Experimental Phasing

Dr. Ed Lowe

edward.lowe@bioch.ox.ac.uk

Department of Biochemistry

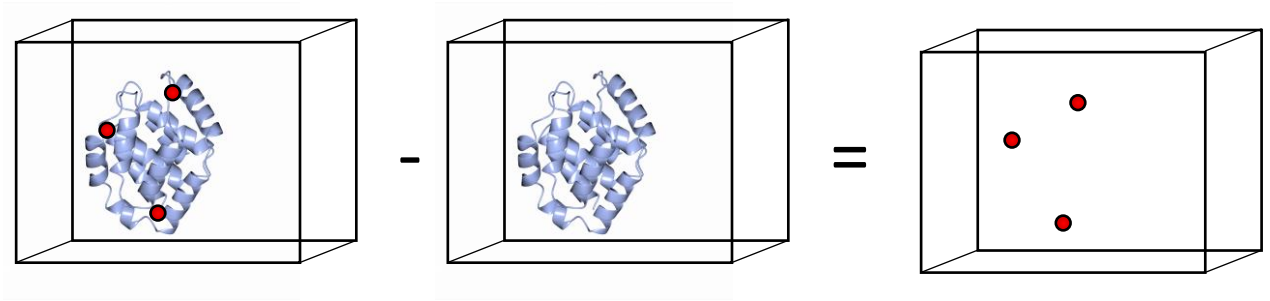
UNIVERSITY OF OXFORD

Isomorphous replacement

To create a substructure the easiest way is to incorporate atoms heavier than the atoms normally found in proteins.

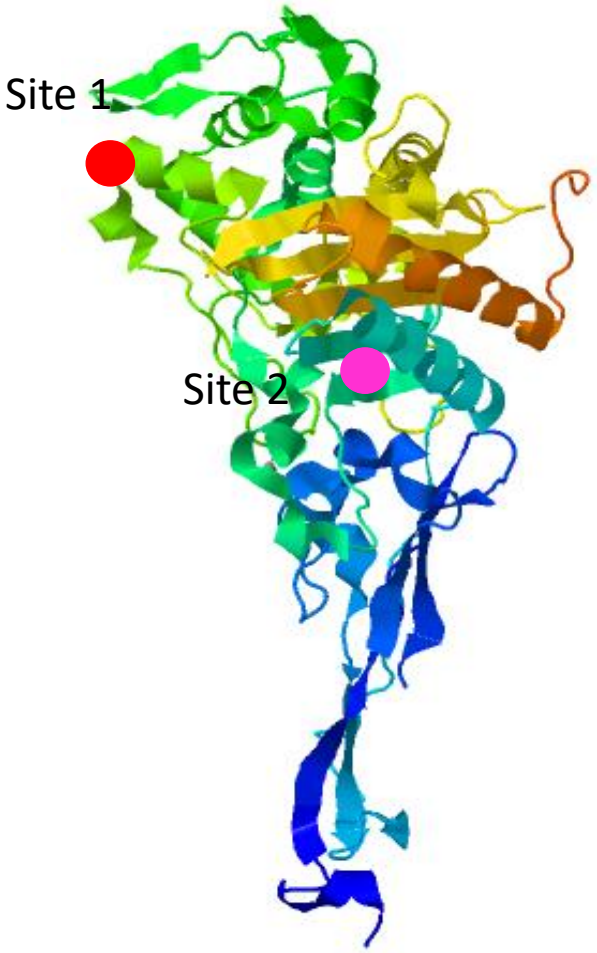
Isomorphous replacement compares the diffraction data from a native protein crystal (nothing bound) to one where a heavy atom has been bound.

The Patterson function (or Direct Methods) is then used to calculate the positions of the heavy atom(s).



$$F_{PH} \quad - \quad F_P \quad = \quad F_H$$

Binding heavy atoms



- To incorporate heavy atoms into the protein crystal the crystals are soaked in solutions containing heavy atom salts.
- The heavy atoms bind to amino acids, metal binding sites or hydrophobic patches.
- Use atoms that are more electron dense but have similar radii to common protein binding ions.
- Common heavy atoms used are mercury, lead, samarium, platinum and gold.
- Uranium is not so common these days!

Data requirements

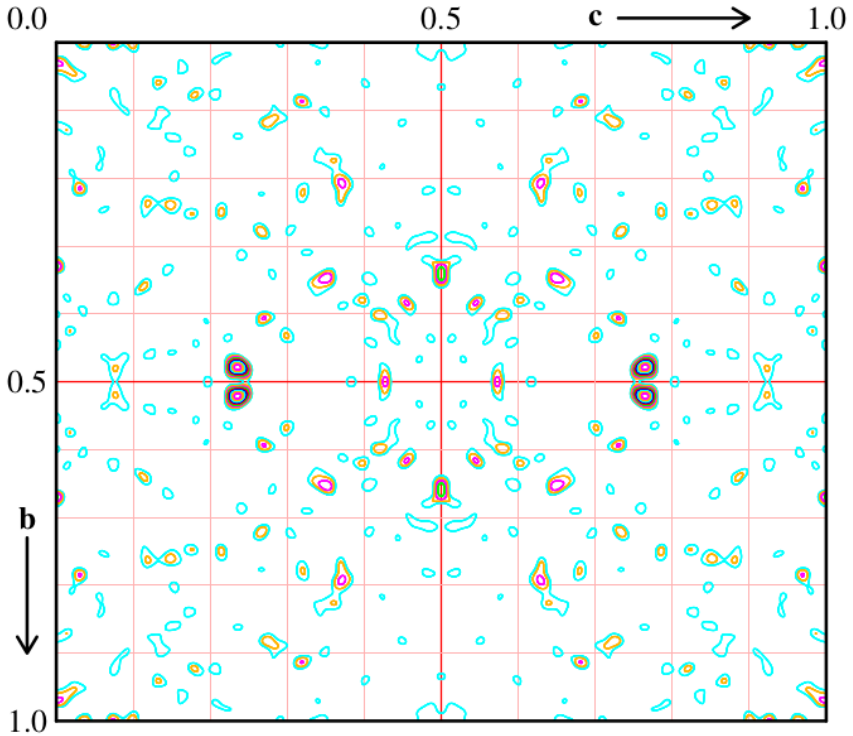
A native data set and an isomorphous heavy atom data set

Finding Our Heavy Atoms

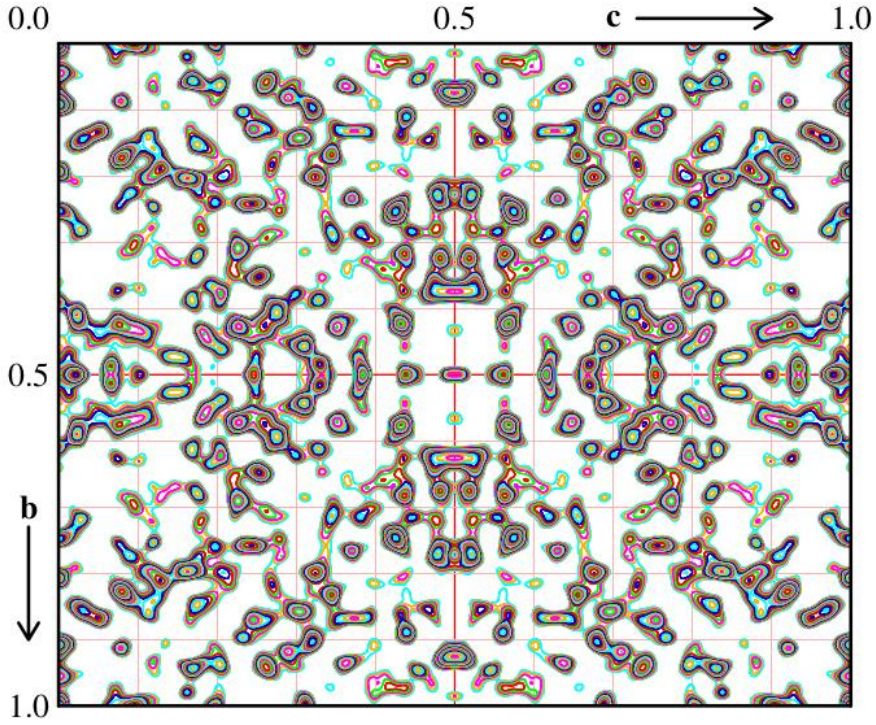
- Patterson Methods
- Direct Methods

- These days, both approaches are typically used in tandem.
- SHELXD, HySS, SnB

Patterson map of simpler substructure from the same protein we saw earlier



Substructure Patterson

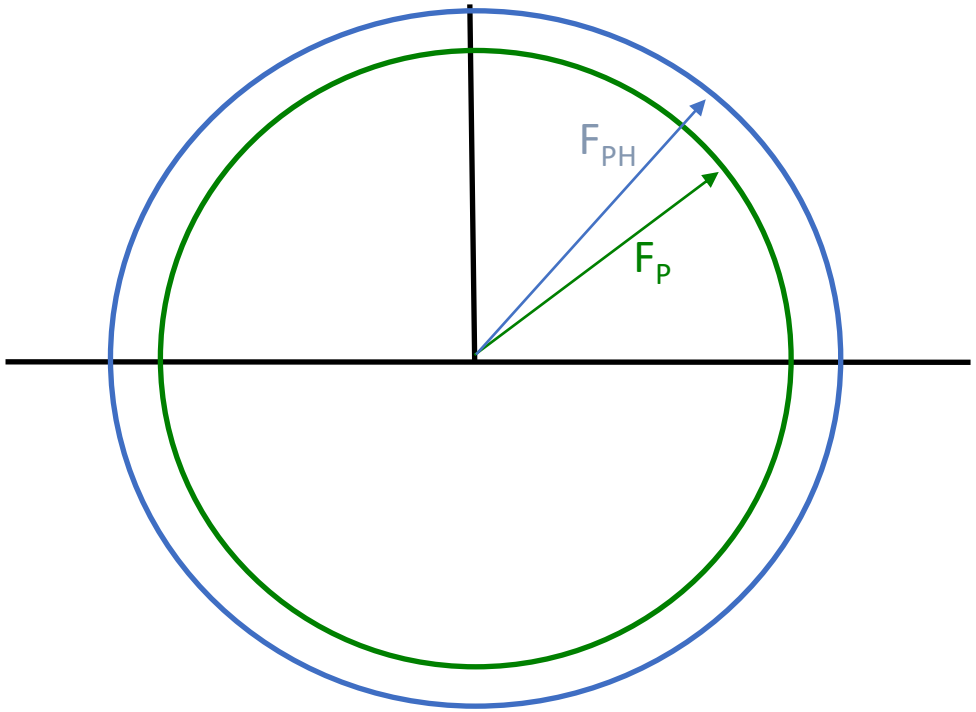


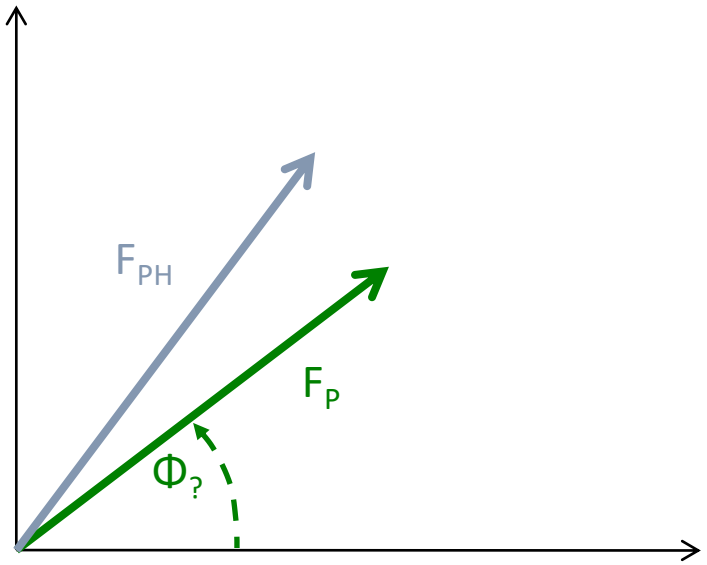
Native Patterson

The story so far…

- At this point we know the positions of a small subset of atoms in our structure
- We refer to this as the substructure
- But how can we use this to calculate phase terms for the entire structure?

Solving the phases using Isomorphous Replacement

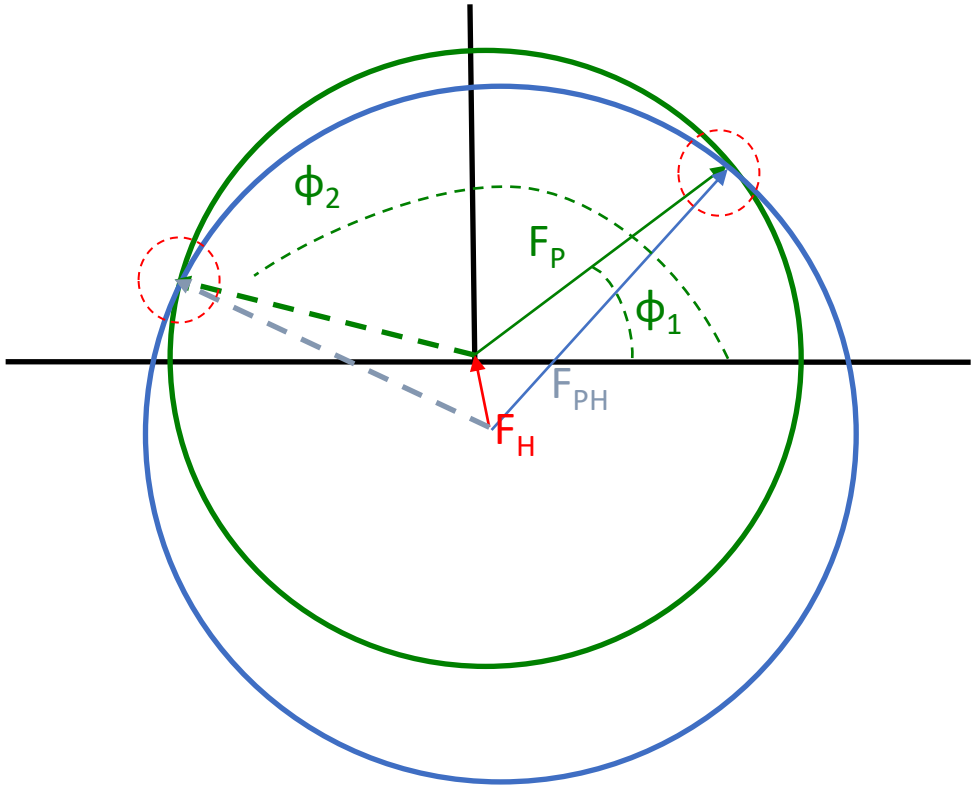




We collect Native (F_P) data

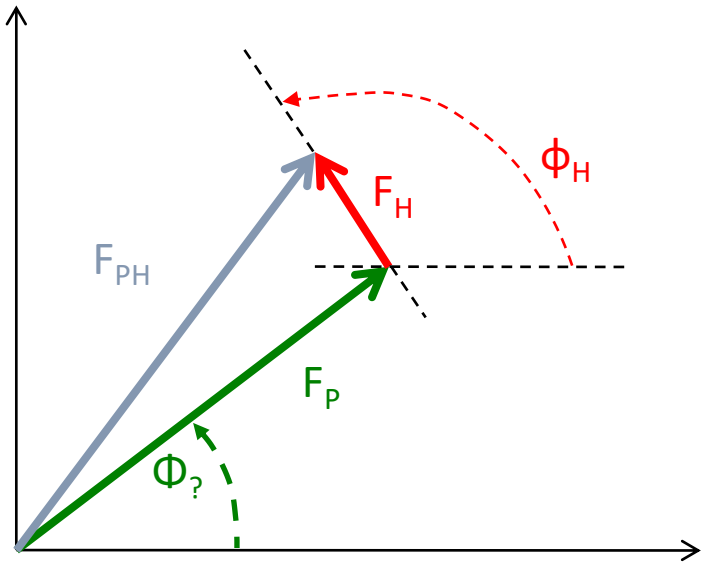
The phase is unknown

Next we collect the Derivative (F_{PH}) data



We offset the F_{PH} term from the origin by the value of the F_H term

The points where the two circles intersect are possible solutions for the phase. This is known as a Harker construction



$F_{PH} = F_P + F_H$

Using the Patterson Function we can solve the position and phase of the heavy atom (F_H)

So, we have two possible solutions but only one can be right – how can we solve this ambiguity?

A: Solve for both solutions and see which one is correct

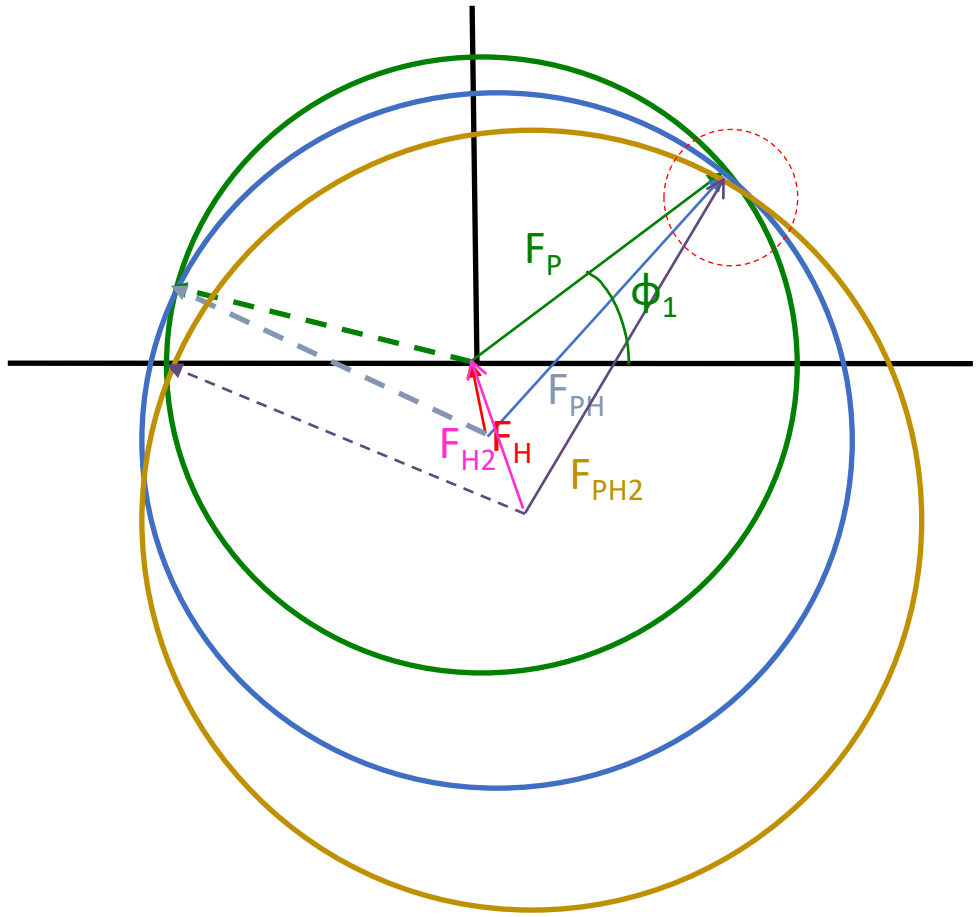
B: Take a guess – one of them has to be right

C: Repeat the whole experiment with a second derivative and look for agreement with one of the first derivative solutions

D: Select the solution that generates a map compatible with a protein composed of L-amino acids

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

Multiple Isomorphous Replacement

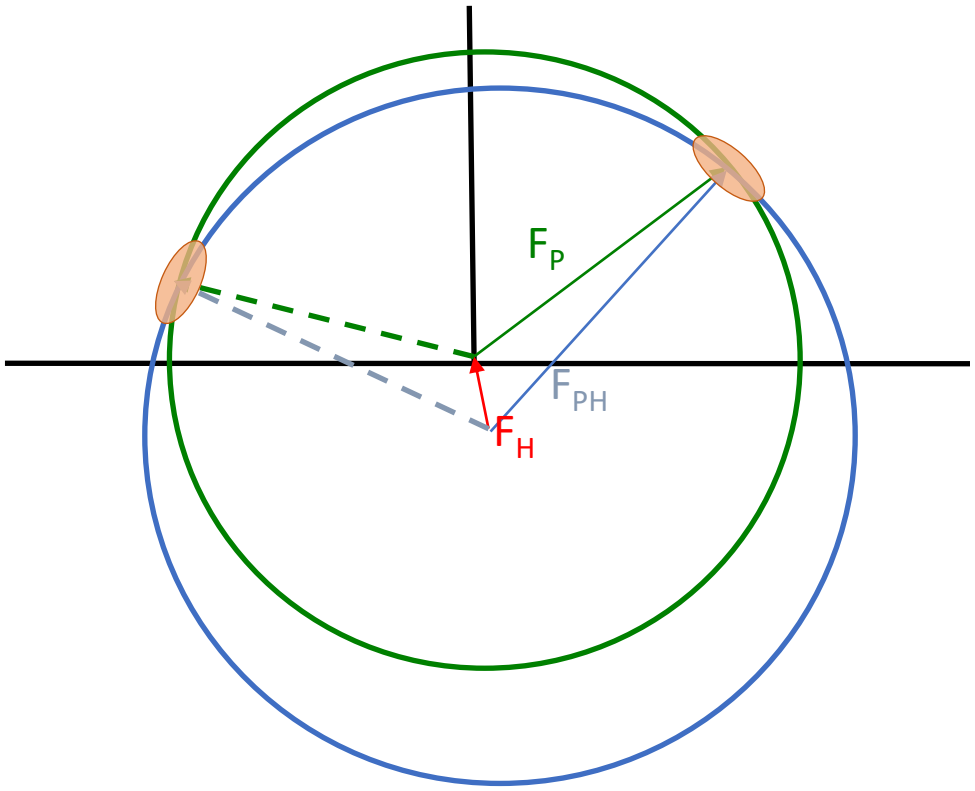


However, there are two possible solutions for ϕ_P, so we need more information.

By using a second derivative binding in a different site on the protein we can potentially solve our problem

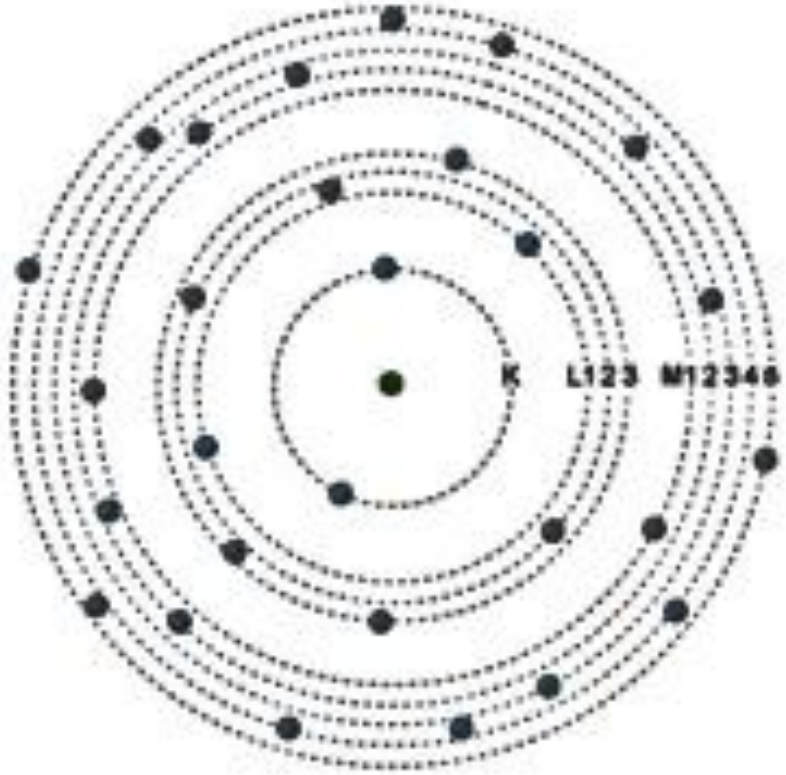
In this case the second derivative suggests that ϕ_1 is correct.

Phase Error



- You may have noticed that the rings in the Harker construction do not overlap perfectly at a point. This is not (just) sloppy draftsmanship on my part!
- In reality, experimental errors in the measurement of structure factors result in there being a range of possible phase values normally described as a phase probability distribution.
- These can be expressed as the four coefficients of a polynomial, the Hendrickson-Lattman coefficients HLA, HLB, HLC, HLD.
- Modern software normally uses Maximum Likelihood methods to derive the phase probability distributions.

- Anomalous scattering is due to the electrons being tightly bound (particularly in K & L shells)
- In classical terms, the electrons scatter as though they have resonant frequencies



When incident photon has relatively low energy:

- The photon is either scattered or not, but is not absorbed
- The photon scatters with no phase delay (imaginary f'' component is 0)

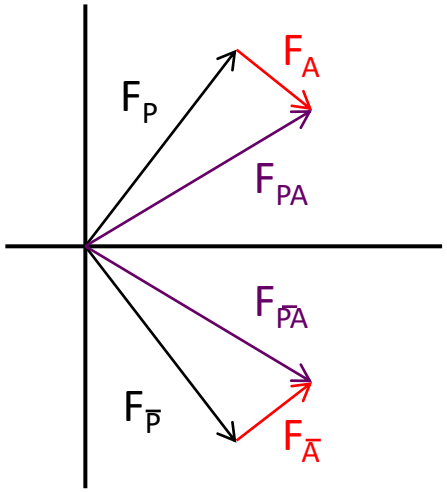
When incident photon has high enough energy:

- Some photons are scattered normally (as above)
- Some photons are absorbed and re-emitted at lower energy (fluorescence)
- Some photons are absorbed and immediately re-emitted at the same energy, gaining an imaginary component to its phase (f'' becomes non-zero)

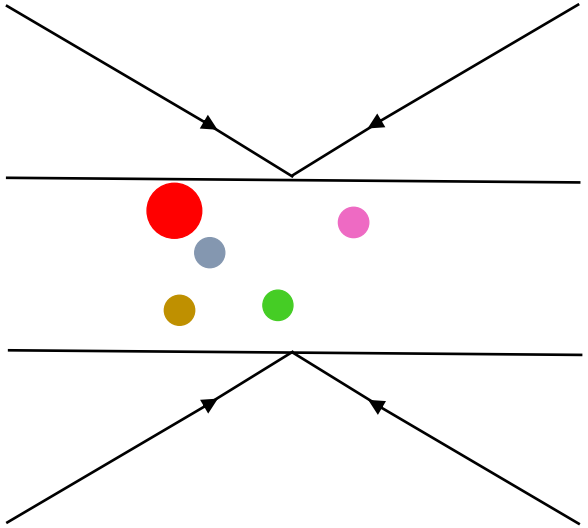
Friedel's Law in normal scattering conditions

Friedel pairs are Bragg reflections related by inversion through the origin

Friedel's Law – A Friedel pair have equal amplitude and opposite phase

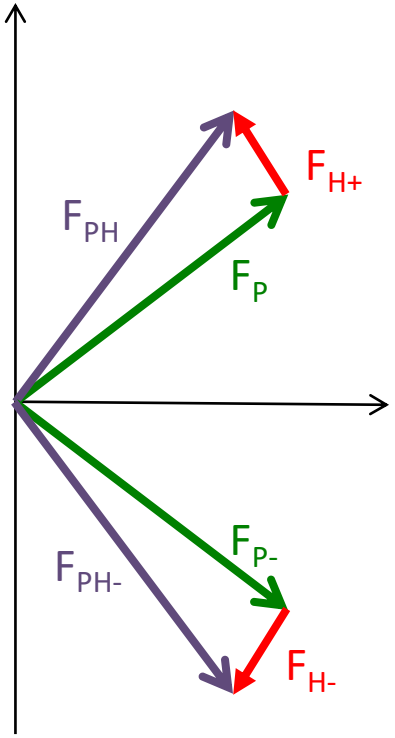


Normal scattering conditions

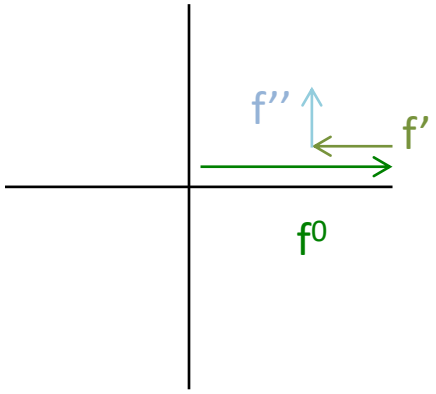


$$|F_{hkl}| = |F_{\overline{hkl}}| \quad j_{hkl} = -j_{\overline{hkl}}$$

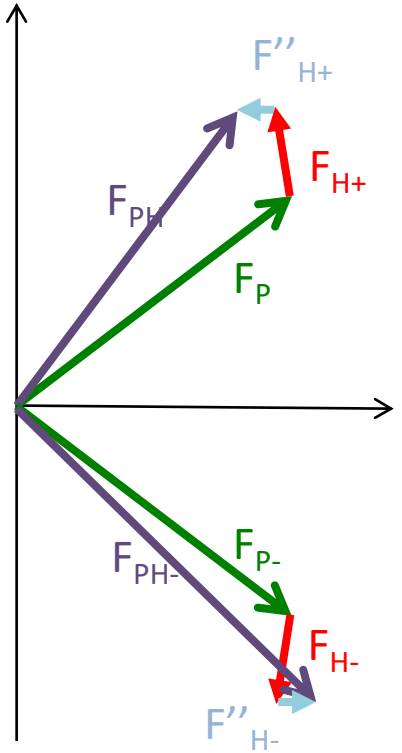
Breaking Friedel's Law



Normal scattering conditions



Under anomalous scattering conditions, the f'' component of atom A lags the phase component of the f^0+f' by 90°. Its phase is always 90° different.



Friedel's Law is broken.

$$|F_{PH}| \neq |F_{\overline{PH}}|$$

How can this help solve the phase problem?

- The atoms normally found in proteins (carbon, nitrogen, oxygen) from do not scatter anomalously at the X-ray wavelengths (energies) we routinely use.

- But heavy atoms do. So we can create a heavy atom substructure again and collect anomalous data.

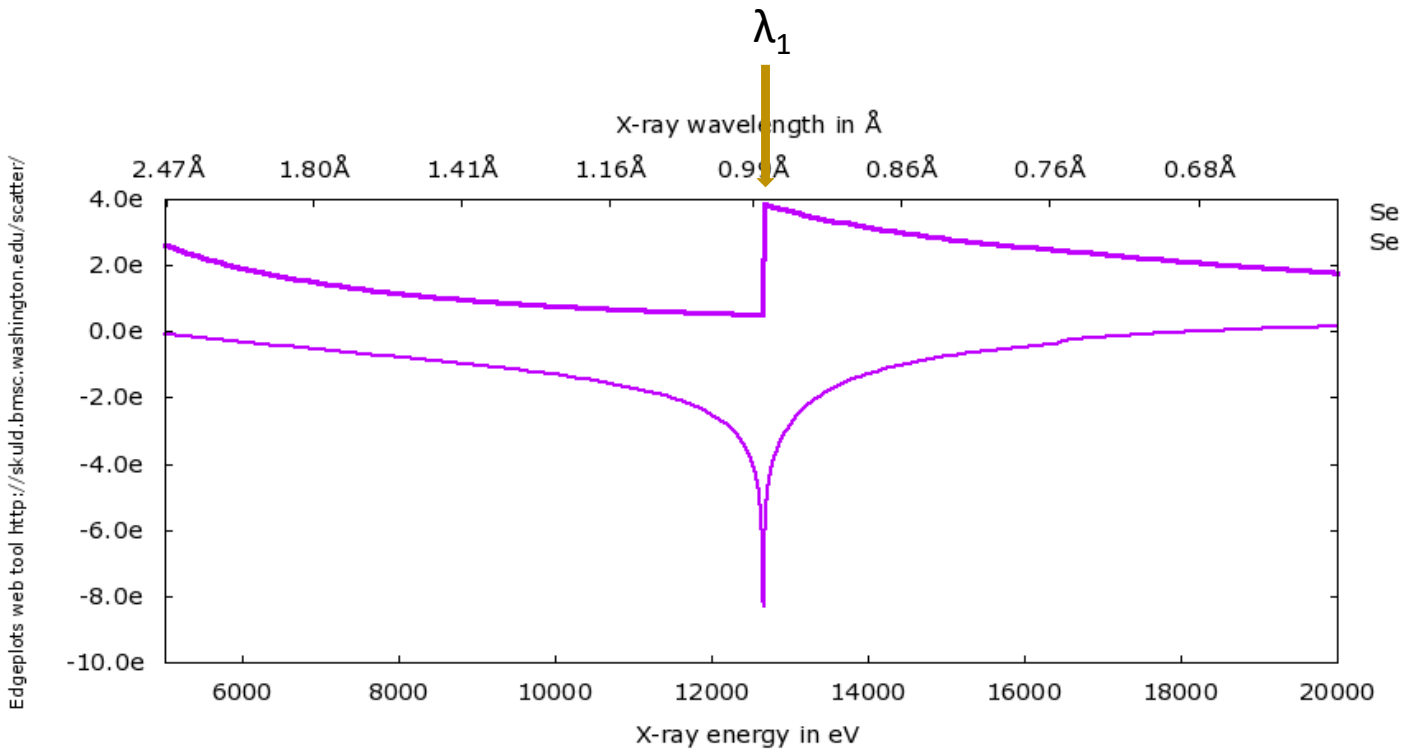
- An additional method of heavy atom incorporation can be used here by incorporating selenomethionine into the protein in place of methionine.

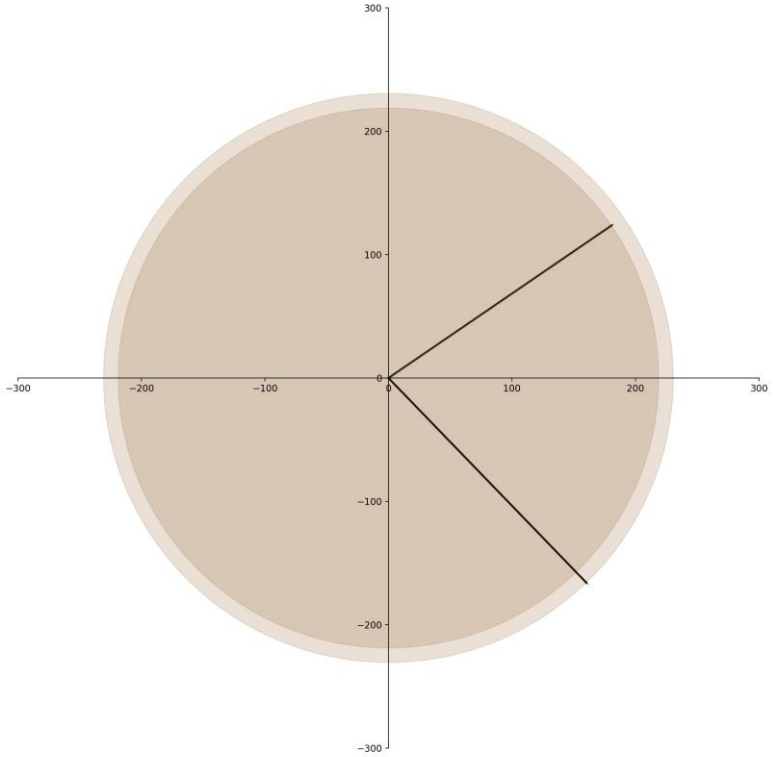
- We need to collect data at a synchrotron as we can select the wavelength and cause our substructure atoms to scatter anomalously.

There are two ways of phasing using anomalous scattering

SAD – Single anomalous diffraction – where we collect a single dataset with the maximum anomalous signal.
MAD – multiwavelength anomalous dispersion – where we collect several datasets with various levels of anomalous scatter and make use of the dispersive differences between wavelengths.

Solving the phases using Single Anomalous Diffraction (SAD)

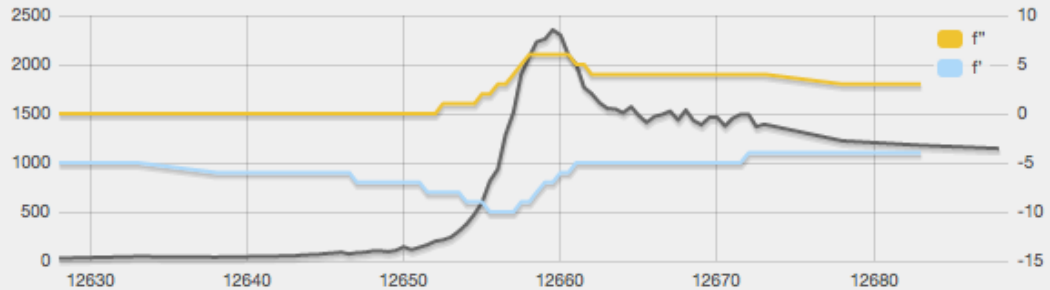




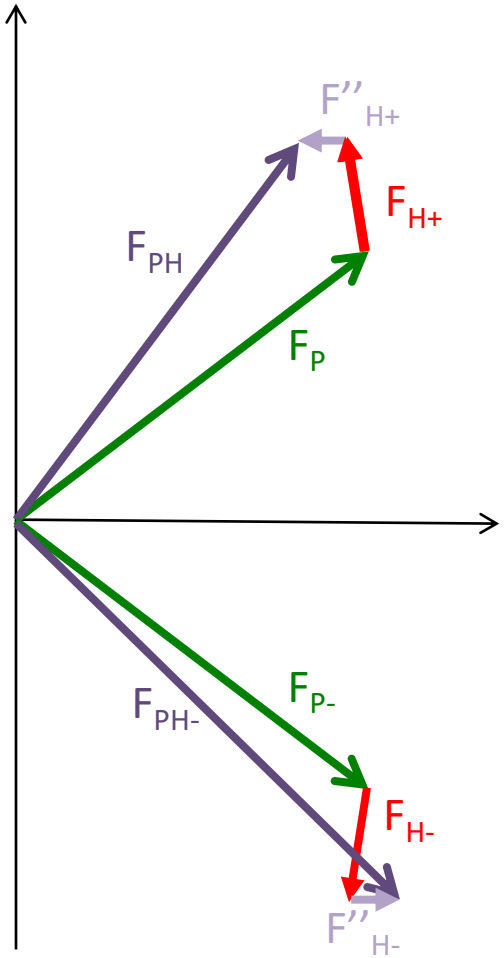
λ_1 = peak = maximum anomalous (f'')

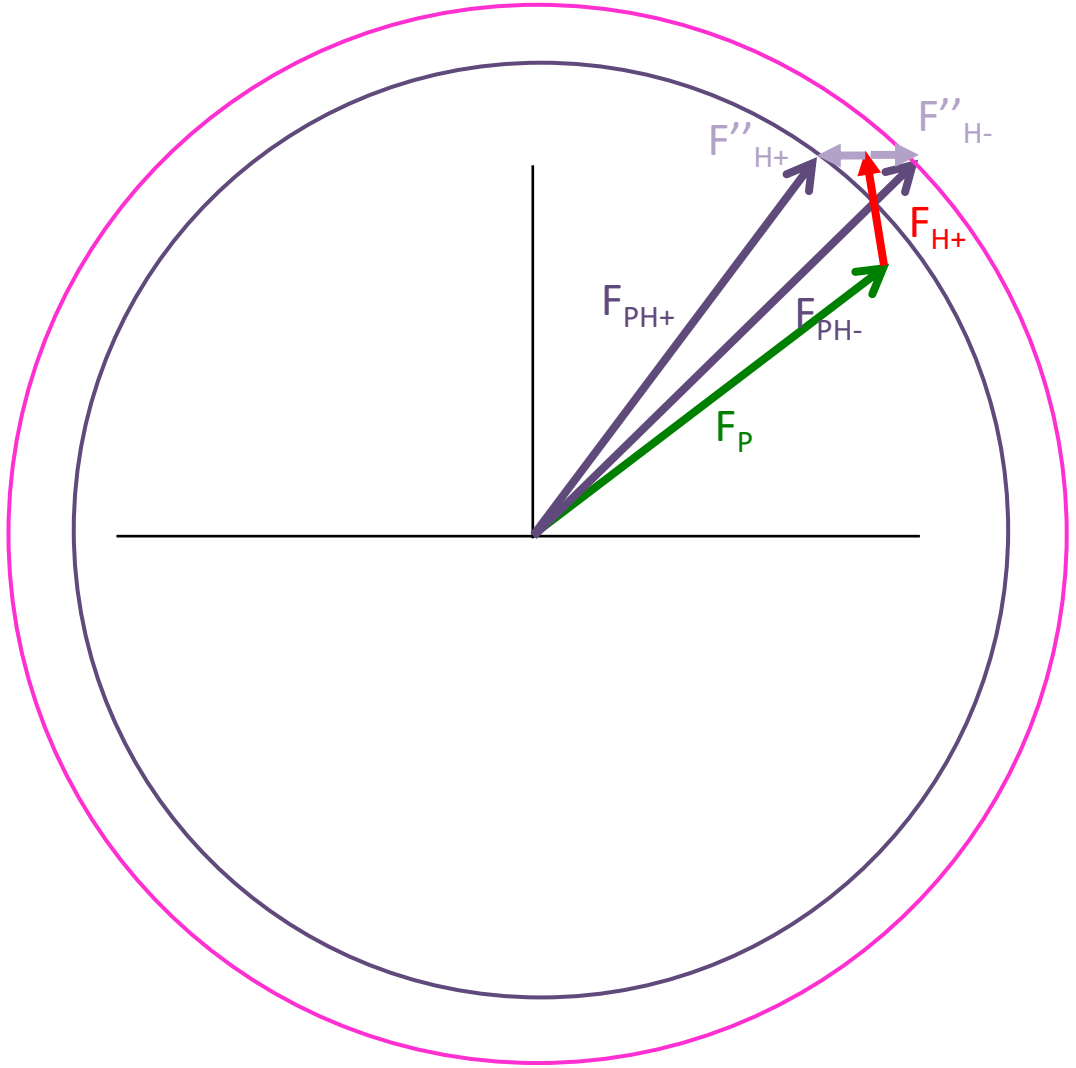
12-05-2016 18:22:26 - Se Edge Scan

Sample: CV_PACT_OPT_C7_A	Scan File: CV_1.fluo
E(Peak): 12659.5eV (0.9794Å)	f'': 6.67 / f': -7.22e
E(Inf): 12656.490234375eV (0.9796Å)	f'': 3.7699999809265 / f': -10.289999961853e
Exposure: 1s	Transmission: 0.40%
Beamsize: 59.46993637085x20µm	
Comment: Click to edit	

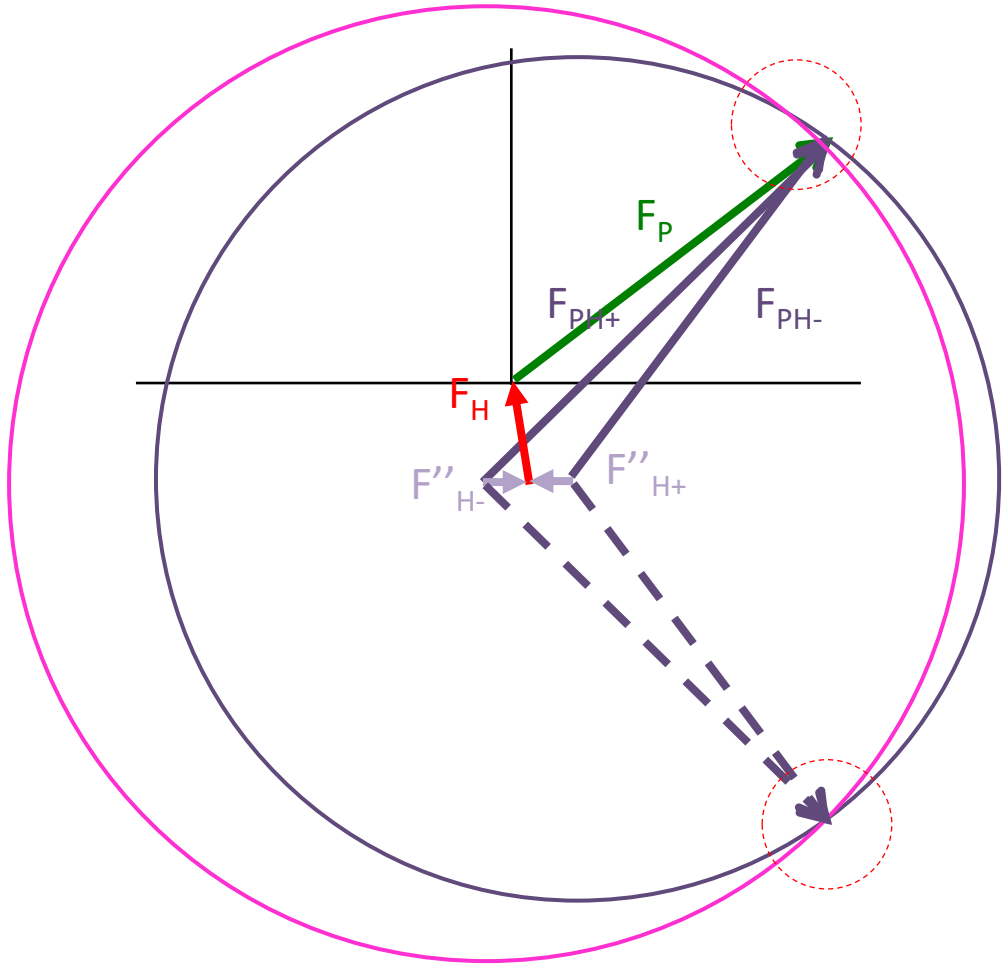


Solving the phases using Single Anomalous Diffraction (SAD)



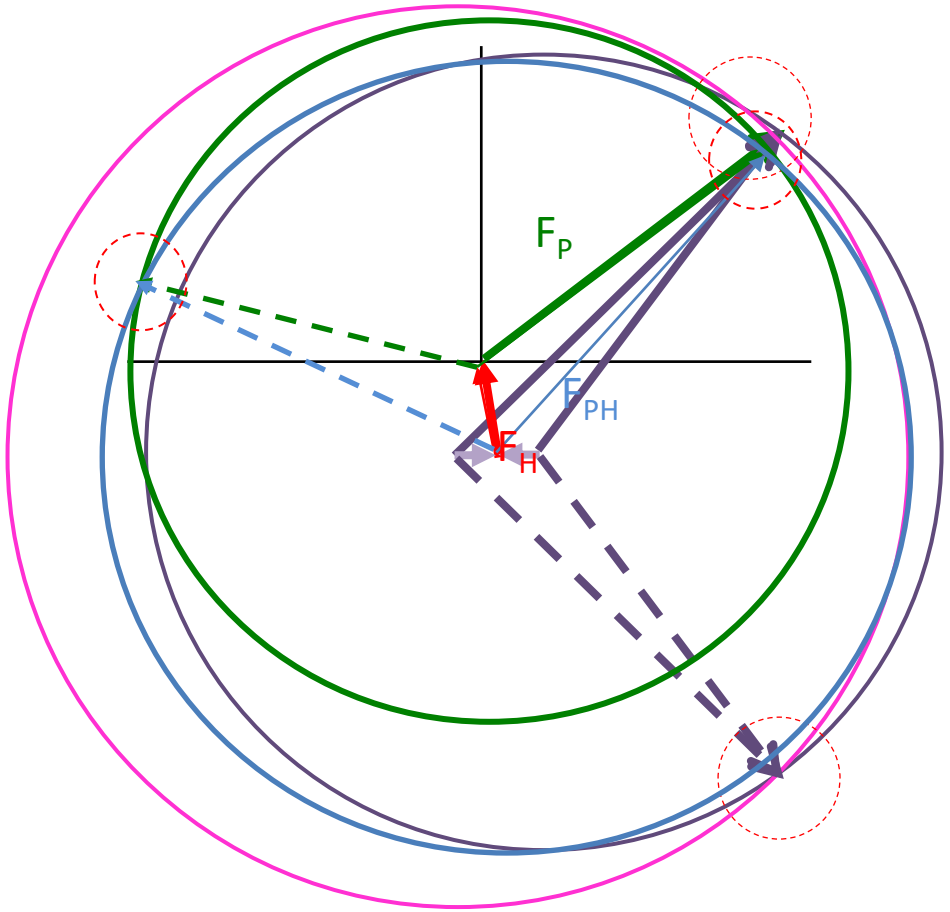


Solving the phases using Single Anomalous Diffraction



- Again we have two possible solutions
- If we had collected data for a MAD experiment we would add the addition wavelengths on to the construction in the same way we added an addition derivative in isomorphous replacement.
- We can also use **density modification** techniques on the electron density maps calculated from both phase solutions.
- In most cases it would be possible to tell which was the correct solution by the fact only one map would look like protein electron density.

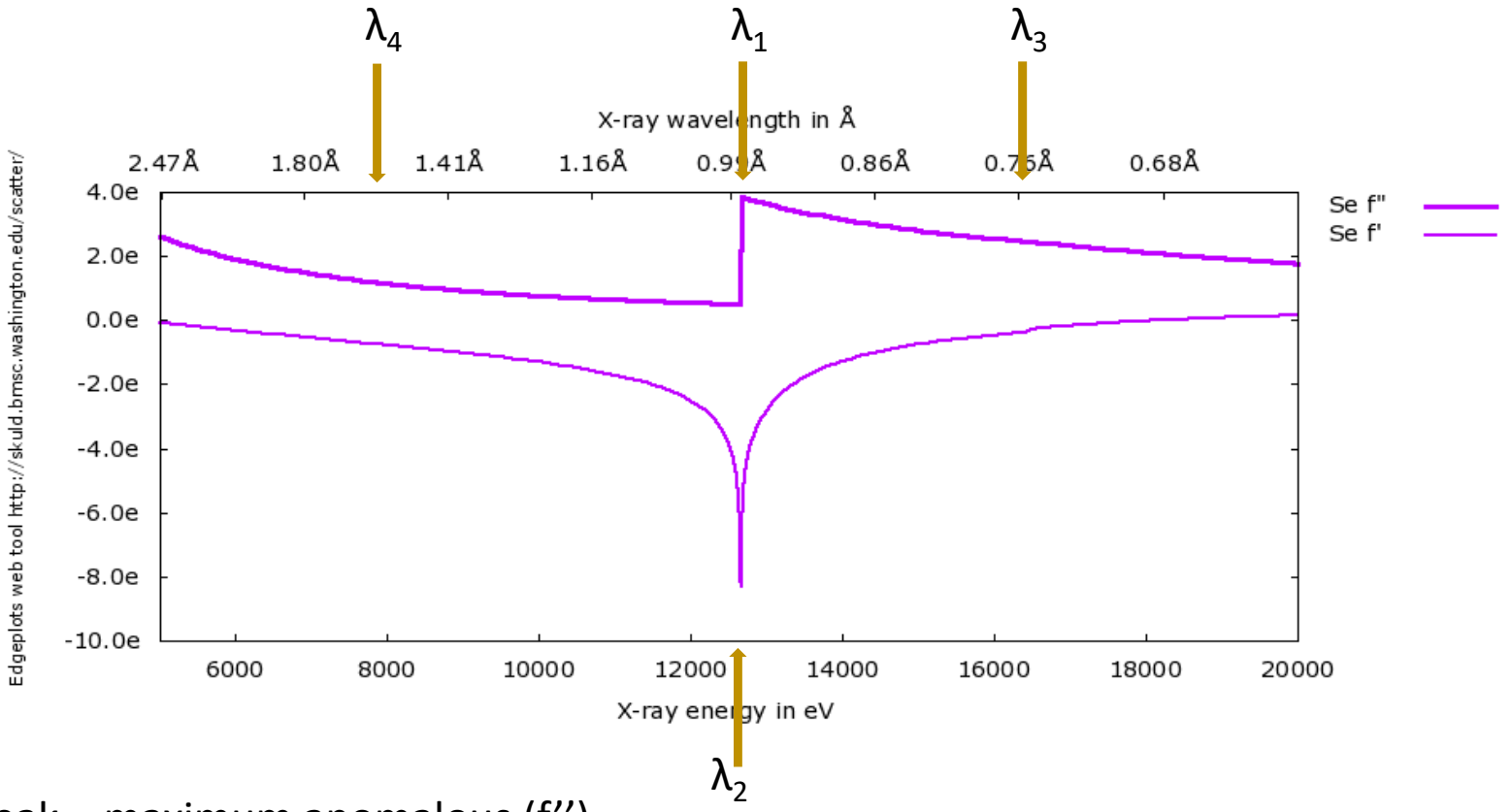
Solving the phases using Single Anomalous Diffraction



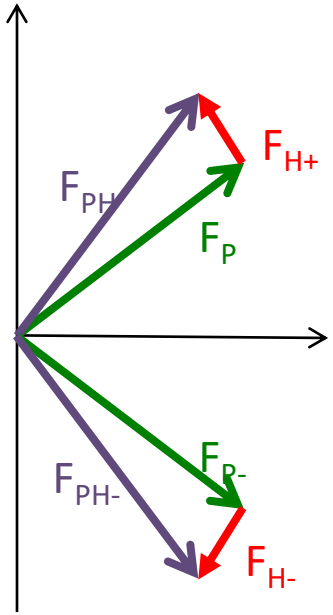
However, we can also combine the phasing experiment we did for the isomorphous with the SAD phasing experiment.

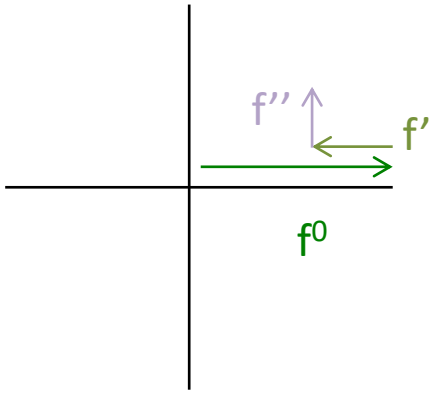
This is called SIRAS – Single isomorphous replacement anomalous dispersion.

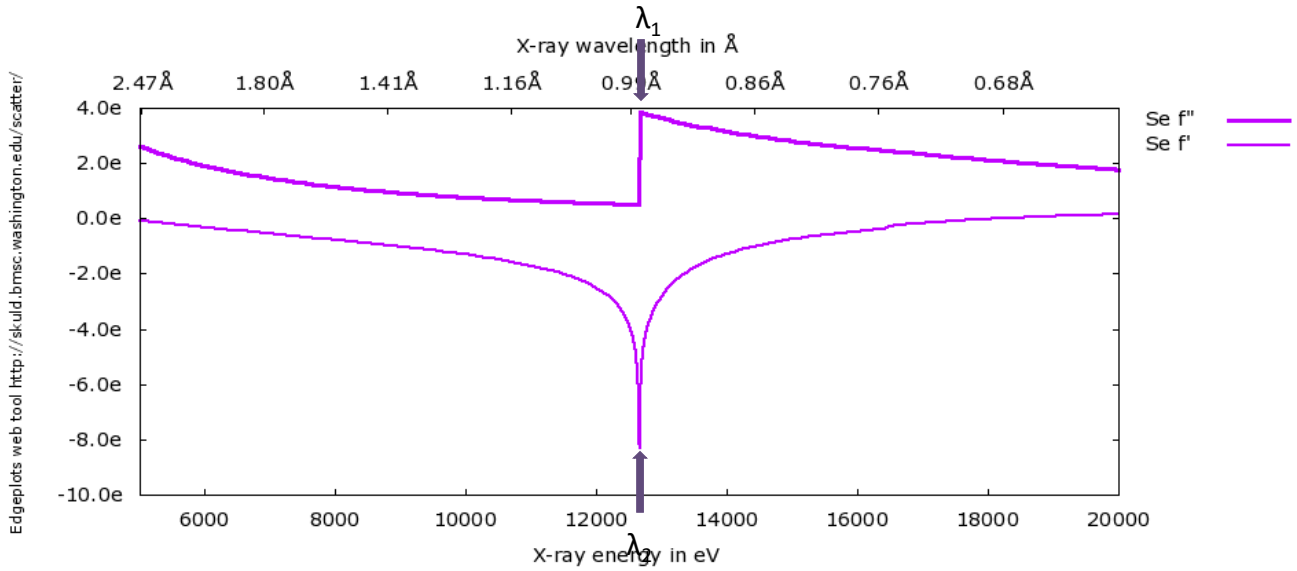
Solving the phases using Multiwavelength Anomalous Diffraction

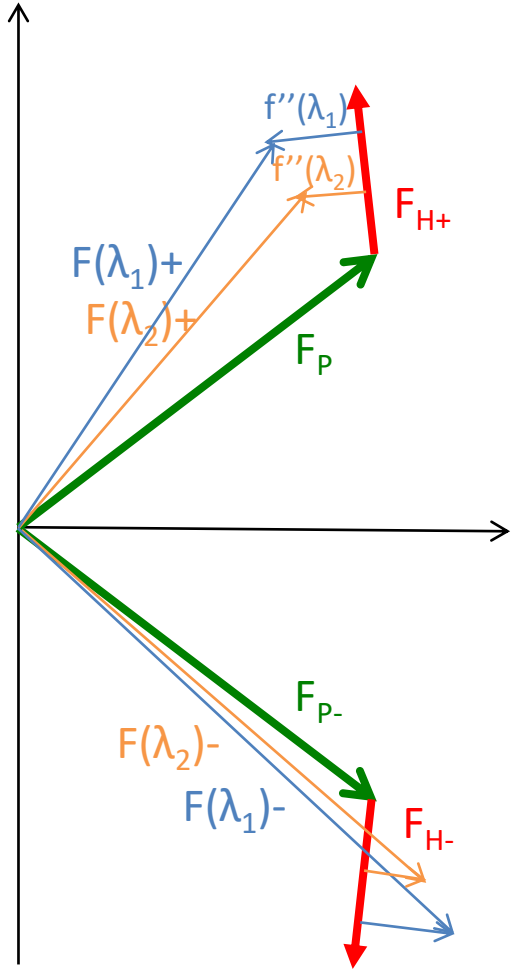


λ_1 = peak = maximum anomalous (f'')
λ_2 = inflection = maximum f'
λ_3 = high energy remote
λ_4 = low energy remote

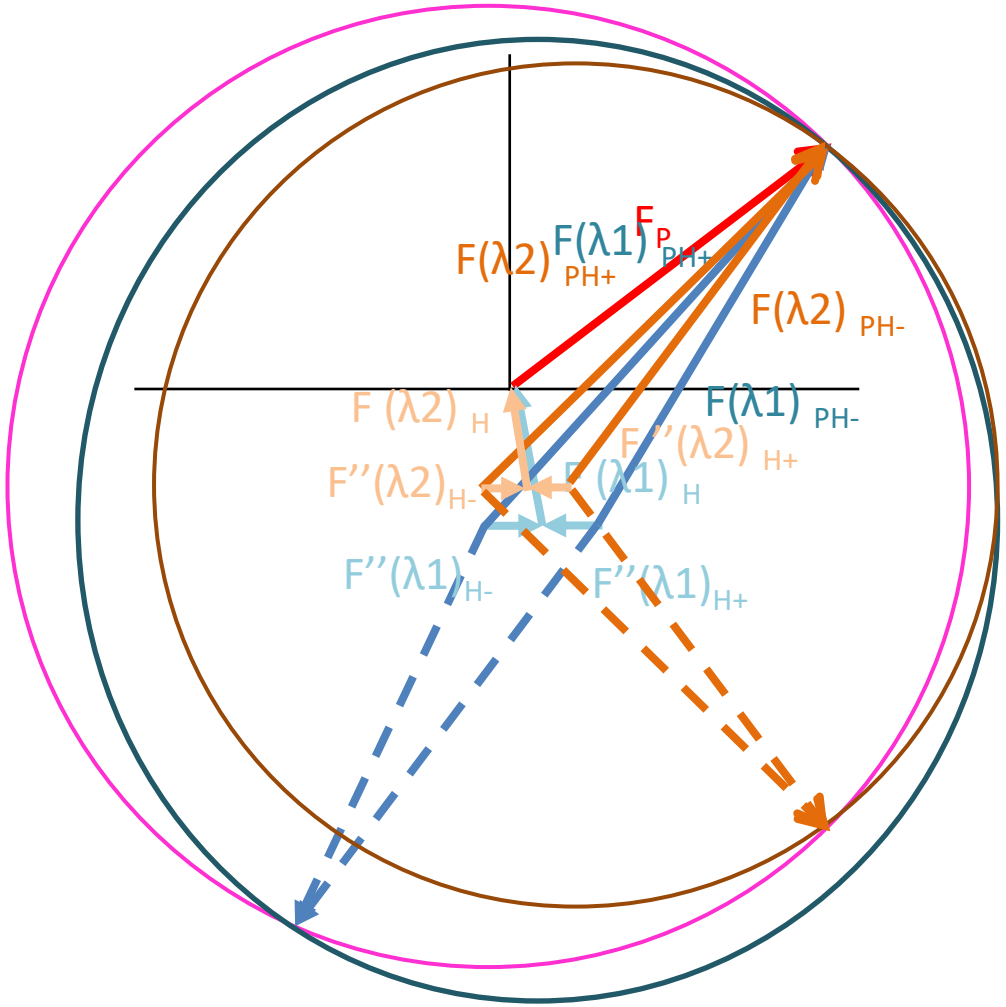

F_{PH}
F_{H+}
F_P
F_{P-}
F_{PH-}
F_{H-}


f''
f'
f^0


λ_1
X-ray wavelength in Å
2.47Å
1.80Å
1.41Å
1.16Å
0.98Å
0.86Å
0.76Å
0.68Å
4.0e
2.0e
0.0e
-2.0e
-4.0e
-6.0e
-8.0e
-10.0e
6000
8000
10000
12000
14000
16000
18000
20000
λ_2
X-ray energy in eV
Se f''
Se f'
Edgeplots web tool http://skuld.bmsc.washington.edu/scatter/


$f''(\lambda_1)$
$f''(\lambda_2)$
F_{H+}
$F(\lambda_1)+$
$F(\lambda_2)+$
F_P
F_{P-}
$F(\lambda_2)-$
$F(\lambda_1)-$
F_{H-}

Solving the phases using Multiwavelength Anomalous Diffraction

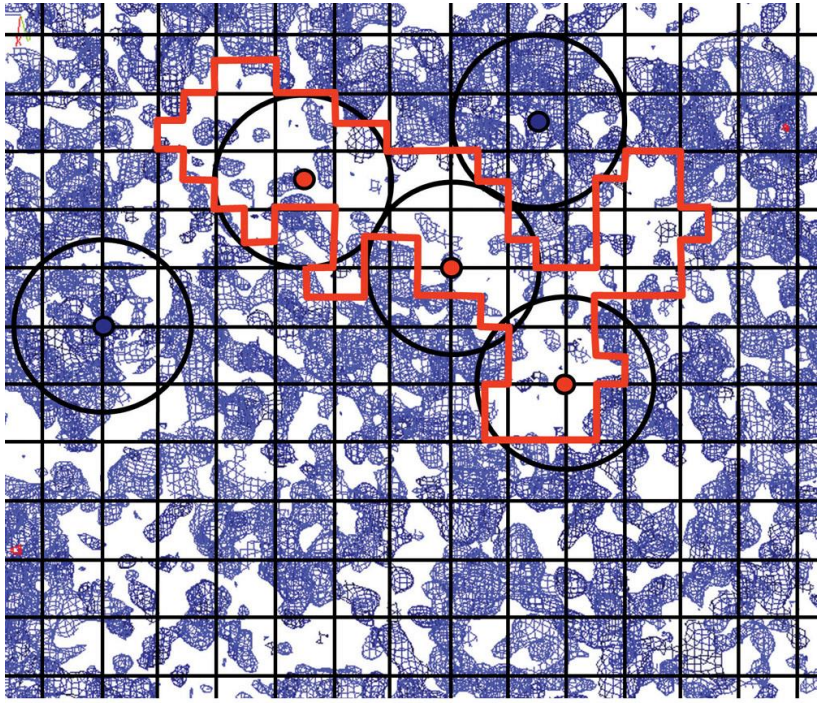


Density Modification (Improvement)

- Initial maps phased purely from substructure atoms are rarely sufficiently interpretable to build a model
- Density modification encompasses a range of techniques to take advantage of prior knowledge in order to improve our electron density maps, generally by improving our estimates of the phases.
- Many of these calculations are carried out in real space.

Solvent Flattening (and phase extension)

- Disordered region of electron density $\rho(x)$ have a constant flat value of around 0.33 $e^-/Å^3$ for pure water – protein has a higher average electron density of around 0.44 $e^-/Å^3$
- A mask is constructed to describe contiguous solvent regions of the initial map (if this is not possible, density modification will not work)
- Within the solvent region the electron density is set to the average value and new structure factors are calculated.
- This procedure is most effective when solvent content is relatively high.



Summary of phasing

- **Patterson maps** can be calculated from the measured intensities of the diffraction.
- They require no phase information as phases are set to zero.
- While this can allow the positions and phases to be calculated for a small number of atoms it is not possible for the 1000's of atoms in a protein.
- But it is possible to solve the positions and from that the phases of heavier atoms

- **Isomorphous replacement** is the where a heavy atom is incorporated in to the protein crystal and the position of the heavy atoms can be calculated using Patterson or direct methods allowing the calculation of phases.
- SIR – One derivative
- MIR – Multiple derivatives
- SIRAS – Single derivative with anomalous scatter
- MIRAS – Multiple derivatives with anomalous scatter
- *Benefits* – Phasing experiments can be done at the home lab.
- *Drawbacks* – It can be difficult to bind a heavy atom and keep the crystals isomorphous and it can be very difficult to get the derivatives anyway.

Summary of Phasing

- **Anomalous dispersion** is the process of using the correct wavelength of X-ray to cause alterations in the scattering properties of heavy atoms. Again the positions of the heavy atoms can be determined allowing phase calculation.
- SAD – Single wavelength
- MAD – Multiple wavelengths
- *Benefits* – Can incorporate selenomethione into the protein, thereby reducing the non-isomorphism. Can use Sulfur for SAD on home source.
- *Drawbacks* – The need for a tunable X-ray source, radiation damage from multiple datasets creates non-isomorphous data.

- **Direct Methods** are not frequently used in X-ray protein crystallography to solve structures due to resolution requirements. However, are often used in combination with Patterson methods to solve substructures.