ESRF | The European Synchrotron

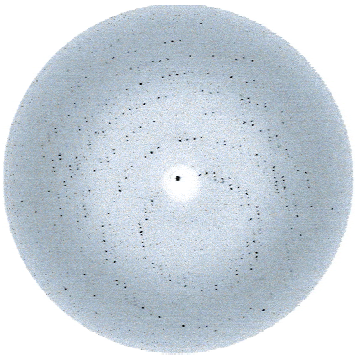
THE PHASE PROBLEM – AND HOW TO SOLVE IT

GORDON LEONARD
ESRF STRUCTURAL BIOLOGY GROUP

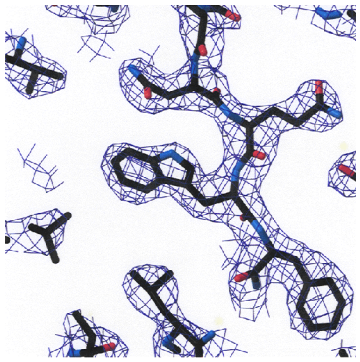
OUTLINE

1. **The Phase Problem**
2. **Isomorphous Replacement**
   1. What is it?
   2. SIR/MIR
   3. Phase probability distributions
3. **Anomalous Scattering**
   1. What is it?
   2. How can we exploit it?
      1. *S(M)IRAS*
      2. *MAD*
      3. *SAD*

STRUCTURE DETERMINATION: FROM INTENSITIES TO ELECTRON DENSITY



Integrated Intensities



Electron density
(= protein structure)

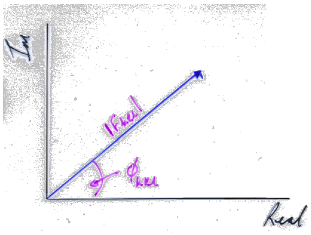
THE ELECTRON DENSITY EQUATION

$$\rho_{(x,y,z)} = \frac{1}{V_c}\sum_h\sum_k\sum_l F_{hkl}\exp(-2\pi i(hx+ky+lz))$$

$$F_{hkl} = \sum_j f_j \exp -2\pi i(hx_j + ky_j + lz_j)$$

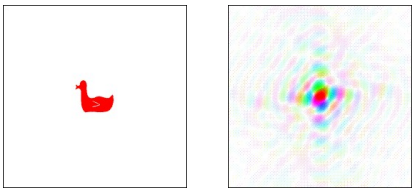
$$F_{hkl} = \sum_j |F_{hkl}| \exp -2\pi i(\phi)$$

Amplitude Phase

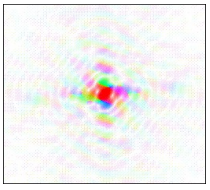


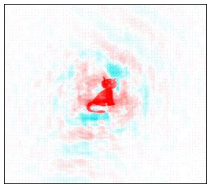
From an x-ray diffraction experiment we can 'measure' the amplitude ($|F| \alpha \sqrt{I}$) but get *no information* about the phase. Deriving the phase is known as **'the phase problem'**

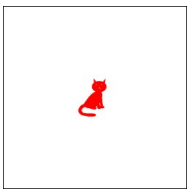
'GOOD' PHASES ARE MORE IMPORTANT THAN 'GOOD' AMPLITUDES

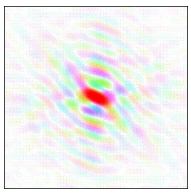


Amplitudes (colours) from duck transform









Phases (brightness) from cat transform

http://www.ysbl.york.ac.uk/~cowtan/fourier/magic.html

METHODS FOR RESOLVING THE PHASE PROBLEM IN MX

1. Direct Methods

- requires high (near atomic; d_{min} ~1.2 Å) resolution data, still quite rare
- However, is often used at the start of experimental phasing procedures (Andrea's talk?).

2. Molecular Replacement

- requires some prior knowledge of the crystal structure you want to solve (homologous protein, domain, etc.)

3. Experimental Phasing

1. Isomorphous Replacement
2. Anomalous Dispersion Methods

MOLECULAR REPLACEMENT?

'**The term 'molecular replacement' (MR) is generally used to describe the use of a known molecular model to solve the unknown crystal structure of a related molecule.** MR enables the solution of the crystallographic phase problem by providing initial estimates of the phases of the new structure from a previously known structure, as opposed to the other two main methods for solving the phase problem, i.e. experimental methods (which measure the phase from isomorphous or anomalous differences) or direct methods….'

'….The use of MR has naturally become more common as the database of known structures expands**. MR is currently used to solve [> 80%] of deposited macromolecular structures….**'

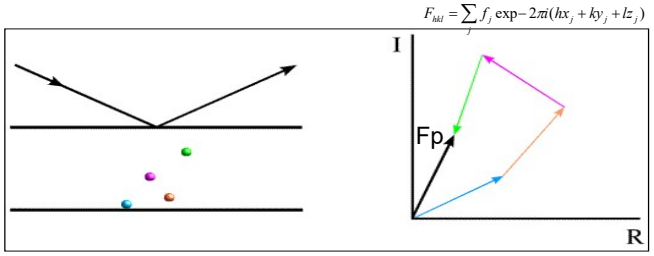
Philip Evans and Airlie McCoy (2008). An introduction to molecular replacement. *Acta Crystallogr* **D64**, 1–10.

When MR doesn't work we need to use experimental phasing techniques based on Isomorphous and/or anomalous differences between diffraction intensities.

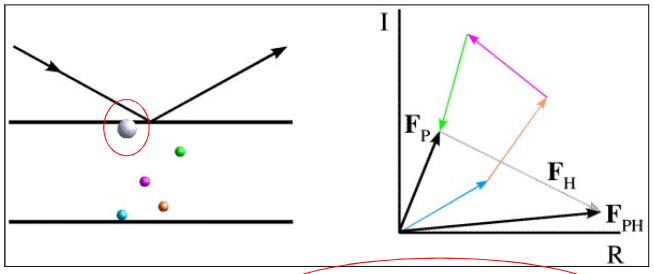
INTRODUCING A NEW (HEAVY) ATOM INTO A CRYSTAL

$$F_{hkl} = \sum_j f_j \exp -2\pi i(hx_j + ky_j + lz_j)$$

(protein) Crystal



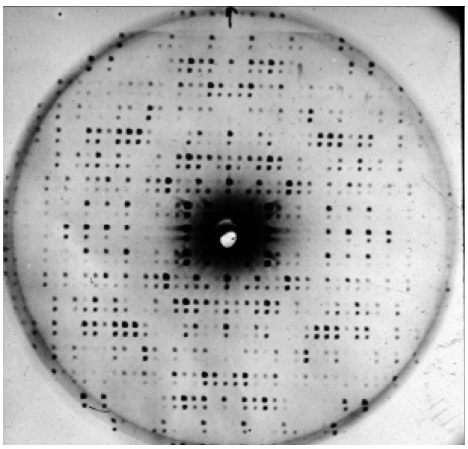
(protein) Crystal
+
(heavy) atom



Get a **different structure factor**, F_{PH}

$$\boldsymbol{F}_{ph} = \boldsymbol{F}_{p} + \boldsymbol{F}_{h}$$

INTRODUCING A NEW (HEAVY) ATOM INTO A CRYSTAL: ISOMORPHOUS DIFFERENCES



Taylor, G. L. (2010). Introduction to phasing. *Acta Cryst.* **D66**, 325-338.

Two protein diffraction patterns shifted vertically relative to one another. One is from native bovine β-lactoglobulin, the other from a crystal soaked in a mercury-salt solution. Note: 1) intensity changes for reflections in the latter; 2) identical unit cells suggesting isomorphism.

$$|\Delta F| = |\boldsymbol{F}_{ph} - \boldsymbol{F}_{p}|$$

Addition of 1 Hg atom to a protein of 1000 atoms will produce an average fractional change of intensity of ~25% so differences should be easy to measure

$$<\frac{\Delta I}{I}> = \sqrt{\frac{N_H}{2N_p}}\, f_H / f_p$$

We can use isomorphous differences to derive phase information via isomorphous replacement

HOW IS ISOMORPHOUS REPLACEMENT DONE?

1. Collect data set from **native** crystal (structure factors, $\boldsymbol{F_p}$)
2. Introduce a small number of heavy atoms (Pt, Hg, Br, I, etc…) into another crystal of the same protein
3. Collect data set from **derivative** crystal (structure factors, $\boldsymbol{F_{ph}}$)
4. Use differences (**isomorphous differences**) to find heavy atom positions in unit cells
5. Use these positions to initiate phase calculations

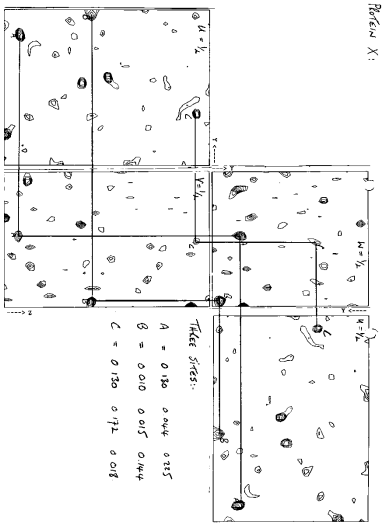
FINDING HEAVY ATOM SITES

Patterson Functions: maps of interatomic vectors

$$P_{(u,v,w)} = \frac{1}{V_c}\sum_h\sum_k\sum_l |F_{ph} - F_p|^2 \exp(-2\pi i(hu + kv + lw))$$

$$P_{(u,v,w)} = \frac{1}{V_c}\sum_h\sum_k\sum_l |\Delta F_{ano}|^2 \exp(-2\pi i(hu + kv + lw))$$

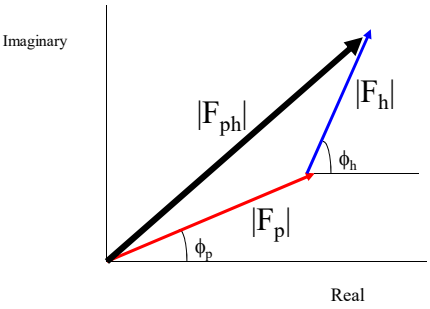
$$P_{(u,v,w)} = \frac{1}{V_c}\sum_h\sum_k\sum_l |\Delta F_{disp}|^2 \exp(-2\pi i(hu + kv + lw))$$



Use of direct methods <u>much</u> more common

HOW DOES SINGLE ISOMORPHOUS REPLACEMENT WORK?

$$\boldsymbol{F}_{ph} = \boldsymbol{F}_{p} + \boldsymbol{F}_{h}$$



Want to calculate ϕ_p:

$$\Phi_p = \Phi_h \pm \cos^{-1}(F_{ph}^2 - F_p^2 - F_h^2 / 2F_pF_h)$$

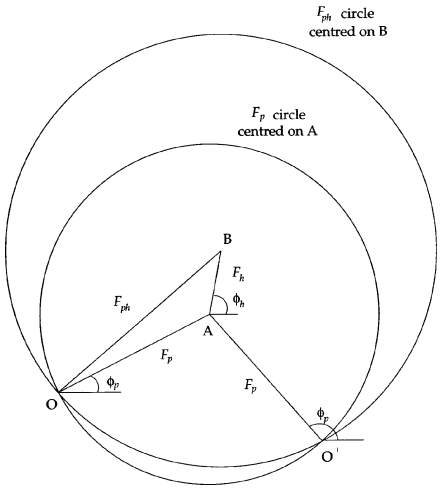
$$\Phi_p = \Phi_h \pm \Phi'$$

Phase Ambiguity

$$|F_h| \approx |F_{ph}| - |F_p|$$

Measure $|F_p|$, $|F_{ph}|$; Calculate $|F_h|$, ϕ_h from heavy atom positions

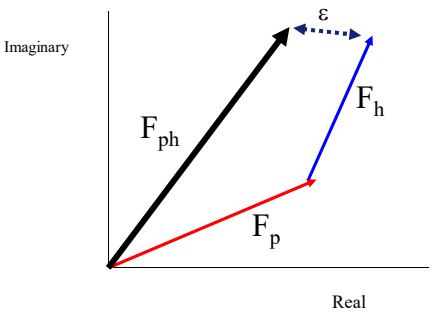
GRAPHICAL ILLUSTRATION OF PHASE AMBIGUITY



No errors shown here. What we actually get are phase probability distributions

LACK OF CLOSURE (ε) AND PHASE PROBABILITIES

1. Measurement errors in $|F_P|$, $|F_{PH}|$
2. Model of heavy atom substructure is incorrect (minor sites not included)
3. Native and heavy atom derivative crystal are not isomorphous



$$\Phi_p = \Phi_h \pm \cos^{-1}(F_{ph}^2 - F_p^2 - F_h^2 / 2F_pF_h)$$

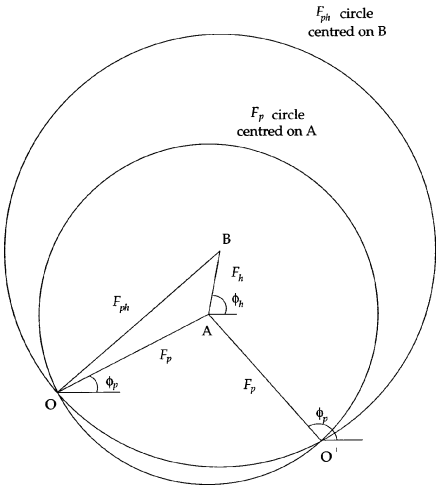
$$\varepsilon = |F_{ph(obs)}| - |F_{ph(calc)}|$$
$$= |F_{ph(obs)}| - |[|F_p| \exp(i\phi_p) + |F_h| \exp(i\phi_h)]|$$

$$P(\alpha_p) \alpha \exp(-\varepsilon^2 / 2E^2)$$

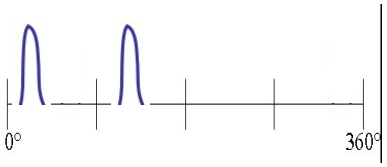
$$\varepsilon = <[F_{ph(obs)} - F_{ph(calc)}]^2>$$

Blow, D. M. & Crick, F. H. C. (1959). Acta Cryst. **12**, 794-802.

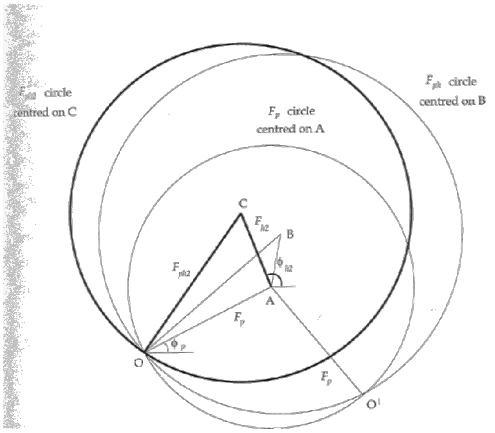
SIR PHASE PROBABILITY DISTRIBUTION

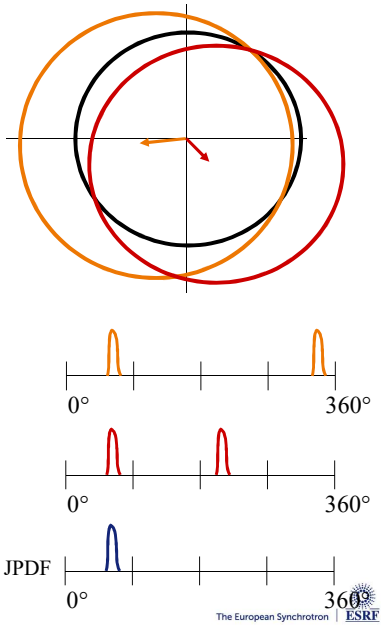


$$\Phi_p = \Phi_h \pm \Phi'$$



A SECOND DERIVATIVE RESOLVES THE SIR PHASE AMBIGUITY

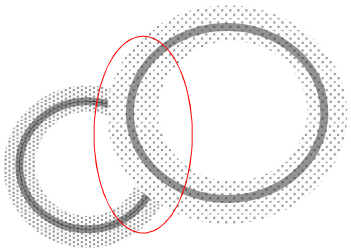


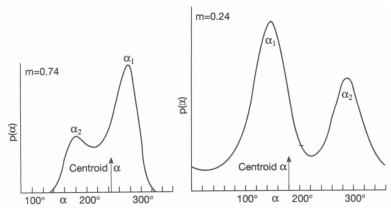


PHASE PROBABILITY DISTRIBUTIONS

1. Measurement errors in $|F_P|$, $|F_{PH}|$
2. Model of heavy atom substructure is incorrect (minor sites not included)
3. Native and heavy atom derivative crystal are not isomorphous

$$F_{ph} \neq F_p + F_h$$



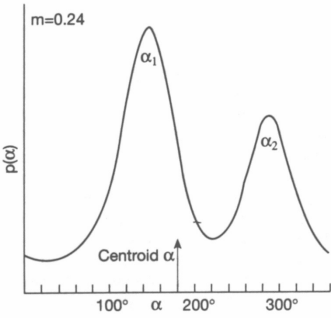


The most probable phase is not be the best phase to use in map calculations. The **centroid phase** (best phase) gives the least error in the Fourier map.

$$\rho_{(x,y,z)} = \frac{1}{V_c} \sum_h \sum_k \sum_l m_{hkl} |F_{hkl}| \exp[-2\pi i(hx + ky + lz) - \alpha_{hkl}]$$

THE FIGURE OF MERIT

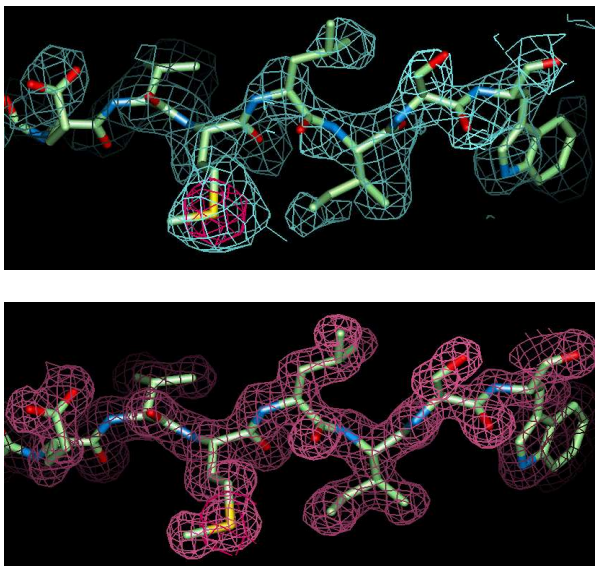
$$m = \frac{\sum P(\Delta\alpha)\cos(\Delta\alpha)}{\sum P(\Delta\alpha)}$$



m = probability that centroid phase is correct
$\Delta\alpha$ = phase angle measured from the centroid phase.
$P(\Delta\alpha)$ = probability that $\Delta\alpha$ is the true phase.

m = 0.866 => phase error = 30°
m = 1.0 => phase error = 0°
m = 0.0 => phase error = 90°
m = 0.5 => phase error = 60°

IMPROVING EXPERIMENTAL PHASES

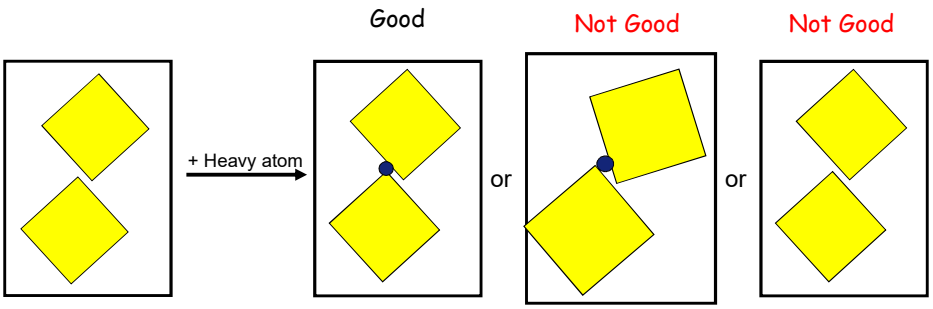


Mean phase error = 53°

Phase Improvement

Mean phase error = 41°

WHAT CAN HAPPEN IN HEAVY ATOM SOAKING EXPERIMENTS?



If heavy atoms don't bind to protein - or are not ordered in the crystal - then there will be no intensity differences

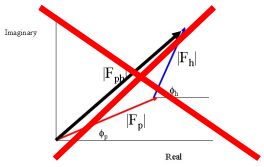
Binding of heavy atom to protein in crystal can cause protein to rotate and/or translate in unit cell and/or unit cell dimensions to change

Non-isomorphism: For a 100 Å cubic unit cell a 0.5% change in unit-cell dimensions or a 0.5° rotation of the molecule within the unit cell would produce an average 15% change in intensity.

NON ISOMORPHISM (AGAIN)

Adding heavy atoms to protein crystals can cause the protein molecules to rotate and/or translate inside the unit cell. This induces non-isomorphism between native and derivative crystals

$$F_{ph} \neq F_p + F_h$$

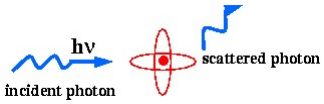


To help get around this take advantage of **anomalous scattering**

NORMAL SCATTERING

X-ray photon interacts with the electrons in the atoms of the sample. Atomic scattering factor *f* proportional to atomic number, Z, of atoms

$$f = f_o$$



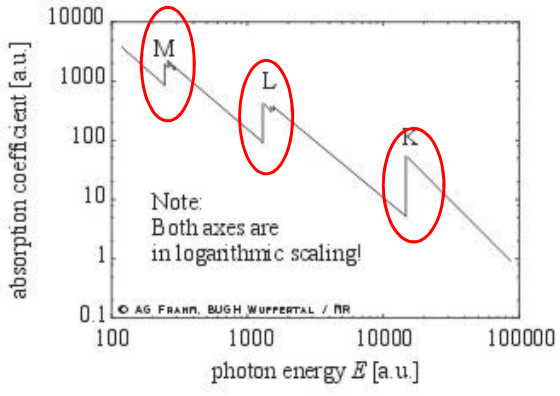
$$F_{hkl} = \sum_j f_j \exp -2\pi i(hx_j + ky_j + lz_j)$$

$$F_{hkl} = \sum_j |F_{hkl}| \exp -2\pi i(\phi)$$

All atoms scatter X-rays in the same way. **Friedel's Law is obeyed**.

$$|F_{hkl}| = |F_{\bar{h}\bar{k}\bar{l}}|$$

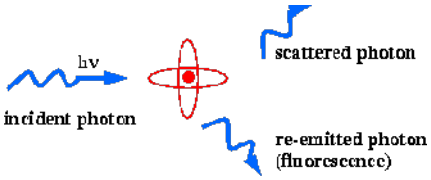
BUT, ELECTRON SHELLS IN ATOMS HAVE X-RAY ABSORPTION EDGES

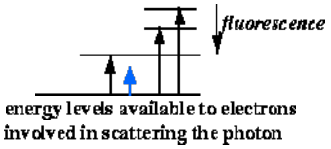


At incident X-ray energies close to absorption edges scattering from atoms is no longer normal

NORMAL SCATTERING IS MODIFIED CLOSE TO ABSORPTION EDGE

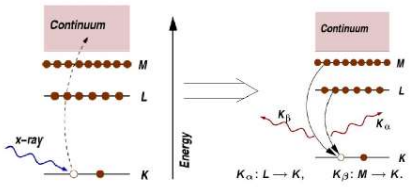
- Photons interact with the electrons in the atoms of the sample.
- Depending on the photon energy the probability of absorption changes.
- At photon energies close to absorption edges of a given atom, some photons are absorbed by the atom and electrons are ejected from its inner shell.
- Electrons from higher energy shells decay to lower energy shells **emitting fluorescence as they do so**

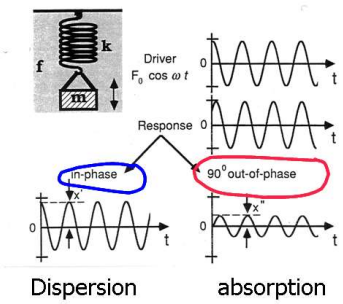




http://www.bmsc.washington.edu/scatter/AS_tutorial.html

SCATTERING FACTORS AT X-RAY ENERGIES CLOSE TO AN ABSORPTION EDGE





Treat electron shells as *harmonic oscillators* with resonance frequencies close to energy of shell absorption edge. Scattering factor modified so that:

$$f = f_o + f'(\lambda) + if''(\lambda)$$

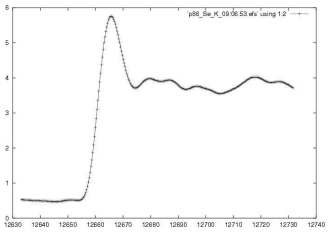
Scattering factor modified in such a way that it has additional, wavelength(energy)-dependent dispersive (*f'*) (real) and absorptive (*f''*) (imaginary) terms

DERIVING ANOMALOUS SCATTERING FACTORS

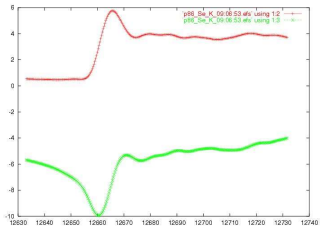
$$f = f_o + f'(\lambda) + if''(\lambda)$$

transmission fluorescence

$\mu(E) = \ln\left(\frac{I_0}{I}\right)$ $\mu(E) = \frac{I_f}{I_0}$



Experimental fluorescence scan

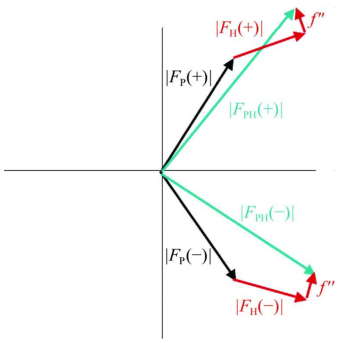


Experimentally determined anomalous scattering factors

$$f''(\omega) = \left(m\omega\ c / 4\pi Ne^2\right)\mu(\omega)$$

$$f'(\omega) = (2/\pi)\int_0^\infty \left[\omega' f''(\omega') / \left(\omega^2 - \omega'^2\right)\right] d\omega'$$

UNDER ANOMALOUS SCATTERING FRIEDEL'S LAW IS NO LONGER OBEYED



$$f = f_o + f'(\lambda) + if''(\lambda)$$

$$|F_{hkl}| \neq |F_{\bar{h}\bar{k}\bar{l}}| \; i.e. \; |F(+)| \neq |F(-)|$$

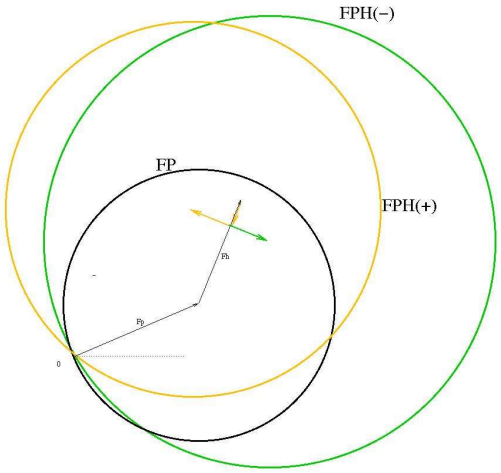
$$\Delta F_{ano} = \| F(+) | - | F(-) \|$$

$$F_{mean} = \frac{\| F(+) | + | F(-) \|}{2}$$

We can use the **anomalous difference** to (help) solve the phase problem.

The larger f", the bigger anomalous difference and the easier (in principle) it will be to solve the phase problem

USING ANOMALOUS SCATTERING - SIRAS PHASING



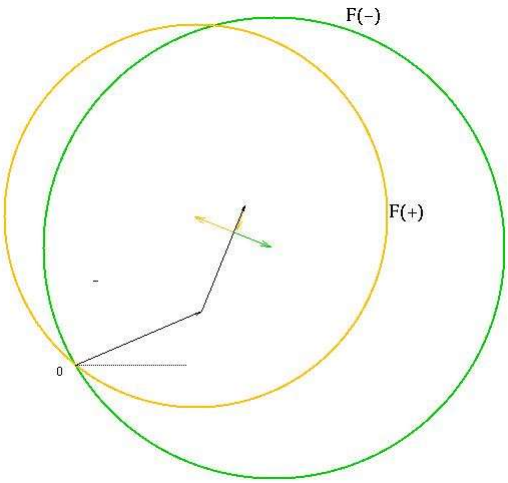
$$f = f_o + f'(\lambda) + if''(\lambda)$$

If in an SIR experiment we measure F_p (native crystal) and *both* F_{ph}^+ and F_{ph}^- for the derivative crystal we get three phasing circles. Hence no phase ambiguity

So, need only one derivative thus less chance of non-isomorphism

PHASING WITHOUT A NATIVE CRYSTAL USING SAD

$$f = f_o + f'(\lambda) + if''(\lambda)$$

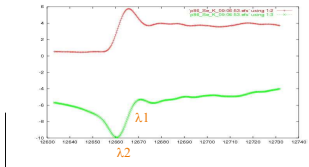


SAD results (like SIR) in phase ambiguity. But can use this technique to solve the phase problem.

Also allows the solution of the phase problem by collecting data from native crystals of macromolecules (i.e. no derivative). Exploits small f" available from S or P at the wavelengths we can routinely access.

WHY DO WE CARE ABOUT *F*''?

$$f = f_o + f'(\lambda) + if''(\lambda)$$



If we at collect data at different wavelengths about absorption edge then F_{hkl} at different wavelengths changes

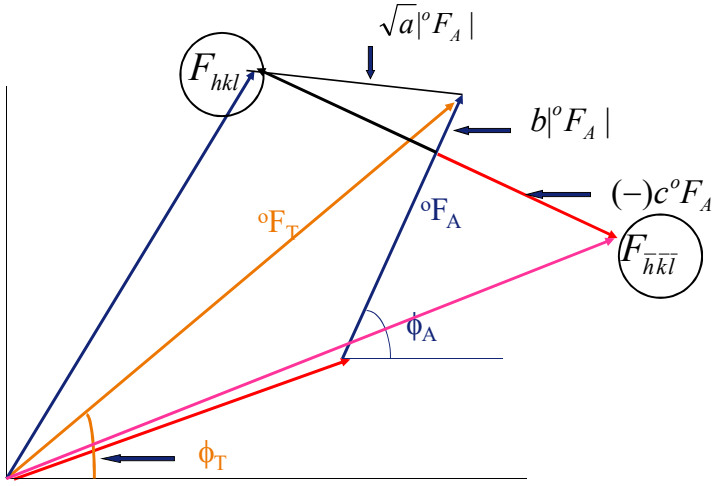
$$^{\lambda 1}F \neq {}^{\lambda 2}F$$

DISPERSIVE difference:

$$\Delta F_{disp} = |{}^{\lambda 1}F_{hkl}| - |{}^{\lambda 2}F_{hkl}|$$

By combining measurement of anomalous and dispersive differences we can solve the phase problem using HA derivatives in MX without measuring data from a native crystal. How?

THE STRUCTURE FACTOR IN THE PRESENCE OF ANOMALOUS SCATTERING



PHASING USING MULTI-WAVELENGTH ANOMALOUS DISPERSION (MAD) - DATA COLLECTED AT SEVERAL WAVELENGTHS AROUND ABSORPTION EDGE

$$|^{\lambda n}F_{hkl,\bar{h}\bar{k}\bar{l}}|^2 = |^oF_T| + a(\lambda)|^oF_A|^2 + b(\lambda)|^oF_T||^oF_A|\cos(^o\phi_T - ^o\phi_A)$$
$$\pm c(\lambda)|^oF_T||^oF_A|\sin(^o\phi_T - ^o\phi_A)$$

$$|^{\lambda n}F_{hkl}|^2 - |^{\lambda n}F_{\bar{h}\bar{k}\bar{l}}|^2 = 2c(\lambda_n)|^oF_T||^oF_A|\sin(^o\phi_T - ^o\phi_A)$$

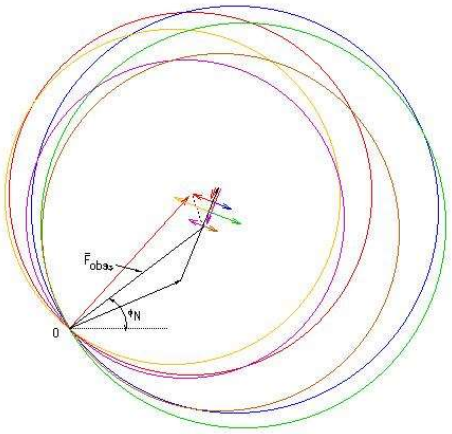
$$|^{\lambda i}F|^2 - |^{\lambda j}F|^2 = [a(\lambda_i) - a(\lambda_j)]|^oF_T|^2 + [b(\lambda_i) - b(\lambda_j)]|^oF_T||^oF_A|\cos(^o\phi_T - ^o\phi_A)$$

$$a = \frac{(f''^2(\lambda) + f'^2(\lambda))}{f_o} \qquad b = \frac{2f'(\lambda)}{f_o} \qquad c = \frac{2f''(\lambda)}{f_o}$$

^oF_A and ^of_A are derived from positions of anomalous scatterers which are elucidated using as anomalous of dispersive differences

Solve simultaneous equations for ^oF_T and ($^o\phi_A$- $^o\phi_T$)

MAD AS QUASI-(M)IR

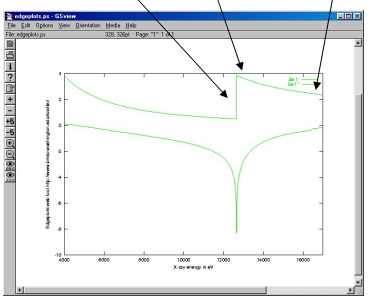


f'at minimum, moderate value of f"

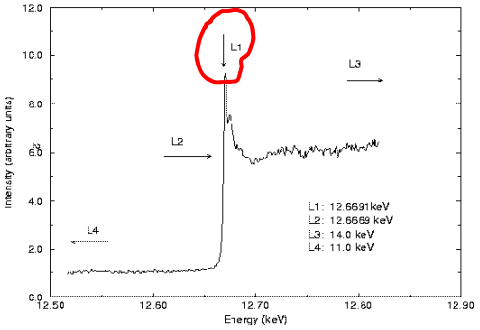
f' moderate, f" moderate

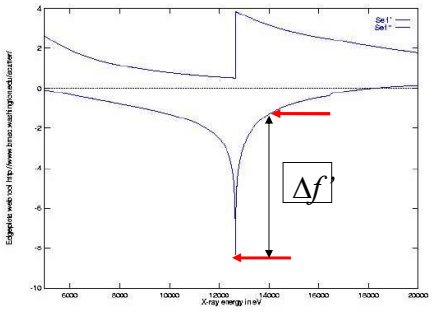
f' moderate,

f' large



MAD - WHICH WAVELENGTHS? - I



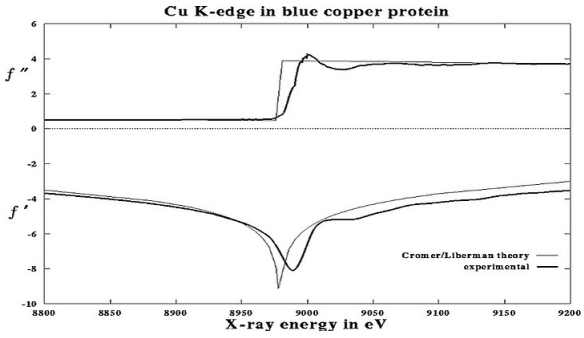


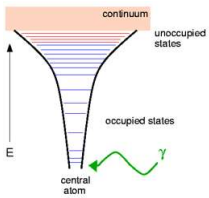
"From a practical point of view a larger f" will give more accurate values for the possible solutions [of the phase]...."
Woolfson & Fan, "Solving Crystal Structures" (1995) Cambridge University Press

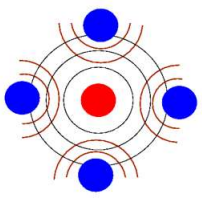
"....while a larger difference in f' will provide better discrimination in resolving the [phase] ambiguity"
Woolfson & Fan, "Solving Crystal Structures" (1995) Cambridge University Press

To maximise *f'* need to collect two data sets: (1) at inflection point. (2) (preferably) high energy remote. So, a typical MAD experiment uses three wavelengths (peak, inflection point, remote)

MAXIMUM SIGNAL: EXPERIMENT DIFFERS FROM THEORY. DON'T RELY ON TABLES







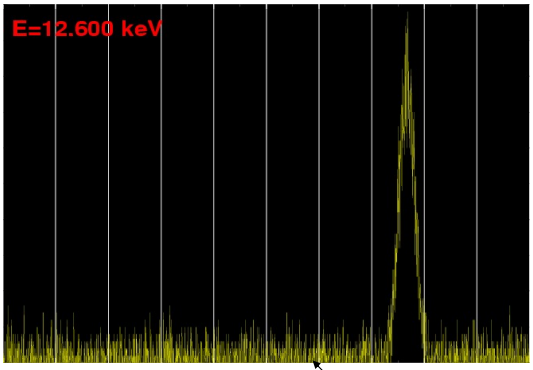
***JM Guss, EA Merritt, RP Phizackerley, B Hedman, M Murata, KO Hodgson, & HC Freeman (1989)**. "Phase determination by multiple-wavelength X-ray diffraction: crystal structure of a basic blue copper protein from cucumbers". Science 241, 806-811.*

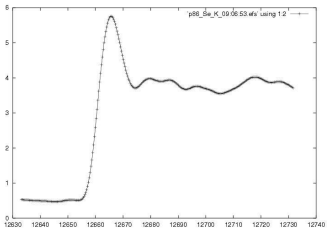
We're not dealing with free atoms. Interaction of anomalous scatterer with other atoms causes deviation from ideality, as do changes in oxidation state. So, we MUST measure absorption edge experimentally if we want to be sure to obtain maximum 'signal'

MEASURING AN ABSORPTION EDGE

'Incident X-ray peak' ~static (i.e. energy change not big). We don't really care about this

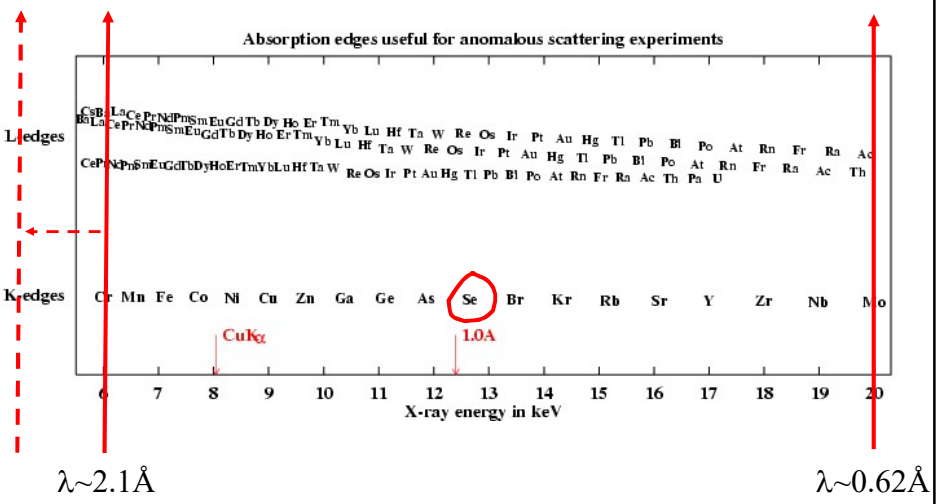
Measure intensity contained in 'region of interest' around expected position (i.e. energy) of emission line as we scan from low to high energy around the absorption edge.





Experimental fluorescence scan

ENERGY OF ABSORPTION EDGES COMMONLY USED IN EXPERIMENTAL PHASING PROCEDURES



PHASING EXPLOITING ANOMALOUS SCATTERING

1. Introduce a small number of anomalously scattering atoms into crystals
2. **Collect both F^+ and F^-** at one (or more) wavelength(s). If more than one wavelength make sure these are chosen using absorption edge scan.
3. Derive anomalous and (if more than one wavelength) dispersive differences.
4. Use anomalous and/or dispersive differences to locate anomalously scattering atoms
5. Use these positions to initiate phase calculations

COLLECTING BOTH F(+) AND F(-) (AFTER DAUTER, Z. *METHODS IN ENZYMOLOGY*, 276)

Crystal System	Laue Group	Total Oscillation required for	
		Standard data collection	**Anomalous data collection**
Triclinic	-1	180°	**360°**
Monoclinic	2/m	180° (b*), 90° (a*,c*)	**180° (a*,b*,c*)**
Orthorhombic	mmm	90°(a*, b*,c*)	**90°(a*,b*,c*)**
Tetragonal	4/m	90°(a*, b*,c*)	**90°(c*), 180°(a*, b*)**
	4/mmm	45°(c*), 90° (a*, b*)	**45°(c*), 90°(a*,b*)**
Trigonal	3/m	60°(c*), 90°(a*,b*)	**120°(c*), 180°(a*, b*)**
	-31m	30°(c*), 90°(a*,b*)	**60°(c*), 180°(a*, b*)**
	-3m1	30°(c*), 90°(a*,b*)	**30°(c*), 180°(a*, b*)**
Hexagonal	6/m	60°(c*), 90°(a*,b*)	**60°(c*), 180°(a*, b*)**
	6/mmm	30°(c*), 90°(a*,b*)	**60°(c*), 90°(a*, b*)**
Cubic	23	~60°	**~70°**
	432	~35°	**~45°**

(S)MAD: GOOD & BAD

1. No 'native' crystal – no problems with isomorphism when using heavy atom derivatives
2. Selenomethionine can be introduced into many protein sequences and can be used as a "magic bullet"
3. The 'method of choice' for *de novo* structure solution in macromolecular crystallography
4. Signals are small in comparison to those from isomorphous replacement – data have to be good!
5. In MAD have to collect a lot of data from one crystal – radiation damage is a problem

• Evans, P. & McCoy, A. (2008). An introduction to molecular replacement, *Acta Cryst.* **D66,** 1–10.

• Taylor, G.L. (2010). Introduction to phasing. *Acta Cryst.* **D66**, 325-338.

• Dauter, Z., Dauter, M. & Dodson, E.J. (2002). Jolly SAD. *Acta Cryst.* **D58**, 494-506.

• Hendrickson W.A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*. **254**, 51-8.

• Sheldrick, G. M. (2008). A short history of SHELX. *Acta Cryst.* **A64**, 112-122.

• Bourenkov, G.P. & Popov, A.N. (2010). Optimization of data collection taking radiation damage into account. *Acta Cryst.* **D66**, 409-419.