Problems – and Answers –
in Macromolecular Crystallography

Douglas C. Rees
California Institute of Technology
dcrees@caltech.edu

Table of Contents

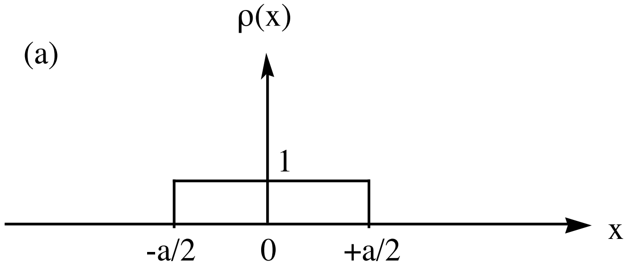
Problem Set 1. Fourier Transforms, Symmetry, Reindexing and Lattices

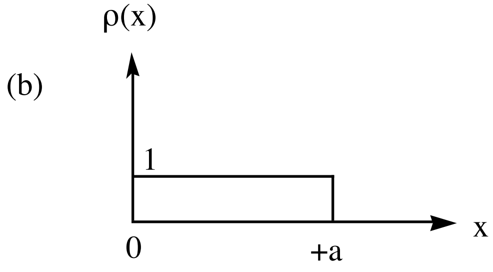
A. Fourier Transforms

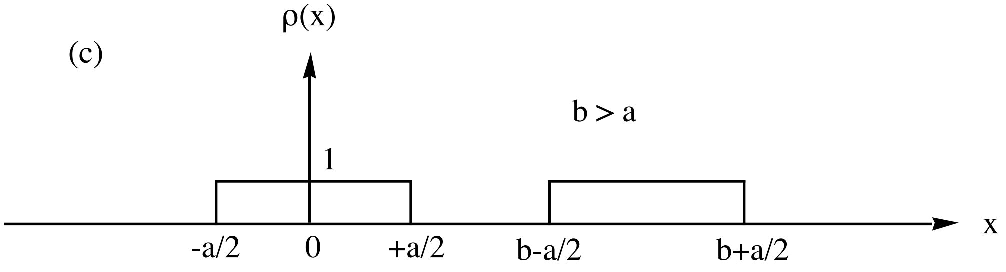
1. Evaluate analytically the scattering pattern for the following one-dimensional objects, using the relationship:

$$F(S) = \int_{-\infty}^{\infty} \rho(x) e^{2\pi i S x} dx$$

where $\rho(x)$ is the one dimensional electron density and S $(= 1/d = 2\sin\vartheta/\lambda)$ is the diffraction vector.







Note: this integral may be helpful: $\int_a^b e^{px} dx = \frac{1}{p}\left(e^{pb} - e^{pa}\right)$

2. The scattering pattern from a spherically averaged, three dimensional object is given by:

$$F(S) = \frac{2}{S}\int_0^{\infty} r\rho(r)\sin(2\pi S r) dr$$

Evaluate the scattering factor for a hydrogen atom as a function of S $(= 1/d = 2\sin\vartheta/\lambda)$, given that the electron density (square of the wavefunction) for a 1s electron is given by:

$$\rho(r) = \frac{e^{-(2r/a)}}{\pi a^3}$$

where a = 0.529Å, the Bohr radius. From this expression, calculate the value of the hydrogen scattering factor at S = 0.0, 0.2, 0.4, and 1.0 $Å^{-1}$. For comparison, the corresponding hydrogen scattering factor values calculated from an accurate Hartree-Fock level wavefunction are 1.0, 0.811, 0.481 and 0.071 electrons (see Table 6.1.1.1. (pg 477) of Intl. Tables Vol. C).

The following definite integral may be helpful:

$$\int_0^\infty x e^{-\alpha x} \sin(\beta x) dx = \frac{2\alpha\beta}{\left(\alpha^2 + \beta^2\right)^2}$$

3. Show that F(h,k,l) = F(-h,k,-l) in space group P2, with equivalent positions (*x,y,z*) and (*-x,y,-z*). What is the relationship between F(h,k,l) and F(-h,k,-l) in space group $P2_1$ (equivalent positions (*x,y,z*) and (*-x,y+1/2,-z*)?

4. Show that F(0,k,0) vanishes when k is odd for a structure in space group $P2_1$.

5. Write out the three rotation matrices R_i corresponding to the symmetry operators of space group P3. Show that these operations form a group by evaluating all the matrix products (R_i*R_j, i,j = 1,2,3) and show that the products all belong to this set.

B. Reindexing and Equivalent Reflections

1. Occasionally it is necessary to reindex a unit cell for consistency with earlier data sets or to take advantage of some particular property, as illustrated by the following examples:

a. It is possible to index cells such as $P3_121$ in two distinct ways, that differ by rotation of the a and b axes by 60° about the c axis. What are the transformations that relate coordinates and reflections in these two indexings?

A related example involves space groups such as $P4_1$ that can be indexed in two distinct ways that differ by a 180° rotation about the diagonal axis between the a and b axes. What are the transformations that relate coordinates and reflections in these two indexings?

b. Native CPA crystals grow in space group $P2_1$ with a = 51.4 Å, b = 60.3 Å, c = 47.2 Å and β = 97.6°, with one molecule in the asymmetric unit. Addition of ligands to these crystals occasionally results in packing changes that increase the asymmetric unit volume by factors of 2 or 3. The new unit cell dimensions in these cases are a = 74.4 Å, b = 60.4 Å, c = 65.6 Å, β = 97.72° for the doubled cell and a = 100.7 Å, b = 60.4 Å, c = 74.4 Å, β = 104.6° for the tripled cell, respectively. All space groups are $P2_1$. What are the transformations that can be used to relate the native cell to these ligand-bound crystals, and what are the transformations that relate the native reflection indices to those measured from soaked crystals?

c. MoFe-protein crystals (space group $P2_1$) can be indexed in two distinct fashions with cell dimensions a = 108.4 Å, b = 130.5 Å, c = 81.5 Å, β = 110.8°, and a = 110.1 Å, b = 130.5 Å, c =

81.5 Å, β = 113.0°, respectively. What is the relationship between these two forms and what is the transformation that interconverts the reflection indices?

d. A DNA crystal was collected in space group C2 with unit cell dimensions a = 66.6 Å, b = 38.6 Å, c = 102.8 Å, β = 102.56°. Self rotation functions indicated that there were very strong two fold axes spaced every 60° in the ab plane, and a three fold axis perpendicular to this plane, suggesting that the crystals actually have higher symmetry. What might be the space group and unit cell dimensions of this higher symmetry form? (Hint: think rhombohedral, and remember there are two types of settings: obverse (the standard setting with allowed reflection condition $h-k-l=3n$) and reverse (an alternative setting with $h-k+l=3n$)).

2. What are the indices of equivalent reflections in space groups P3_1 and I422?

3. What are the phase relationships between equivalent reflections in space groups P3_1 and I422?

4. What classes of reflections are centric in space groups P2_1, P$2_12_12_1$ and P$3_1$21? Calculate the two phase possibilities for centric reflections in these space groups.

C. Lattices and Miller Indices

1. Assign Miller indices to the various rows in the following "crop lattice" along the I-5 in California's Central Valley.

D. Crystal and Molecular Symmetry

1. Unit cell and space group information are occasionally useful for deducing aspects of molecular symmetry. An example of this type of reasoning is described in the following (slightly edited) excerpt from a paper. Does the stated conclusion necessarily follow?

Unit cell dimensions are a=122.0 Å and c=142.1Å (hexagonal indexing for space group P321). From the unit cell volume of 1.83×10^6Å^3, the molecular weight of about 3.1×10^5, the assumed density of 1.2 gm cm^{-3}, and assumed water content of about 40 per cent, we find that there are three molecules of the X complex in this unit cell. Accordingly, the space group requires that this molecular complex has at least one two fold axis.

E. Membrane and Solution Scattering

1a. Interpretation of the diffraction pattern from a stack of membrane bilayers (a multilayer) of repeat distance d= 50 Å gives the following values for the amplitude and phases for Miller indices h = 0 to 4:

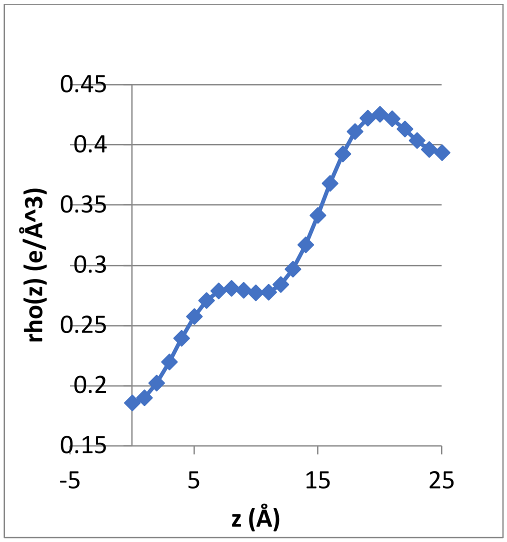
F(0) = +954, F(1) = -156, F(2) = 0, F(3) = 0, and F(4) = -43, in units of electrons.

(Since the one dimensional electron density profile is centrosymmetric, the phases correspond to either "+" (0 degrees) or "-" (180 degrees).

The center of the bilayer corresponds to z = 0, and so one full bilayer would extend from -25 < z < 25 Å. Using an Excel spreadsheet, Mathematica, or other such program, calculate the electron density profile, ρ(z), in units of e/Å^3 for 0 < z < 25 Å (in intervals no larger than1 Å) from the following equation:

$$\rho(z)=\frac{1}{V}\left[F(0)+2\sum_{h=1}^{4}F(h)\cos\left(\frac{2\pi hz}{d}\right)\right]$$

V is the volume of the two phospholipids spanning a bilayer = 3000 Å^3 .
(hint: it should look like the following graph)



1b The electron density of a bilayer should have the minimum value at the center of the bilayer (z = 0). Show that if the sign of either F(1) or F(4) is changed from "-" to "+", the calculated electron density at z = 0 increases. These types of considerations can be used to assign phases in the initial stage of a structural analysis of this type (or check that the experimental phases are consistent with expectations (assuming the expectations are correct)).

2 From a Guinier plot calculated from the following small angle x-ray scattering data, calculate the radius of gyration of the scattering object in Å. The Guinier plot is discussed in the section on Small Angle Scattering in the Crystallographic Appendices.

S (Å^{-1})	I(S)/I(S=0)
0.186530000E-02	0.884620011
0.205149991E-02	0.869170010
0.223929994E-02	0.834320009
0.242549996E-02	0.791580021
0.261169998E-02	0.814920008
0.279790000E-02	0.782050014
0.298570003E-02	0.731760025
0.317199994E-02	0.723540008
0.335819996E-02	0.692969978
0.354439998E-02	0.664690018
0.373060000E-02	0.643660009
0.391840003E-02	0.615379989

Problem Set 1 Answers

A. Fourier Transforms

1a

$$\begin{aligned} F(S) &= \int_{-a/2}^{a/2} (1)e^{2\pi iSx}dx \\ &= \frac{1}{2\pi iS}\left[e^{2\pi iSx}\right]_{-a/2}^{a/2} = \frac{1}{2\pi iS}\left[e^{\pi iSa} - e^{-\pi iSa}\right] \\ &= \frac{2i}{2\pi iS}\sin\pi Sa = \frac{\sin\pi Sa}{\pi S} \end{aligned}$$

1b

$$\begin{aligned} F(S) &= \int_{0}^{a} (1)e^{2\pi iSx}dx \\ &= \frac{1}{2\pi iS}\left[e^{2\pi iSx}\right]_{0}^{a} = \frac{1}{2\pi iS}\left[e^{2\pi iSa} - 1\right] \\ &= e^{\pi iSa}\frac{\sin\pi Sa}{\pi S} \end{aligned}$$

This answer is equivalent to the transform calculated in part (a) translated by a/2, which introduces a phase shift of exp(2πiS(a/2)) = exp(πiSa).

1c This integral can be calculated directly, or more easily evaluated by seeing that it will consist of the sum of two terms: the Fourier transform of the object in (1a) plus the same Fourier transform multiplied by the phase shift corresponding to a translation of "b" along x:

$$\begin{aligned} F(S) &= \frac{\sin\pi Sa}{\pi S} + e^{2\pi iSb}\frac{\sin\pi Sa}{\pi S} \\ &= \frac{\sin\pi Sa}{\pi S}\left\{e^{\pi iSb}\left(2\cos\pi Sb\right)\right\} \end{aligned}$$

2

$$
\begin{aligned}
F(S) &= \frac{2}{S}\int_0^\infty r\rho(r)\sin 2\pi Sr\,dr \\
&= \frac{2}{S}\frac{1}{\pi a^3}\int_0^\infty re^{-2r/a}\sin 2\pi Sr\,dr \\
&= \frac{2}{\pi a^3 S}\frac{2(2/a)(2\pi S)}{\left((2/a)^2+(2\pi S)^2\right)^2} \\
&= \frac{1}{\left(1+(\pi Sa)^2\right)^2}
\end{aligned}
$$

S (=1/d) ($Å^{-1}$)	$\sin\vartheta/\lambda$ =1/2d ($Å^{-1}$)	F(S) calc electrons	F(S) Intl. Tables Vol C. table 6.1.1.1 electrons
0.0	0.0	1.000	1.000
0.2	0.1	0.811	0.811
0.4	0.2	0.481	0.481
1.0	0.5	0.071	0.071

This expression for ρ(r) is equivalent to the Hartree-Fock calculation for a 1 electron atom, and so the scattering factors agree exactly.

3 In P2, with N atoms in unit cell, there are N/2 pairs of atoms related by crystallographic symmetry.

$$
\begin{aligned}
F(hkl) &= \sum_{j=1}^{N/2} f_j\left(e^{2\pi i(hx_j+ky_j+lz_j)}+e^{2\pi i(-hx_j+ky_j-lz_j)}\right) \\
F(\bar{h}k\bar{l}) &= \sum_{j=1}^{N/2} f_j\left(e^{2\pi i(-hx_j+ky_j-lz_j)}+e^{2\pi i(hx_j+ky_j+lz_j)}\right) \\
&= F(hkl)
\end{aligned}
$$

P2$_1$:

$$
\begin{aligned}
F(hkl) &= \sum_{j=1}^{N/2} f_j\left(e^{2\pi i(hx_j+ky_j+lz_j)}+e^{2\pi i\left(-hx_j+ky_j+\frac{k}{2}-lz_j\right)}\right) \\
F(\bar{h}k\bar{l}) &= \sum_{j=1}^{N/2} f_j\left(e^{2\pi i(-hx_j+ky_j-lz_j)}+e^{2\pi i\left(hx_j+ky_j+\frac{k}{2}+lz_j\right)}\right) \\
&= e^{\pi ik}F(hkl)
\end{aligned}
$$

4 from problem 3, $F(\bar{h}k\bar{l}) = e^{\pi ik}F(hkl)$

for the (0k0) reflection, $F(0k0) = e^{\pi ik}F(0k0)$

for k = even, this expression is always true, since $e^{\pi i(2n)} = 1$

for k = odd, this expression is in general not true, since $e^{\pi i(2n+1)} = -1$, so $F(0k0) = -F(0k0)$

this expression can only be true if F(0k0)=0 when k is odd.

5. Rotation matrices in space group P3:

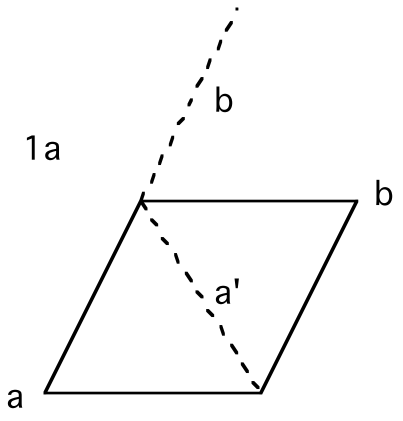
$$R_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad R_2 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad R_3 = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

by the rules of matrix multiplication, the products of any two rotation matrices correspond to a rotation matrix in this group.

R1*R1=R1 R2*R1=R2 R3*R1=R3
R1*R2=R2 R2*R2=R3 R3*R2=R1
R1*R3=R3 R2*R3=R1 R3*R3=R2

B. Reindexing and Equivalent Reflections

1a for the $P3_1$ cell:



$$\begin{pmatrix} a' & b' & c' \end{pmatrix} = \begin{pmatrix} a & b & c \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} h' & k' & l' \end{pmatrix} = \begin{pmatrix} h & k & l \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} h+k & \bar{h} & l \end{pmatrix}$$

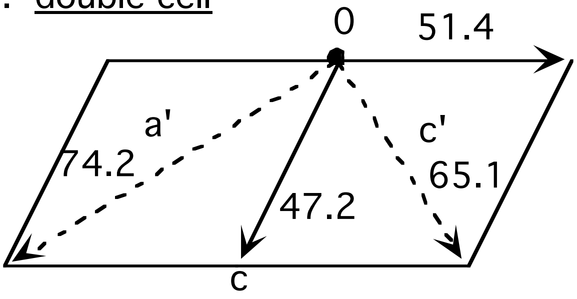
1a, part 2 for the P4$_1$ cell

$$\begin{pmatrix} a' & b' & c' \end{pmatrix} = \begin{pmatrix} a & b & c \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

$$\begin{pmatrix} h' & k' & l' \end{pmatrix} = \begin{pmatrix} h & k & l \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} k & h & \bar{l} \end{pmatrix}$$

b. double cell



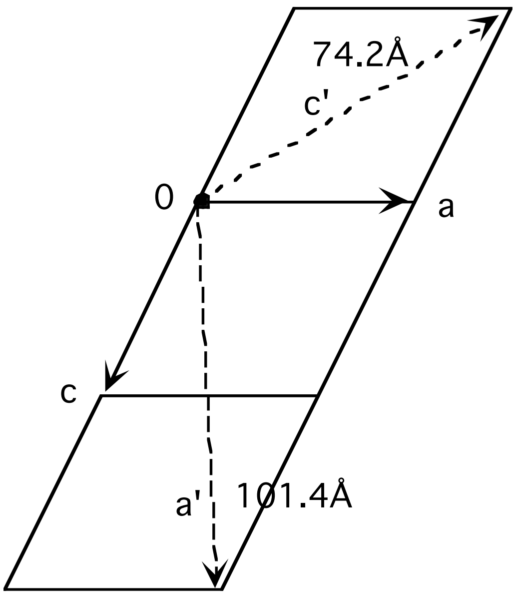
$$(a'b'c') = (abc) \begin{pmatrix} -1 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

$$(h'k'l') = (l-h, -k, h+l)$$

This is a pseudo B-centered cell, since $h' + l' = 2l = \text{even}$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}' = \frac{1}{2} \begin{pmatrix} -1 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

triple cell



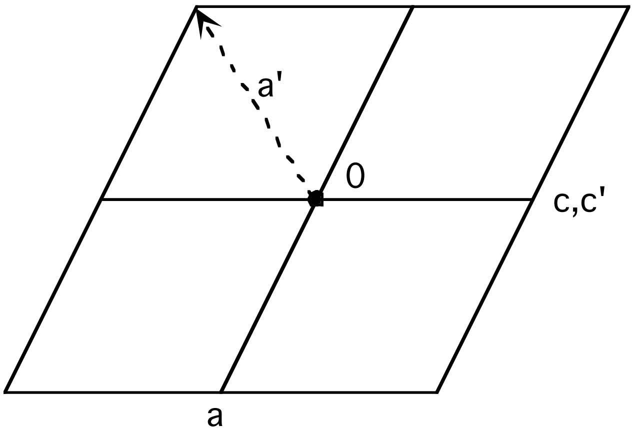
$$(a'b'c') = (abc)\begin{pmatrix} 1 & 0 & 1 \\ 0 & -1 & 0 \\ 2 & 0 & -1 \end{pmatrix}$$

$$(h'k'l') = (h+2l, -k, h-l)$$

In this cell, $h' - l' \approx 3l$ for allowed reflections

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}' = \frac{1}{3}\begin{pmatrix} 1 & 0 & 1 \\ 0 & -3 & 0 \\ 2 & 0 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

1c

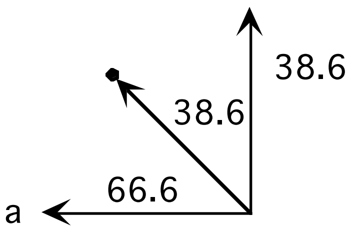


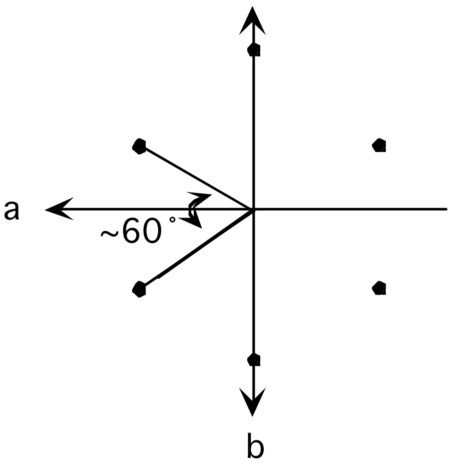
$$(a'b'c') = (abc)\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

$$(h'k'l') = (-h-l, -k, l)$$

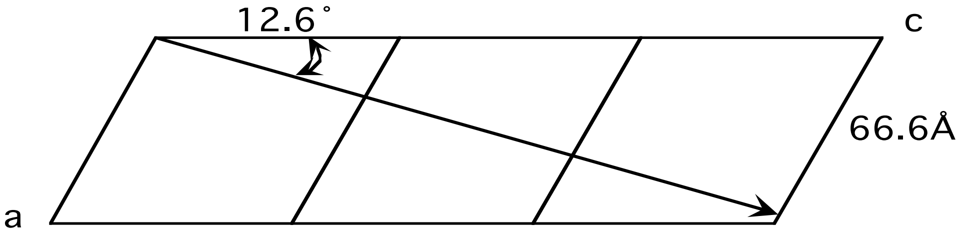
$$a' = 110\text{Å},\ b' = 130.5\text{Å},\ c' = 81.5\text{Å},\ \beta = 113.0^\circ$$

1d





Transform to R32, with a ~38.6 Å, c ~ 305 Å



$$\frac{66.6}{\sin 12.6} \cong 305 \text{ Å}$$

reindex $(a'b'c') = (abc)\begin{pmatrix} \frac{1}{2} & 0 & 1 \\ -\frac{1}{2} & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$

$(h'k'l') = \frac{h-k}{2}, k, h+3l$ $\qquad$ $h'-k'+l' = 3n$ $\qquad$ reverse

reindex $(a'b'c') = (abc)\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 3 \end{pmatrix}$

$(h'k'l') = \frac{h+k}{2}, \frac{-h+k}{2}, h+3l$ $\qquad$ $h'-k'-l' = 3n$ $\qquad$ obverse

reindex $(a'b'c') = (abc)\begin{pmatrix} 0 & -\frac{1}{2} & 1 \\ 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 3 \end{pmatrix}$

$(h'k'l') = k, \frac{-h-k}{2}, h+3l$ $\qquad$ $h'-k'+l' = 3n$ $\qquad$ reverse

$$\text{reindex}\,(a'b'c') = (abc)\begin{pmatrix} -\frac{1}{2} & 0 & 1 \\ \frac{1}{2} & -1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

$$(h'k'l') = -h+k, -k, h+3l \qquad\qquad -h'+k'+l' = 3n \qquad\qquad \text{obverse}$$

Problem 2.
Space group P3$_1$

$$C_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \ (h\,k\,l);\ C_2 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \ (k, -h-k, l);$$

$$C_3 = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \ (-h-k, h, l)$$

plus Friedel mates $\left(\overline{hkl}\right)$

Space group I422

$$\begin{array}{cccccccc} xyz & \overline{xy}z & \bar{x}y\bar{z} & x\overline{yz} & yx\bar{z} & \overline{yxz} & \bar{y}xz & y\bar{x}z \\ hkl & \overline{hk}l & \bar{h}k\bar{l} & h\overline{kl} & kh\bar{l} & \overline{khl} & \bar{k}hl & k\bar{h}l \end{array}$$

plus Friedel mates $\left(\overline{hkl}\right)$

Problem 3. $\alpha\left(h_j\right) = \alpha\left(h\right) - 2\pi\left(h_j \bullet t_j\right)$

Space group P3$_1$

$$t_2 = \begin{pmatrix} 0 \\ 0 \\ 1/3 \end{pmatrix} \quad \alpha\left(k, -h-k, l\right) = \alpha\left(h\,k\,l\right) - 2\pi\frac{l}{3}$$

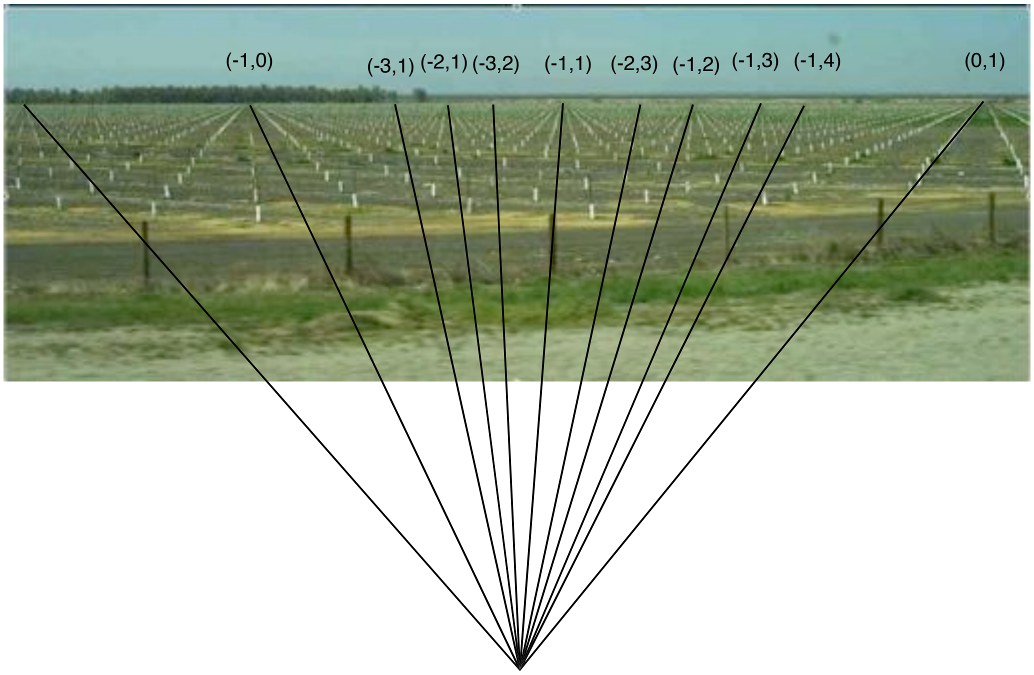
$$t_3 = \begin{pmatrix} 0 \\ 0 \\ 2/3 \end{pmatrix} \quad \alpha\left(-h-k, h, l\right) = \alpha\left(h\,k\,l\right) + 2\pi\frac{l}{3}$$

Space group I422 - all t's = (000) and $\alpha(h_j) = \alpha(h)$

Problem 4. Indices of centric reflections satisfy the relationship: $h^T C_j = -h^T$
with phase choices given by $\alpha\left(h_j\right) = \pi\left(h_j \bullet t_j\right)$ modulo(π)

space group	centric reflection	phase choice	equiv. position
$P2_1$	h0l	0	-x, 1/2+y, -z
$P2_12_12_1$	hk0	πh/2	1/2-x, -y, 1/2+z
	h0l	πl/2	-x, 1/2+y, 1/2-z
	0kl	πk/2	1/2+x, 1/2-y, -z
$P3_121$	h -h l	0	y, x, -z
	h0l	πl/3	-x, y-x, 1/3-z
	0kl	2πl/3	x-y, -y, 2/3-z

C. Lattices and Miller Indices



from the perspective of August 2021, I'm not sure why I indexed the principal rows as (-1,0) and (0,1)….

D. Crystal and Molecular Symmetry

This problem was based on an analysis of packing in ATCase crystals (Steitz, Wiley and Lipscomb, *PNAS* **58**, 1859-61 (1967)). While it is not clear how the number of ATCase molecules in a P321 unit cell was originally calculated, it is clear that values for several critical parameters were not experimentally obtained. The original analysis may have developed as follows, defining the various variables as follow:

Symbols:

n	number of molecules in unit cell
N_A	Avogadro's number
M	molecular weight of ATCase ($=3.1\times10^{5}$)
V_C	unit cell volume ($=1.83\times10^{-18}$ cm^3 for ATCase)
ρ_c	density of crystal (~1.15 gm/cm^3)
ρ_s	density of water and crystallization buffer
$\bar{v}_P$	partial specific volume of protein (0.737 cm^3/gm)

ρ_p	density of protein (= $1/v_p$)
V_p	volume of unit cell occupied by protein
V_s	volume of unit cell occupied by water
X_p	volume fraction of unit cell occupied by protein (= V_p/V)
X_s	volume fraction of cell occupied by water (= V_w/V)

From the conservation of mass:

$$N_A\rho_c V_C = N_A\rho_p V_p + N_A\rho_s V_s$$

mass of crystal = mass of protein + mass of water

Now,

$$N_A\rho_p V_p = nM$$
$$= N_A\rho_c V_c - N_A\rho_s V_s$$

if the <u>volume</u> fraction of water in the cell is X_w, then:

$$n = [N_A\rho_c V_c - N_A\rho_s V_c X_s] / M$$

for ATCase, if X_S is taken to be 0.4 (assumed by Steitz, *et al*), then n is calculated as ~2.7 ~ 3, as originally reported.

However, the crystal density and solvent content are not actually independent, and it can be shown that $X_S = \frac{\rho_P - \rho_c}{\rho_P - \rho_S}$ and $X_P = \frac{\rho_c - \rho_S}{\rho_P - \rho_S}$ Hence, if the $\rho_c = 1.15$ and $\rho_P = 1.35$ gm/cm^3, then $X_S \sim 0.57$ and $X_p \sim 0.43$. In terms of <u>mass</u> fractions, the crystal is almost exactly half protein and half solvent. Confusion on this point may have resulted in the erroneous conclusion.

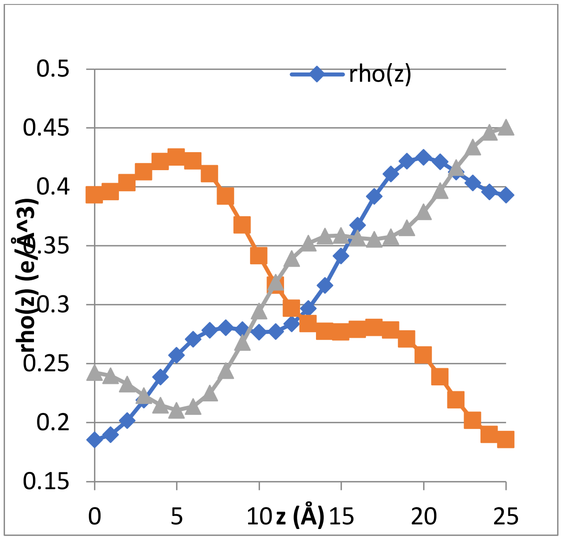
In a subsequent paper (Wiley and Lipscomb, *Nature* **218**, 1119-1121 (1968)), ρ_c, $\bar{v}_P$ and ρ_S were measured more accurately. Since the crystallization buffer and water have similar densities:

$$n = \frac{N_A V_C}{M}\frac{(\rho_C - \rho_S)}{(1 - \bar{v}_P \rho_S)}$$

inserting the relevant values gives n = 2.02. This implies that ATCase has a 3 fold axis in order to crystallize in space group P321. At the time, this result would have been striking, since ATCase was originally believed to have the composition $(RC)_4$.

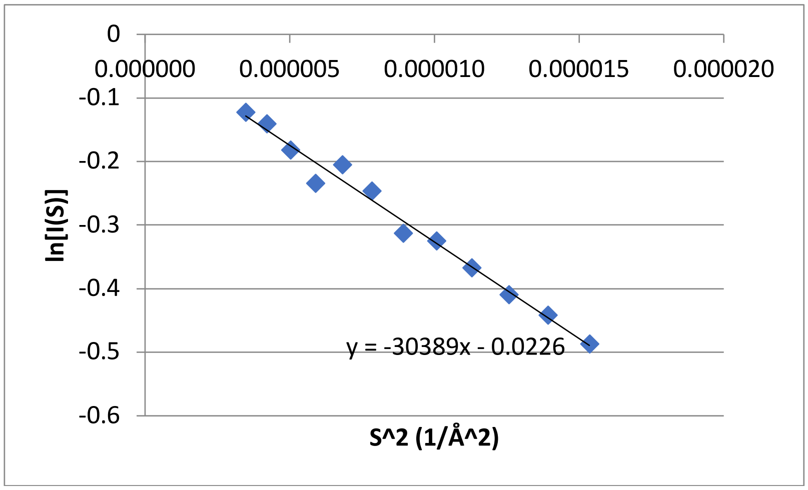
E. Membrane and Solution Scattering

1. Electron density profile for the correct phases, and changing the signs of F(1) and F(4)



Changing the sign of either F(1) or F(4) (or both, although I didn't calculate that) increases the calculated electron density at z $= 0$.

2. Guiner plot (this is for the nitrogenase ADP-AlF stabilized complex)



$$\ln\frac{I(S)}{I(0)} = -\frac{4\pi^2}{3}R_G^2S^2$$

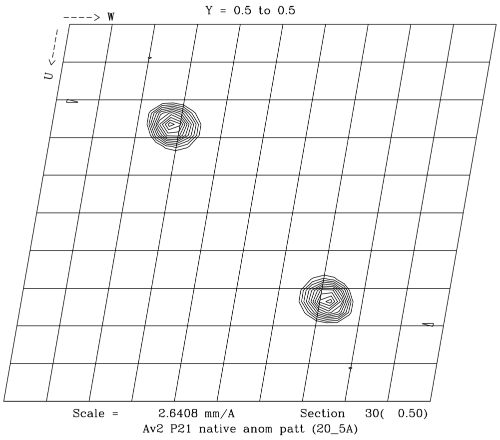
$$\text{slope} = -\frac{4\pi^2}{3}R_G^2 = -30389 \;\Rightarrow R_G = 48\mathring{A}$$

Problem Set 2. Patterson Problems

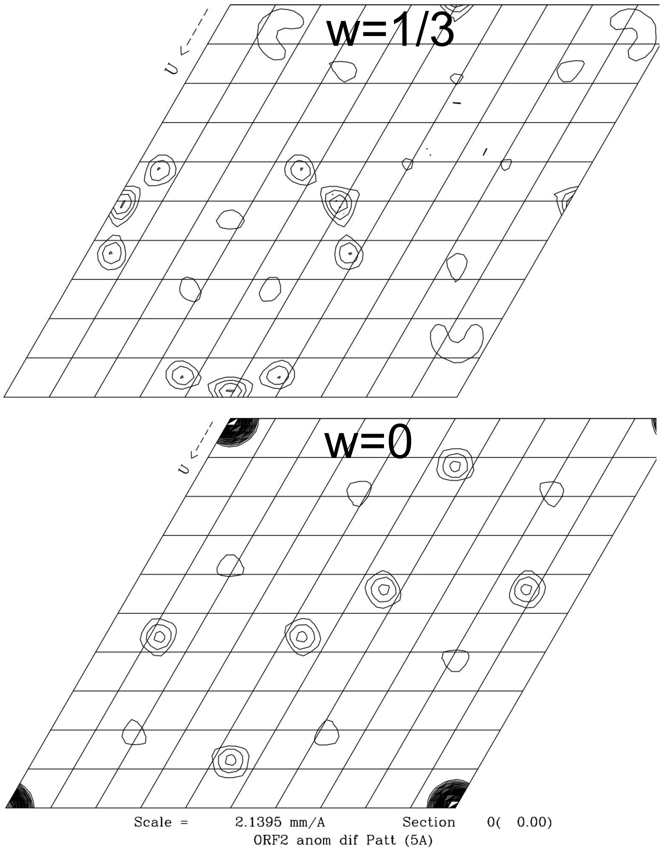
1. For the space groups $P3_1$ and $P4_2$, determine the (u,v,w) positions of self-vectors in the Patterson map in terms of the (x,y,z) coordinates for an atom in these space groups. What is (are) the Harker section(s) in these cases? Notice that these self-vectors do not uniquely define an origin. What are the (x,y,z) coordinates of the possible origins in these space groups?

2. find <u>single</u> site solutions from the following Patterson sections (the horizontal and vertical axes in each extend from 0 to 1)

a. $P2_1$; v=1/2 Harker section
(Georgiadis *et al. Science* **257**, 1653 (1992))



b. $P3_121$, w=1/3 Harker section
plus w = 0 (non-Harker) to get z (hint)
(Peters *et al. NSB* **3**, 991 (1996))

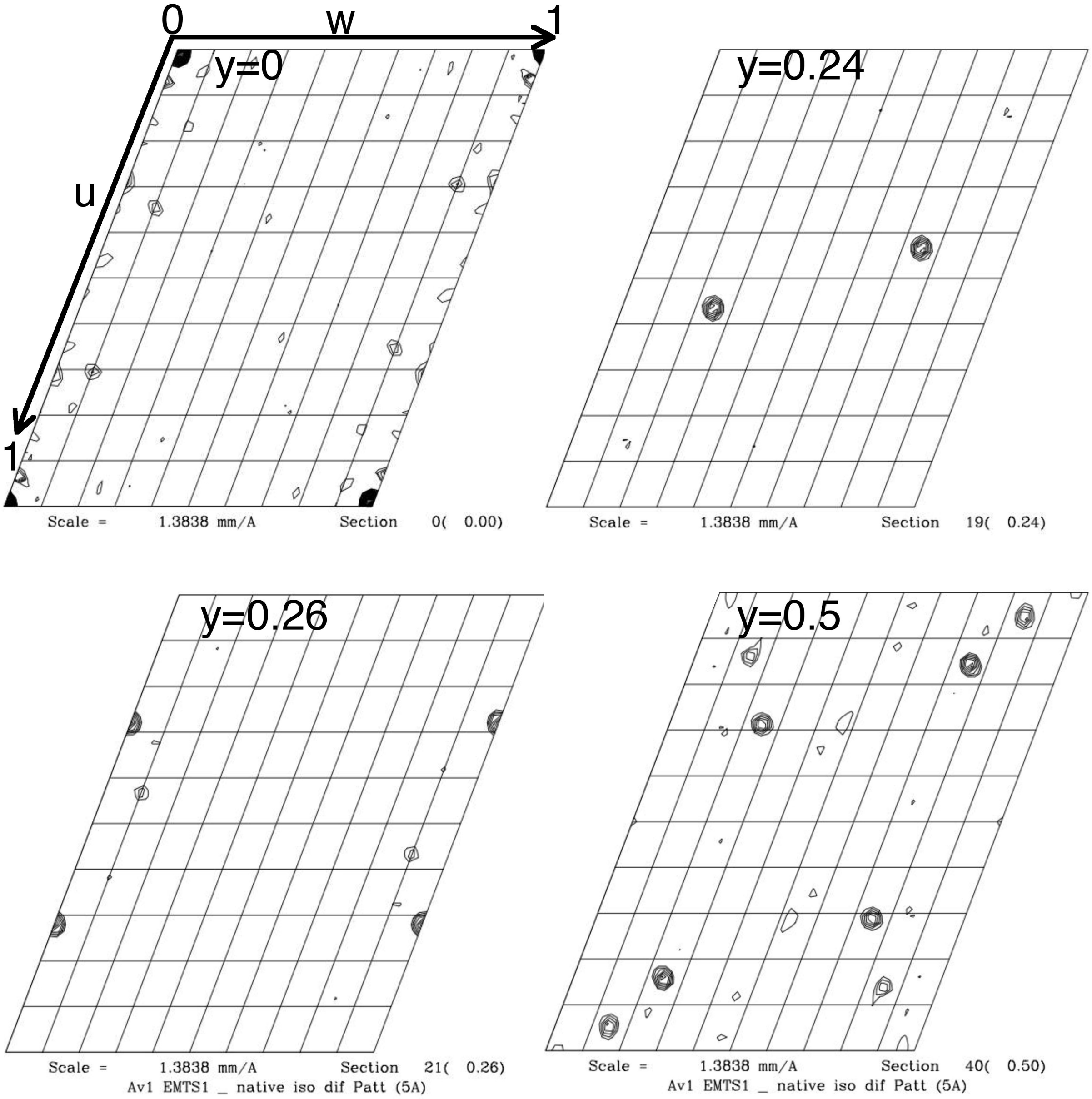


c. $P6_5$ (u,v) coordinates of Harker section peaks (neocarzinostatin, Pb1 derivative)
(Kim *et al. Science* **262**, 1042 (1993))

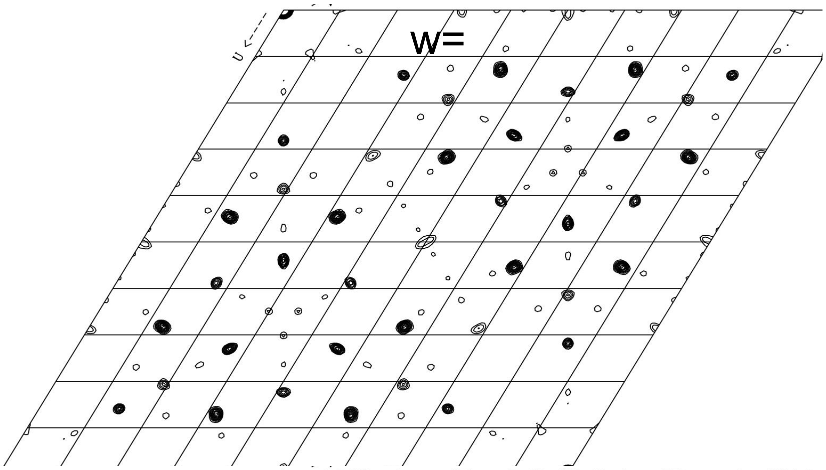
w=1/6	w=1/3	w=1/2
0.04, 0.65	0.26, 0.56	0.09, 0.34
0.35, 0.39	0.30, 0.74	0.25, 0.91
0.39, 0.04	0.44, 0.70	0.34, 0.25
0.61, 0.96	0.56, 0.30	0.66, 0.75
0.65, 0.61	0.70, 0.26	0.75, 0.09
0.96, 0.35	0.74, 0.44	0.91, 0.66

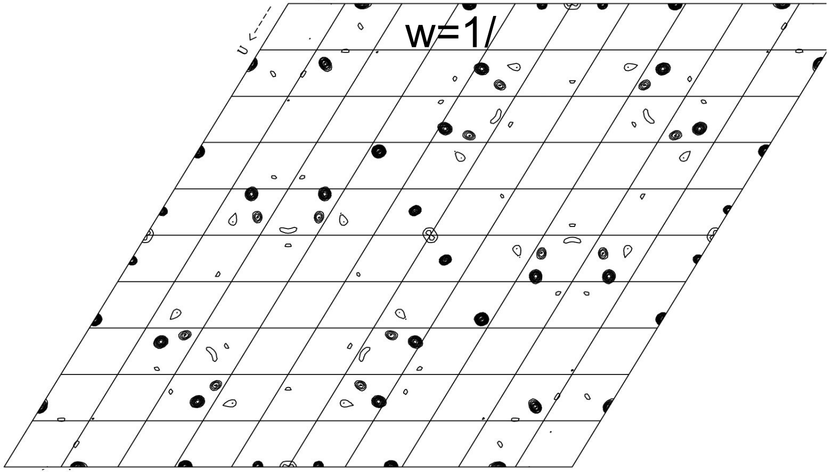
3. The 3-dimensional peak listing and selected sections of a heavy atom difference Patterson map of the nitrogenase MoFe-protein (Kim *et al. Science* **257**, 1677 (1992)), are illustrated below in space group $P2_1$. Interpret this Patterson map in terms of two major sites and two minor sites. The two minor sites have the approximately the same y coordinates as the two major sites.

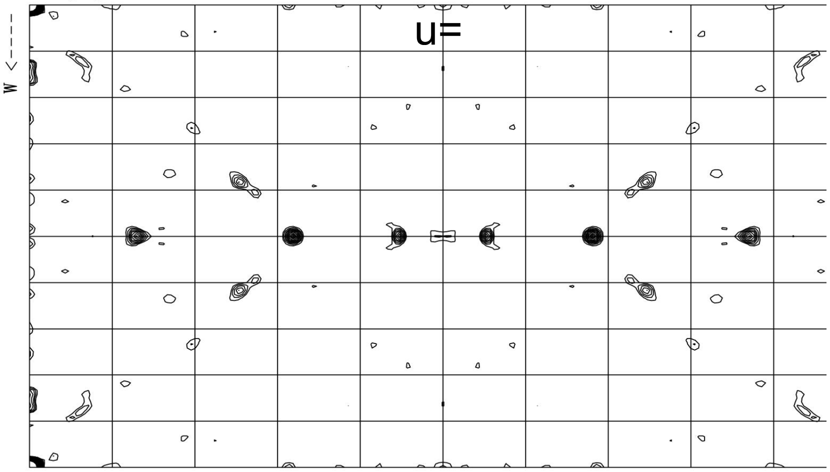
peak height	u	v	w
4038.	0.000	0.000	0.000
304.	0.068	0.000	0.000
202.	0.290	0.000	0.000
202.	0.709	0.000	0.000
325.	0.936	0.000	0.007
206.	0.704	0.000	0.096
206.	0.295	0.000	0.903
325.	0.063	0.000	0.992
255.	0.000	0.030	0.071
255.	0.000	0.030	0.928
435.	0.565	0.238	0.248
435.	0.434	0.238	0.751
200.	0.014	0.238	0.850
449.	0.280	0.262	0.000
449.	0.719	0.262	0.000
200.	0.620	0.262	0.100
200.	0.434	0.262	0.100
221.	0.139	0.500	0.147
324.	0.945	0.500	0.145
376.	0.288	0.500	0.249
450.	0.844	0.500	0.246
450.	0.155	0.500	0.753
376.	0.711	0.500	0.750
324.	0.054	0.500	0.854
221.	0.860	0.500	0.852



4. The w=0, w=1/2 and u=0 Harker sections for a heavy atom difference Patterson map calculated in space group $P6_322$ are illustrated below (Joshua-Tor *et al. Science* **271**, 1116 (1996)). Interpret these Harker sections in terms of one major site. A second minor site with the same x coordinate is also present.







Scale = 1.2900 mm/A Section 0(0.00)
GAL6 EMTS _ native iso dif Patt (2.2A)

Problem Set 2 Answers

1. Space group $P3_1$ (#144)
Harker peaks u,v,w = -x-y, x-2y, 1/3; 2y-x, y-2x, 1/3; 2x-y, y+x, 1/3.
Equivalent solutions can be obtained by adding either (1/3, 2/3) or (2/3, 1/3) to (x, y)
The z coordinate for the first site is arbitrary.

Space group $P4_2$ (#77)
Harker peaks u,v,w = 2x, 2y,0; 2y, 2x,0; -y-x, x-y, 1/2; x-y, x+y, 1/2
Equivalent solutions can be obtained by adding 1/2 to both (x, y).
The z coordinate for the first site is arbitrary.

2a. Harker peak in $P2_1$ (u, v, w) = (2x, 1/2, 2z) = 0.04, 1/2, 0.12, which yields (x,y,z) = (0.02, 0, 0.06). The y coordinate is arbitrary; 1/2 can be added to either or both x and z; and the signs of both x and z can also be switched.

2b. Space group $P3_121$ (#152) Harker peaks:
w = 1/3 u, v = -y-x, x-2y; x-2y; 2y-x, -2x+y; 2x-y, x+y; -x-y, y-2x; 2x-y, x-2y; 2y-x, x+y
w = 2z u, v = x-y, y-x; x-y, 2x-2y; -2x+2y, -x+y;
w = 1/3+2z u, v = -2y, -y; y, -y; y, 2y; (inverse positions for 2/3-2z)
w = 1/3-2z u, v = -2x, -x; x, -x; x, 2x; (inverse positions for 2/3+2z)

Let's arbitrarily assign the peak at (0.42, 0.05, 1/3) to –y-x, x-2y.
summing u and v = 0.42 + 0.05 = 0.47
this also equals –y-x+x-2y = -3y or y = -0.16; together with –y-x = 0.42 gives x = +0.74 so the trial solution is (0.74, -0.16) = (0.74, 0.84).

Now, there are three origins (+(0, 0), +(1/3, 2/3) or +(2/3, 1/3)) that are equivalent with the 3_1 screw, but not all of these are compatible with the two-fold Harker sections at 2z, 1/3-2z, etc. Adding these origin shifts to (0.74, 0.84) gives the trial solutions of (0.74, 0.84); (0.07, 0.51); (0.41, 0.17).

The two-fold Harker sections have peaks along the long diagonal such as (x-y, y-x, 2z); (x, -x, 1/3-2z); (y, -y, 1/3+2z); the corresponding peaks in the 2z section for the three trial solutions are: (x-y, y-x) = (0.90, 0.10), (0.56, 0.44), or (0.24, 0.76)

Of these, the solution with (0.56, 0.44) corresponds to peaks in the w=0 section, which gives z=0; this is consistent with (y, -y, 1/3+2z) = (0.51, 0.49, 1/3). Hence, the solution is (0.07, 0.51, 0). The five other sites related by crystallographic symmetry are (0.59, 0.56, 1/3); (0.44, 0.93, 2/3); (0.51, 0.07, 0); (0.93, 0.44, 1/3); (0.56, 0.49, 2/3).

In the original analysis, the coordinates used were (0.44, 0.93, 0.16), which is equivalent to (0.44, 0.93, 2/3) with an allowed z origin shift of 1/2.

2c. Space group $P6_5$ (#170) Harker peaks:

w = 1/6 u, v = -y, x-y and 5 other peaks
w = 1/3 u, v = -y-x, x-2y and 5 other peaks (can add 1/3, 2/3 or 2/3, 1/3 to x, y)
w = 1/2 u, v = 2x, 2y and 5 others (can add 1/2 to either or both of x, y)

From the peaks u,v,w = (0.04, 0.65, 1/6) get x,y,z = (0.61, 0.96, 0) (z is arbitrary).
5 other equivalent solutions can be generated by applying a 6 fold rotation:
x,y = (0.61, 0.96); (0.04, 0.65); (0.35, 0.39); (0.39, 0.04); (0.96, 0.35); (0.65, 0.61)

3a. Space group $P2_1$, equivalent positions x,y,z and -x,y+1/2,-z

Harker section vectors ± (2x,1/2,2z)

peak heights in a Patterson are ~ $f_i f_j$ for the vector between atoms i and j. So, the largest peaks will involve the two major sites, the next largest will be between the major and minor sites, and the weakest will involve only the minor sites.

Structure solution strategy: Start with self vectors? This is the hard way to approach the problem (at least by hand), due to complications of hand and origin ambiguities. Instead, start with cross vectors, which we will call site 1 and site 2:

site 1 1a: x1, y1, z1 1b: -x1, y1+1/2, -z1
site 2 2a: x2, y2, z2 2b: -x2, y2+1/2,- z2

crosspeaks: 2a-1a: x2-x1, y2-y1, z2-z1
1a-2b: x2+x1, 1/2-(y2-y1), z2+z1
⇒ these peaks are located in v sections symmetric about v=1/4!!

so, if 2a-1a is assigned to the peak 0.28, 0.26, 0.0
and if 1a-2b is assigned to the peak 0.57, 0.24, 0.25

these equations can be solved to give:

x2, y2, z2 0.43, 0.26, 0.12
x1, y1, z1 0.15, 0.0, 0.12

checking the self peaks for these two sites shows that they are the two largest peaks on the Harker section. These sites also have the same z coordinate, which reflects the fact that the NCS twofold in this structure is perpendicular to the crystallographic a and b axes; ie, it is along c*.

Finding the minor sites is harder - in a way, this is done by trial and error. In this particular case, we'll use the fact that pairs of sites have the same y coordinates (of course, one doesn't know this when starting to search for heavy atoms!). Hence, one minor site has y = 0 and the other has y = 0.26. The crosspeak from the major site at 0.15, 0.0, 0.12 to the minor site with y = 0.26 will be found in the v = 0.26 section of the Patterson map. There are 4 possibilities for this crosspeak:

0.43, 0.26, 0.10; .57, 0.26, 0.90; and 0.62, 0.26, 0.10; .38, 0.26, 0.10

By adding the coordinates of this crosspeak to 0.15, 0.0, 0.12, the possibilities for the coordinates x3, y3, z3 of site 3 are:

0.58, 0.26, 0.22; 0.72, 0.26, 0.02; 0.77, 0.26, 0.22; 0.53, 0.26, 0.22

Now, these sites can be checked by looking for crosspeaks to the major site with 0.43, 0.26, 0.12:

0.58, 0.26, 0.22 and symmetry equivalent 0.42, -0.24, 0.78 will have crosspeaks to 0.43, 0.26, 0.12 of 0.15, 0.0, 0.10 and .01, 0.5, 0.34; these peaks are not present. This rules out this choice for x3.

0.72, 0.26, 0.02 and symmetry equivalent 0.28,-.24, 0.98 will have crosspeaks to 0.43, 0.26, 0.12 of 0.29, 0.0, 0.90 and .15, 0.5, 0.14, which are present. Thus, this site is probably x3, which is confirmed by testing out the other two possibilities.

The coordinates for the 4th site can be found using the strongest unaccounted for peak on the Harker section (0.05, 1/2, 0.85 and 0.95, 1/2, 0.15). This should then represent a crosspeak involving the minor site with y = 0, with the major site at 0.15, 0.0, 0.12 (because this pair will have a crosspeak at y = 1/2). Hence the possibilities for x4 are:

0.20, 0.5, 0.97 = 0.80, 0.0, 0.03 and 0.10, 0.5, 0.27 = 0.90, 0.0, 0.73.

The first possibility would have crosspeaks to the site 0.43, 0.26, 0.12 of 0.63, 0.26, 0.09 (observed), while the second would have crosspeaks at 0.53, 0.26, 0.39 (unobserved). Hence the final solutions are

major sites:	0.43, 0.26, 0.12 and 0.15, 0.0, 0.12
minor sites	0.72, 0.26, 0.02 and 0.80, 0.0, 0.03

4. Space group $P6_322$ (#182) Harker sections w = 0, 1/2 and u = 0

w = 0   x+y, 2y-x; 2y-x, y-2x; y-2x, y-x; y+x, 2x-y; 2x-y, x-2y; 2y-x, y+x (+inverses)

w = ½ 2x, 2y; x-y, x; y, y-x; -2y, 2x-2y; -x,-y; 2y-2x,-2x (+inverses)

u = 0   0, x-2y, ½-2z; 0, y-2x, ½-2z; 0, y+x, ½-2z   (+inverses)

The simplest section to study is the u = 0. It is expected that there should be lines of constant w, with 3 peaks. Note that the only line of constant w that contains 3 (or more) peaks is w = ½, with 6 peaks at v = ±0.13, ±0.32, ±0.45. This implies that z = 0, so that the peaks at w = ½ -2z and ½ + 2z merge.

Since z = 0, the relative complexity of this Patterson reflects the coalescence of the w=0 and w=2z Harker sections.

Peaks can be assigned on the u = 0, w = 1/2 line by noting that the sum of the v coordinates for two peaks equal the negative v coordinate of the third peak. For example, (y-2x) + (x-2y) add up to -(y+x). Three peaks with u = 0, w = 1/2 that satisfy this are at v = 0.13, 0.32, -0.45. With y-2x = 0.13 and x-2y = 0.32, these can be solved to give x, y, z = (0.14, 0.41, 0). (NOTE: alternative solutions related by crystallographic and inversion symmetry are also possible).

[The w = ½ section has some interesting properties, since it has peaks at x, y, ½ and 2x, 2y, ½. Peaks with this relationship are found at uvw = (0.14, 0.41, ½) and (0.28, 0.82 ½), for example; again giving (0.14, 0.41, 0) as the solution for xyz.]

The minor site with the same x coordinate as the major site (x_1, y_1, z_1) can be found from the positions of cross vectors in the u = 0 section, by noting that they occur at (0, v_1, w_1) = (0, y_2-y_1, z_2-z_1) and (0, v_2, w_2) = (0, x-y_2-y_1, ½-z_2-z_1) with $x \equiv x_1 = x_2$ and $z_1 = 0$. Hence there should be a pair of peaks with $v_1+v_2 = x-2y_2 = +0.32$ and at z_2 and $\frac{1}{2} \pm z_2$. A pair of peaks satisfying these relationships are (0, 0.07, 0.12) and (0, 0.26, 0.38) that may be solved for the coordinates of the minor site (x_2, y_2, z_2) = (0.14, 0.48, 0.12).

[The self vectors for the minor sites do not appear among the more significant peaks in the Harker sections, illustrating the general point that minor sites are often identified through their cross vectors with major sites]

Problem Set 3. Phasing: isomorphous replacement and anomalous scattering

1a. Native and derivative data are measured from a centric structure. For a specific reflection $|F_p|$ and $|F_{ph}|$ are measured to be 10 and 8 electrons, respectively. The heavy atom structure factor, f_h, is calculated to be -2 electrons. What is the native phase? Assume perfect data and a perfect phasing model.

1b. Native and derivative data are measured from a centric structure. For a specific reflection $|F_p|$ and $|F_{ph}|$ are measured to be 10 and 8 electrons, respectively. The heavy atom structure factor, f_h, is calculated to be -3 electrons. The average lack of closure <E> is calculated to be 2. What are the best native phase and figure of merit? What are the best native phase and figure of merit when <E>=6?

2. Native and derivative data are measured from an acentric structure. For a specific reflection, the amplitudes from crystals of the native, derivative 1 and derivative 2 are measured to be 15 electrons, 14.4 electrons and 10 electrons, respectively. The calculated heavy atom contribution to derivative 1 is 5 electrons, with a phase of 37°, while for derivative 2 the corresponding quantities are 5 electrons and 323°. What is the native phase? Assume perfect data and model.

3. For the hkl = 3 2 1 reflection of a structure in the centric space group $P\bar{1}$, the following information is available for the native and two derivative data sets. Each derivative has one heavy atom site:

data set	|F|	heavy atom (xyz)			lack of closure <E>
native	10				
derivative 1	13	.2	.2	.0	1
derivative 2	8	.0	.0	0.16667	3

Assume the heavy atom form factor is 1 electron (including the temperature factor, occupancy, etc.)

a. Determine the most probable phase for the native structure, using each derivative data set individually.

b. Determine the best phase and figure of merit using both derivatives.

4. Show $m\cos\alpha_{best} = \tanh A$ for a centric structure with the inversion center at the origin, where A is the Hendrickson-Lattman coefficient.

5. A certain heavy atom derivative has a value of k (the ratio of the real to imaginary scattering factors) of 10. For the hkl = 5 10 2 reflection, Δ_{ano} = -1.7 electron and Δ_{iso} = 10 electrons. The heavy atom phase for this reflection is 120°. Assuming the data is perfect, what is the phase (α_{ph}) of the heavy atom derivative?

6. The real (A) and imaginary (B) components for the structure factors from a protein (F_p) and heavy atom (normal scattering f_h and anomalous scattering δ_h) are listed below:

	A	B
F_p	15.0	5.0
f_h	1.0	3.0
δ_h	-0.3	0.1

a. Calculate the phases of (i) F_p and (ii) the vector sum of the normal scattering components F_p+f_h.

b. Δ_{ano} is observed to have a positive value. By calculating the expected value of Δ_{ano} from the information above, decide whether the hand (enantiomorph) of the structure has been correctly assigned.

7 It is essential to the calculation of MAD phases that the correct hand of the heavy atom solution be established, as the incorrect hand will lead to garbage phases and electron density maps. In a space group such as P1, it is straightforward to change the hand; if one solution has coordinates (*x y z*), then the other hand has coordinates (*-x –y –z*). For space groups that occur in enantiomorphic pairs, such as $P4_12_12$ and $P4_32_12$, things are somewhat more complicated, because changing the heavy atom hand also requires changing the space group.

(a) suppose the heavy atom site (*xyz*) represents a solution to a Patterson map interpreted in space group $P4_32_12$. What are the coordinates/space group of the heavy atom solution corresponding to the opposite hand?

(b) suppose the heavy atom site (*xyz*) represents a solution to a Patterson map interpreted in space group $I4_1$. What are the coordinates/space group of the heavy atom solution corresponding to the opposite hand? (Hint: there is no space group $I4_3$)

8a. A crystal consisting entirely of one type of anomalous scattering atom crystallizes in an acentric space group. Is Friedel's law maintained?

b. A crystal consisting entirely of one type of anomalous scattering atom crystallizes in a centric space group defined with the inversion center at the origin. Is Friedel's law maintained? If so, are the two phase possibilities 0° and 180°?

9. The heavy atoms in an isomorphous derivative of a protein are related by an inversion center at the origin. [An example would be a single site heavy atom derivative in space group $P2_1$, located at x,+1/4,z and -x,-1/4,-z]. SIR phases for the protein are calculated using this derivative. What are the consequences of this situation for the native electron density map calculated with these SIR protein phases?

10. The positions of a native anomalous scatterer in a protein are related by an inversion center at the origin. [An example would be a single scatterer in space group $P2_1$, located at x,+1/4,z and

-x,-1/4,-z]. Phases for the protein are calculated using this anomalous scattering information. What are the consequences of this situation for the native electron density map calculated with these phases?

Hints:

For isomorphous differences

$$\Delta_{\mathrm{iso}} \cong f_h \cos\left(\alpha_p - \alpha_h\right) \equiv f_h \cos(\vartheta)$$

Solutions:

$$\alpha_{p_1} - \alpha_h = \vartheta \text{ or } \alpha_{p_2} - \alpha_h = -\vartheta$$

$$\alpha_{p_2} = 2\alpha_h - \alpha_{p_1}$$

For anomalous differences

$$\Delta_{\mathrm{ano}} \cong 2\delta_h \sin\left(\alpha_{ph} - \alpha_h\right) \equiv 2\delta_h \cos\left(\alpha_{ph} - \alpha_h + \frac{\pi}{2}\right)$$

Solutions:

$$\alpha_{ph_1} - \alpha_h + \frac{\pi}{2} = \vartheta \text{ or } \alpha_{ph_2} - \alpha_h + \frac{\pi}{2} = -\vartheta$$

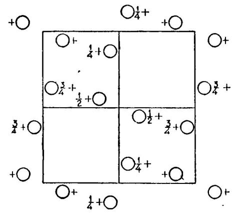
$$\alpha_{ph_2} = 2\alpha_h + \pi - \alpha_{ph_1}$$

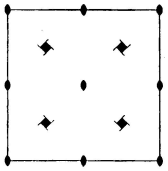
With one source of phase information, and two choices (in general) for the protein phase, it is necessary to use both in calculating the native electron density map:

$$F_p e^{i\alpha_p} = F_p\left[e^{i\alpha_{p1}} + e^{i\alpha_{p2}}\right]$$

One of the phase choices will be correct, and will lead to the true electron density for the structure, while the other is incorrect, and will lead to the artefactual features that are addressed in the last two problems.

$I4_1$ C_4^6 No. 80 $I\,4_1$ 4 Tetragonal

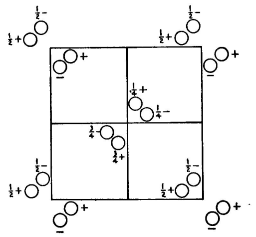


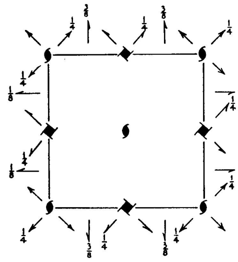


Origin on 2

Number of positions, Wyckoff notation, and point symmetry	Co-ordinates of equivalent positions	Conditions limiting possible reflections
	$(0,0,0;\ \tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2})+$	General:
8 b 1	$x,y,z;\ \bar{x},\bar{y},z;\ \bar{y},\tfrac{1}{2}+x,\tfrac{1}{4}+z;\ y,\tfrac{1}{2}-x,\tfrac{1}{4}+z.$	hkl: $h+k+l=2n$ $00l$: $l=4n$

$P4_12_12$ D_4^4 No. 92 $P\,4_1\,2_1\,2$ 4 2 2 Tetragonal

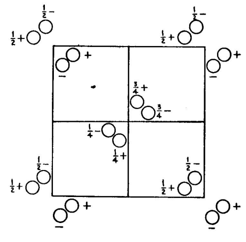


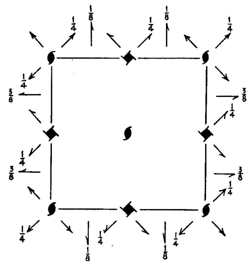


Origin at 2_12 ([001] [110])

Number of positions, Wyckoff notation, and point symmetry	Co-ordinates of equivalent positions	Conditions limiting possible reflections
		General:
8 b 1	$x,y,z;\ \bar{x},\bar{y},\tfrac{1}{2}+z;\ \tfrac{1}{2}-y,\tfrac{1}{2}+x,\tfrac{1}{4}+z;\ \tfrac{1}{2}+y,\tfrac{1}{2}-x,\tfrac{3}{4}+z;$ $y,x,\bar{z};\ \bar{y},\bar{x},\tfrac{1}{2}-z;\ \tfrac{1}{2}-x,\tfrac{1}{2}+y,\tfrac{1}{4}-z;\ \tfrac{1}{2}+x,\tfrac{1}{2}-y,\tfrac{3}{4}-z.$	hkl: No conditions $00l$: $l=4n$ $h00$: $h=2n$

$P4_32_12$ D_4^8 No. 96 $P\,4_3\,2_1\,2$ 4 2 2 Tetragonal





Origin at 2_12 ([001] [110])

Number of positions, Wyckoff notation, and point symmetry	Co-ordinates of equivalent positions	Conditions limiting possible reflections
		General:
8 b 1	$x,y,z;\ \bar{x},\bar{y},\tfrac{1}{2}+z;\ \tfrac{1}{2}-y,\tfrac{1}{2}+x,\tfrac{3}{4}+z;\ \tfrac{1}{2}+y,\tfrac{1}{2}-x,\tfrac{1}{4}+z;$ $y,x,\bar{z};\ \bar{y},\bar{x},\tfrac{1}{2}-z;,\ \tfrac{1}{2}-x,\tfrac{1}{2}+y,\tfrac{3}{4}-z;\ \tfrac{1}{2}+x,\tfrac{1}{2}-y,\tfrac{1}{4}-z.$	hkl: No conditions $00l$: $l=4n$ $h00$: $h=2n$

Problem Set 3 Answers

1a. $\Delta_{iso}=|F_{ph}|-|F_p| = -2 < 0$ so that F_p has opposite sign from f_h; f_h=-2 electrons, so that F_p=+10

1b. There are two phase choices for Fp: (+) and (-). For E=2

P(+). $F_{ph,calc} = F_p + f_h = 10-3=7$
$\varepsilon(+) = |F_{ph,obs}|-|F_{ph,calc}| = 8-7=1$
$P(+) = \exp[-\varepsilon^2/2E^2] = 0.88$
P(-) $F_{ph,calc} = F_p + f_h = -10-3 = -13$
$\varepsilon(-) = |F_{ph,obs}|-|F_{ph,calc}| = 8-13 = -5$
$P(-) = \exp[-\varepsilon^2/2E^2] = 0.044$
$m\cos\alpha = [\cos(0)P(+) + \cos(180)P(-)]/[P(+)+P(-)] = 0.84/0.93 = 0.91$
$m = 0.91$; $\alpha = 0$

For E = 6; P(+) = 0.99, P(-) = 0.71; m = 0.17, $\alpha = 0$

2. $|F_p| = 15$, $|F_{ph1}| = 14.4$, $|f_h| = 5$, $\alpha_{h1} = 37°$, $|F_{ph2}| = 10$, $|f_{h2}| = 5$, $\alpha_{h2} = 323°$

$$\alpha_p = \alpha_h \pm \vartheta$$

$$\vartheta = \cos^{-1}\frac{F_{ph}^2 - f_h^2 - F_p^2}{2 f_h F_p}$$

derivative #1, $\theta = 106.5°$; $\alpha_p = 143.5°, -69.5°$
derivative #2, $\theta = 180°$; $\alpha_p = 143°, 143°$ (note: heavy atom and protein Fs are collinear)
protein phase = 143°

3. $f_h = 2\cdot 1\cdot \cos 2\pi\left(3x+2y+1z\right)$; derivative 1, f_h = +2.0; derivative 2, f_h = +1.0.

		α=0			α =180		
Derivative	$\lvert F_{ph,obs}\rvert$	$\lvert F_{ph,calc}\rvert$	$\lvert F_{ph,o}\rvert - \lvert F_{ph,c}\rvert$	P(+)	$\lvert F_{ph,calc}\rvert$	$\lvert F_{ph,o}\rvert - \lvert F_{ph,c}\rvert$	P(-)
1	13	10+2=12	1	.607	|-10+2|=8	5	$3.7\text{x}10^{-6}$
2	8	11	-3	.607	9	-1	.946

The most probable phases for derivative 1, 2 = 0, 180°, respectively.
Combined phase, P(0) = .607x.607=.368; P(180)=$3.7\text{x}10^{-6}$x.946=$3.5\text{x}10^{-6}$
$m\cos\alpha = [.368-3.5\text{x}10^{-6}]/[.368+3.5\text{x}10^{-6}]=1.0$. m=1.0; α=0

4. $P(\alpha) \sim \exp(A\cos\alpha)$
$m\cos\alpha_{best} = [P(0)-P(180)]/[P(0)+P(180)] =$

$$\frac{e^A - e^{-A}}{e^A + e^{-A}} = \tanh A$$

5.

$$\frac{\Delta_{ano}}{\Delta_{iso}} = \frac{-1.7}{10} = 2\left[\frac{\delta}{f}\right]\frac{\sin\left(\alpha_{ph}-\alpha_h\right)}{\cos\left(\alpha_{ph}-\alpha_h\right)}$$

$$\tan\left(\alpha_{ph}-\alpha_h\right) = 5\left[\frac{-1.7}{10}\right]$$

$$\left(\alpha_{ph}-\alpha_h\right) = -40^o$$

$$\alpha_{ph} = 80^o$$

6a

$$\alpha_p = \tan^{-1} B/A = \tan^{-1} 5/15 = 18.4^\circ$$

$$\alpha_{ph} = \tan^{-1} B/A = \tan^{-1} 5+3/15+1 = 26.6^\circ$$

6b

$$\begin{aligned}
\left|F^+\right| &= \sqrt{(15+1-.3)^2+(5+3+.1)^2} = 17.67 \\
\left|F^-\right| &= \sqrt{(15+1+.3)^2+(5+3-.1)^2} = 18.11 \\
\Delta_{ano} &= 17.67-18.11 = -0.44 \\
&\approx 2\left|\delta_h\right|\sin\left(\alpha_{ph}-\alpha_h\right) = 2(.32)\sin(26.6-71.6) \\
&= -0.45
\end{aligned}$$

the wrong hand was assigned

7 Changing the enantiomer involves inverting the crystal structure (-x, -y, -z) through the appropriate origin such that the symmetry operators (of opposite hand) remain in the same place with respect to the crystal axes.

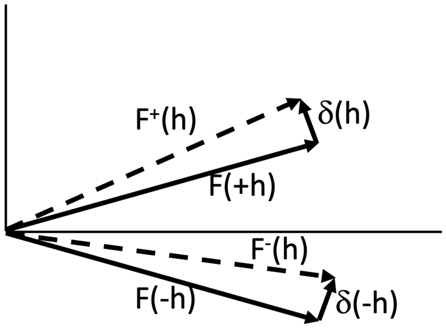
7a. The 4_1 axes in $P4_12_12$ are positioned at (½, 0, 0) and (0, ½, 0), the same as the 43 axes in $P4_32_12$. Hence changing the enantiomer from x, y, z in $P4_32_12$ involves changing to –x, -y, -z in $P4_32_12$.

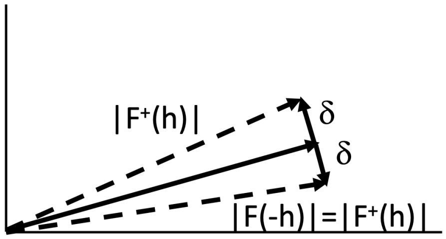
7b. The 4_1 and 4_3 axes in $I4_1$ are located at ±(¼, -¼, 0) and ± (¼, ¼, 0), respectively, which means that they are related by an inversion center at (½, ¼, 0) Hence, the x,y,z coordinates must also be inverted through this point, so that the enantiomeric structure has coordinates

$$(x,y,z) = (1/2,1/4,0)-(x-1/2,y-1/4,z) = (-x,1/2-y,-z)$$

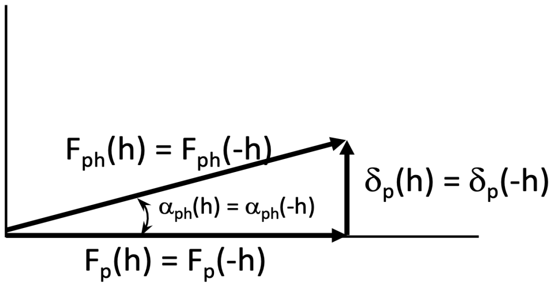
in space group $I4_1$.

8a. Yes, Friedel's law is maintained : $\left|F_{hkl}\right| = \left|F_{\overline{hkl}}\right|$





8b. Friedel's law is maintained, but the phases are no longer 0° or 180°.



9. Since $\alpha_h=0^\circ,180^\circ$

$$\begin{aligned} F_{hkl} &= \left|F_{hkl}\right|\left\{e^{i\alpha_{p1}}+e^{i\alpha_{p2}}\right\} \\ &= \left|F_{hkl}\right|\left\{e^{i\alpha_{p1}}+e^{i\left(2\alpha_h-\alpha_{p1}\right)}\right\} \\ &= \left|F_{hkl}\right|\left\{e^{i\alpha_{p1}}+e^{i\left(-\alpha_{p1}\right)}\right\} \end{aligned}$$

The net result is the real structure is superimposed on the inverse structure.

10. Since $\alpha_h=0^\circ,180^\circ$

$$\begin{aligned} F_{hkl} &= \left|F_{hkl}\right|\left\{e^{i\alpha_{p1}}+e^{i\alpha_{p2}}\right\} \\ &= \left|F_{hkl}\right|\left\{e^{i\alpha_{p1}}+e^{i\left(2\alpha_h+\pi-\alpha_{p1}\right)}\right\} \\ &= \left|F_{hkl}\right|\left\{e^{i\alpha_{p1}}-e^{i\left(-\alpha_{p1}\right)}\right\} \end{aligned}$$

The net result is the real structure is superimposed on the negative of the inverse structure. In space group $P2_1$, points with coordinates (x, 1/4,z) and (-x,-1/4,-z) are equivalent. Since they are also related by an inversion center, the only way all these conditions can be satisfied is if all the densities in the y=1/4 and -1/4 sections are zero; ie. there is nothing in the y sections with the heavy atoms.

Problem Set 4. Molecular replacement

1. Packing peaks: Even fold non-crystallographic axes parallel to even-fold crystallographic axes give rise to large packing peaks in the native Patterson map, since the non-crystallographic symmetry in these cases can also be described as a translation. Assume a non-crystallographic two fold axis is parallel to the b axis in space group $P2_1$, and is defined by x and z coordinates x_o and z_o. What are the coordinates of the packing peak?

2. Klug peaks: One dimer per asymmetric unit crystallizes in space group $P2_1$, with cell angle $\beta = 90.00°$. The dimer twofold axis is parallel to the a axis, and is defined by y and z coordinates $y_o=0$ and z_o. What does the two-fold rotation ($\kappa=180°$) section of the self rotation function look like for this crystal? The two dimers are related by both the crystallographic 2_1 screw operation, and a non-crystallographic symmetry operation. What is the non-crystallographic symmetry relationship between the two dimers? Is the non-crystallographic symmetry reflected by the presence of any packing peaks in the native Patterson map for this crystal?

3. Show that a translation of all atoms in a structure in real space corresponds to a phase shift of all reflections in reciprocal space. The relationship $F(h) = \sum_{\text{atoms } j} f_j e^{2\pi i h x_j}$ might be useful.

4. A trimeric molecule crystallizes in space group $P3_1$, with one trimer per asymmetric unit. The trimer three fold axis is parallel to the unit cell c axis (the 3_1 screw axis direction). It is noted that reflections with h and k indices both even obey the rhombohedral extinction rule h-k-l=3n (ie, reflections with both h and k even and h-k-l ≠ 3n are systematically weak). Based on this information, determine the locations of the trimer three-fold axes. This problem is loosely based on crystals of the lambda repressor DNA binding domain.

5. Real Space Translation Functions: Assume you have:

An experimentally determined electron density map, calculated with complex structure factors $F_o(h) = A_o(h) + iB_o(h)$.

A properly oriented (but perhaps incorrectly positioned) model structure, which when placed in a P1 cell of the same dimensions as the unknown structure has the corresponding complex structure factors $F_m(h)= A_m(h) + iB_m(h)$.

The appropriate translation, u, needed to properly position the model in the unknown map may be determined by finding the u that maximizes the function T(u), which is the integral over the entire cell of the product between ρ_o and ρ_m, the observed and model electron densities, respectively:

$$T(u) = \int \rho_o(x)\rho_m(x+u)dx$$

Although T(u) could be evaluated by a brute force calculation (i.e. translate the model to u_1, calculate $T(u_1)$; translate to u_2, calculate $T(u_2)$; etc.) show that a more elegant and faster way to calculate T(u) for all u is with the following Fourier transform:

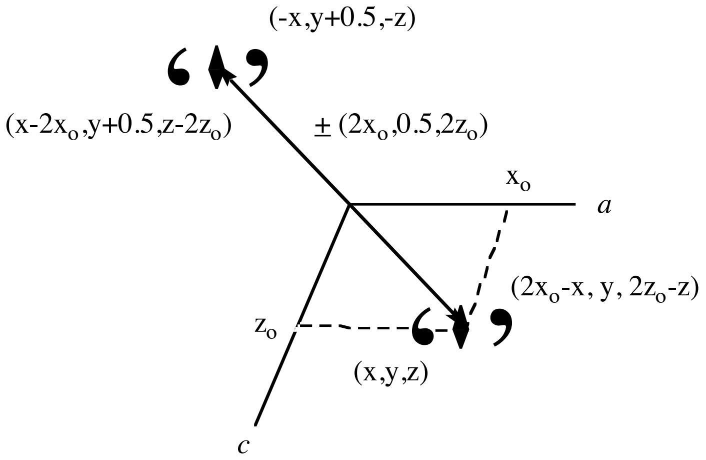
$$T(u) = \frac{1}{V^2}\sum_h \left[F_o(\bar{h})F_m(h)\right]e^{-2\pi i h\cdot u}$$

$$= \frac{1}{V^2}\sum_h \left[\left(A_oA_m + B_oB_m\right) + i\left(A_oB_m + B_oA_m\right)\right]e^{-2\pi i h\cdot u}$$

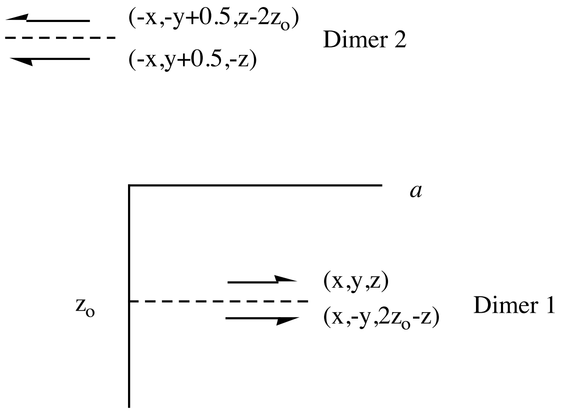
It may be helpful to use the relationship: $\rho(x) = \frac{1}{V}\sum_h F(h)e^{-2\pi i h\cdot x}$

Problem Set 4 Answers

1. The ncs twofold along y that passes through (x_o,z_o) converts a point with (x,z) to $-(x-x_o)+x_o,-(z-z_o)+z_o$. The packing peak has coordinates ± ($2x_o$,0.5,$2z_o$), which is the same as the self vector coordinates in the Harker section of a point at x_o,z_o.



2. The ncs two-fold parallel to a with y=0 and $z=z_o$ relates the point (x,y,z) to the one with ($x,-y,2z_o-z$).



Applying the crystallographic symmetry generates Dimer 2 from Dimer 1. From the coordinates of the equivalent positions, the relationship between Dimer 1 and Dimer 2 corresponds to a non-crystallographic 2_1 screw axis parallel to c with a translational component along z of $-2z_o$, and that passes through the point (x,y) = (0,0.25):

$$\begin{bmatrix} -x \\ 0.5-y \\ z-2z_0 \end{bmatrix}_{\text{Dimer 2}} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} 0 \\ 0.5 \\ -2z_o \end{bmatrix}$$

In this case, a self rotation function will show three perpendicular two fold axes, even though the molecule does not have 222 point group symmetry.

3.

$$F(h) = \sum_{\text{atoms } j} f_j e^{2\pi i h x_j} \quad \text{replace } x_j \text{ by } x_j + \Delta x$$

$$\begin{aligned} F'(h) &= \sum_{\text{atoms } j} f_j e^{2\pi i h (x_j + \Delta x)} \\ &= F(h) e^{2\pi i h \Delta x} \\ &= F(h) e^{i \Delta \alpha_h} \end{aligned}$$

4. The key point of this problem is that if a trimer with a three-fold symmetry axis is rotated by ± 120˚ about the three fold axis, it looks the same. Hence, the trimers that are related by the three-fold screw axis in space group $P3_1$ all have the same orientation. Consequently, the screw axis operations are equivalent to translational relationships between the trimers spaced by 1/3 along z. Let T(*h*) be the molecular transform of a trimer positioned with the three-fold axis along z at x,y = (0,0). Then, the transform of a trimer with the three-fold axis at x_o,y_o is given by:

$$T'(h) = T(h) e^{2\pi i (h x_o + k y_o)}$$

And the transform of the $P3_1$ cell is given by:

$$\begin{aligned} F(h) &= T(h)\left[e^{2\pi i (h x_o + k y_o)} + e^{2\pi i \left(k x_o + (-h-k) y_o + l/3\right)} + e^{2\pi i \left((-h-k) x_o + h y_o - l/3\right)} \right] \\ &= T(h) e^{2\pi i (h x_o + k y_o)} \left[1 + e^{2\pi i \left((k-h) x_o + (-2k-h) y_o + l/3\right)} + e^{2\pi i \left((-2h-k) x_o + (h-k) y_o - l/3\right)} \right] \end{aligned}$$

For an R3 cell, x_o and y_o must be such that the exponential terms equal 1 when h-k-l=3n. This will occur when:

$$\begin{aligned} &(k-h)x_o + (-2k-h)y_o + l/3 = \text{integer, and} \\ &(-2h-k)x_o + (h-k)y_o - l/3 = \text{integer} \end{aligned}$$

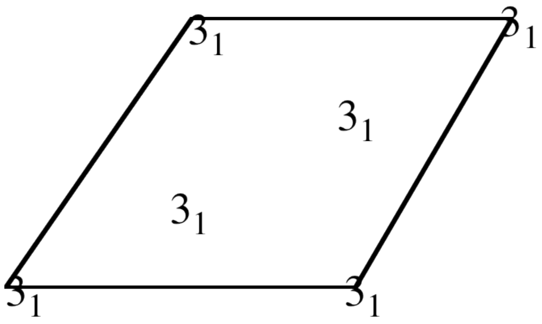
with h-k-l=3n, then k-h = 3n’-l and -2k-h = k-h - 3k = 3n”-l

$$\left(3n' - l\right)x_o + \left(3n'' - l\right)y_o + l/3 = \text{integer, etc.}$$

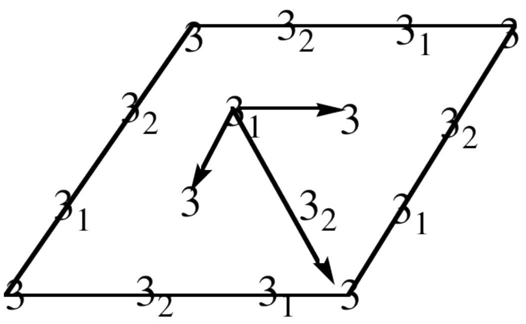
If x_o and y_o are multiples of 1/3, then the terms involving 3n’ and 3n” will automatically equal integers. Hence, the equation to be solved is:

$$\left(-x_o - y_o + 1/3\right) = \text{integer}$$

which has solutions (x_o,y_o) = (1/3,0), (0,1/3), (2/3,2/3), etc. It should be checked that these solutions actually do satisfy the extinction rule that observed reflections occur only when h-k-l=3n. Hence, to create an R3 cell by placing trimers with the three fold axis along *c* in a $P3_1$ cell, the three fold axes should be placed at any of these points. The same conclusion can be determined by inspection from the positions of the various symmetry axes in $P3_1$ and R3:



$P3_1$



R3

However, the actual problem is not to make an R3 cell from a $P3_1$ cell, but to see rhombohedral extinctions when h and k are both even. The way to achieve this is to translate the trimers from the R3 position by 1/2 along either x and/or y. Consequently, the three-fold axes in this case should be positioned at points such as (5/6,0), or (1/2,1/3), or (1/6,2/3), etc.

5. This problem will be solved in one-dimension, although the extension to three-dimensions is “straight-forward”.

$$T(u) = \int_0^1 \rho_o(x)\rho_m(x+u)\,dx$$

with $\rho(x) = \frac{1}{V}\sum_h F(h)e^{-2\pi i h x}$

$$T(u) = \int_0^1 \left[\frac{1}{V}\sum_k F_o(k)e^{-2\pi i k x}\right]\left[\frac{1}{V}\sum_h F_m(h)e^{-2\pi i h(x+u)}\right]dx$$

$$= \frac{1}{V^2}\sum_k\sum_h F_o(k)F_m(h)e^{-2\pi i h u}\int_0^1 e^{-2\pi i(h+k)x}\,dx$$

the integral equals 0, unless $h=-k$, and then equals 1

$$\therefore T(u) = \frac{1}{V^2}\sum_h F_o(\bar{h})F_m(h)e^{-2\pi i h u}$$

If $F_o(h) = A_o(h) + iB_o(h)$ and $F_m(h) = A_m(h) + iB_m(h)$

then $F_o(\bar{h})F_m(h) = A_c(h) + iB_c(h)$

where $A_c(h) = A_o(h)A_m(h) + B_o(h)B_m(h)$

and $B_c(h) = A_o(h)B_m(h) - B_o(h)A_m(h)$

So, *T(u)* can be calculated by a P1 Fourier series using coefficients A_c and B_c.

This problem can also be solved by the convolution theorem (the Fourier transform of the product of two functions is the convolution of the Fourier transforms of the two functions).

References to this so-called phased translation function can be found in:

- G.A. Bentley and A. Houdusse, *Acta cryst.* **A48**, 312-322 (1992)
- P.M. Colman, H. Fehlhammer and K. Bartels in *Crystallographic Computing Techniques* (F.R. Ahmed, K. Huml and B. Sedlacek, eds), pp. 248-258, Copenhagen:Munksgaard (1976)
- X. Zhu, *et al. Science* **251**, 90-93 (1991)

Problem set 5. Least squares

1. N pairs of points (u_i,v_i), i = 1, N are measured. You expect a linear relationship of the form: $v_i = m\,u_i + b$ to be satisfied by these observations. Derive the least squares expressions for the slope m and intercept b.

2. What is the least squares expression for the scale factor α that multiplies an intensity set F_h to minimize the deviation to a reference data set G_h?

3. The y origin is arbitrary in space group P1 (along with the x and z origins). Suppose that you have two sets of structure factors, F_1 and F_2, that were calculated from coordinate sets that used different y origins. Recall that the amplitudes of these two F sets will be identical, but the phases will be different due to the relative translation of the two structures along y. You wish to determine the difference in y origins to minimize $(\alpha_1-\alpha_2)^2$ by shifting one structure along y. In terms of a Taylor series expansion, the shift in y may be obtained from the following equation:

$$\alpha_1 = \alpha_2 + \frac{\partial \alpha_2}{\partial y}\Delta y$$

or $\Delta\alpha = \alpha_1 - \alpha_2 = \dfrac{\partial \alpha_2}{\partial y}\Delta y$

Solve for the y shift, using a non-linear least squares method based on minimizing $(\alpha_1-\alpha_2)^2$.
Hint: the following relationships may be helpful:

$$F = A + iB$$

$$\alpha = \tan^{-1}\frac{B}{A}$$

$$\frac{\partial\alpha}{\partial y} = \frac{\partial \tan^{-1}\frac{B}{A}}{\partial y} = \frac{1}{1+\frac{B^2}{A^2}}\frac{\partial\frac{B}{A}}{\partial y}$$

4. In the centric space group $P2_1/c$

$$F_c(hkl) = A(hkl) = 4\sum_{i=1}^{N} f_i e^{-B_i\frac{\sin^2\vartheta}{\lambda^2}}\cos 2\pi\left(hx + lz + \frac{k+l}{4}\right)\cos 2\pi\left(ky - \frac{k+l}{4}\right)$$

Evaluate $\dfrac{\partial |F_c|}{\partial p_j}$ for p = x, y, z and B.

Problem Set 5 Answers

The key to all these problems is to convert them to the form F=AX, where F is the vector of observables; A is a matrix of known coefficients; and X is a vector with the variables to be determined. The least squares solution to this equation is given by

$$X = (A^T A)^{-1} A^T F$$

$A^T A$ is also known as the normal matrix.

1. The slope m and intercept b in the equation $v_i = mu_i + b$ can be determined from pairs of (u_i,v_i), i=1,N using the correspondence: $F_i \rightarrow v_i$; $A_{i1} \rightarrow u_i$; $A_{2i} \rightarrow 1$; $X_1 \rightarrow m$ and $X_2 \rightarrow b$. With this, one can show:

$$A^T A = \begin{bmatrix} \sum u_i^2 & \sum u_i \\ \sum u_i & N \end{bmatrix} \quad \text{and} \quad A^T F = \begin{bmatrix} \sum u_i v_i \\ \sum v_i \end{bmatrix}$$

After suitable effort, the explicit solution can be obtained:

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} m \\ b \end{bmatrix} = \frac{1}{N\sum u_i^2 - \left(\sum u_i\right)^2} \begin{bmatrix} N\sum u_i v_i - \sum u_i \sum v_i \\ \sum v_i \sum u_i^2 - \sum u_i \sum u_i v_i \end{bmatrix}$$

2. One wants to minimize $\Phi = \sum_h \left(G_h - \alpha F_h\right)^2$ with respect to α. Following the derivation outlined in the problem set, one finds:

$$\alpha = \frac{\sum_h F_h G_h}{\sum_h G_h^2}$$

3. In this problem, structure 1 (and hence α_1) is kept fixed, while α_2 is varied while moving structure 2 along the y axis by an amount δ. The correspondence to F=AX is given by:

$$F_i = \Delta\alpha_i = \left(\alpha_1 - \alpha_2\left(\delta\right)\right)$$

$$A_i = {\partial \alpha_{2,i}}\big/{\partial \delta} \text{ so that } A^T A = \sum_i \left({\partial \alpha_{2,i}}\big/{\partial \delta} \right)^2$$

$$X = \delta = \text{ shift in y origin for structure 2}$$

The key derivative is ${\partial \alpha_{2,i}}\big/{\partial \delta}$:

$$\frac{\partial \alpha}{\partial \delta} = \frac{1}{1 + B^2\big/A^2} \frac{\partial B/A}{\partial \delta} = \frac{1}{A^2 + B^2}\left[A\frac{\partial B}{\partial \delta} - B\frac{\partial A}{\partial \delta} \right]$$

$$A = \sum_{\text{atoms } j} f_j \cos 2\pi\left(hx_j + k\left(y_j + \delta\right) + lz_j\right)$$

$$\frac{\partial A}{\partial \delta} = -2\pi k \sum_{\text{atoms } j} f_j \sin 2\pi\left(hx_j + k\left(y_j + \delta\right) + lz_j\right) = -2\pi k B$$

$$\frac{\partial B}{\partial \delta} = 2\pi k B$$

$$\frac{\partial \alpha}{\partial \delta} = 2\pi k$$

$$\therefore \delta = \frac{\sum_i \left(\frac{\partial \alpha_i}{\partial \delta} \right) \Delta\alpha_i}{\sum_i \left(\frac{\partial \alpha_i}{\partial \delta} \right)^2} = \frac{\sum_i k_i \Delta\alpha_i}{2\pi \sum k_i^2}$$

4. For space group P2$_1$/c, $\frac{\partial |F_c|}{\partial p_j} = \cos\alpha \frac{\partial A_c}{\partial p_j}$, with cosα= ± 1.

The derivatives can be evaluated in straight-forward, if not tedious, fashion:

$$\frac{\partial A}{\partial x_j} = -8\pi h f_j e^{-\left(B_j \sin^2\vartheta/\lambda^2\right)} \sin 2\pi\left(hx_j + lz_j + \frac{k+l}{4}\right) \cos 2\pi\left(ky_j - \frac{k+l}{4}\right)$$

$$\frac{\partial A}{\partial y_j} = -8\pi k f_j e^{-\left(B_j \sin^2\vartheta/\lambda^2\right)} \cos 2\pi\left(hx_j + lz_j + \frac{k+l}{4}\right) \sin 2\pi\left(ky_j - \frac{k+l}{4}\right)$$

$$\frac{\partial A}{\partial z_j} = -8\pi l f_j e^{-\left(B_j \sin^2\vartheta/\lambda^2\right)} \sin 2\pi\left(hx_j + lz_j + \frac{k+l}{4}\right) \cos 2\pi\left(ky_j - \frac{k+l}{4}\right)$$

$$\frac{\partial A}{\partial B_j} = -\frac{4\sin^2\vartheta}{\lambda^2} A$$