Computer Lab 6 – Collective Motions and Dynamical Networks

In the typsin inhibitor simulation we concentrated on the conformational dynamics that are behind the average structure that we see as the X-ray crystal structural model. The amplitudes of conformational fluctuations were relatively small and were all part of the distribution describing a single native average structure. (One exception was LYS_15 that had different dynamics in the simulation of isolated inhibitor compared to the apparent dynamics when bound to trypsin in the X-ray complex structure.)

In contrast to these small amplitude, very rapid conformational fluctuations many proteins undergo slower, large backbone conformational fluctuations during their respective functions. Standard MD does not usually simulate for long enough times to observe these large conformational changes and if they are there they are difficult to visualize with all the other fluctuations going on.

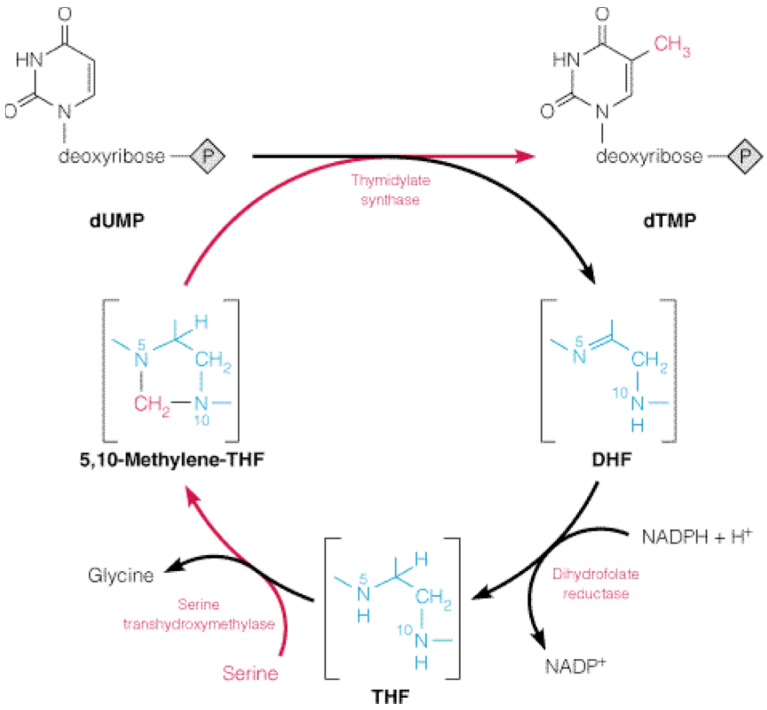
One way to obtain large collective motions is to perform what is called a normal mode simulation as described in lecture. A second way to obtain information on large scale collective motions is to analyze the packing of the average protein structure. From this packing, one can estimate the energetic penalty of moving a protein atom (or group of atoms) in either the x, y or z direction. A third way to obtain an indication of which segments of a protein will be involved in large or collective conformational change is to identify those atoms that move together. This means that the atoms will have correlated movement, or collective motion, and lead to concerted movement of the backbone chain. This third method is the focus of today's lab.

In this lab, you will analyze a trajectory of the protein dihydrofolate reductase (DHFR) to identify those segments involved in *correlated motion*. To help us visualize these segments, the protein will be modelled as a network of nodes connected by edges, each node representing an amino acid residue. Groups of residues that have correlated motion will be clustered into communities and the allosteric linkage between different communities will be investigated.

In its functional form dihydrofolate reductase (DHFR) contains two prosthetic groups, NADPH and folate. You will need a trajectory of only the apoenzyme without the prosthetic groups.

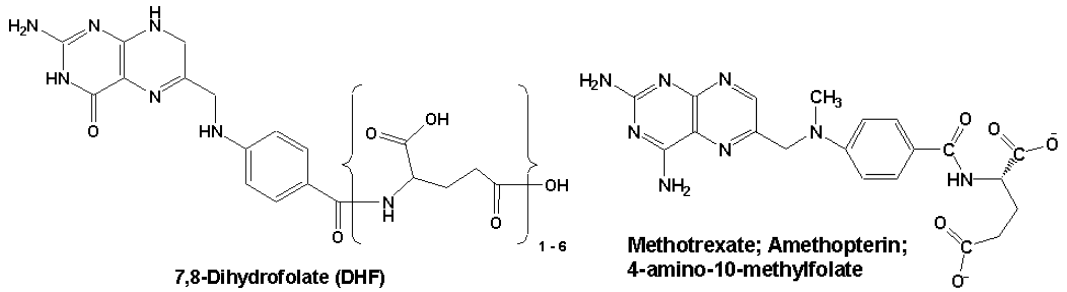
Answer the questions highlighted in yellow and send the answers to **achin14@jhu.edu** either in the body of an email or as an attached file with your JHEDID in the name (e.g. *JHEDID_lab5c.txt* if made with vi or *JHEDID_lab5c.docx* if made with MSWord).

DHFR Background. This enzyme is in the pathway that regenerates a small organic molecule necessary for the synthesis of DNA and RNA. The organic molecule is the vitamin folate that is present in fruits and vegetables in an unusable form called dihydrofolate (DHF). The figure below depicts the role of DHF in methyl group donation and the reactions involved in its regeneration.



Dihydrofolate reductase is an allosteric enzyme in the sense that binding of NADPH promotes binding of DHF and *vice versa*.

Since inhibiting DHFR stops cell growth it is a target for anti-cancer drugs such as methotrexate. From the structures below you can see the resemblance of DHF and methotrexate that allows methotrexate to bind to DHFR in the DHF binding site.



[Note: If you already have your trajectory in a *prod.dcd* file skip to part II below.

I. MD trajectory of apo-dihydrofolate reductase. In its functional form dihydrofolate reductase (DHFR) contains two prosthetic groups, NADPH and folate. You should have a trajectory of only the apoenzyme without the prosthetic groups. The instructions below were included at the end of the last lab and are here just to remind you what you should have done by now.

The DHFR PDB file *1RA2.pdb* is available in the */home/compbio2/Shared/* directory on the cluster or you can download it from the PDB (www.rcsb.org). On a Mac, make a *network/* directory under your permanent home directory, change there and put the file *1RA2.pdb*

there. To remove the prosthetic groups and waters you can grep out only the ATOM records to create a "clean" polypeptide PDB file. For example,

```
grep ^ATOM 1RA2.pdb > 1ra2_cln.pdb
```

This would be equivalent to the *inh.pdb* file in your previous simulation.

Copy over the *prot.pgn* and *top_all27_prot_lipid.inp* files from your *protsim/* directory to your *network/* directory. Edit the *prot.pgn* file appropriately (just change the file names) and make both the *start.psf* and *start.pdb* files as you did for lab 5.

Solvate, ionize and center the system as previously. Write down the minimum and maximum system coordinates so you can calculate the primary cell size for use in the **.conf* files.

On the cluster you should create a new directory called *network/* under your *compbio2/[JHEDID]* directory. Then gather the files needed to do an MD simulation exactly similiar to the trypsin inhibitor simulation you previously carried out. You will need the **.conf*, **.inp* and *run.namd* files from that exercise. These should be edited for the approriate file names, primary cell sizes, and PME grid sizes.

After the minimization and equilibration run, make a trajectory of 1000 frames as you did before. Fetch the final *prod.dcd* trajectory and the back to the Mac in the *network/* directory there.

Hint: After submitting your job to the queue watch the log file grow for about 5 minutes to make sure the job doesn't crash right away. This will give you time to make corrections and restart the simulation (and you will have a trajectory to analyze!). You can watch the log file grow with the following command,

```
tail –f min_equil.log
```

or

```
tail –f production.log
```

You can stop the **tail** session with ctrl-**c**.

II. Network Analysis of Atomic Collective Motions. Visualize a simple network as a set of nodes with edges connecting pairs of nodes. For this analysis, we define a node as an amino acid residue and use the CA atom as the center of each node. Then we draw an edge between nodes whose residues are within 4.5 Å for at least 75% of the MD trajectory. We then calculate the correlations of the fluctuations of nodes during the trajectory and use the correlation data to weight the edges. This is all done in the background when you call the program **networkSetup** from within VMD.

In your *network/* directory on the Mac create a file called *network.config* containing the following lines,

```
>Psf
ionized.psf

>Dcds
prod.dcd

>SystemSelection
(chain P) and (not hydrogen)

>NodeSelection
(name CA)

>Restrictions
notSameResidue
notNeighboringCAlpha
```

This file is a command file for the program **networkSetup** and tells it to read the *ionized.psf* and *prod.dcd* files, select the protein (P as opposed to nucleotide, N), make the CA atoms the centers of the nodes, and restrict the edges to two nodes that are not adjacent in the chain. You should have fetched the *prod.dcd* from the cluster back to the Mac. Now from the *network/* directory on the Mac fetch three files from */home/compbio2/Shared/* on the cluster: **carma**, **gncommunities**, **subopt, catdcd**. Then in the *network/* directory on the Mac enter the following,

vmd
vmd> **networkSetup network.config** (this will take a minute)
VMD Main | Display | Orthographic
VMD Main | Slide the frame selection slider all the way left to frame zero
VMD Main | Extensions | Analysis | RMSD Trajectory Tool
    Selection Modifiers | Backbone
    Leave the Reference mol | Top button activated
    Click ALIGN (Close the RMSD Trajectory Tool) (This aligns the protein molecule in all trajectory frames to the protein molecule in frame 1).
VMD Main | Graphics | Colors | Display | Background | 8 white
VMD Main | Graphics Representations...
    Selected Molecule | 0: ionized.psf (Double click the Lines rep to deactivate)
    Selected Molecule | 1: adjancency.psf
        Selected Atoms | **protein** <enter return>
        Drawing Method | New Cartoon

You should see a ribbon drawing of DHFR. You will display the network below.
One of the output files from the networkSetup program is called, *contact.dat* and is in your pwd. Now you will use that node contact data to run a program that clusters the nodes into communities as follows,

vmd> **gncommunities contact.dat communities.out**

And now view the network,

VMD Main | Extensions | Analysis | NetworkView
File | Load Network | contact.dat
File | Load Community Data | communities.out
Display Parameters | Edge Size | weight
Click Apply
Click Draw

You should see the CA atoms as nodes in a network with the edge thickness proportional to the correlation between nodes. The network is overlaid on the ribbon drawing. 1. Are residues linked by backbone hydrogen bonds highly correlated? It appears that residues *i* are correlated to *i*+2. 2. Why don't we see high correlation between residues *i* and *i*+1? Turn off edges.

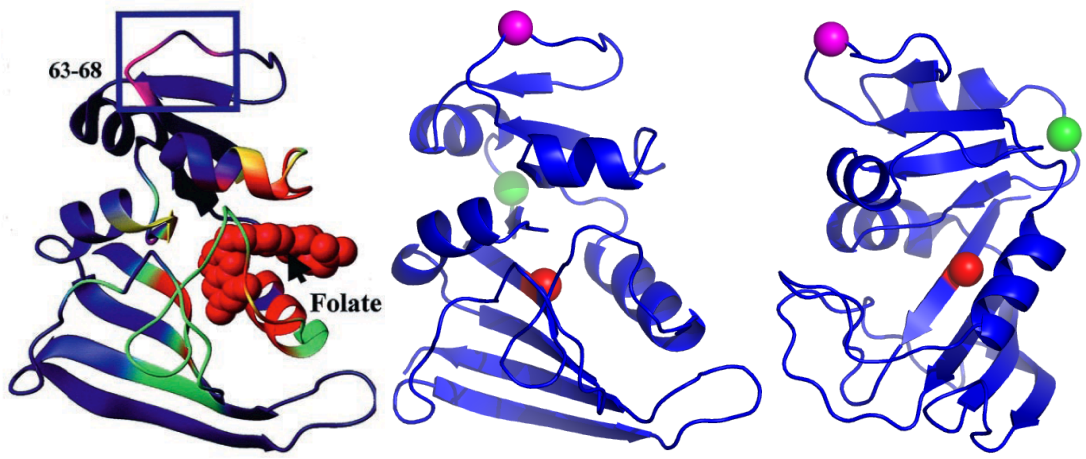
Display Parameters | Network Fragment | edges
Action | deactivate
Click Apply
Click Draw

Now color the communities (clusters) of correlated nodes,

Display Parameters | Network Fragment | network
Node Selection | Community | All
Action | Color Communities
Click Apply
Click Draw

Now you see communities (clusters of nodes) of similar colors that have correlated motion. You should have 5-10 communities. 3. Does each secondary structural element (helix, sheet) have its own unique cluster? 4. What other results did you obtain in this class that showed heterogeneity of secondary structure dynamics? Run the movie and observe the underlying motion that gave rise to these correlated clusters of motions. Kind of hard to see how they relate! Stop the movie and put the frame slider back to frame zero.

In the figure below I have identified a residue in each of three segments that exhibited cooperativity in the study we discussed in lecture. The single red sphere (resid 5) is the main binding interaction with folate; the green sphere (resid 88) was positively cooperative with residue 5; and the magenta sphere (resid 66) was negatively cooperative with residue 5.



5. Are these three residues in different communities in your VMD display?

Hint:

VMD Main | Graphics Representations...
- Selected Molecule | 1:adjacency.psf
- Create Rep
- Selected Atoms | **`name CA and (resid 5 or resid 66 or resid 88)`** <enter return>
- Drawing Method | VDW
- Sphere Scale | **`1.5`**
- Material | Transparent

Let's calculate and display the linkage paths between residue 5 and each of these other residues. You have to identify the community node number associated with a residue to proceed. Use the following commands to get the community node numbers of these residues in your network,

```
vmd > ::NetworkView::getNodesFromSelection "chain P and resid 5"
vmd > ::NetworkView::getNodesFromSelection "chain P and resid 88"
vmd > ::NetworkView::getNodesFromSelection "chain P and resid 66"
```

Make a note of the community numbers returned. Calculate the linkage paths for your system by substituting the community numbers you obtained above in the following command (just the numbers obtained above, with no bracket),

```
vmd > subopt contact.dat path 20 [resid 5 node] [resid 88 node]
```

Load the path data,

Extensions | Analysis | NetworkView
- File | Load Suboptimal Path Data | path.out
- Display Parameters | Edge Size | global
- Node Selection | Suboptimal Path | All
- Action | Activate
- Click Apply
- Click Draw

You should see the linkage path(s) between these two residues as thick sticks. This path represents the covalent and/or non-covalent bonds that allow communication between the two residues chosen (5 and 88 in this case). 6. Did you find only covalent, only non-covalent, or both types of connection between correlated residues? Residues along these paths are likely to be important for allostery between the two residues – mutations along these paths may render the protein non-functional even though these residues are not in the binding sites.

Now do the same for the path between residue 5 and residue 66.

vmd > **subopt contact.dat path 20 [resid 5 node] [resid 66 node]**

Read the messages in the terminal window. Did the program find a path? If the network finds no path between these residues try between residue 5 and residue 67 or 65. (You have to find the node number of these residues as above).
Load the path data,

Extensions | Analysis | NetworkView
File | Load Suboptimal Path Data | path.out
Display Parameters | Edge Size | global
Node Selection | Suboptimal Path | All
Action | Activate
Click Apply
Click Draw

This is the path of linkage between the main folate binding interaction and the loop at the top of the protein. 7. Did you find only covalent, only non-covalent, or both types of connection between correlated residues? Compare with your neighbor - did they get the same path? Let the instructor know if yours was different from your neighbor.

