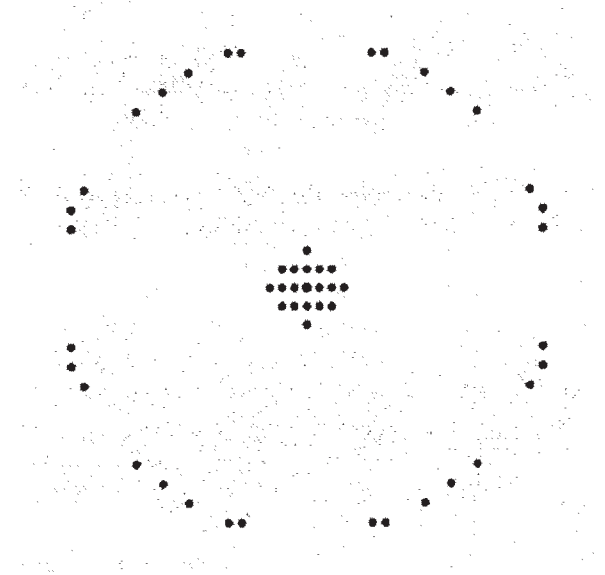
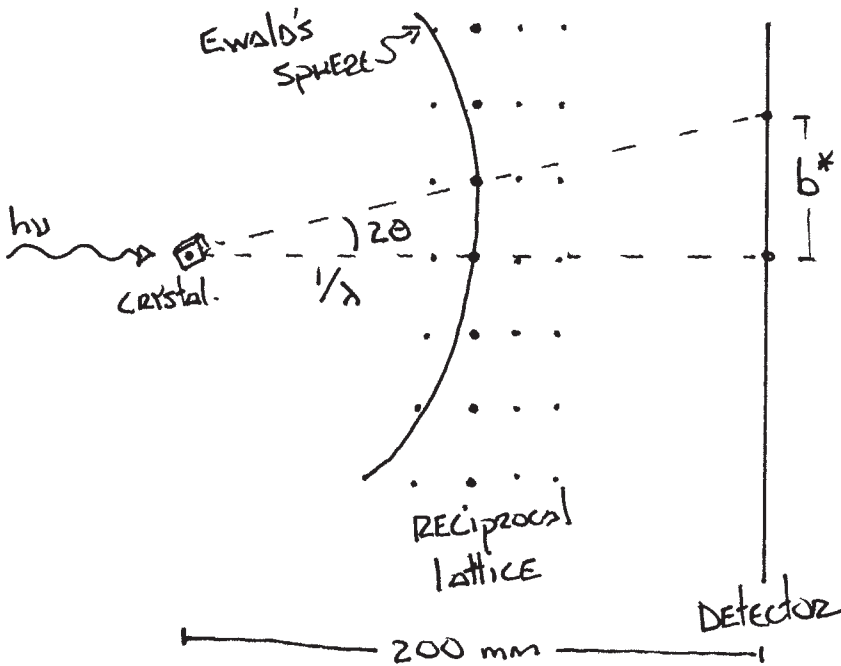


Answers to Problem Set #5:

1. In a few sentences, explain why the x-ray diffraction pattern obtained from a protein crystal appears as discrete reflections instead of a continuum. What information is encoded by the positions of the reflections? What is encoded by the relative intensities of reflections?

The observed diffraction pattern is a convolution of the scattering by atoms within a protein molecule and the molecular arrangement of proteins in a crystalline lattice. The continuous scattering of x-rays by atoms is sampled only at lattice points corresponding to the in-phase addition of waves obeying the von Laue conditions for scattering by a crystal. The positions of reflections are determined by the dimensions and shape of the crystal lattice, which serves to align molecules with respect to the incident x-ray beam. The relative intensities of reflections provide information about the physical arrangement of atoms within the unit cell, including the protein molecular structure and the symmetry operations relating different protein molecules within the unit cell.

2. Bragg's law states that for 2 x-rays to scatter in phase, the condition $\lambda = 2d \sin\theta$ must be satisfied. For the diffraction image below, the detector was positioned 200 mm from the crystal and $\lambda = 1.5 \text{ \AA}$. What unit cell dimensions can be derived from the spacings of reflections within the central zone on this diffraction image ($a^* = 7/4 \text{ mm}$ [horizontal direction]; $b^* = 10/4 \text{ mm}$ [vertical direction])?



$$\text{for } a^*: \tan 2\theta = \frac{7/4 \text{ mm}}{200 \text{ mm}}$$

$$2\theta = 0.5013^\circ$$

$$\lambda = 2d \sin(\theta)$$

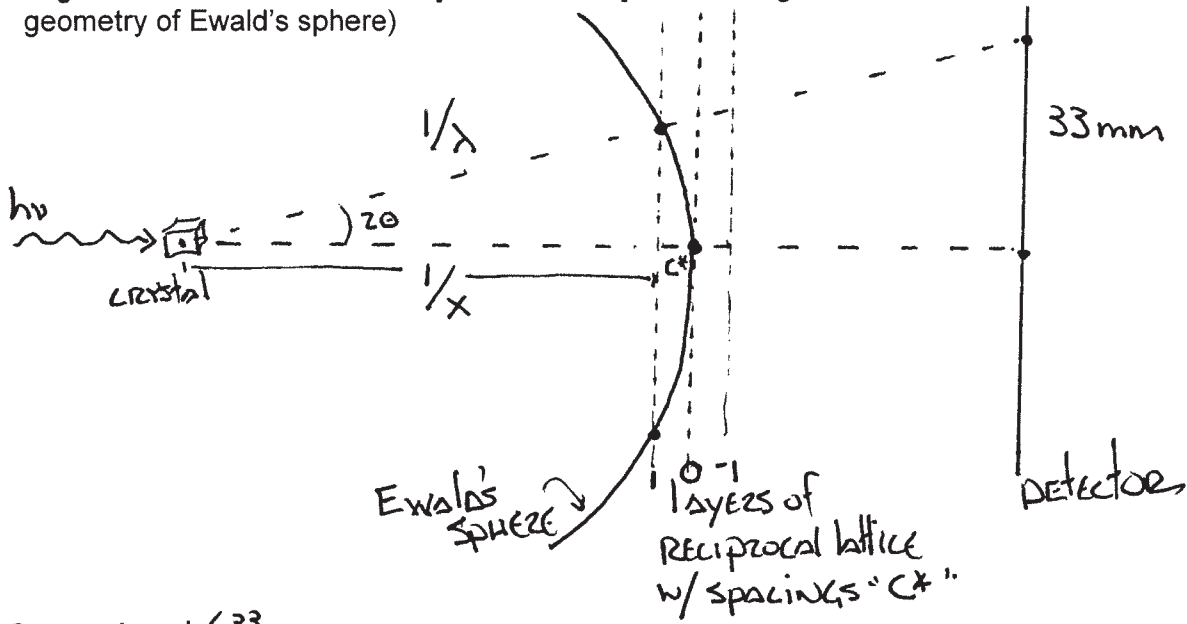
$$d = \lambda / (2 \sin(\theta)) = \underline{171 \text{ \AA}}$$

$$\text{for } b^*: \tan 2\theta = \frac{10/4 \text{ mm}}{200 \text{ mm}}$$

$$2\theta = 0.7162^\circ$$

$$d = \underline{120 \text{ \AA}}$$

3. What is the unit cell dimension ($c = ?$) parallel to the x-ray beam? (hint: calculate the 2θ angle for the lune of reflections [33 mm radius] surrounding the central zone then consider the geometry of Ewald's sphere)



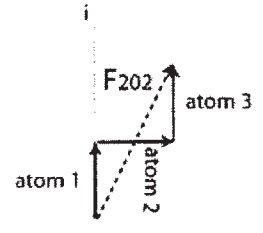
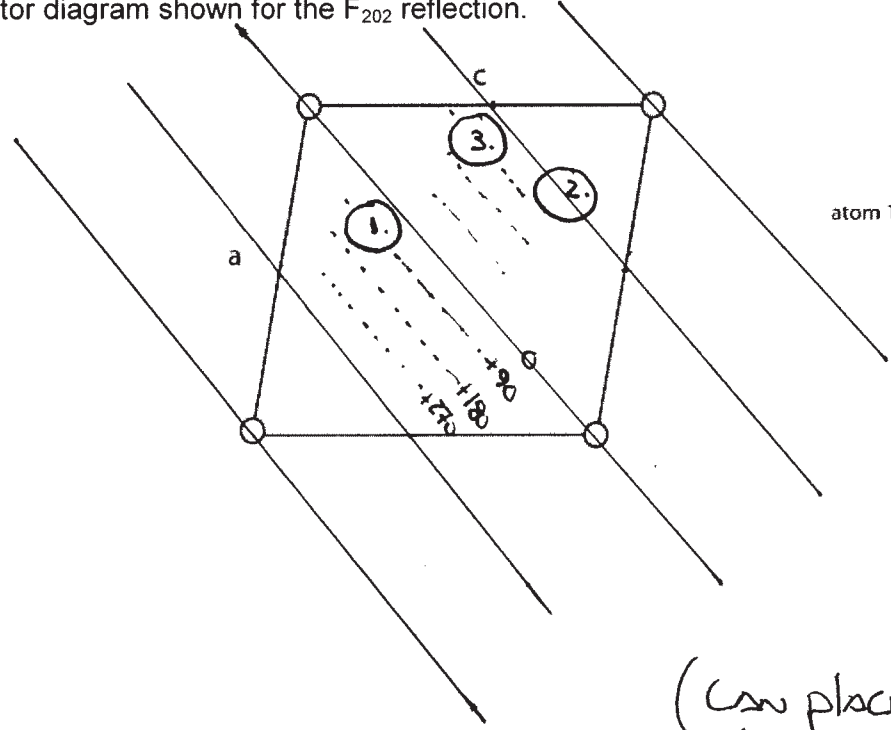
$$2\theta = \tan^{-1}\left(\frac{33\text{ mm}}{200\text{ mm}}\right) = 9.369^\circ$$

$$\cos(2\theta) = \frac{1/x}{1/\lambda} ; \frac{1}{x} = \left(\frac{1}{\lambda}\right) \cos(2\theta)$$

$$c = 112 \text{ \AA}$$

$$\frac{1}{\lambda} = c^* + \frac{1}{x} ; c^* = \frac{1}{\lambda} - \left(\frac{1}{\lambda}\right) \cos(2\theta) = \frac{1}{1.5} - \left(\frac{1}{1.5}\right) \cos(9.369) = \frac{1}{112 \text{ \AA}}$$

4. On the following drawing of a unit cell, draw the Bragg reflection planes corresponding to Miller indices ($hkl = 202$). Draw 3 atoms with positions in the unit cell that are consistent with the vector diagram shown for the F_{202} reflection.



atom 2: $\phi = 0^\circ \rightarrow$ on the plane of (202) layer

atoms 1, 3: $\phi = 90^\circ$
located $1/4$ spacing off the (202) layer

(can place atoms anywhere in unit cell satisfying conditions above.)

6. Describe in physical terms what is meant by "crystal mosaic spread" (mosaicity) and explain how it affects the diffraction pattern of a crystal, leading to inaccurate x-ray intensity measurements.

Mosaic spread refers to the misalignment of separate domains within a crystalline lattice. We can think of blocks of perfectly ordered crystalline lattice that are fused together in a slightly haphazard arrangement to form a single crystal. This misalignment causes the diffracted x-rays to be slightly divergent (not perfectly collinear). As a result, the recorded reflection intensities (signal) are blurred over a larger area of the detector that has intrinsic noise (diffuse x-ray scatter in air, noise in the x-ray detection photochemistry and electronics, etc.). The signal-to-noise ratio is lower for a mosaic crystal in comparison to a well ordered crystal.

5. The space group $P2_1$ has 1 Harker section located at $v = \frac{1}{2}$. Using the symmetry operators below, write the Harker vector equation(s) and calculate the atomic coordinates ($x = ?$; $z = ?$) for the heavy atom site corresponding to the peak shown below.

HARKER VECTORS:

$$(u, v, w) = (2x, \frac{1}{2}, 2z) \stackrel{OR}{=} (-2x, \frac{1}{2}, -2z)$$

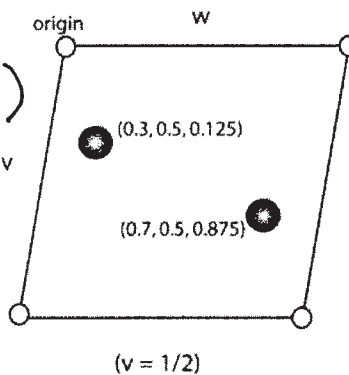
HARKER SECTIONS AT $+\frac{1}{2}$ AND $-\frac{1}{2}$ ARE EQUIVALENT.

POSSIBLE SOLUTIONS:

$$(.15, \frac{1}{2}, .0625) ; (.35, \frac{1}{2}, .4375)$$

$$(-.15, \frac{1}{2}, -.0625) ; (-.35, \frac{1}{2}, -.4375)$$

NOTE THAT PAIRS OF SOLUTIONS ARE RELATED BY $(x \pm \frac{1}{2}, z \pm \frac{1}{2})$. THIS AMBIGUITY RESULTS FROM AN ALTERNATIVE CHOICE OF ORIGIN FOR SPACE GROUP $P2_1$: $(x + \frac{1}{2}, 0, z + \frac{1}{2})$.



Symmetry for $P2_1$:
 (X, Y, Z)
 $(-X, Y + 1/2, -Z)$

7. The protein phase estimate (ϕ_p) for each reflection calculated from a single heavy atom derivative (F_{ph}) is ambiguous (more than 1 possible choice of phase value). Describe 2 different types/sources of phase ambiguity in the SIR (single isomorphous replacement) experiments, and at least 2 methods (experimental or computational) to identify the correct phase value.

A single heavy atom derivative yields 2 possible choices of phase value for each reflection because the phase triangle can be constructed in 2 different ways that both satisfy the known parameters: F_h , $|F_p|$, and $|F_{ph}|$ (i.e., the amplitude and phase of the heavy atom structure factor, and only the amplitude of the structure factors for native protein and the protein heavy atom derivative). A second source of phase ambiguity results from the centrosymmetric nature of Patterson space—the coordinates of heavy atoms located by Patterson methods are either (x, y, z) or $(-x, -y, -z)$, corresponding to phase values of (ϕ_h) or $(\phi_h + 180)$. We can fix this second ambiguity by "inverting the phases" of all reflections. Both types of phase ambiguities are eliminated by obtaining additional phasing information from 1) several independent heavy atom derivatives, 2) anomalous scattering data, or 3) a molecular replacement solution.

8. Give 2 reasons why the rotation function for molecular replacement is typically calculated at low resolution (10-4 Å) instead of over the full resolution range of the x-ray data for the unknown crystal. For an unknown crystal with multiple protein molecules in the asymmetric unit, does limiting the resolution improve the odds of obtaining a correct rotation solution using a single protein molecule for the search model? Justify your answer.

Adequate sampling of candidate orientations is computationally feasible only at modest resolution. Coarse sampling of the Patterson-based rotation function will skip over the correct solution that aligns interatomic vectors of the search model with those of the unknown crystal. This sampling problem is especially severe for the high order terms (large (hkl) values) representing high resolution/high precision information about interatomic distances. The second reason for limiting resolution is the limited structural homology between the search model and the unknown crystal. Even when properly oriented, the search model is only a rough facsimile of the unknown protein. Errors in the positions of atoms (structural divergence) are less significant when calculations are done at low resolution. Calculations done at inappropriately high resolution introduce random noise into the overall optimization scheme because the high order terms of the Fourier cannot be accurately represented by the (unrefined) model.

Limiting the resolution does not help the problem of “missing atoms” caused by an incomplete search model (1 molecule in the search model vs. several molecules in the crystal asymmetric unit). The effects of missing atoms are constant for high and low resolution terms of the Fourier/Patterson synthesis.