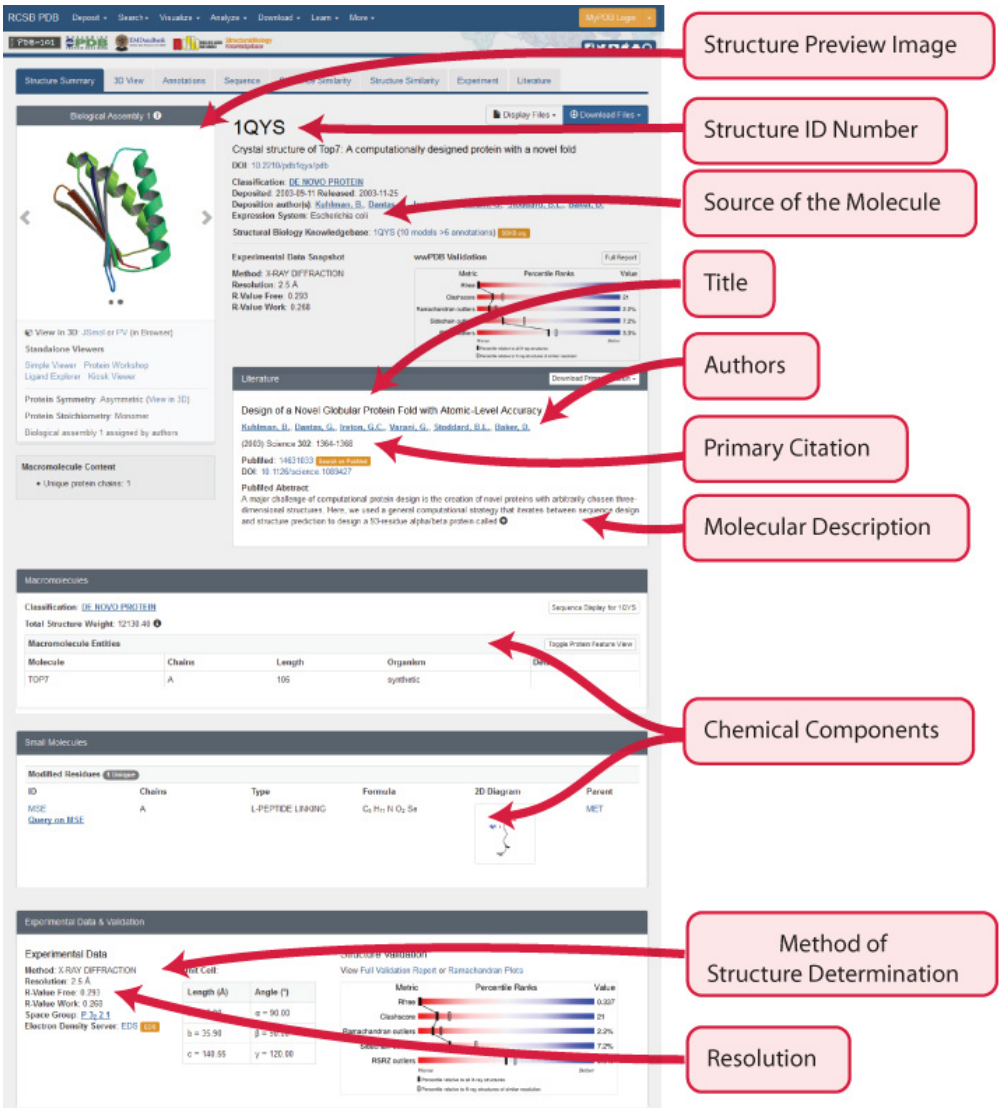
Finding Structures on the Protein Databank

Each structure hosted on the Protein Databank has a unique four character long alpha-numeric identifier, referred to as the structure's **PDB ID**.

Often more than one .pdb file will exist for a specific type of protein. For example, there are hundreds of .pdb file entries for the relatively common protein **Hemoglobin**. It is often a good idea to use specific information about a structure listed below to help determine if you have found the best possible file.

- Who are the authors of the PDB file?
- In which journal was the primary citation published?
- On what date was the file deposited into the PDB?
- How many chains are in this file?
- Are there any heterologous groups within this PDB file? If so, which ones?
- From what source was this molecule isolated?

When you click on a specific PDB ID, you will initially see the ***Structure Summary*** page for the structure. This page includes a variety of useful in formation about the structure.



- **Structure Preview Image** - Provides a quick overview of what the molecule or protein looks like.
- **Structure ID Number** - This 4 letter/number ID is a unique identifier that is assigned to the crystal data file upon deposition into the database.
- **Source of the Molecule** - From which species was the molecule isolated, such as human, bacterium, virus, mouse, etc..
- **Title** - Title of the .pdb file
- **Authors** - These are the researchers who were involved with the crystallization of the molecule. The senior author or principal investigator is usually the last author in science publications.
- **Primary Citation** - The journal article that accompanies the .pdb file. This is usually an excellent research resource for understanding the function of the molecule.
- **Molecular Description** – The abstract associated with the primary citation.
- **Chemical Component** - This will tell you the number of chains within the molecule and the chain identity. For example, in the hemoglobin file 1a3n.pdb, the chains A and C are the alpha-globin molecules and chains B and D are the beta-globin molecules. This section also tells you if there are any heterologous groups that were crystallized with the molecule. Not all .pdb files will have this section.
  - The 2-3 letter identifier used to designate the chemical components contained within the file listed are recognized by Jmol and can be used to select these molecules with the Jmol Console.
  - For example, if this section stated that there was NAG (N-acetyl-glucosamine) contained within the molecule, RasMol would recognize “NAG” and you could therefore “select NAG” and RasMol would be able to select the atoms within that chemical component of the PDB file.
- **Method of Structure Determination** - The method that was used to obtain the structural data (NMR, X-ray diffraction).
- **Resolution** - How accurate the data is; the smaller the number, the better the data.

The Sequence Page

Just above the .pdb file **Title** should be a series of tabs, the fourth of which is the **Sequence** tab. This section of the .pdb file page provides specific sequence information as well as secondary structure information about the molecule. **You can identify the alpha helices or beta sheets as well as the amino/carboxyl termini, which are the first and last amino acids of the protein.**