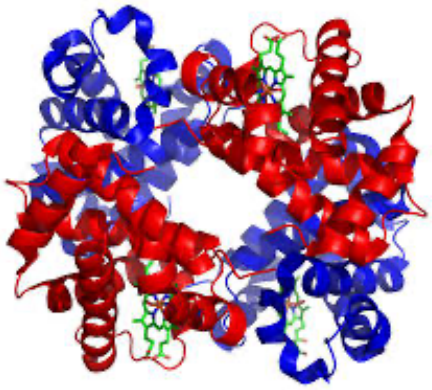
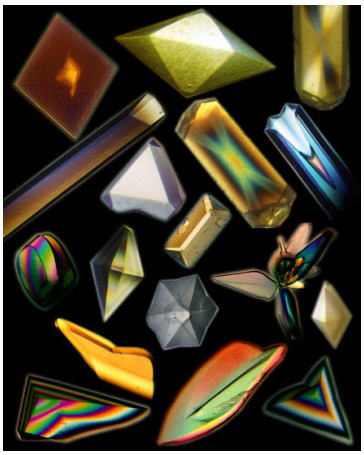
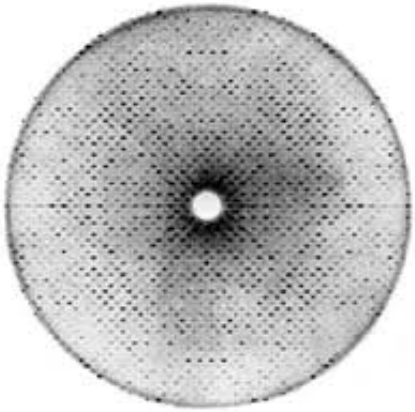
Lecture 20: 3D analysis

https://biopython.org/wiki/The_Biopython_Structural_Bioinformatics_FAQ

PDB format

You should be familiar with the idea that genes encode proteins. While it's easy to predict the sequence of a protein it takes a lot of work to solve the structure. Protein structures are typically solved from X-ray diffraction patterns taken from protein crystals. To the left you see a solved structure of hemoglobin (one of the first proteins to have it's crystal structure solved) as well as pictures of a number of different protein structures.

This data (the location of all non-hydrogen atoms) is typically stored in a PDB file. PDB is one of the worst formats in the world so hopefully you never have to write your own parser for it.

There are many programs that exist to visualize protein structure: https://en.wikipedia.org/wiki/Software_for_protein_structure_visualization

You are welcome to use what you are most comfortable with, but I tend to use UCSF Chimera. You can download it here (http://www.rbvi.ucsf.edu/chimera/) by registering for an account. As we discussed before, protein structures are stored in the protein data bank (http://www.rcsb.org/pdb/home/home.do). To start, let's download a representative PDB file – 3RIF. This is the structure of the glutamate-gated chloride channel. This protein is of interest as it's a known target if ivermectin (an important anthelmintic drug used to kill parasitic nematodes). Open this in chimera. Some things to note:

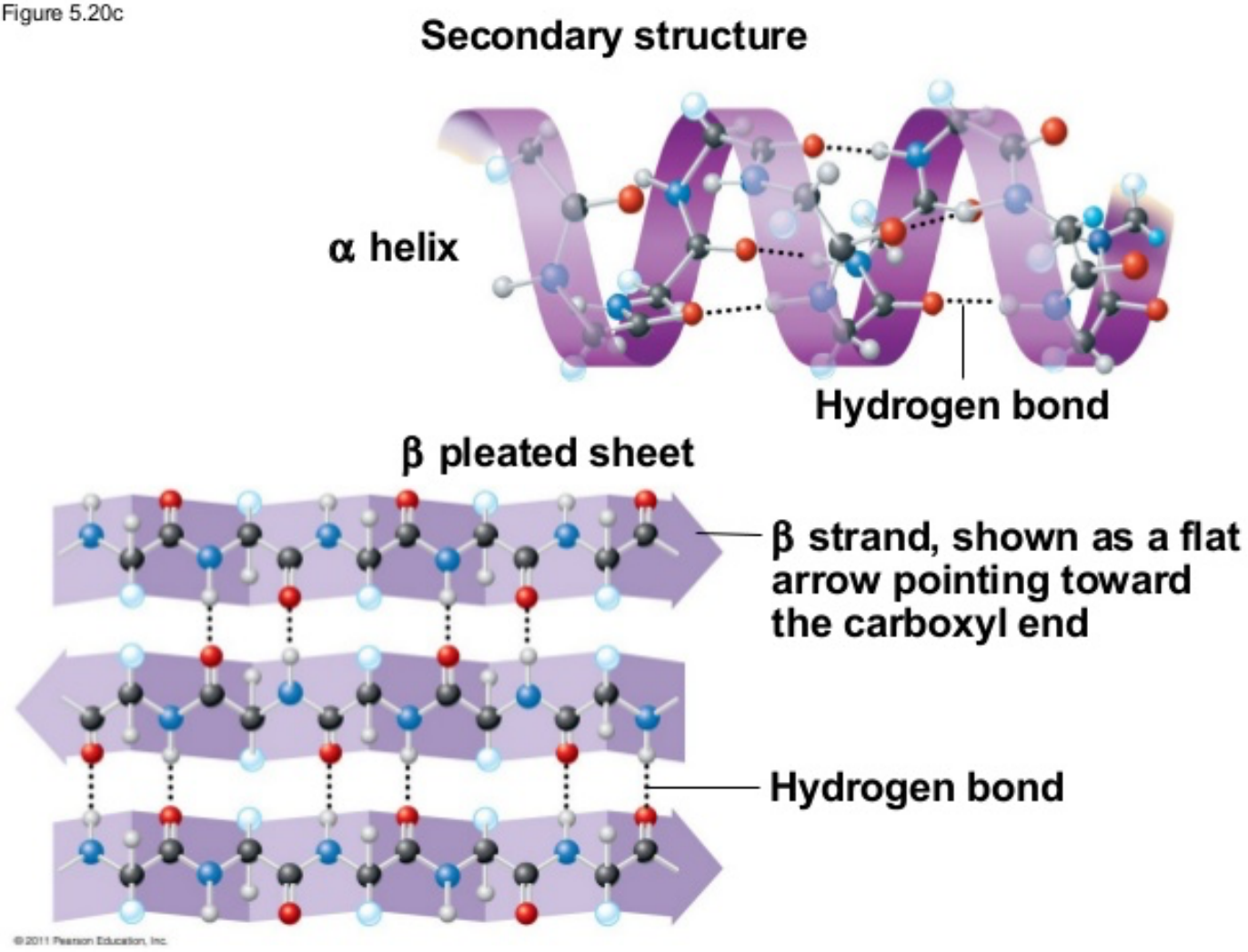
1. Chains
   In general, most structures contain multiple chains. These can be identical proteins that that form dimers (or higher ordered homomers). Sometimes multiple proteins are solved at the same time. For example, in membrane proteins, it is often necessary to include antibodies to stabilize the crystal structure. So don't be surprised if there are additional proteins that you are uninterested in within the pdb file.
2. Ribbon structure
   Showing all the atoms of a protein is in general too much info. Instead most side chains of amino acids are hidden completely, and the backbone is replaced with a ribbon representation. In the ribbon representation, proteins are represented by their secondary structure. Distinct widths and looks are used for alpha helices, beta sheets, and unstructured coils.
3. Display

   Almost all the features for each model/chain/residue/atom can be modified. You can change the color, ribbon display, atoms displayed, etc independently

4. Commands

Chimera accepts commands to change any of the display features. Further you can place these commands in a file. This can be very useful for using python to analyze the structure and then create a file you can use to modify the visualization of the protein.

Figure 5.20c



Commands

Some common commands are to color specific residues/chains different colors or display the side chain of certain residues. `color` is the command used to change certain colors. This works like so:
Change the color of all residues number 337 to red:

```
color :337 red
```

Change all arginines to orange

```
color :ARG orange
```

Color ligand yellow

```
color ligand yellow
```

color chain A blue

```
color /A blue
display /B:337
```