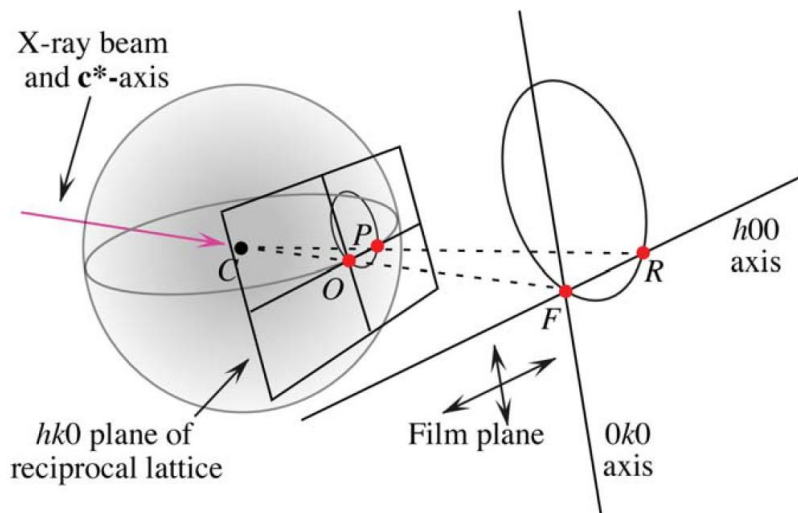


1. Determining Distances (5 points)

During office hours today, a wise student pointed out that in a given crystallography experiment we start out without knowing how big the unit cell is, and we need this to know the inter-plane spacing for any (hkl) planes. It turns out that you can figure this out with Ancient Greece-level science (a solid understanding of triangles, a ruler, and an X-ray home source). So, let's figure out the length of one dimension in our hypothetical crystal. Below is a diagram of the Ewald Sphere (now in its full spherical glory) from Crystallography made Crystal Clear (Gale Rhodes, 2006).



Assume the following conditions:

- $\lambda = 1.5$ Angstroms
- F is the spot 000 (beam stop)
- R is the spot 100
- The distance from the center of the crystal to the film plane is 1 m
- The spacing between the beam stop and point R is 1 mm

Hints:

- What is the radius of the Ewald sphere?
- If the plane shown in contact with the sphere at point P is a reciprocal lattice plane, then P must be the reciprocal lattice point with the same index as R on the diffraction pattern
- How many times do the 100 Bragg planes cut the unit cell along the horizontal axis?
- What is the relationship between RF and PO?

RF = the distance from the beam stop to reflection 100 = 10^{-3} m.

CF = the distance from the beam stop to the crystal center = 1 m.

Because P is the lattice point corresponding to reflection 100, then the distance PO is $1/d$ for this plane

Because the radius of the Ewald sphere is $1/\lambda$, the distance CO is $1/\lambda$

Triangles CPO and CRF are similar triangles so:

$$\frac{RF}{CF} = \frac{PO}{CO} = \frac{PO}{\frac{1}{\lambda}} = PO * \lambda = \frac{1}{d} * \lambda = \frac{\lambda}{d}$$

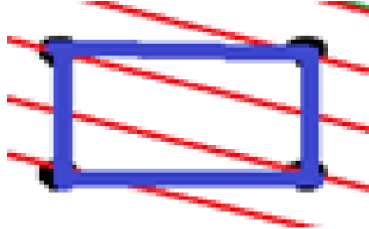
$$d = \lambda * \frac{CF}{RF} = (1.5 * 10^{-10} \text{ m}) * \frac{1 \text{ m}}{10^{-3}} = 1.5 * 10^{-7} \text{ m} = 1500 \text{ \AA} = 150 \text{ nm}$$

2. Ruminating upon Reconstruction

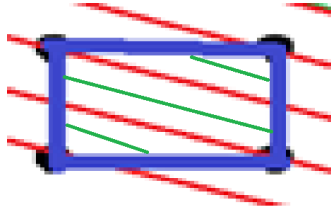
- a. When we're looking at a diffraction pattern to perform the Fourier summation in order to reconstruct the electron density, how do we know how much relative density is produced by each wave? How do we know what the frequency of each wave is? (5 points).

The relative density is determined by the wave's structure factor, or equivalently the relative darkness of the spot on the detector. The frequency in each direction is determined by the (hkl) index of the spot (just like each pixel location in a 2D or 3D Fourier transform determines the frequency of its representative wave).

- b. If we say that the reflection (120) has a phase of π (180°), what does it indicate about the location of the contributing electron density? On the unit cell diagram below, draw the position of the set of atoms that contribute to the density and reflection. Note that you don't know where each individual atom is; only the plane which they contribute to. (5 points).



The Bragg planes (120) occupy a given spacing along the unit cell. However, the atoms which physically have the same spacing in the crystal don't actually have to sit on these planes. The phase describes the offset of the atomic positions from the "ideal" case where they sit on the Bragg plane ($\theta = 0$). A phase of π indicates that the contributing atoms actually sit half way between the corresponding ideal Bragg planes.



3. Solving the Phase Problem (35 points)

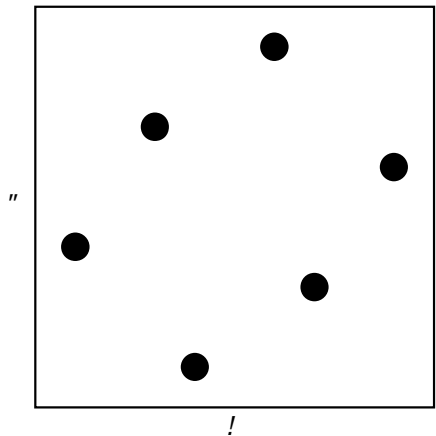
In this problem, you will use isomorphous replacement to determine possible phases of a single X-ray reflection from a protein crystal. For simplicity, you will only consider a two-dimensional lattice and the unit cell will not have any symmetry operators.

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, $|\mathbf{F}_{PH}|$, and the native protein crystal, $|\mathbf{F}_P|$. In other words, it is a map of $|\mathbf{F}_{PH}| - |\mathbf{F}_P|$. Assume that the mercury atoms do not perturb the protein structure, so that the peaks correspond only to mercury inter-atomic vectors.

- a) Why must the extra atom(s) for isomorphous replacement be heavy atoms? What must they not do to the protein structure? (2 points).

The extra atom(s) only appear at one or a few places in the unit cell. They must therefore diffract strongly to produce a detectable change in the intensities of the diffraction pattern. Since diffraction intensity depends on the size of the electron cloud, heavy atoms are ideal for this purpose. However, the heavy atoms must not change the structure of the protein or else of course the new diffraction pattern is incompatible with the original – hence, the protein-metal crystal must be isomorphous with the original protein crystal.

- b) (10 points)



The coordinates of the six peaks on the map are:

- $u=0.3, v=0.7$
- $u=0.9, v=0.6$
- $u=0.7, v=0.3$
- $u=0.4, v=0.1$

- $u=0.1, v=0.4$
- $u=0.6, v=0.9$

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 $x = 0, y = 0$. Note that the choice of atoms 1, 2, and 3 is also arbitrary. The work you show will depend on your starting peak, but you will get the same three atomic positions regardless of your choice.
- Designate a peak on the Patterson map to be the vector between the first and second atom ($a_2 - a_1$). 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. 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).

There are many solutions to this Patterson map. They all describe an equivalent lattice (or its enantiomer) and differ by which atom is treated as the origin.

Some possible solutions:

<u>Atom 1 (x y)</u>	<u>Atom 2 (x y)</u>	<u>Atom 3 (x y)</u>
0 , 0	0.3 , 0.7	0.4 , 0.1
0 , 0	0.3 , 0.7	0.9 , 0.6
0 , 0	0.7 , 0.3	0.1 , 0.4
0 , 0	0.4 , 0.1	0.1 , 0.4
0 , 0	0.1 , 0.4	0.4 , 0.1

- c) (8 points) Calculate the amplitude and phase of $\mathbf{F}_{\text{Hg}}(5\ 6)$, the scattering from the heavy atom substructure alone for the reflection at $h = 5$, $k = 6$ in the crystal

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

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

diffraction pattern. 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, where f_j is the atomic scattering magnitude (70), 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!):

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.

Each solution from part A will result in the same calculated value of the heavy atom structure factor magnitude, but they will have different phases. Full credit is given based on consistency.

Example:

$$\mathbf{F}(\mathbf{hk}) = f_j \exp[2\pi i(hx_1 + ky_1)] + f_j \exp[2\pi i(hx_2 + ky_2)] + f_j \exp[2\pi i(hx_3 + ky_3)]$$

$$\mathbf{F}(5\ 6) = 70 \cos[2\pi(5x_1 + 6y_1)] + 70 \cos[2\pi(5x_2 + 6y_2)] + 70 \cos[2\pi(5x_3 + 6y_3)] \\ + 70i \sin[2\pi(5x_1 + 6y_1)] + 70i \sin[2\pi(5x_2 + 6y_2)] + 70i \sin[2\pi(5x_3 + 6y_3)]$$

For the first solution from part A:

$$\mathbf{F}(5\ 6) = 70 \cos[2\pi(5 \times 0 + 6 \times 0)] + 70 \cos[2\pi(5 \times 0.3 + 6 \times 0.7)] + 70 \cos[2\pi(5 \times 0.4 + 6 \times 0.1)] \\ + 70i \sin[2\pi(5 \times 0 + 6 \times 0)] + 70i \sin[2\pi(5 \times 0.3 + 6 \times 0.7)] + 70i \sin[2\pi(5 \times 0.4 + 6 \times 0.1)]$$

$$\mathbf{F}(5\ 6) = -8.26 - 107.7i$$

$$|\mathbf{F}(5\ 6)|^2 = (-8.26)^2 + (107.7)^2$$

$$|\mathbf{F}(5\ 6)| = 108$$

$$\alpha(5\ 6) = \tan^{-1}(107.7/8.26) + \pi = 4.64 \text{ radians}$$

- d) (7 points) The amplitude of the native structure factor at $h = 5$, $k = 6$ is given by $|\mathbf{F}_P(5\ 6)| = 66.57$. The mercury derivative amplitude is $|\mathbf{F}_{PH}(5\ 6)| = 60$. 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:

$$\alpha = \alpha_{Hg} \pm \cos^{-1} [(|\mathbf{F}_{PH}|^2 - |\mathbf{F}_P|^2 - |\mathbf{F}_{Hg}|^2) / (2|\mathbf{F}_P|(|\mathbf{F}_{Hg}|))]$$

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 $\mathbf{F}_{PH}$, the structure factor of the mercury derivative protein.

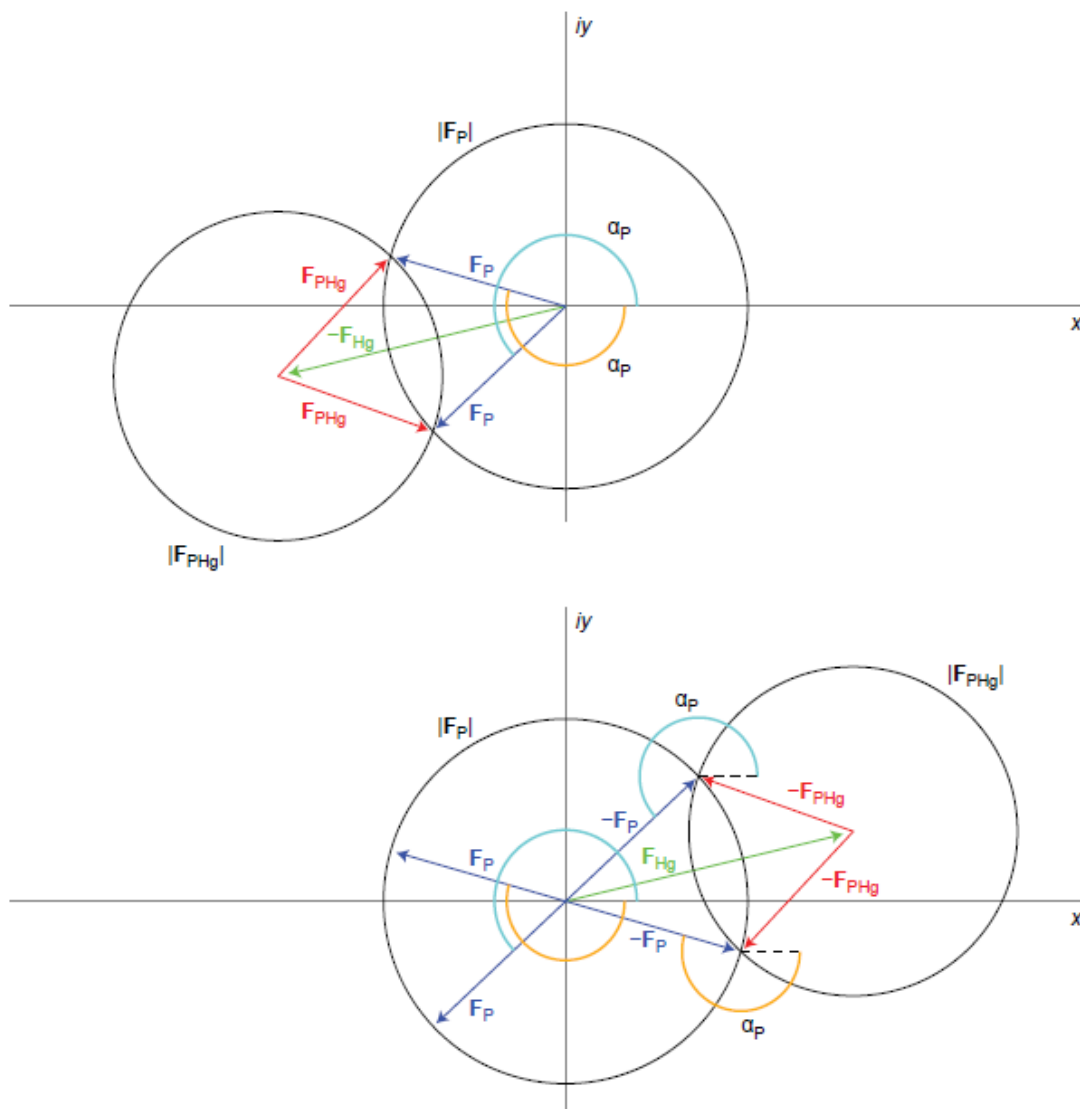
Example

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

$$\alpha_p = 4.64 \pm \cos^{-1}[(60^2 - 66.57^2 - 108^2) / (2 \times 66.57 \times 108)]$$

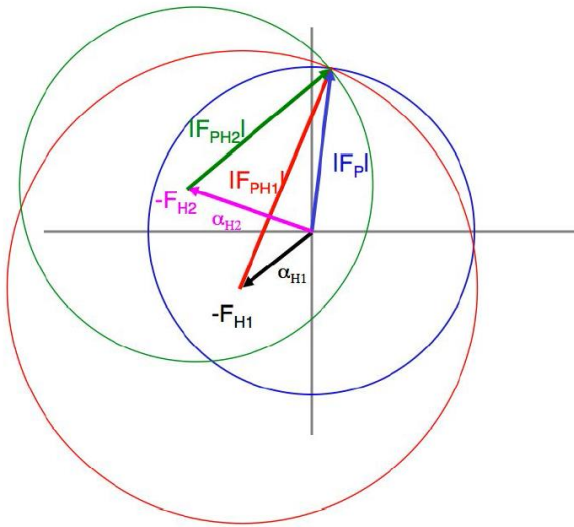
$$\alpha_p = 4.64 \pm 2.62 = 7.26 \text{ or } 2.02$$

The geometric meaning of the Harker construct depends on whether you plot $+F_{Hg}$ or $-F_{Hg}$. The following diagrams are drawn to scale for the example solution.



- e) (3 points) In part c, you solved for two possible phase angles, represented in your Harker construct as the intersection of two circles. Draw a Harker construct that represents the resolution of this phase ambiguity by using a second heavy atom derivative.

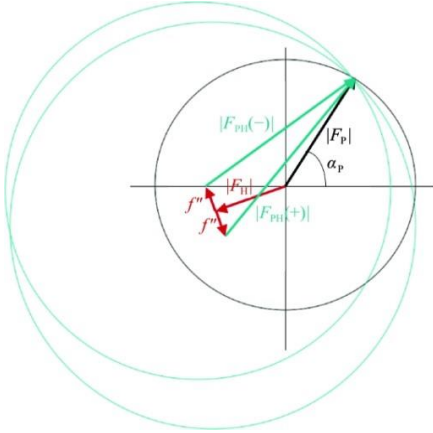
Answer



- f) We have made the assumption above that if we subtract the diffraction spots of the protein crystal from those of the protein-metal crystal, we get only the diffraction spots from the metal (heavy atom). Mathematically, this was written $|F_{PH}| - |F_P| = |F_H|$. Recall that the structure factor is a vector in the complex plane. Is the general statement $|A + B| - |A| = |B|$ true for all vectors? Under what circumstances would it be true for the three structure factors? Is it reasonable to make this assumption? (5 points).

The general statement is false (you can demonstrate to yourself by coming up with two random vectors, taking their magnitude, and then taking the magnitude of their vector sum – unless you were very lucky, the magnitudes should be different). The only condition by which the equation is true is if the two vectors are parallel (they would then add directly end to end) – in other words, if the phase between them is zero. However, if the magnitude of one vector is small relative to the other, then the sum will not deflect significantly from the original, and we can make the *approximation* that the angle is zero, keeping in mind that this is not strictly speaking true.

- 4 (5 points) Phase ambiguity can be resolved by taking advantage of anomalous scattering by heavy atoms in your crystal, as represented in the Harker construct shown below. Describe the additional $\mathbf{F}_{PH(+)}$, $\mathbf{F}_{PH(-)}$, and $\mathbf{f}'$ terms, and under what experimental conditions they can be expected to be observed.



Anomalous scattering is a phenomenon which occurs when electrons absorb and reemit X-rays having an energy close to that of the electron's nuclear binding energy. The re-emitted X-rays have the same wavelength as the incident radiation phase shifted by 90° . Anomalous scattering causes a breakdown in Friedel's law ($\mathbf{F}_{P+} \neq \mathbf{F}_{P-}$), which can be utilized to provide phase information. $\mathbf{F}_{PH+}$ and $\mathbf{F}_{PH-}$ are the structure factors for the derivative crystal Friedel pair, and $\mathbf{f}'$ the anomalous scattering component. Experimentally, X-rays need to be tuned to appropriate absorption energies for the specific heavy-atom in order to obtain sufficient anomalous scattering.

- 5 (5 points) While completing your experiments, the structure of a homolog with 90% sequence identity to your protein is published. What phasing technique could be performed with this information to solve your structure?

Molecular Replacement

6. Model Building (25 points)

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

- Navigating in Coot:
- Detailed instructions on using Coot are available at
- <http://www.ysbl.york.ac.uk/~emsley/coot/doc/user-manual.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 → Fetch 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 $2|F_o| - |F_c|$ and a $|F_o| - |F_c|$ map. Hide the $|F_o| - |F_c|$ 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_2015_PS6.zip from <http://saf.bio.caltech.edu/bi173/>. 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 β -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 helix. Note: In previous versions of Coot, the program could automatically recognize helical density and build a polyaniline 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 β -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

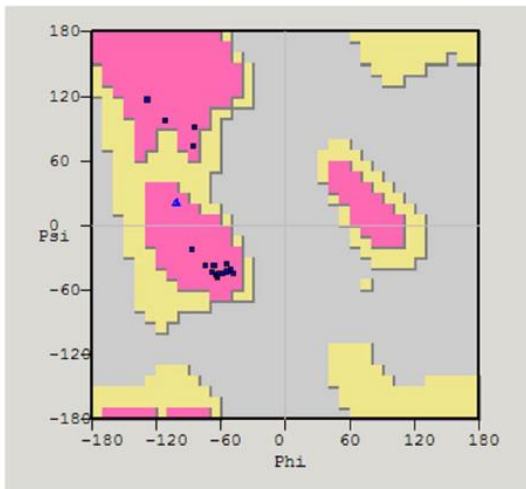
The resulting graph will plot the errors in density fit. Manipulate your structure to minimize errors in the density fit.

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 danpiraner.pdb). Email me your coordinate file (dpiraner@caltech.edu).

Answer

The correct sequence is MAAGDVFRAAASNELQIWAAR.

Ramachandran Plot:



Density Fit Analysis:



4. Model Validation/Interpretation (15 points)

a) (5 points) What is the R-factor? How is it calculated?

The R-factor is a measure of agreement between a model and data

$$R_{\text{cryst}} = \frac{\sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum_{hkl} |F_{\text{obs}}|}$$

b) (5 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?

R_{free} ensures that the R-factor is not being artificially reduced by only being calculated on 5-10% of the data that has been omitted from refinement. An approximate target would be 25%

c) (5 points) State the resolution required to distinguish the following features in an electron density map:

- Individual atoms
- All side chains
- Larger side chains
- Secondary structure elements

- Individual atoms: $<1\text{\AA}$
- All side chains: $1-3\text{\AA}$
- Larger side chains: $3-4\text{\AA}$
- Secondary structure elements: $4-8\text{\AA}$