LAB 5: MORE ABOUT PROTEIN STRUCTURES.

We began using the basic functions of the molecular viewer PyMol in an earlier laboratory. Now we want to move on to some more advanced functions.

To begin with, let's **look at the .pdb data files**. These files, like .sdf or .mol2 files, are text files. They can be opened, viewed, edited with any text editor, like notepad, WordPad, Word, etc. However, you must save any changes as text; if you save a molecular file in Word format, it will not function. Let's open Ahelix.pdb in Notepad. (Recall, you can download the necessary pdb files at: http://courses.cm.utexas.edu/jrobertus/ch204/)

This file is unusually simple. There is a one-line REMARK as the header, and then it begins a list of the atoms in the molecule. Those for the first residue, Ala 1 are:

```
ATOM        1   N    ALA      1        1.470     0.000     0.000   1.00 25.00          N
ATOM        2   CA   ALA      1        0.000     0.000     0.000   1.00 25.00          C
ATOM        3   C    ALA      1       -0.514     0.000    -1.440   1.00 25.00          C
ATOM        4   O    ALA      1       -1.338    -0.835    -1.807   1.00 25.00          O
ATOM        5   CB   ALA      1       -0.517     1.209     0.769   1.00 25.00          C
```

The first column says this is an ATOM, but files may also contain HETATM (heteroatoms) which belong, not to the protein, but are ligands. Column 2 is the atom number, then the atoms type (N, CA, C, O, or CB), then the residue name (ALA), and number (they all belong to residue 1). The next three columns are floating point numbers that define the x, y, and z coordinates of the atom. The next 2 columns define the occupancy (1.0 for the protein, but ligands might be less), and then the temperature factor (here all set to a reasonable value of 25.00); the temperature factor is a measure of how much a given atom moves around from its central position. The last column has the atomic symbols.
 When you download full protein .pdb files, the headers will be much more complicated, with authors, methods, data collection, sequences, references and more. They will eventually come to the business part, though, which is the data we see here.

Now that we know what the files look like, let's open PyMol and the Ahelix.pdb file in it (top panel: File – open).

Stereo Viewing
We want to introduce you to viewing models in **stereo**. In this mode two images of the same model are displayed side by side; the two views are separated by a rotational difference generally of 6°. There are two kinds of stereo, divergent (or "wall-eyed") in which the left image is meant for your left eye and the right for your right and then convergent (or "cross-eyed") stereo in which the left image is viewed by the right eye and the right by the left. Divergent stereo requires a device to view with and we don't have that. Most structural biologists cultivate the skill of viewing cross eye stereo. It works on the computer screen, in large lecture hall, and at poster sessions. Not everyone can do it so don't feel bad if it is problematic. However, if you CAN make it

work your experience will be much richer and it will be much simpler to explore protein structures, which are rather complicated.

It is best to practice with the simplest possible models. You may therefore want to HIDE the models side chains (H tab, hide side chains). Indeed you may want to HIDE everything and the SHOW cartoon to give a smooth ribbon of helix. To go into stereo, go to the menu, click Display-Stereo. Two images should appear side by side. Stereo is a **toggle**, so by clicking in again, you can go back to mono. Note that beneath the Stereo toggle is a choice on handedness; you can fool with them, but keep cross eyed mode for practice. You can now rotate, translate, zoom the stereo image as you could mono. You can also use line commands, stereo and stereo off, to change the display.

To see stereo, though, at first keep the images still and centered in their frames (put the helix axis up and down the screen). Now focus between the two images, and slowly cross your eyes, bring two copies together in the center (you will at first be aware of two dimmer images off to each side). If all goes well, the central image should spring into 3D, like an old view master slide! BE PATIENT. This skill takes a while, but it is worth practicing a bit over a period of days. With the smooth ribbon you may not feel stereo has accomplished much, but wait until you pull up the entire model with side chains, or a whole protein in future work, and the power will be evident.

Rotamers and mutations

All side chains except Gly and Ala, have rotamers. There is a rotatable bond between the central $C\alpha$ and the first atom in the side chain ($C\beta$); there may be more rotatable bonds along the side chain as well. For a variety of reasons, not all possible rotations are allowed or favorable. Theory and observation have shown that certain configurations are preferred, and these are stored in PyMol as **rotamers**.

 Go back to a mono display of the helix, and display the side chains. Let's use Phe 4 to demonstrate. Select resi 4, go to the A tab of the selection and center on it. Now go to the top panel Wizard pool down and select **mutagenesis.** A new working window appears in the right hand side of the viewing panel, the top of that menu should say No Mutation by default. If not open it and select no mutation. The viewing screen will say "pick a residue" so click on the $C\alpha$ of Phe 4. The side chain will change color and jump to a new location; this is Phe rotamer 1. Now use the left/right arrow keys to scroll through the 5 allowed Phe rotamers. Note that position #5 clashes with the N terminal part of the helix and PyMol puts up some red balloons to show the collision.

You can also change your model. With F4 still selected go to the top bar of the mutagenesis window and click on No change. A fly-out menu appears that allows you to convert F4 to any other residue. Try switching to met, and scroll through its 13 rotamers; to make the change click on "apply" in the mutagenesis window. If you like you can mutate a structure and save it. After applying a mutation go to the top panel: File- save molecule-Ahelix, and a save window appears that allows to save the mutated file, under a new name, where ever you want to put it.

This option can be useful in drug screening and many other structural biology tasks. Often you will be interested in a protein from some species, but that structure is unknown. A strongly related one may be known and, with care, you can "**convert**" the known experimental model to your hypothetical one, using this mutagenesis option.

To clear the Ahelix file out, go to File – **reinitialize** and it will clear the screen.

VIEWING A PROTEIN
We now want to look at a real protein. We'll use a small, but very important enzyme called **dihydrofolate reductase**, or **DHFR**. The job of this enzyme is to transfer hydrogen from a carrier, called NADPH, to a "substrate" called dihydrofolate, and thereby convert it to tetrahydrofolate. That is: NADPH + DHF $\rightarrow$ NADP$^+$ + THF.

We will discuss this more later, because we need to know how enzymes work to design inhibitors and drugs. For now, we will just say that they are relatively large ligands, and we will say that the drug methotrexate is an analog, a look-alike, for DHF and THF, which are "folates".

Go to the class web site and **upload** the file called **DHFR-hs.pdb**, this is the human form of DHFR.

A. Look at the pdb file
Use WordPad, or other text editor, to look at this file. Notice how long the header information is. You can scroll down to find the ATOM list, which includes 186 amino acid residues. Following them you will see a HETATM list. It begins with something called NDP (all these names must be 3 letters), which stands for the nucleotide cofactor abbreviated NADPH. Further down is the ligand MTX, which stands for the drug methotrexate. Then there is a long list of water molecules (named HOH) that solvate the protein. Near the bottom of the file, the CONECT lines tell PyMol (or other viewers) how to connect up the atoms of the ligands so they display properly (the viewers have a built in understanding of how to connect protein atoms).

B. Load the file
Open the file DHFR-hs.pdb in PyMol; it will look like the image to the right, which is pretty difficult to make any sense of. Let's break it down into more manageable, and useful, components. As before, the yellow arrow points to the molecule file, and it's A, S, H, L, and C tabs. We know that this molecule contains the protein, but also three heteroatom groups. In fact, the display shows a bunch of red squares or crosses surrounding the protein; these are the solvent water molecules.


It is useful to **break out all the groups** within this
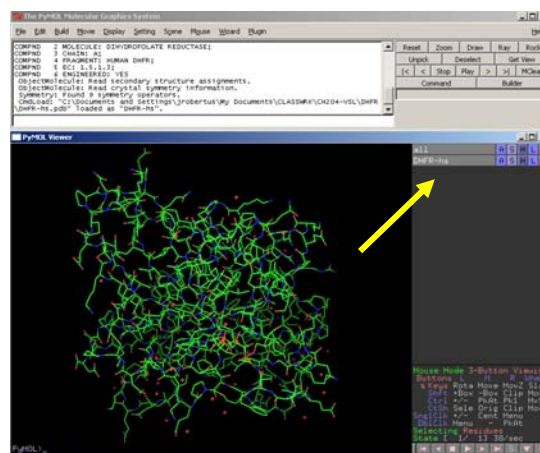
model so they can be operated on independently. To do that, let's use the line commands.

1. type: select wat, resn hoh   → this will select the waters and create a new object (row) in the right hand window. You can go to H, and hide all the waters.

2. select ndp, resn ndp → select the NADPH, go to S and display this as sticks.

3. select mtx, resn mtx → select the MTX, go to S and display this as sticks.

4. It also useful to create a group that is just the protein so type:

select pro, resi 1-186 Go to S and choose cartoon.

The display should look like the picture to the right; note that your 4 named objects are all in the right hand panel, ready for use. (We will add more later in this exercise)
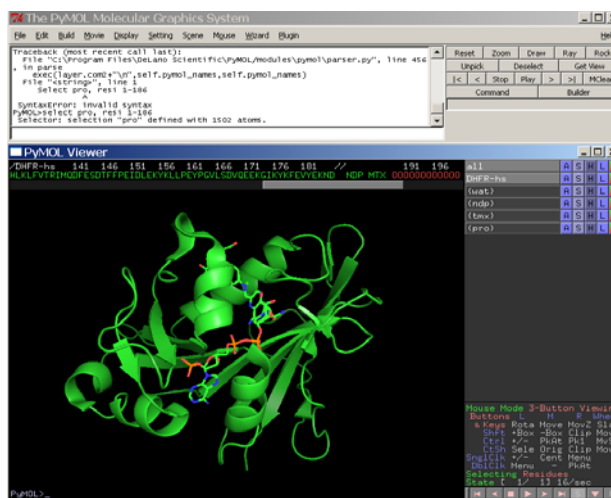
You can now rotate the model, spin it around and get a feel for its beauty and the way the ligand bind in their specific pockets.

→In general, it is a good idea to **hide all hydrogen atoms** in your display. You can clear them from everything, at any time, by going to the top "ALL" object, click on H (hide) and scroll down to hydrogens. This makes viewing easier by reducing clutter.

Save your work

When you work on an extensive PyMol project like this, it is prudent to save it. Go to the upper panel, click on File – **save session** as and you will be prompted for a name and location. Be sure to save this as a **.pse** file so your objects are all saved. You can later save updated sessions as you add more things.

Look at the substrates

Go to the pro selection, and click H everything (hide all). All that is left is NADPH, the hydrogen donor for the enzyme reaction, and MTX, posing as the substrate folate. Note that the business ends of each are very close in space. Although it is not obvious to the casual viewer, they are also optimally aligned for hydrogen transfer. This illustrates a key way in which enzymes work. By providing perfect little nests for them, they bring 2 chemical close together in space so they can react easily. A typical enzyme can increase reaction rates 10^{12} fold (a million million times). We will come back and look at these ligands more later.

Look at the protein.

Use the H tab to hide NDP and MTX. Go to the "pro" object row, and use the S tab to **show ribbon** (you can Hide other representations you may have brought up, eliminate all but the ribbon for now)**.** This produces a so called alpha carbon trace of the

backbone. To enhance our investigation, lets label the CAs of a few residues. We can select every 10th CA with this line command:
select trace, resi 10+20+30+40+50+60+70+80+90+100+110+120+130+140+150+160+170+180 & n. ca

A new object called "trace" appears in the molecule list. Go to the "trace" row, click on the L tab and scroll down to choose "residue identifier". Every 10th residue is now labeled and you can follow the backbone path. If you prefer, go to the "pro" object and Show cartoon; you can also hide the ribbon so only the fancy presentation of the protein chain is now labeled. Notice that the path of the folded protein appears almost random. The tracing can be made even easier by **color**. Go to the C tab of "pro" , color by chain- rainbows. The amino terminus is blue and as the chain progresses the colors move along the rainbow, winding up with the C terminus colored red.
As you view the protein in this simple form you can see the presence of α-helices as spiral structures. You can also see β-sheets which in the cartoon mode are represented as broad arrows. We looked at helices in a prior lab so let's look at the sheets now.

Beta sheet
In order to explore a β-sheet it may be convenient to isolate one and temporarily blank out the rest of the structure. Make this new object:
select beta, resi 4-10+130-138+176-185

Now Hide the "pro" and Show "beta" as lines. It still looks messy so Hide the beta side chains; as usual, you can zoom in with the right hand mouse button for a bigger view. You can Show beta as a cartoon to remind you of the arrows, which show this is a Mixed sheet with 2 strands parallel and 2 antiparallel. Now Hide the cartoon.

We want to display the hydrogen bonds within this mixed sheet. Go to the "beta" A tab, scroll to "find", then "polar contacts" and then choose "just intra-main chain". The hydrogen bonds are displayed as dashed lines and hold this together as a basically flat, 2D structure.

Now go to "beta" and show the side chains as lines. It is a complicated picture, but if you attend to the first strand (residues 4-10) you can see that the side chains project above and below the plane of the sheet in an alternating fashion: above, below, above, etc. This is true of β-sheets in general.

Protein folding
We said earlier that proteins tend to fold with hydrophobic residues inside, and charged and polar residues outside. Let's see that idea at work.
Clear your screen by Hiding everything for beta and for pro
Show the ribbon form of pro and Color it green.

Create an object for the nonpolar side chains with this command set:
        select hyfo, resn val+leu+ile+met+phe+trp+tyr
        Show sticks – Color yellow

Hide main chain

Create an object for the electrostatically charged side chains with this command set:
select ion, resn asp+glu+lys+arg
Show sticks – Color blue
Hide main chain

Create an object for the polar side chains with this command set:
select pol, resn ser+asn+gln
Show sticks – Color blue
Hide main chain

We have treated charged and polar residues the same
way here, but you can view them separately if you like.
In any case, the model will look something like the picture
at the right. The nonpolar, hydrophobic groups (yellow)
tend to be in the center, while the hydrophilic charged and
polar groups are on the surface. There are exceptions to
this, especially in active site regions where nonpolar
residues are held to interact with certain substrates on the
surface, but you can see the general pattern.
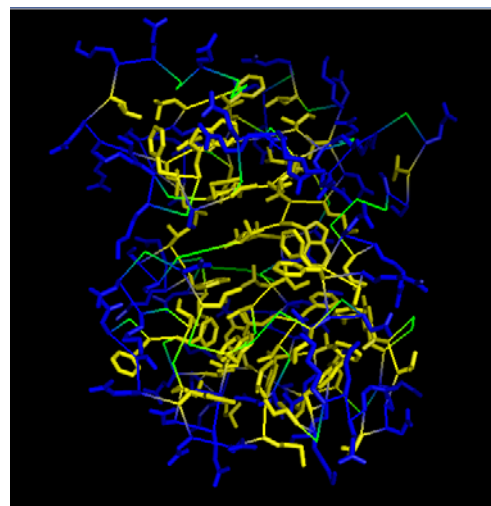
More on substrate binding

Clear the screen, Hide everything under "pro". Now Show "pro" ribbon and color it
green.
Show ndp and tmx as sticks.
Both of these substrates (MTX is substrate analog) have hydrophobic moieties, but are
also negatively charged, poly anions. DHFR provides cations to pair with these part of
the substrates. To see this: select resi 54+55+68+70, then Show them as sticks, color
them blue, and hide the main chain. Rotate the model around; you should see the way
in which these positively charged side chains pair up with the negatively charged
carboxylates of the substrates.

Now clear the screen as before. Go to "pro" and **Show surface,** then go to the C tab
and **Color it by atom.** This combination shows the "space filling" view of the protein.
You can see that the areas where the substrates bind are deep pocket.

Go to the "ndp" object and show sticks. Do the same for "mtx". You can see where the
polar oxygens (red) and nitrogens (blue) are with respect to the substrate molecules.
This reinforces the idea of the active site cleft. As your rotate the model around, note
how the "business ends" of the substrates (the nicotinamide and folate pterin groups)
are buried deep inside the protein, away from water. This is where the hydrogen will be
transferred.

DHFR is a drug target protein. MTX is a substrate analog, but it binds to the protein more tightly than the real substrate (DHF) and blocks the enzyme up acting as an inhibitor. This is the general way most drugs act (though of course they bind to different proteins). We need an understanding of these principles to identify, or design, novel inhibitors and drugs.

Things to do

1. DHFR is an example of a "globular" protein, but it is clearly not a sphere.
    a. Measure the size of DHFR using 3 pairs of atoms to mark a height, width, and breadth. Now let's estimate its volumn by assuming it is a sphere of a radius that is ~average of your 3 measurements. Convert this volume to cm^3.
    b. The MW of DHFR is ~ 20,000 (daltons). Given that a dalton corresponds to 1 gram/($6x10^{23}$), compute the molecules weight of DHFR in grams.
    c. Compute the rough density of the protein; recall the density of water is 1 gm/cm^3.

2. Measure the distance between several of the oxygen of the ndp phosphates and the nitrogens of the cationic side chains of DHFR which ion pair with them. (record your choices). In the same way, measure the distances between the carbonyl oxygenss of several peptides in the β-sheets. Compare these two ranges of values.

3. The hydrogen (hydride ion really) that is carried by NADPH is on atom C4N. Hide the protein, select ndp by clicking on it. Then right click and choose "label" then atom name. (You may need to display ndp as lines, not sticks to see the atom names clearly. Remember you can zoom in to enlarge the image by depressing the right mouse button and dragging toward yourself). Do the same thing for mtx. Now measure the distance between C4N on ndp and N5 of MTX. This is approximately there the reduction would occur with a real folate substrate. You can remove the labels for the objects by clicking on Hide labels.

4. Hide all, and then display the protein as a surface. Blue and red atoms are polar nitrogen and oxygens respectively, but carbon is green. Make 2 pictures of the protein which show what you think is the highest and lowest fraction of polar atoms. estimate what fraction of the molecular surface is carbon. Some proteins fold so that there are large nonpolar areas; these fascilitate binding with other molecules to form dimers or higher aggrgates.