LAB 6: Dihydrofolate Reductase (DHFR)

As described earlier, **dihydrofolate reductase**, or **DHFR** is an enzyme that transfers hydrogen from a carrier, called NADPH, to a "substrate" called dihydrofolate, and thereby convert it to tetrahydrofolate. That is: NADPH + DHF → NADP$^+$ + THF. The biochemistry, and biological significance, of this reaction is too complicated for us to describe in detail here, but suffice it to say that DHFR action is required for DNA synthesis (it helps make one of the bases). DHFR has been a favorite target for **drug design**.

Cancer cells divide much more rapidly than ordinary cells, and have a higher requirement for DNA synthesis. Inhibitors of DHFR harm all replicating cells, but tumors are hit proportionately harder than healthy cells. As mentioned above, DHFR has two substrates, NADPH and DHF. Inhibitors, drugs, tend to mimic substrates and bind preferentially to the substrate site, thereby slowing enzyme action. Inhibitors could be designed to mimic NADPH, but a great many enzymes, form many metabolic systems, use NADPH so such an inhibitor would have many undesirable results. Fewer enzymes use folates as substrates, so this is a better choice, and the family of drugs that mimic folates are called anti-folates. Methotrexate, MTX, is one of the most commonly used, and we saw it as a folate analog in the previous lab.

In addition to cancer treatment, antifolates are also used as antibiotics. Nearly all organisms have DHFR activity, but over the course of evolution, the structure of the enzymes have diverged. For example, humans and bacteria have not had a common ancestor in over a billion years. Over this immense time the DHFR from mammals and bacteria have drifted apart until only the most important functional parts remain intact. As a result, it should be possible to find drugs that can bind to the bacterial DHFR, but not to the human. (In fact one drug, trimethoprim, does just this.) What we want to do in this lab is to compare human DHFR with a representative bacterial one, and we will use the well characterized enzyme from *Lactobacillus casei* as an example.

To begin, lets look at the amino acid sequences of DHFR from humans and L. casei. The human enzyme has 186 residues, and L. casei just 162. Over time mutations in the genes generated different amino acids, but also insertions and deletions. The analysis

```
DHFR-hs      1 VGSLNCIVAVSQNMGIGKNGDLPWPPLRNEFRYFQRMTTTSSVEGKQNLV     50
                ...:.|.::|..|||:|.||| .|.::..||:..|.      || ::
dhfr-lc      1    TAFLWAQNRNGLIGKDGHLPW-HLPDDLHYFRAQTV-----GK--IM      39

DHFR-hs     51 IMGKKTWFSIPEKNRPLKGRINLVLSRELKEPPQGA---HFLSRSLDDAL     97
                ::|::|:.|.|: |||..|.|.:||:.:.....|||  |.::.....|.
dhfr-lc     40 VVGRRTYESFPK--RPLPERTNVVLTHQEDYQAQGAVVVHDVAAVFAYAK     87

DHFR-hs     98 KLTEQPELANKVDMVWIVGGSSV---YKEAMNHPGHLKLFVTRIMQDFES    144
                :..:| ||.        |.||:.:   :|:.::     .|.|||:...||.
dhfr-lc     88 QHLDQ-ELV-------IAGGAQIFTAFKDDVD-----TLLVTRLAGSFEG    124

DHFR-hs    145 DTFFPEIDLEKYKLLP----EYPGVLSDVQEE--KGIKYKFEVYEKND     186
                ||           |::|      ::..|.|...|:  ..:.:.:||::|..
dhfr-lc    125 DT---------KMIPLNWDDFTKVSSRTVEDTNPALTHTYEVWQKKA     162
```

of these sequences is a rigorous study in its own right, so we won't describe how it is done, but simply present a reasonable alignment of the two. Identical residues are joined by a bar, very conservative changes with 2 dots (like Leu and Ile etc), and conservative changes with a dot. Insertions and deletions are shown with horizontal dashes. The 2 sequences are 27% identical and 47% "similar".

Now we want to look at their structures.
1. Bring up PyMol and load your human DHFR project, or reconstruct it if necessary.
That is, load DHFR-hs, and break it into protein, NDP, MTX, and wat as before. To facilitate viewing, show the ligands as sticks, but shown only the ribbon mode for the protein.

2. Now download and display dhfr-lc. Show only the ribbon form initially. From the top panel click **display** and **sequence.**

Your screen should resemble the picture at the right. DHFR-hs is in green and dhfr-lc is cyan. The coordinates of the two were independently measured in their respective crystals, and so they are randomly positioned with respect to one another. At the top are the 2 sequences. We will use this list to create objects for the bacterial enzyme. (If we use line commands, like select resi 1-162, they will be ambiguous and select those residues from **both** proteins.).
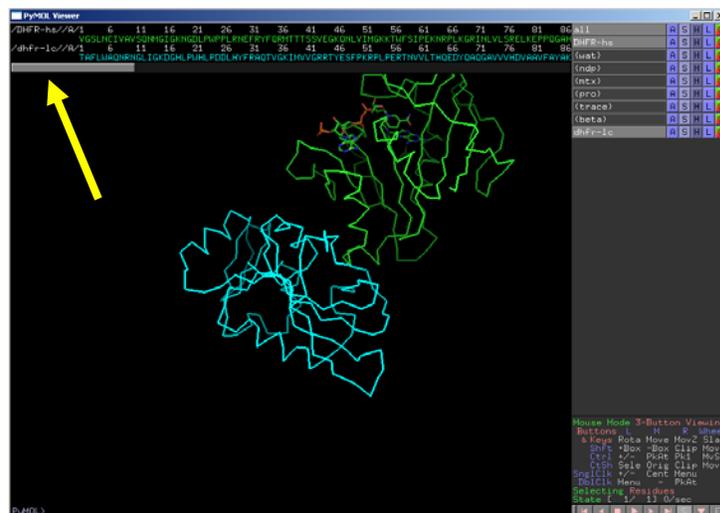
3. Make new objects with the sequence. Click on T #1 in the L. casei row, hold the shift key, scroll to the end (A 162) using the gray bar (yellow arrow), and click on the A. You should select the L. casei protein and it will appear as a selection in the object list. Go to the A tab, and **rename it pro2**. Slide the bar along the L casei sequence and select NDP; rename it ndp2. Similarly make a mtx2, and wat2 (the waters are shown as red "o"s in the list). Your list should resemble that at the right; now you can easily control all the major aspects of both proteins independently.

4. In order to compare the human and bacterial enzymes we would like to superimpose them. This can be a tricky business depending on the proteins involves, say if only one domain of a complex pair is evolutionarily related. However, here it is straightforward despite the great evolutionary distance. PyMol has several ways of analyzing the relation, but the simplest will do just fine. It is common in structural biology to align two proteins in space based on minimizing the square of the distance (rmsd) between corresponding atoms. (The alignment algorithms generate a 3x3 rotation matrix and 3

element translation vector to be applied to one set of atom coordinates, but we don't need to delve into that for our work.) In general only the CAs are used unless there is some reason to use other main or side chain atoms. The PyMol line command we will use is this:
align dhfr-lc & n. ca, DHFR-hs & n. ca

it says "move the bacterial protein onto the human, using only atoms named CA". PyMol carries out the minimization algorithm, and when complete it writes key information into the top panel window, right above where the line commands are entered. It will probably say "RMS =   0.850 (92 to 92 atoms)". That is 92 atoms figured in the superposition calculation, and they had an rms deviation of 0.85 Å. **Write that down for your report.**

RECALL: 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.

 Notice the cyan molecule now lies over the green molecule, but they are no longer centered. You can drag the to the center using the middle button (depress the wheel), or select (right click) and atom near the middle of the protein, go to the A tab, and click on center.

5. **Compare the human and bacterial enzymes.**
A. The first thing you notice is that the folding pattern, or "fold", of the two proteins is astonishingly similar, despite the differences in amino acid sequence. You have just learned a key rule of molecular biology: **tertiary structure** (the fold) **is more conservative than primary structure** (the sequence).

Go back to the figure with the two sequences. Note that in the alignment the first insertion we see in the human (or deletion in the bacteria) begins at human residue 40 and extends about 8 residues. To see this insertion on the model **select the inserted residues**, using the sequence selection tool. That is, on the human sequence, click on S#40 and shift click on N48. This will add a selection object to the molecule list and highlight these residues on the model. By rotating the screen images you can see that the amino acid sequence insertions form a structural loop that is absent in the bacterial structure. That is, the insertion has been added to the outside, flexible part of the protein, or alternatively, the bacterial protein deleted a surface loop that it did not need with the information presented here, we can not say which is likely to be correct. To get rid of the selection markers from the screen, just click the selection name (sele), or right click on a blank region of the display window, and the highlighting dots disappear.

To further analyze this insertion/deletion, go to the sequence selector. Click on the human T38, now CONTROL click on L49, now control click on L casei (lc) T34, and control click on lc I38. Go to the sele object and Show sticks. These 4 emphasized residues mark the fixed part of the DHFR core fold. You can rotate and zoom your image to lay out these anchor points which are clearly homologous residues in the

ancient structure, and see the different folding paths that have evolved. The bacteria makes a short turn and the human has a longer loop. **Make a picture of this for your report.** You can analyze the other insertion/deletions as well, but you should have the idea with this clear example.

B. Another key point in this species comparison is to note that the active site area is the most conservative part of the protein. Once nature has selected for an efficient catalytic mechanism, it tends to keep it. Remote scaffolding parts of the proteins, like the loop we just saw, can change without to much effect, but the business end is much more conservative. Be sure that the substrates, ndp, ndp2, mtx, and mtx2, are shown as sticks to facilitate viewing. Notice how closely they overlap, especially the rings where the hydride is transferred. Also note that the β-sheet lying under these molecules is virtually superimposable between the two proteins. (**If your screen starts to get messy at any point in the following, just hide everything, and re-display what you want, say the two protein ribbons, the 4 ligands as sticks, and then whatever else you need**)

Let's look at some aspects of the conserved active sites. For ease of viewing, be sure the proteins are just ribbons, and the substrates are sticks. Now on the sequence selector, click on human A9 and control click on L. casei (lc) A6. On the sele object, Show sticks. Zoom in on this area and note that these two alanines are virtually superimposable. The amide group of NADPH is forming a pair of hydrogen bonds with the carbonyl oxygen and backbone amide of the alanines. **Measure these distances and include it in your write up** (the measurements appear as objects in your list but you can delete them to clean up your work space)**.** In this instance **backbone** atoms are binding the substrate.

Now go to the sequence selector and click on hs E30 and control click on lc D26. Show as sticks if you like, These residues superimpose and they point at the MTX group (folate mimic). The carboxylic acid **side chains point at the MTX** exocyclic amine (N2 if you label it) and at the MTX ring N1. This bonding is very important for correctly orienting the substrate ring. **Measure these O to N distances for your report.**

We explored the role of ion pairs previously, but lets note them again. Select hs R70 and lc R57. Show them as sticks, and notice how these positively charged side chains ion pair with the negatively charged carboxylate of the MTX "Glu tail".

Finally, we should note that the DHFR active site, like many, many enzyme active sites, is largely **hydrophobic** and shaped to fit the substrates. To get a feel for this lets display a few conserved nonpolar residues. Clear things up showing the ribbons and the 2 ndp molecules (sticks). Now select from hs I7, F31 and F34 and from lc select L4, L27, and F30; display them as sticks. Note how the side chains make a nonpolar nest for the pterin ring of MTX to rest on. To make this point more strongly, go to the sele line and Show them as **spheres**. Notice one side of the ring is "open", not cradled by hydrophobic residues; this is where the orienting carboxylic acid goes. To remind yourself, select E30 from human and show it as sticks.

One last object lesson: notice that the side chains which abut one another in space to form this pocket may be close in sequence number, like F31 and F34, but may include remote residues like Ile7. The complex chain folding can bring remote parts of the sequence close in space.

C. One last comparison. Clear the screen. Show human pro as a surface, color it blue. Also display mtx as sticks. If you toggle the name, DHFR-hs, the image comes and goes.

Now Show the L casei pro2 as a surface, color it red, and display mtx2 as sticks. You can toggle it too. When both are displayed the color differences show the subtle differences in the structure of the two proteins. They are basically similar, but there are many small differences. Put the mtx in the middle of the screen and zoom in a bit. **Make a picture or two for your report that you think makes this point**.

An interesting question for antibiotic design is "can I find, or build, a compound that will bind to the bacterial enzyme, but not to the human?". It doesn't look easy, but this species specific binding is a big part of modern structure based drug design. There is a compound, methotrexate (TRR) that binds about 3000 times more tightly to bacterial DFHR than to human. This allows it to be an effective antibiotic. It is not easy to understand exactly why the drug binds more strongly to bacteria, and the deep analysis is complicated, delving into the conformations elicited by binding NADPH and the drug.

Download the tiny file called TRR-k2.pdb. This file shows the binding of TRR to human and L. casei DHFR; it has been oriented into the DHFR-hs system so you don't need to go through another alignment to see the drug binding.

Clean up you screen and show the human and lc proteins as ribbons. Show the human ndp and mtx molecules as a wires to mark the folate binding site. Now open the TRR-k2 file to display the two modes of TRR binding.

You can see the TRR binds in the folate pocket but is smaller than the folates and replaces the bicyclic folate ring with a single 6-membered ring. You can see the orientation of TRR is subtlely different between the two species. There is no striking interaction to explain why TRR binds 3000 times more strongly to bacteria than to humans. The effect appears to arise from a number of small, subtle, interactions. To investigate some possible contributors, open the sequence selecting tool.
Select from DHFR-hs: D21, F31, F34, and P61
Select from dhfr-lc: H18, L27, F30, and P50
Display them as sticks.

As one example of how the the drug binding differs between species, note how the bulky F31 of the human enzyme forces the drug into a slightlly different orientation compared to the bacterial enzyme which has L27 at this place. This reorientation may contribute to the observed difference in binding strength.