BMB 173 – Winter 2017
Homework Set 6 (200 points) – Assigned 2/16/17. Due 2/23/17 Revision 1

TA: Emily Wyatt, ewyatt@caltech.edu
**Office Hours:**
Tuesday   2/21/17 from 1-2pm in Spalding B123 or by appointment
Wednesday   2/22/17 from 11am-12pm in Spalding B123 or by appointment

1. **Solving the Phase Problem** (80 points)

   In this problem set, you will use isomorphous replacement to determine possible phases of a single X-ray reflection from a protein crystal. For simplicity, you will only consider a two-dimensional lattice and the unit cell will not have any symmetry operators.

   You have collected the structure factor amplitudes of a two-dimensional protein crystal. You have also collected the structure factor amplitudes of a derivative crystal containing three mercury atoms bound to the protein. The Patterson map shown below is a difference map between the structure factor amplitudes for the mercury derivative, $\left|\mathbf{F}_{\mathrm{PHg}}(hk)\right|$, and the native protein crystal, $|\mathbf{F}_{\mathrm{P}}(hk)|$, for the reflection at $h = 5, k = 6$ in the crystal diffraction pattern. In other words, it is a map of $\left|\mathbf{F}_{\mathrm{PHg}}(5\ 6)\right| - |\mathbf{F}_{\mathrm{P}}(5\ 6)|$. Assume that the mercury atoms do not perturb the protein structure, so that the peaks correspond only to mercury inter-atomic vectors.
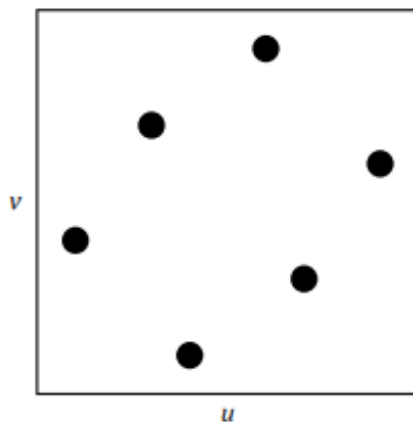
   The coordinates of the six peaks on the map are:
   - $u = 0.3 , v = 0.7$
   - $u = 0.9 , v = 0.6$
   - $u = 0.7 , v = 0.3$
   - $u = 0.4 , v = 0.1$
   - $u = 0.1 , v = 0.4$
   - $u = 0.6 , v = 0.9$

a. (20 points) Following the steps below use this Patterson map to calculate the real space positions of all three mercury atoms in terms of x and y, the fractional unit cell coordinates.

Things to know about Patterson maps:

- If an atom a_1 is present at the position (x, y) in real space, and a second atom a_2 is present at position $(x + u, y + v)$, then in the Patterson function there will be a peak at (u, v), representing the vector $a_2 - a_1$, and a second peak at $(-u, -v)$, representing the vector $a_1 - a_2$.

- Everything is considered in fractional coordinates. For example, a value of -0.3 is equal to 0.7 in an adjacent unit cell.

How to solve the Patterson map:

- There are no symmetry operators in this problem, so there are no peaks between symmetry related atoms. The choice of origin is arbitrary, which means the position of the first atom, a_1, can be set to $x = 0, y = 0$.

- Designate a peak on the Patterson map to be the vector between the first and second atom $(a_2 - a_1)$. Note that the choice of atoms 1, 2, and 3 is also arbitrary. For this assignment, let $(a_2 - a_1) \equiv u = 0.3, v = 0.7$.

- If a_1 is at $x = 0, y = 0$, then a_2 is at $x = u, y = v$. Then, from the remaining peaks, determine which one represents the vector from the second atom back to the first $(a_1 - a_2)$.

- There are four peaks remaining: two relating the first atom to the third atom $(a_3 - a_1$ and $a_1 - a_3)$ and two relating the second atom to the third atom $(a_3 - a_2$ and $a_2 - a_3)$. Guess and check to assign the remaining peaks. To do this, guess which peak corresponds to the first and third atom vector $(a_3 - a_1)$, and then predict the coordinates for the third to first atom vector $(a_1 - a_3)$. If no peak is present at the predicted location, guess again.

- Repeat this process for the second to third $(a_3 - a_2)$ and third to second $(a_2 - a_3)$ vectors.

- The model is complete when all Patterson peaks are accounted for. Please show your assignments for what each peak corresponds to $(a_3 - a_1, a_2 - a_3$, etc).

Make sure to show all work for full credit. Box your assignments for Patterson peaks as well as the positions of the atoms in real space that you calculate.

b. (20 points) Calculate the amplitude and phase of $|F_{Hg}(5\ 6)|$, the scattering from the heavy atom substructure alone. For this reflection, suppose the atomic scattering magnitude of mercury (f_j) is 70. You can use the equation below to calculate scattering from j number of atoms,

$$F(hk) = \sum_j f_j e^{2\pi i (hx_j + ky_j)}$$

2

where f_j is the atomic scattering magnitude, h and k are the Miller indices (5,6) and x and y will be the real space positions determined in part 'a'. To add up the exponential terms, you can use the following relationship (in radians!):

$$re^{i\theta} = r(\cos\theta + i\sin\theta)$$

Your sum will be a vector in the form of $x + iy$. To solve for the amplitude and phase, plot this vector on a graph where the x-axis represents the real component and the y-axis represents the imaginary component. The phase of $\mathbf{F}(hk)$ is the angle between the x-axis and the vector, and the amplitude of $\mathbf{F}(hk)$ is the length of the vector.

c. (20 points) The amplitude of the native structure factor at $h = 5$, $k = 6$ is given by $|F_P(5\,6)| = 66.57$. The mercury derivative amplitude is $|F_{PHg}(5\,6)| = 60$. Draw a Harker construct to demonstrate the two possible phase angles for the $h = 5$, $k = 6$ reflection. Calculate the two possible angles using the following equation, where α is the phase angle in radians:

$$\alpha = \alpha_{Hg} \pm \cos^{-1}\left[\left(\left|F_{PHg}\right|^2 - |F_P|^2 - \left|F_{Hg}\right|^2\right)/(2|F_P||F_{Hg}|)\right]$$

Your Harker construct should illustrate $\mathbf{F_P}$, the structure factor of the native protein, $\mathbf{F_{Hg}}$, the structure factor of the mercury atoms, and $\mathbf{F_{PHg}}$, the structure factor of the mercury derivative protein.

d. (10 points) In part 'c', you solved for two possible phase angles, represented in your Harker diagram as the intersection of two circles. What additional experiment can you do to determine which of these two phases is correct?

e. (10 points) While you are solving the structure of your protein, the structure of a homolog is published. The homolog has 90% similarity to your protein's monomeric structure. What phasing technique could be performed with this information to solve the structure of your protein?

2. **Mosaicity** (20 points)

What is crystal mosaicity? How does crystal mosaicity affect a diffraction pattern? In brief, explain why mosaicity typically affects the collection of higher resolution reflection more than low resolution reflections.

3. **Fitting Protein Segment to Electron Density Map** (80 points)

In this problem, you will use Coot to fit a protein segment to an electron density map.

Installation:

Mac (10.6 or higher):
Go to http://scottlab.ucsc.edu/xtal/wiki/index.php/Stand-Alone_Coot get the standalone precompiled Coot version that is compatible with your version of OS X.

Windows:
Go to http://www.ysbl.york.ac.uk/~lohkamp/coot/wincoot-download.html and download the .exe file under latest stable release WinCoot-0.8.2.The console may give several warnings, but these can be ignored for this problem set.

Navigating in Coot:

Detailed instructions on using Coot are available at
https://www2.mrc-lmb.cam.ac.uk/Personal/pemsley/coot/web/docs/coot.html

Once a PDB file has been opened, the most common navigation methods are:

Left-mouse drag	Rotate view
Ctrl + left-mouse	Translate view
Shift + left-mouse	Label atom
Right-mouse drag	Zoom in and out
Ctrl + shift + right-mouse drag	Rotate view around screen
Middle-mouse	Center on atom
Scroll wheel	Increase and decrease map contour level

Coot can directly download coordinates and electron density maps from the PDB (Note: If this does not work for you, please just proceed to the problem set.) View the X-ray crystal structure of GFP to familiarize yourself with the controls:

File ➔ Get PDB & Map using EDS. Enter 1GFL and click 'Get it'. Wait a few seconds for it to download. You will be presented with coordinates and two maps, a 2|Fo| - |Fc| and a |Fo| - |Fc| map. Hide the |Fo| - |Fc| map.

Draw ➔ Display Manager. Uncheck the display box for the second map, labeled FOFCWT. Try zooming in and out, centering on various atoms, and changing the contour level of the map. The contour level sets the minimum electron density that must be present at an xyz point for the gridded map to be drawn around it. The sigma value indicated in the top right corner as you change contour level is signal over noise, so be careful interpreting maps with sigma less than 1.

Model Building:

Before proceeding, delete the GFP map and coordinates:

File ➔ Close Molecule/Map. Select everything and click 'Delete Marked Molecules & Maps'.

Download the electron density map. You are given a 2.0 Å resolution electron density map (Download BMB173_2017_PS6.zip from http://saf.bio.caltech.edu/bmb173/. Extract the files somewhere on your computer to use for this problem set). It was generated using experimental intensities and calculated phases from a refined model.

File ➔ Open Map. Select map.map and click OK.

Edit ➔ Map Parameters. Change Map radius from 10.0 to 20.0 and click OK.

File ➔ Open Coordinates. Select center.pdb, make sure 'Recentre on Molecule' is selected, and click OK.

File ➔ Close Molecule/Map. Check center.pdb and click Delete Marked Molecules and Images.

Zoom in and out of the fragment to familiarize yourself with its shape. Increase the contour level until you can clearly identify the chain. This region is a 21-residue sequence of an α- helix and ß-strand that is located within the following sequence:

MTLGLDLKGGLRIMAAGDVFRAAASNELQIWAARTGCEIVVIRQPNDQYQF

You will need to determine its identity. You will start building your structure from an idealized polyalanine helix. Note: In previous versions of Coot, the program could automatically recognize helical density and build a polyalanine helix into it. This doesn't always work, so you have been provided with a helix.

File ➔ Open Coordinates. Select helix.pdb, make sure 'Recenter' is checked, and click OK. There will now be a helix of 10 alanines modeled into the density. At this point you will need to identify which range in the sequence given above that this density represents. Look at the side chain density. There is at least one unambiguous side chain present in the helix. As you identify residues, change them in your model.

Calculate ➔ Model/Fit/Refine. This will bring up a tool box. Select Mutate & Autofit and then click on a residue (sometimes it needs to be clicked twice) and select the new amino acid from the list that appears. Mutate all the residues in the helix as you identify them. Afterwards, you will refine the model.

Calculate ➔ Model/Fit/Refine. Select Real Space Refine Zone. Click on a residue and then a second residue 4–5 amino acids down the chain. Check the presented model and

make any changes if necessary by clicking on and pulling the white model. Click OK when you are done.

Calculate ➔ Model/Fit/Refine. Select Regularize Zone. Click on a range of residues as before to idealize geometry.

Use these two tools to refine the entire helix. All the side chains should be modeled into the density, and there should be no major errors in geometry.

Now build the connecting ß-strand. Make sure that the side chain of the last residue of your helix is in side chain density and not main chain density. Change your model if required.

Calculate ➔ Model/Fit/Refine. Select Add Terminal Residue. Click the last residue of your helix going towards the sheet.

A white alanine residue will appear in the window. You can position the residue by clicking and dragging, and when you are happy with its location hit **OK** in the pop-up window. Use the **Rotate/Translate** command in the **Model/Fit/Refine** window to roughly orient the residue so that the side chain is pointing in the right direction, and not into main chain density. **Rotate/Translate** brings a window with sliders that you can use to move and twist the residue. You can also move the residue by clicking and dragging. Use the **Mutate & Autofit** command to change the alanine to the appropriate residue, and use the Real Space Refine Zone and Regularize Zone tools to fit. Complete the beta strand.

There are also two more residues at the other end of the helix. Be sure to build those too.

Validating the Model:

To check your work, you can evaluate the geometry of your building by examining the Ramachandran plot:

Validate ➔ Ramachandran Plot. Select the pdb file for your helix. Coot will automatically generate a plot. Check to make sure your amino acids have phi and psi angles that fall in the preferred regions of the map. Amino acids that fall in the allowed regions are acceptable, but outliers should be fixed. The map will expand to allow more phi and psi angles if the cursor is pointed at a glycine. Conversely, the map will contract to exclude certain phi and psi angles if a proline is selected (there are no prolines in this sequence).

Coot also features tools that evaluate how well the model fits the density:

Validate ➔ Density Fit Analysis

The resulting graph will plot the errors in density fit. Manipulate your structure to minimize errors in the density fit.

You will be evaluated on the quality of your model's primary sequence, Ramachandran plot, and Density Fit Analysis (these graphs can be generated using only your submitted pdb file and the density map, so there is no need to turn these in). When you are finished with your structure, go **to File >> Save Coordinates**, click on **Select Filename**, and save your model as your firstnamelastname.pdb (I would be emilywyatt.pdb).

Email me your coordinate file (ewyatt@caltech.edu).

4. **Validating and Interpreting a Model** (20 points)
   a. (10 points) What is the R-factor? How is it calculated?

   b. (10 points) Why is R_{free} superior to the R-factor? What would be the approximate target R_{free} value during refinement for a 2.5 Å crystal structure?