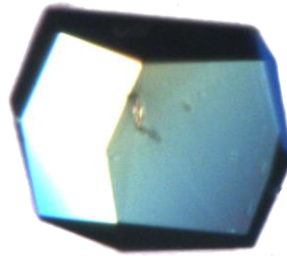


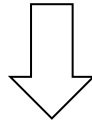
The Phase Problem

Dr. Ed Lowe –
edward.lowe@bioch.ox.ac.uk

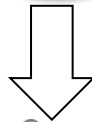
Phasing



Grow Crystals

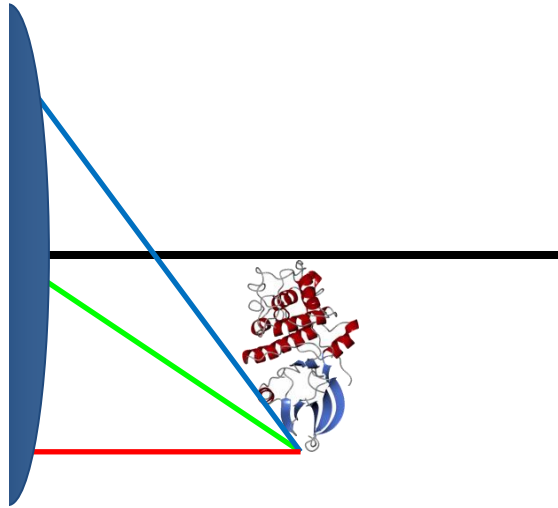


Magic happens...



Structure

Electron Density Equation



- Normally a lens will
 - Bend the scattered light pattern
 - Apply a phase shift
- Each “pixel” of the image formed is therefore a sum of many scattered waves. These will add up in a way that depends upon their phase and on the position of the pixel.
- It is therefore not surprising that the electron density equation looks the way that it does.

All reflections



$$r(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F(h, k, l)| \exp[-2\pi i(hx + ky + lz) + i\alpha(h, k, l)]$$



Electron density at a point



Amplitude

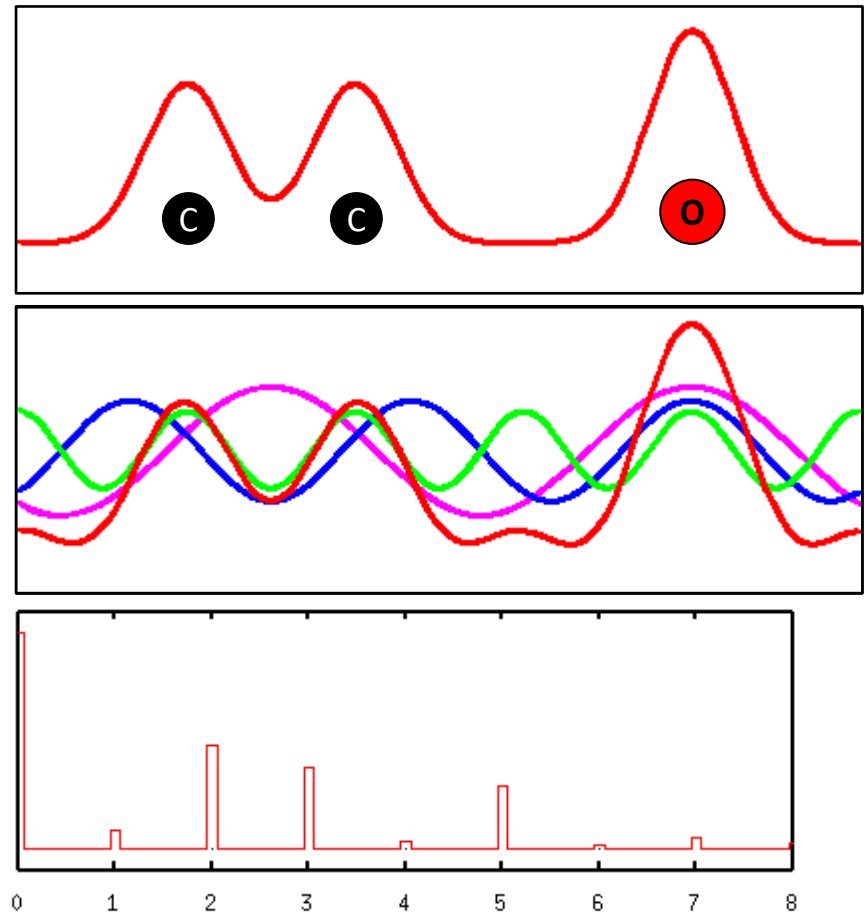


Phase

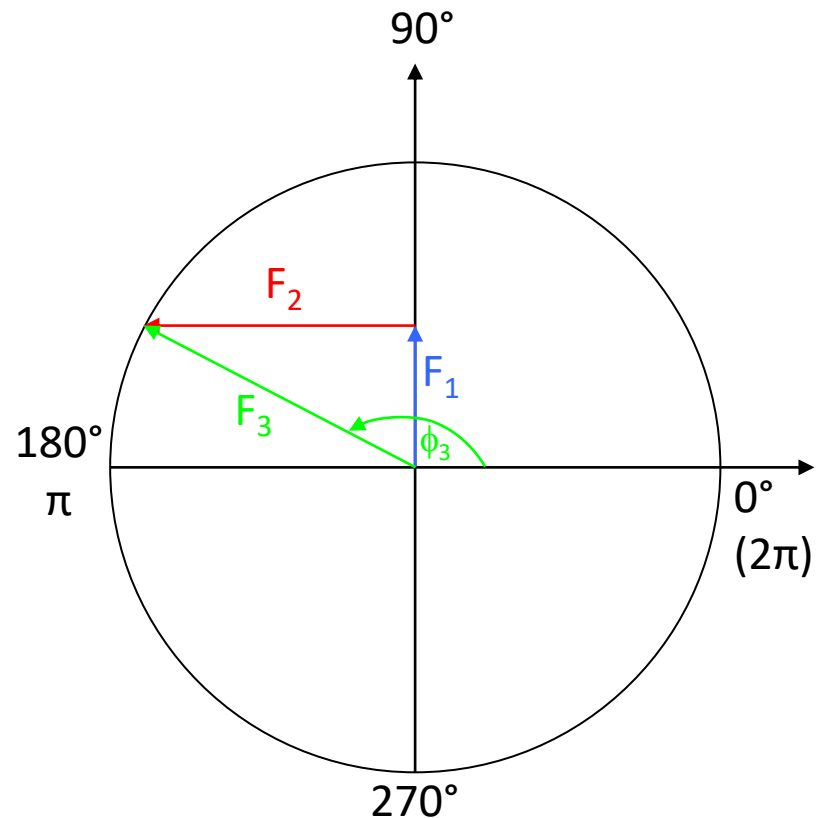
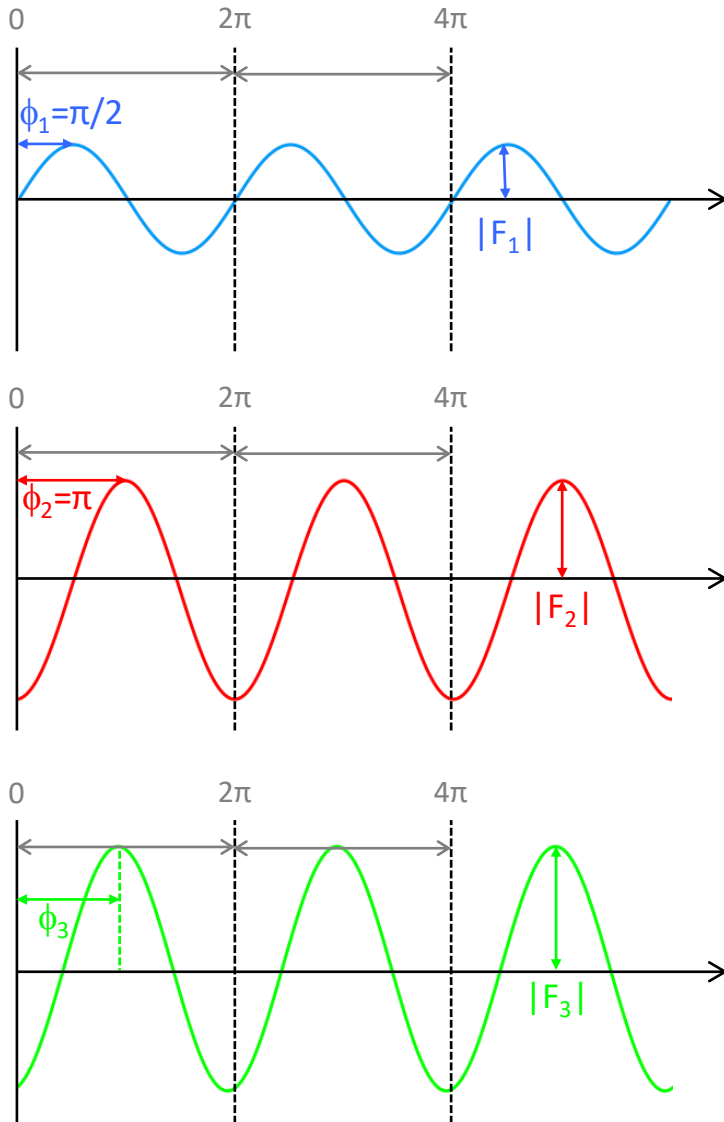
Fourier Syntheses and Transforms

A Fourier synthesis is the representation of a function in terms of a set of sine waves

- A simple crystal with a unit cell containing one oxygen and two carbon atoms
- Trying to represent the unit cell in terms of sine waves, we start by adding a wave of frequency 2
- Note that one peak crosses the oxygen atom whilst the other spans the two carbons
- Now we add a wave of frequency 3. It has a different phase, meaning that we start at a different point in the wave. The amplitude is also different
- Finally we add a wave of frequency 5. Two of the peaks correspond to the carbon atoms
- The sum of the waves is a good approximation of the contents of the unit cell (given appropriate choices of frequency, amplitude and phase)
- The Fourier transform of the unit cell consists of a series of peaks, the largest of which are at 2, 3 & 5 on the x-axis
- These correspond exactly to the sine-wave frequencies used to reconstruct the unit cell. The peak height also corresponds to the sine-wave amplitudes
- For more Fourier theory and some excellent figures:
<http://www.ysbl.york.ac.uk/~cowtan/fourier/fttheory.html>

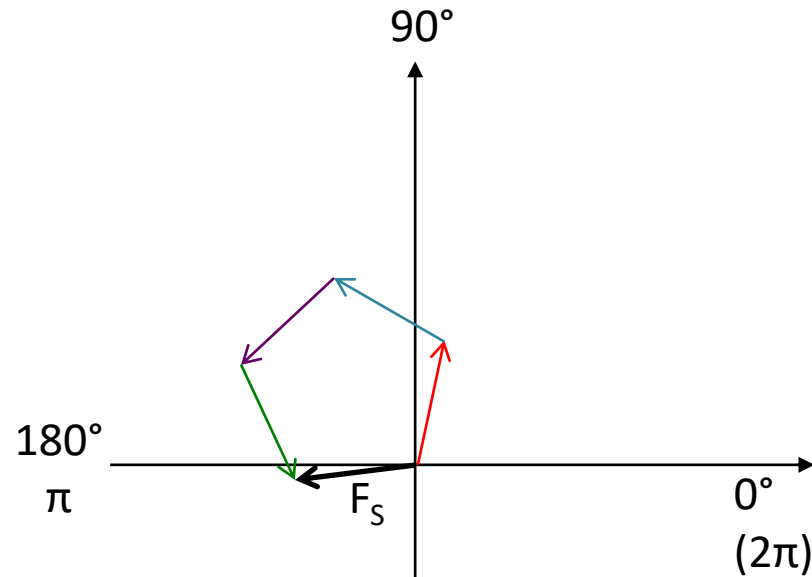
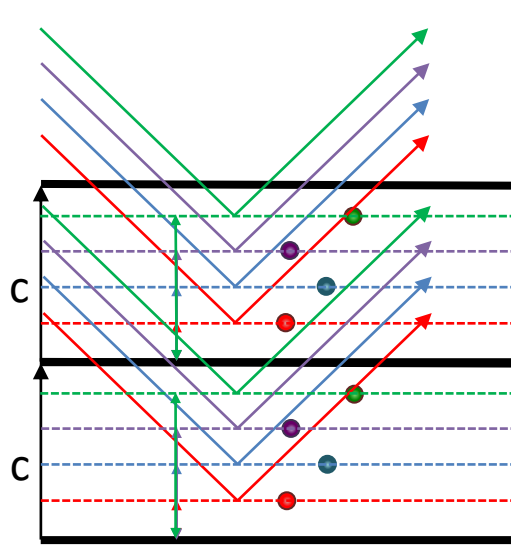


Addition of waves in the complex plane



NB. Experiments are monochromatic, so all waves have the same wavelength

Addition of atomic scattering vectors to produce a structure factor



- When the diffraction condition is satisfied for a particular set of planes (drawn in black), each atom contributes some amount to the total scattering.
- The relative phase of that contribution depends on the position of the atom (fractional z-coordinate shown).
- The total scattering is the vector sum of the individual atomic scattering vectors.
- This is referred to as a structure factor because it is dependant on the arrangement (or structure) of atoms in the unit cell

How will the structure factor change if our atoms remain in the same positions but we consider a different set of Bragg planes?

A: The same phase but different amplitude

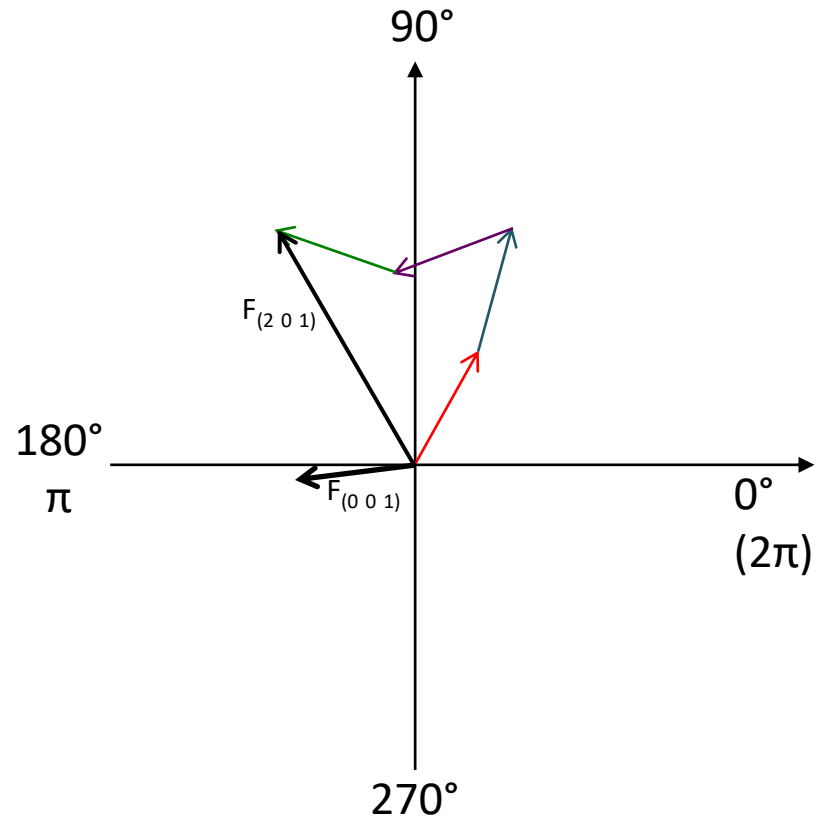
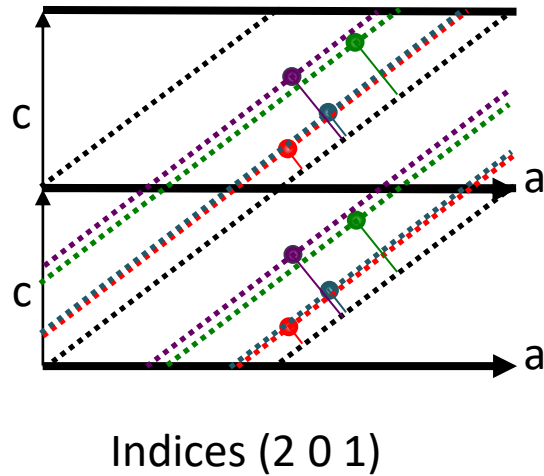
B: The same amplitude but different phase

C: The same phase and the same amplitude

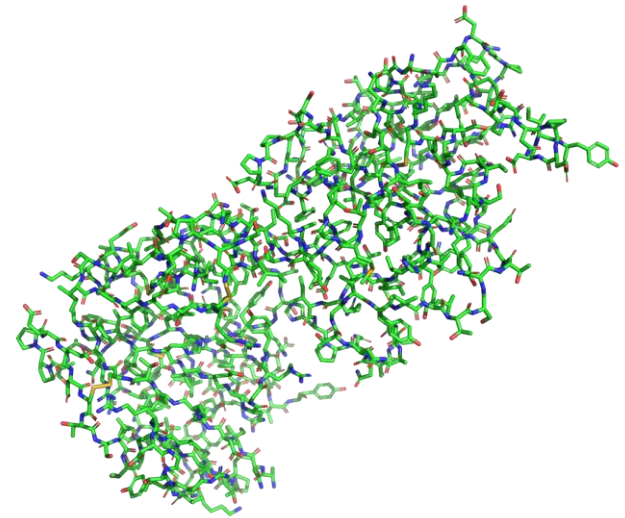
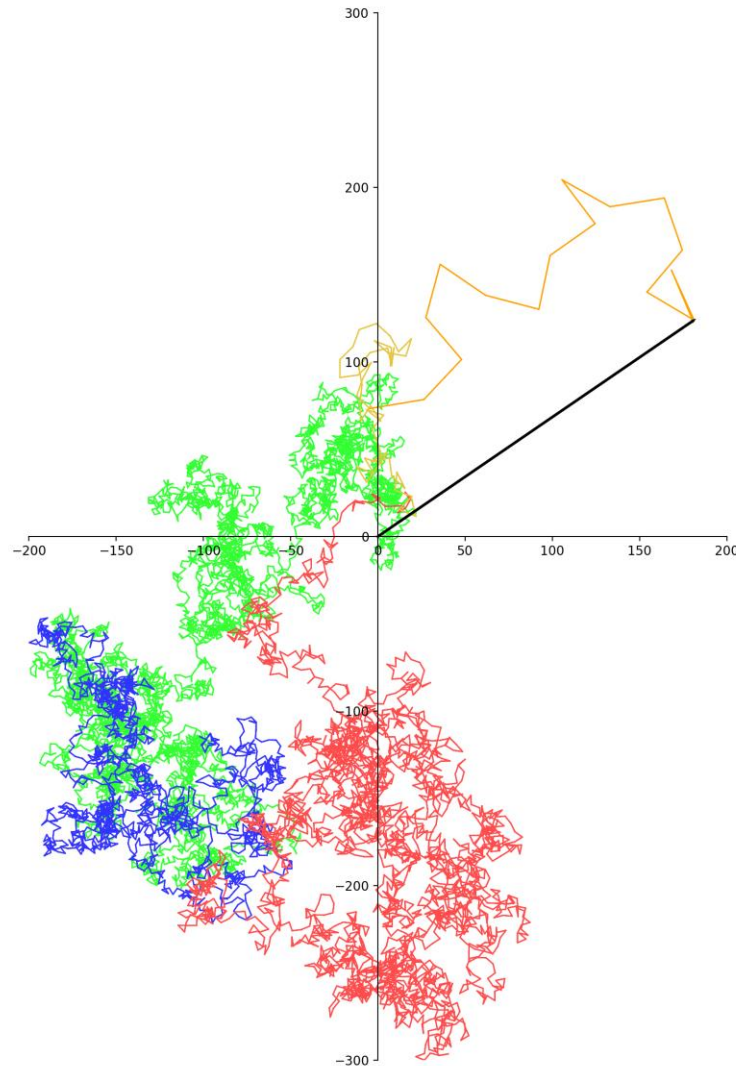
D: A different phase and a different amplitude

E: It will have a different name

How does this look with a different set of Bragg planes?



A bit more complicated for a protein...



What can we achieve if phases are known?

Convolution

- Convolution of two functions mixes them together. For functions f and g this can be expressed by the formula:

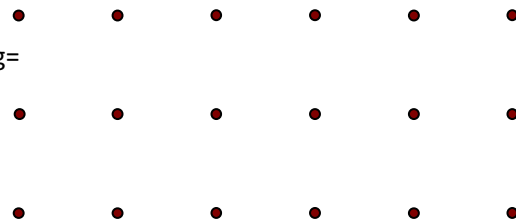
$$C(x) = \int_{h=0}^1 f(h)g(x-h)dh$$

- It might be helpful to illustrate this:

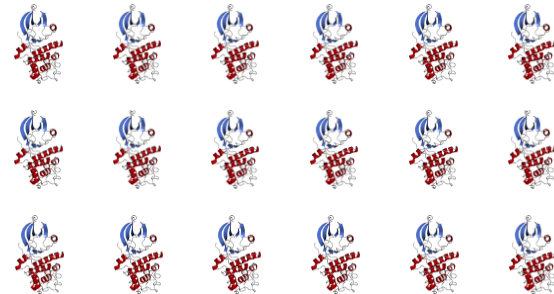
If $f=$



and $g=$



- The convolution of f and g (written $f*g$) is:



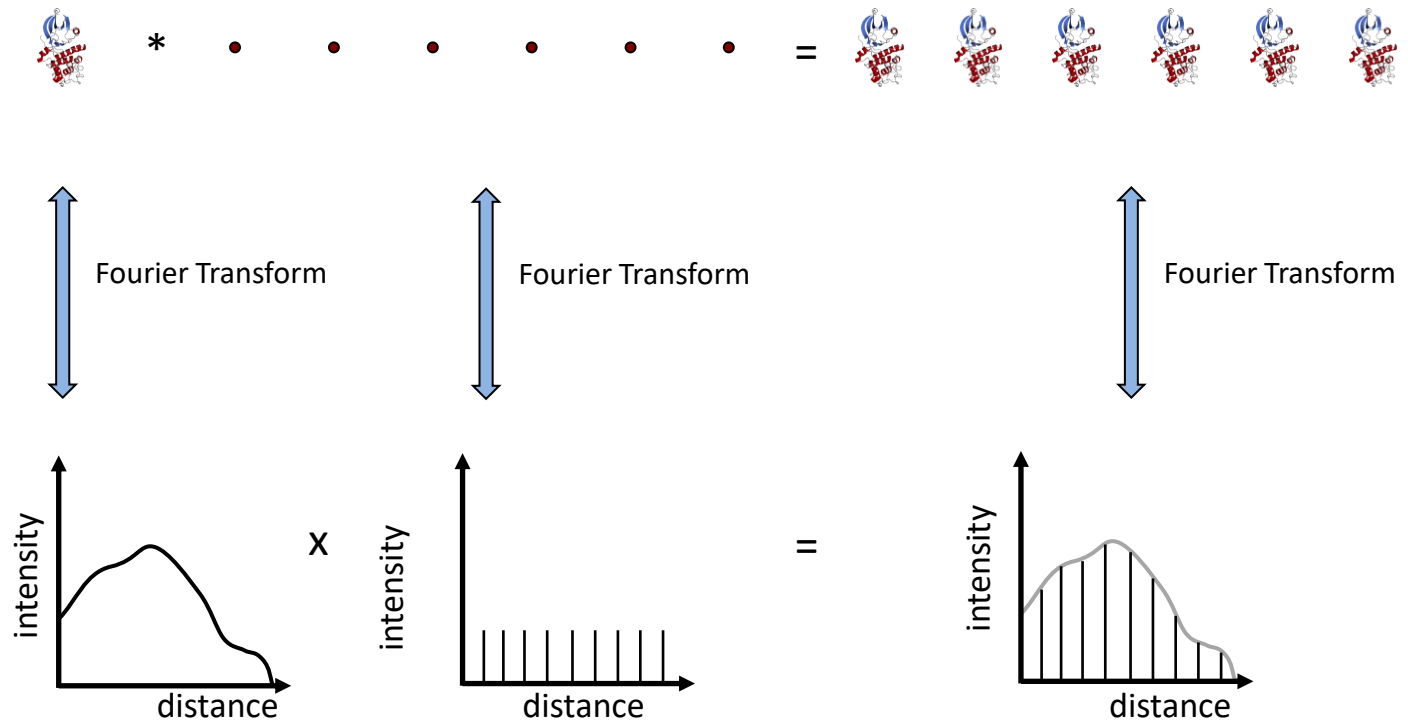
- The convolution theorem states that the Fourier transform of the convolution of two functions is equal to the product of the Fourier transforms of the two separate functions.

- We can see how this relates to crystals, since a crystal is the convolution of a molecule with an infinite three dimensional grid
- Its Fourier transform (and hence its diffraction pattern) is therefore the product of the Fourier transform of one molecule, and the Fourier transform of an infinite grid.
- The Fourier transform of the crystal lattice (i.e. the infinite grid) is the reciprocal lattice (i.e. another infinite grid)
- The Fourier transform of a molecule is a smooth continuous function.
- **The product of these two functions is the diffraction pattern of a protein crystal: a smooth function sampled at reciprocal lattice points.**

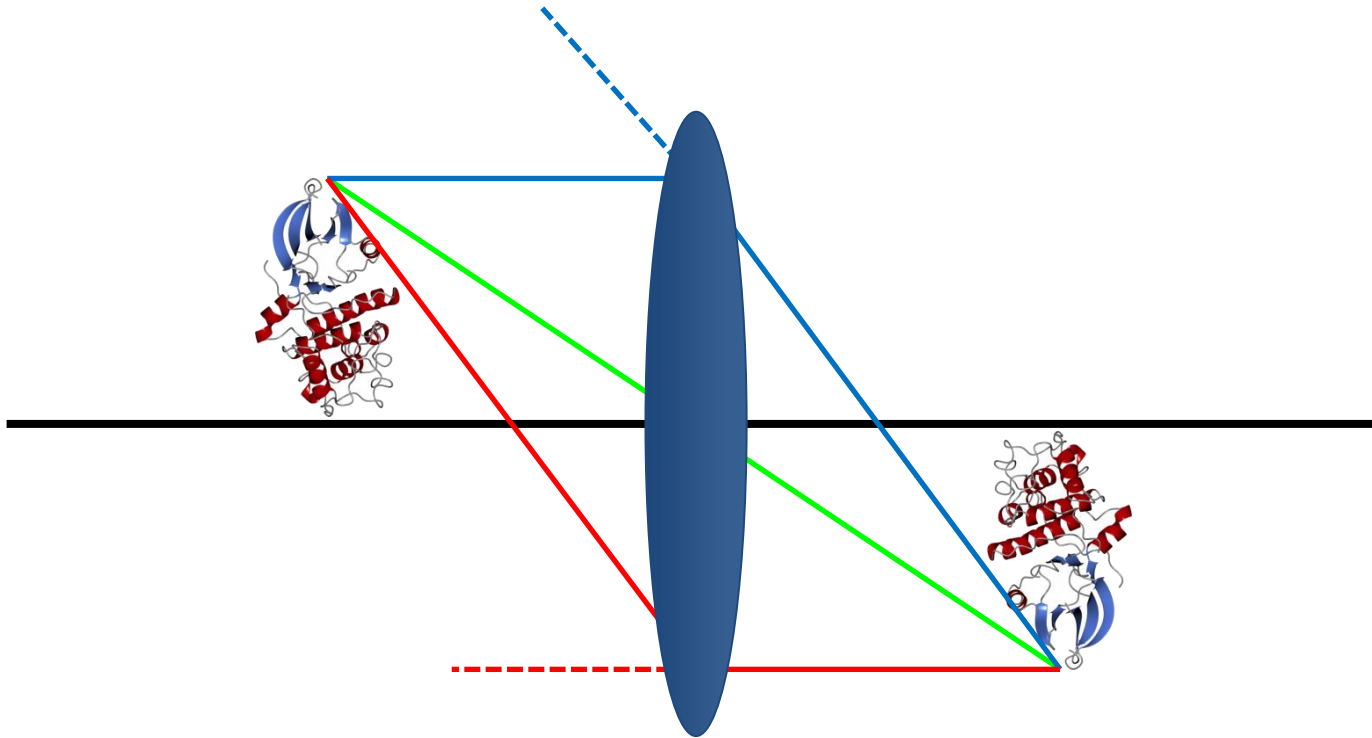
Inversion

- The inversion theorem states that:
If function F is the Fourier transform of function G , then function G is the Fourier transform of function F
- This is important because:
In the same way as we can use Fourier transforms to calculate the diffraction pattern of a protein crystal, we can use Fourier transforms to calculate the electron density distribution in a crystal from an observed diffraction pattern

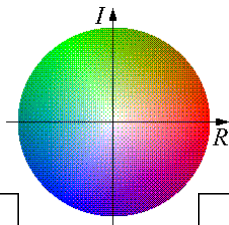
Putting all of this together:



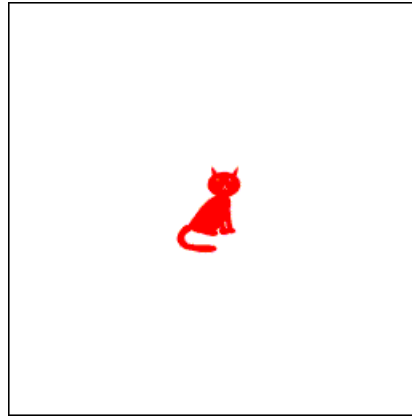
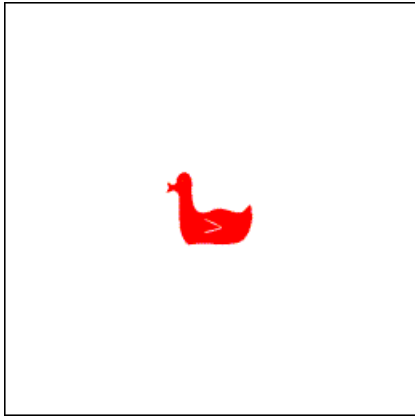
The Phase Problem



Our problem is how to achieve the recombination of scattered waves without having to invent an X-ray lens

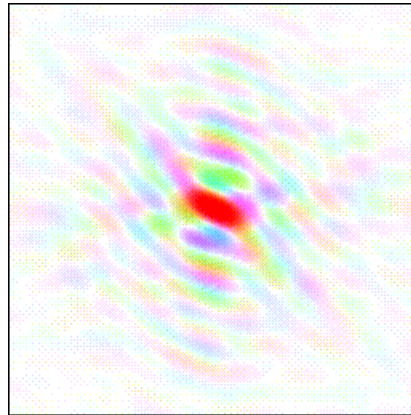
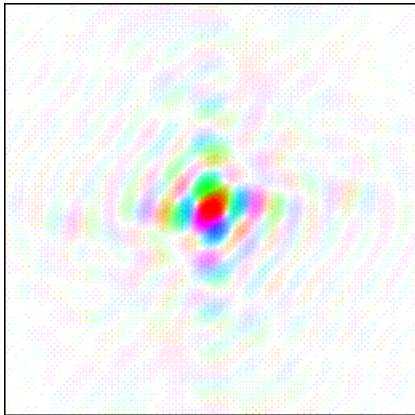


Representing intensity



Fourier transform

Fourier transform



$F(\text{Duck}), \phi(\text{Duck})$

$F(\text{Cat}), \phi(\text{Cat})$

What will the reconstructed image most resemble?

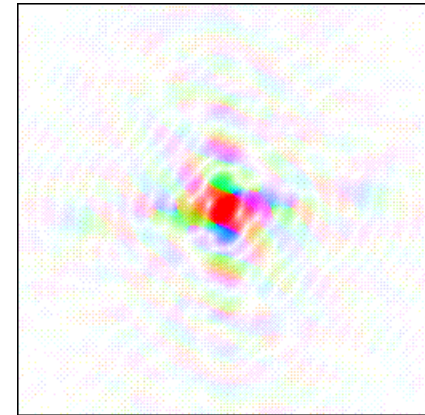
A: A Duck

B: A Cat

C: Mixture (more duck)

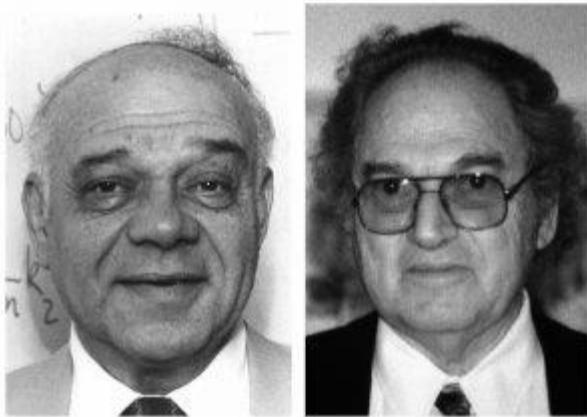
D: Mixture (more cat)

E: A Tin of Sardines



$F(\text{Duck}), \phi(\text{Cat})$

This even applies to Nobel prize winners!



- Jerome Karle (left)
- Herb Hauptman (right)



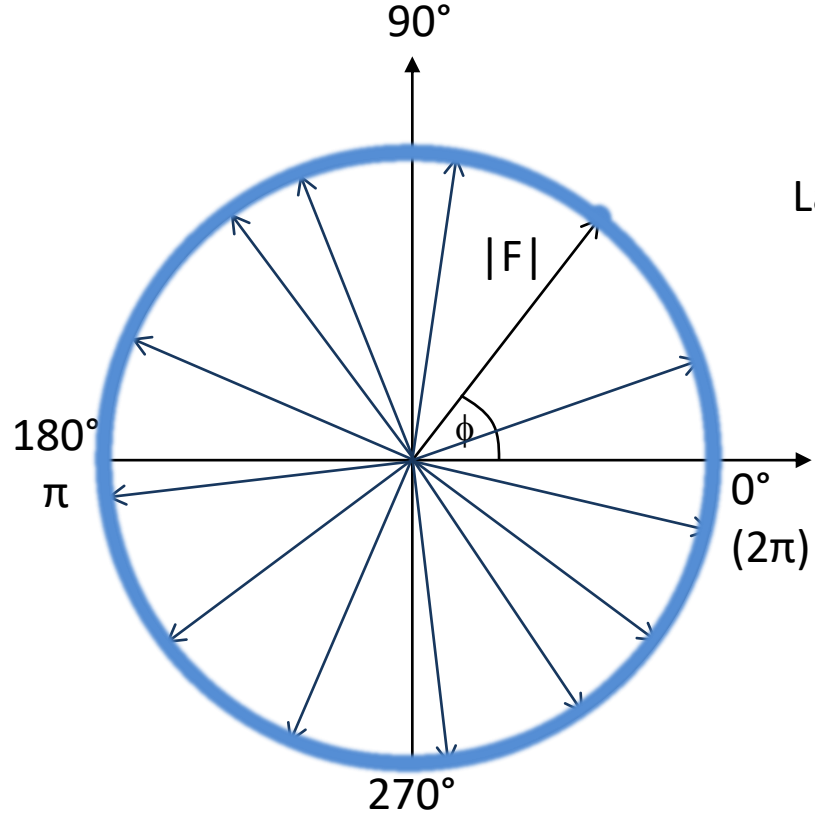
- $|F|_{\text{Karle}}, \phi_{\text{Hauptman}}$
- $|F|_{\text{Hauptman}}, \phi_{\text{Karle}}$

This sounds pretty hopeless – surely we can never trust anything we see in a crystal structure unless experimental phases were measured?

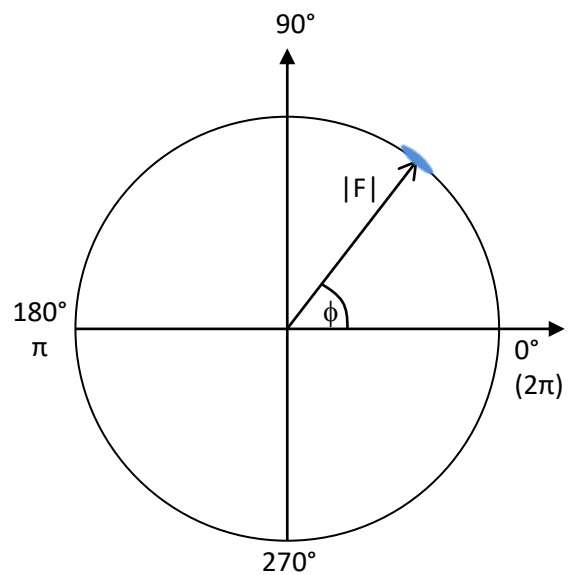
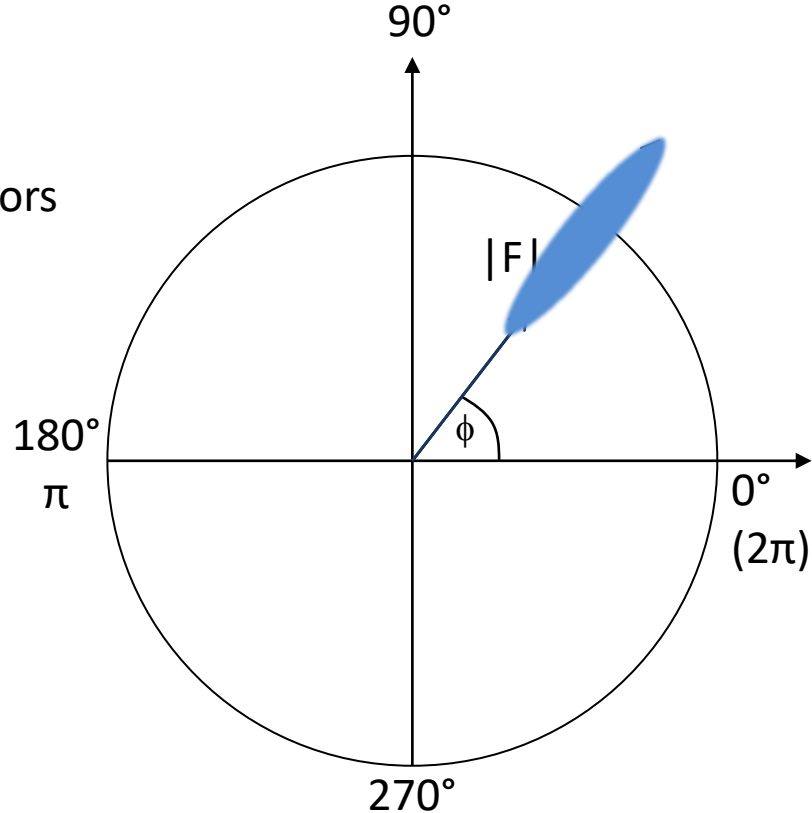
True or False?

But... there is no need to panic just yet!

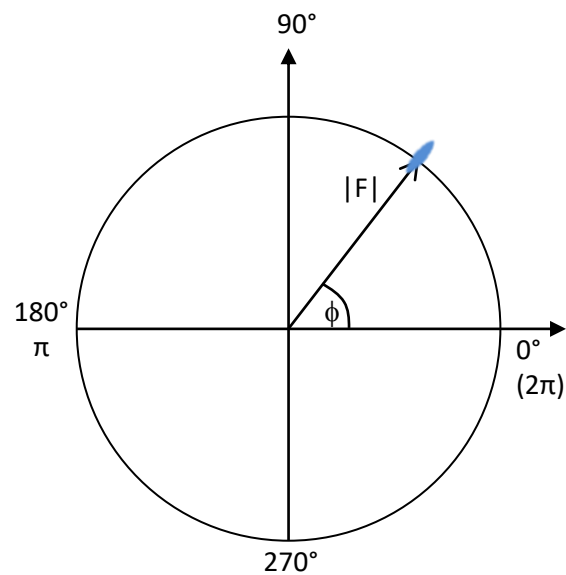
- This disastrous situation is greatly exaggerated when the two sets of phases are completely different.
- It relates to the fact that the vector you get by pointing a correct length vector in the wrong direction is (on average) further away from the true result than that obtained by pointing a vector of any random length in the right direction.
- Where the phase is close to the correct value, this ceases to be the case. This is why we can see features in difference maps $(F_o - F_c)$, ϕ_{calc}
- Despite this, we would achieve a result much closer to the truth if we were able to always record phases with our amplitudes.

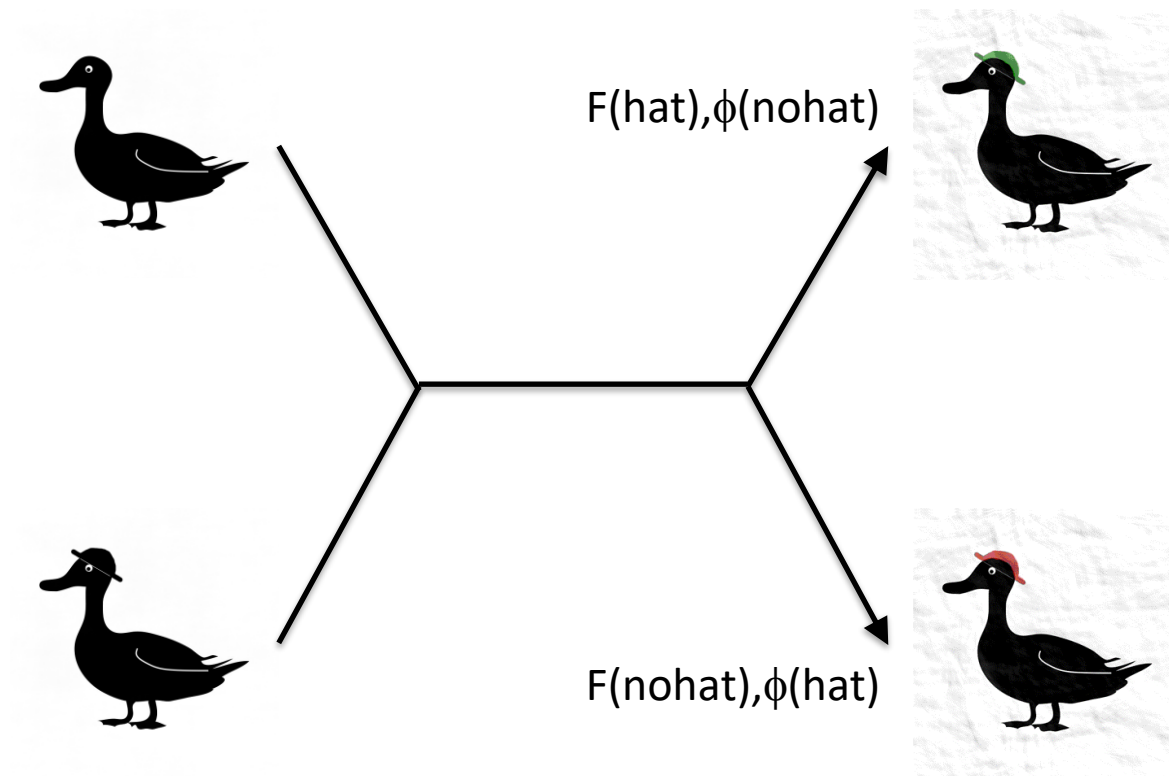


Large errors



Small errors





Phasing

When we talk about solving protein structures we mean we want to calculate the electron density.

The structure factors are calculated from the intensities of the measured reflections, these in turn can be described as a function of electron density.

Refers to the
volume of
the unit cell

As a result of
discrete spots

$$\rho(xyz) = \frac{1}{V} \sum_{hkl} |F(hkl)| e^{-2\pi i(hx + ky + lz) + i\phi(hkl)}$$

Electron density

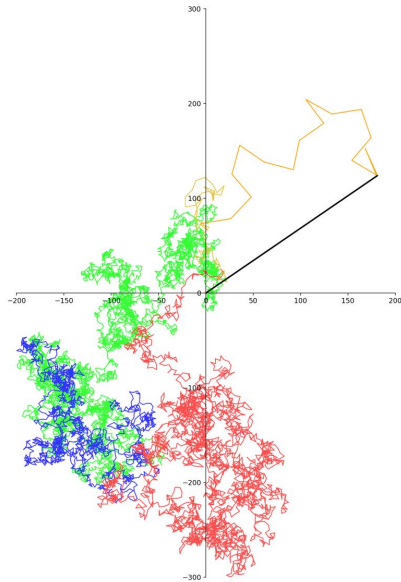
Sum of Structure
Factors for all hkl

Phase contribution

Solutions to the Phase Problem

- "Copy" the phase contributions from a sufficiently similar known structure
- Calculate the phase contributions from experimental data

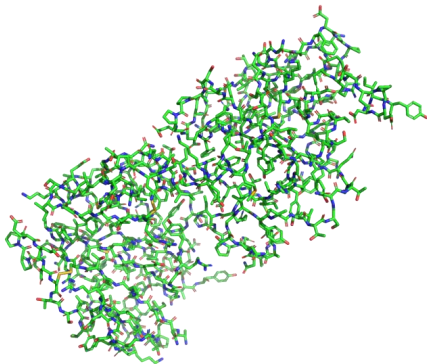
Think back to this Argand diagram...



- If the constellation of atoms is sufficiently similar.
- And if it can be placed in an appropriate position and orientation in the unit cell
- Then the structure factors calculated from these atoms will be similar to those derived from our experimental structure
- Meaning that the phase terms can be borrowed to allow the calculation of an initial electron density map

This is Molecular Replacement

Thanks to AlphaFold and RosettaFold we now have access to a suitable model structure in most cases



What can we achieve if phases are not known?

Direct methods

Structure factors are not independent of each other – they are related through the structure.

All atoms on a lattice plane scatter in phase, additionally those in parallel planes scatter in phase.

This leads to the derivation of the triplet phase relationship. This gives a reasonable estimation of phase for strong reflections.

$$j_{-h} + j_k + j_{h-k} \gg 0$$

- This works only if:
 - The magnitudes of the structure factor amplitudes are large – this relationship applies to strong reflections.
 - The atoms within the structure must be fully resolved from each other (for a protein structure this would require resolution of better than 1.0Å)
 - Relatively few atoms in the unit cell (approx. up to 1000)
- This method can be useful in solving the heavy atom substructures.
- There are fewer heavy atoms in the unit cell and they are further apart, so high resolution data is not needed.

Patterson function

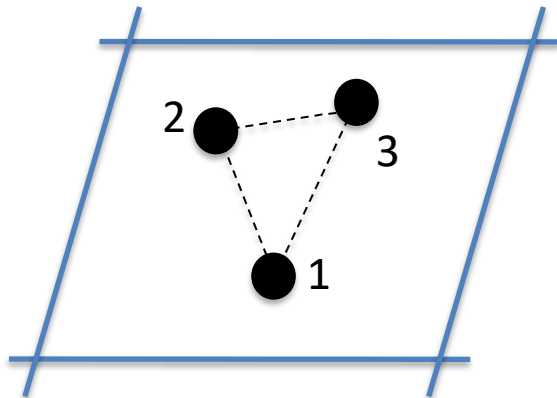
This is a Fourier summation of intensities (as opposed to Structure factors) with phase angles set to zero.

$$P(uvw) = \frac{1}{V} \sum_{hkl} |F(hkl)|^2 \cos[2\pi(hu + kv + lw)]$$

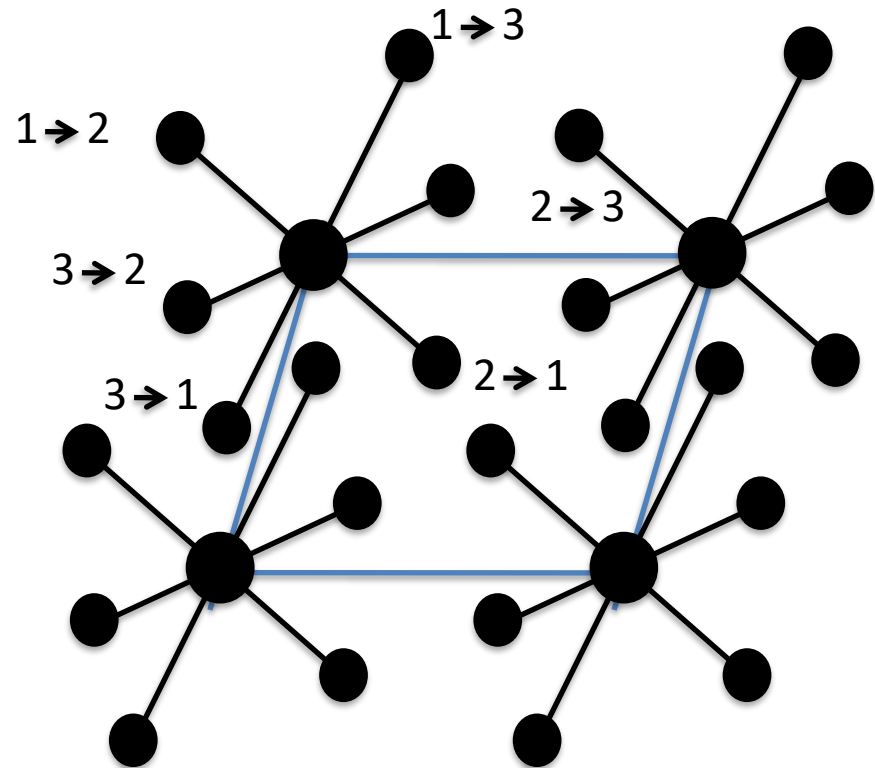
Where u , v and w are the coordinates in the Patterson cell to avoid confusion with x , y and z from the real unit cell (although both unit cells have the same dimensions).

By solving the Patterson equation for the measured intensities you can generate a vector map between the atoms.

Patterson Map



Unit cell containing 3 atoms



Patterson map for the 3 atoms. At the origin we see larger peaks as every atom can be placed at the origin in calculating the map. The peaks in the map are the vectors of the original 3 atoms

How many peaks will there be in the Patterson function of a small protein with 1000 atoms?

A: 9990

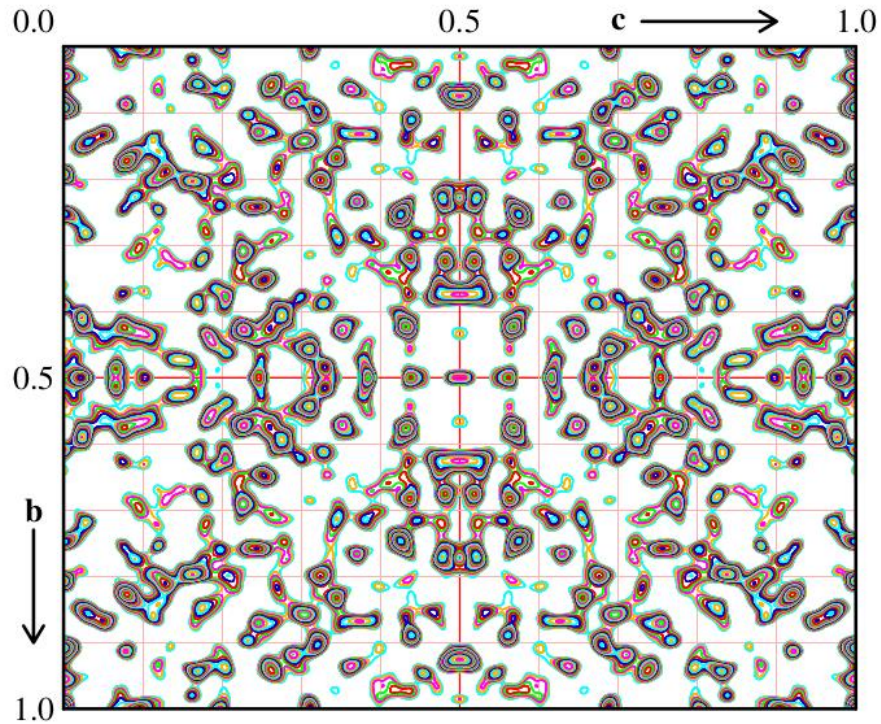
B: 99,900

C: 999,000

D: 9,990,000

E: 42

Patterson map



A protein contains 1000's of atoms that all contribute to its diffraction. The Patterson becomes too complex and you can't solve the atomic positions. A unit cell containing N atoms will generate a Patterson map containing $N^2 - N$ non-origin peaks.

- Is there a way we can utilise Patterson functions to help us phase the protein diffraction data?

So, what methods can we apply to a protein structure containing 5000 atoms at a resolution of 2Å?

A: Direct Methods

B: Patterson Methods

C: A combination of Direct and Patterson Methods

D: Neither, we need a simpler structure for either to work.

E: Throw all of our data at a computer and hope it works it out for us!

A simpler structure...

- We need some way by which we can identify the positions of a small number of atoms – from this we can estimate the phases for all reflections and bootstrap our way up.
 - Atoms placed at random positions
 - Atoms with much higher atomic number than the C, O, N and S making up most of protein molecules
 - Atoms that scatter ‘anomalously’ in a wavelength dependent manner