CHEM 6572
Assignment 6
Manipulating PDB files and Analysis of Phi and Psi, thermal factors, etc

For this assignment, quickly read over the Wikipedia page for MHETase (pdb entry = 6QZ2, https://en.wikipedia.org/wiki/MHETase. We are not terribly interested in the function of the protein but are working with it primarily because it is large and complex.

Go to the PDB entry for this protein.

What is the resolution of this structure?
From what species is this protein derived?
In what organism was this protein expressed?

Set up Pymol as previously by copying and pasting the following into the command line

```
#--------------------------------------------------------------------------
## Header: General Commands ##

# delete all objects and reset pymol
reinitialize
# set the background color to white
bg_color white
# make the background transparent for ray trace
set ray_opaque_background, 0
# set the ray trace mode
# normal color
set ray_trace_mode, 0
# normal color + black outline
#set ray_trace_mode,  1
# black outline only
#set ray_trace_mode,  2
# turn off shadows during ray trace
set ray_shadows, 0
# set the mouse mode for laptop.
config_mouse one_button
# get rid of double bonds and skinny bonds to H
set stick_h_scale, 1
set valence, 0
# high quality surfaces
set surface_quality, 3
## End of Header: General Commands ##
#---------------------------------------
```

Open the coordinate file (6QZ2) in pymol using fetch or load.

This coordinate file has multiple copies of the protein. Remove all of them except one. Also remove the water molecules. Use:

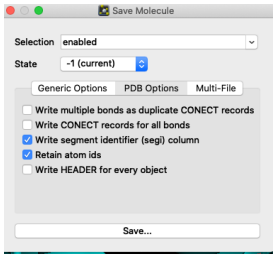
```
remove 6qz2 and not chain I
remove resname HOH
center chain I
```

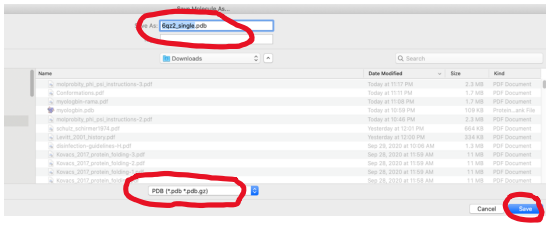
With these commands you have removed all the water molecules, all copies of the protein except chain I, and have centered the display on chain I

Make an image of the protein. Do not forget to ray trace before you save the image. How would describe the global fold? Make a topology diagram of the b-sheet. This is complex protein, so omit the alpha helices (pretend they are parts of loops).

Save the coordinate file of your protein to your disk.

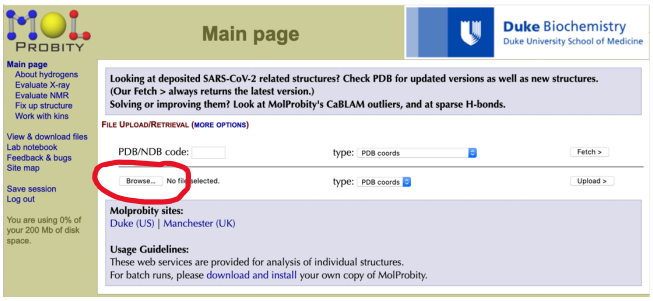
File→export molecule





With this you have created a coordinate file with a single copy of the protein. Only chain I is contained in the coordinate file.

Use Molprobity to generate a Ramachandran plot of your single protein (http://molprobity.biochem.duke.edu/). Follow the instructions posted previously on the course page (https://ww2.chemistry.gatech.edu/~lw26/course_Information/6572/misc/molprobity_phi_psi_instructions.pdf), with the exception that you should use the browse function to upload the coordinate file that you have saved to your disk.



Run "analyze geometry without all-atom contacts", using the following parameters.

Choose the outputs you want:

Default options have been selected based on the content of the submitted file.
Follow the symbols for more information on the validation options.

- ☐ **3-D kinemage graphics**
- ☑ **Charts, plots, and tables**
  - **Universal**
    - ☐ Clashes & clashscore
    - ☑ Geometry evaluation
  - **Protein**
    - ☑ Ramachandran plots
    - ☑ Rotamer evaluation
    - ☑ Cβ deviations
    - ☑ Cis-Peptide evaluation
      - ☐ Show cis-nonPro and twisted peptide statistics even if the model has none
    - ☑ CaBLAM backbone evaluation
  - **RNA**
    - ☐ RNA sugar pucker analysis
    - ☐ RNA backbone conformations
  - **Other options**
    - ☐ Horizontal chart with real-space correlation data
    - ☐ Chart for use with Coot (may take a long time, but should take less than 1 hour)
    - ☐ Suggest / report on automatic structure fix-ups
    - ☑ Create html version of multi-chart
      - ☑ List all residues in multi-chart, not just outliers
      - ☑ Remove residue rows with ' ' altloc when other alternate(s) present

Compare the Ramachandran plot to that shown in Figure 1c of Hovmöller et al., *Acta Cryst*. 2002 (shown on the next page).
Does the 6QZ2 Ramachandran plot look reasonable?

Look carefully at the "Ramachandran distribution Z-score analysis" and find the worst phi/psi pairs (highest z scores).
Create pymol objects for at least two of those residues. For each object include several resides on each side of the residue of interest.
Use:

create myobject, resi 80-88 (but with different object name and resi numbers than I have used here)

Save images of these objects and pull them into your word doc.

Are these residues outside the allowed regions?
Mark them on the Ramachandran plot of your protein.

Look at the caBLAM analysis and find residues with high B-factors.
Create pymol objects with at least two of those residues, include several resides on each side.
Include images of these objects in your document.
What do you notice about the residues with high B-factors?

Look at the caBLAM analysis and find one of the cis peptides.
Create a pymol object with this residue and include several resides on each side.
Include a images of this object in your document.

Save the pymol session file.

File → save session (as pse)

Upload the document file, including the narrative and the images. Also upload the session file.

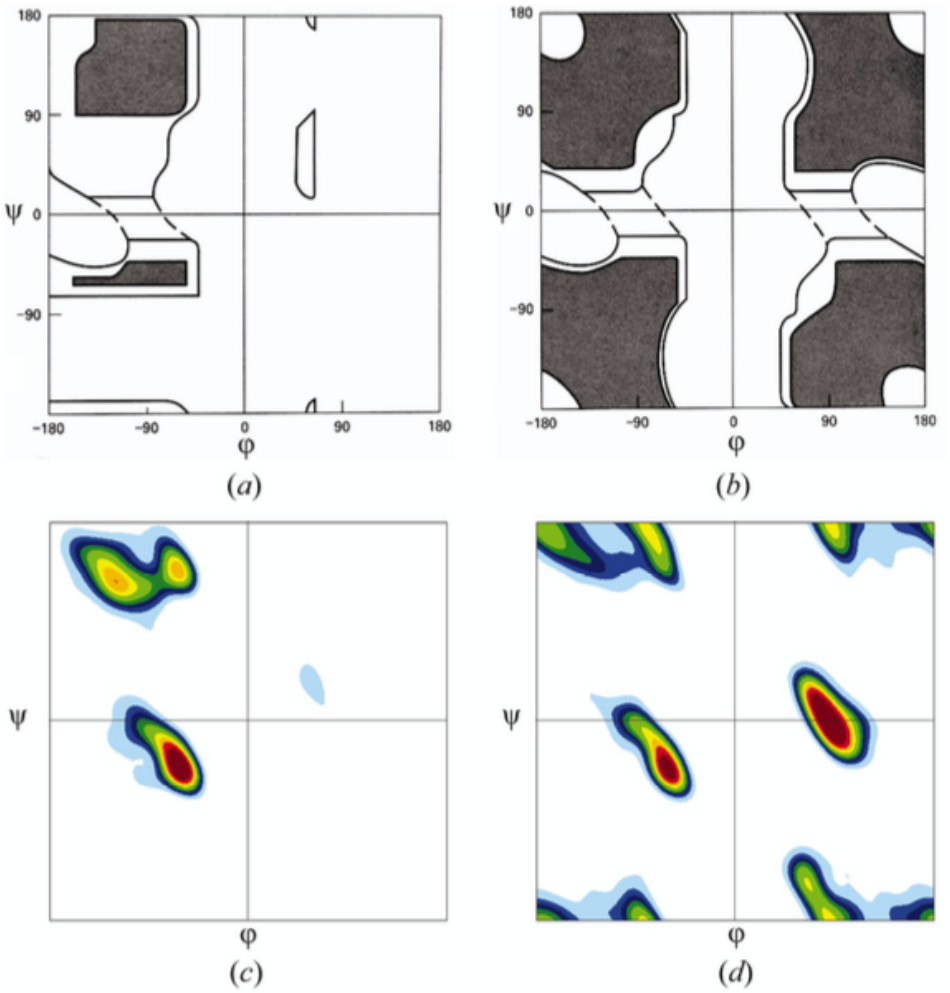


Figure 1
The classical version of the Ramachandran plot for (*a*) alanine (but often taken as typical for all non-glycines) and (*b*) glycine according to Ramachandran & Sasisekharan (1968). The fully allowed regions are shaded; the partially allowed regions are enclosed by a solid line. The connecting regions enclosed by the dashed lines are permissible with slight flexibility of bond angles. These plots were arrived at by computer modelling. Although some overall features of these plots are correct, the details differ from the experimentally observed Ramachandran plots for alanine (see Fig. 5) and (*c*) all 19 non-glycines and (*d*) glycine. The most remarkable differences are that most regions show a 45° slope rather than being parallel to any of the axes, the β-sheet region is split into two distinct maxima and the two most populated regions for glycine seen in (*d*) were predicted to be only just permissible as shown in (*b*). There are five areas in the glycine plot; two with $\psi \simeq 0°$ and three with $\psi \simeq 180°$. [(*a*) and (*b*) Reproduced from Creighton (1996) with permission.]