Haupt/Masterstudiengang Physik
Methoden moderner Röntgenphysik II:
Streuung und Abbildung
SS 2014

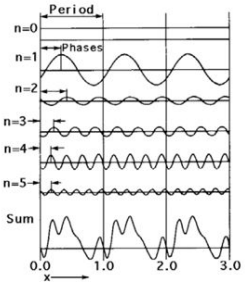
Biology III: Crystallographic phases

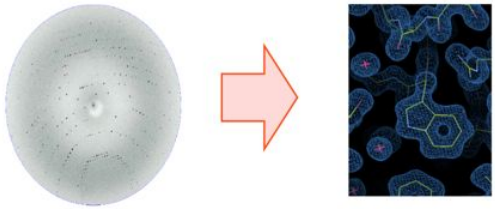
Thomas R. Schneider, EMBL Hamburg
8/7/2013
thomas.schneider@embl-hamburg.de

EMBL

Recap

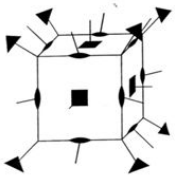
- Real and reciprocal space
- The crystallographic phase problem
- Symmetry in real and reciprocal space

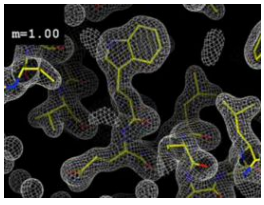


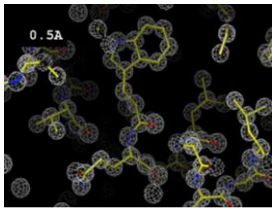


$$\rho_{xyz} = \sum_{hkl} |F_{hkl}| e^{-i\varphi_{hkl}} e^{-2\pi i(hx+ky+lz)}$$

$$F_{hkl} = \sum_{j=1}^{N} f_j e^{2\pi i(hx_j+ky_j+lz_j)}$$







For chiral objects forming crystals, only 65 space groups (lattice + point group) are possible. That is all!

Today

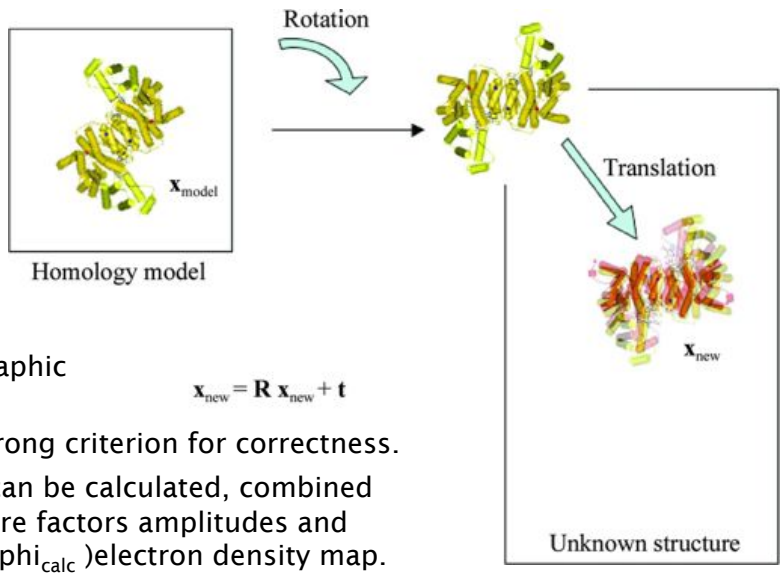
- Phasing methods
- Experiments

Phase Determination in Practise

- **Ab Initio**
  - INPUT: F_{hkl} only
  - Needs data to a resolution where atoms are separated, i.e. 1.1 Å
- **Molecular Replacement**
  - INPUT: F_{hkl} / homologues structure model / position and orientation
  - Needs homologues model
  - Model Bias
- **Multiple Isomorphous Replacement (MIR)**
  - INPUT: F_{hkl} for native, F_{hkl} for derivative 1, F_{hkl} for derivative 2
  - Derivatization
  - Non-isomorphism
- **Anomalous Diffraction (MAD, SAD)**
  - INPUT: F_{hkl} from one crystal

Molecular Replacement

- Assuming that the structure of the molecule of interest and the structure of a molecule with similar amino-acid sequence (the search model) are similar, an approximate model of the content of the crystal can be made by placing the search model into the crystallographic unit cell via a 6D-search.



$$\mathbf{x}_{new} = \mathbf{R}\,\mathbf{x}_{new} + \mathbf{t}$$

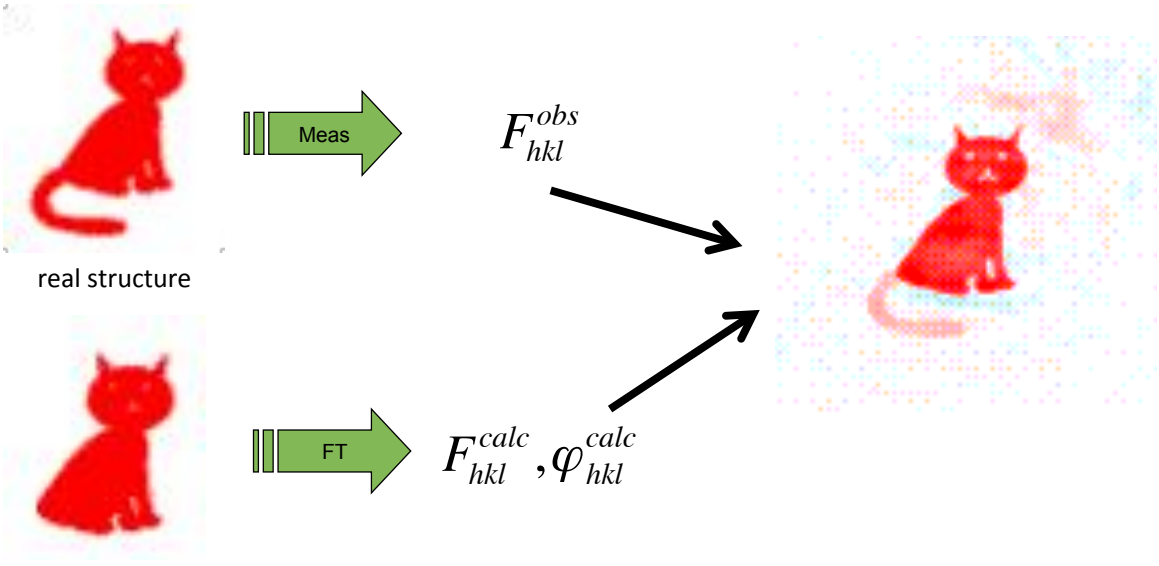
- Consistent packing is a strong criterion for correctness.
- From this model, phases can be calculated, combined with the measured structure factors amplitudes and used to calculate an (F_{obs}, phi_{calc})electron density map.

$$F_{hkl,calc} = \sum_{j=1}^{N} f_j e^{2\pi i(hx_j+ky_j+lz_j)}$$

Model bias I

Adapted from xtal.ohsu.edu/teaching/lewis_clark/Models.pptx

- The first electron density map after molecular replacement is calculated with phases derived from the potentially/likely incorrect search model.



Model bias II

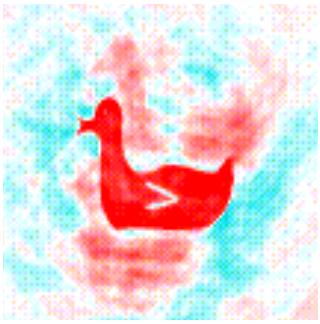
http://www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html

- If the search model is very different from the target model, the image resulting from the (F_{obs}, phi_{calc}) synthesis can be misleading. 'phase bias'.

real structure

Meas

F_{hkl}^{obs}



positioned search model

FT

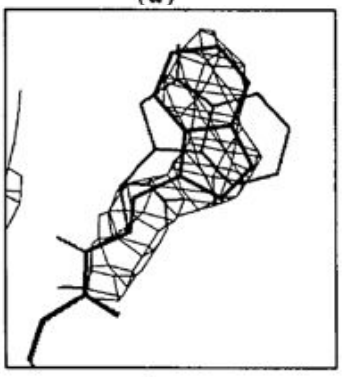
$F_{hkl}^{calc}, \varphi_{hkl}^{calc}$

In reality, distinguishing between I and II is difficult

Model Bias

Fig. 6. The effects of various omit-map techniques around Trp 91 located in the light chain of the AN02 Fab fragment. The correct conformation of Trp 91 is shown in black, the initial incorrect conformation is shown in gray. All maps are of the σ_A-weighted $2F_o - F_c$ type (see *Methods*) at 2.8 Å resolution. Residues 89–97 in the light chain, 93–102 from the heavy chain, and the hapten molecule were omitted for all map calculations and refinements. (*a*) Ordinary omit map of the initial incorrect structure shown at a contour level of 1.2σ. (*b*) The fourth iteration of density modification on the initial incorrect structure shown at a contour level of 0.7σ. The region of modification was defined by a 2.0 Å cushion around the omitted region. (*c*) Minimized omit map of the initial structure shown at a contour level of 1.2σ. The partial structure was refined through 120 cycles of conjugate-gradient minimization at 8.0–2.8 Å resolution. (*d*) Randomized and minimized omit map of the initial structure shown at a contour level of 1.2σ. (*e*) SA omit map of the initial structure shown at a contour level of 1.2σ. The partial structure was refined using a SA slow-cooling protocol (Brünger *et al.*, 1990) with a starting temperature of 3000 K at 8.0–2.8 Å resolution.

(*a*)



(*e*)

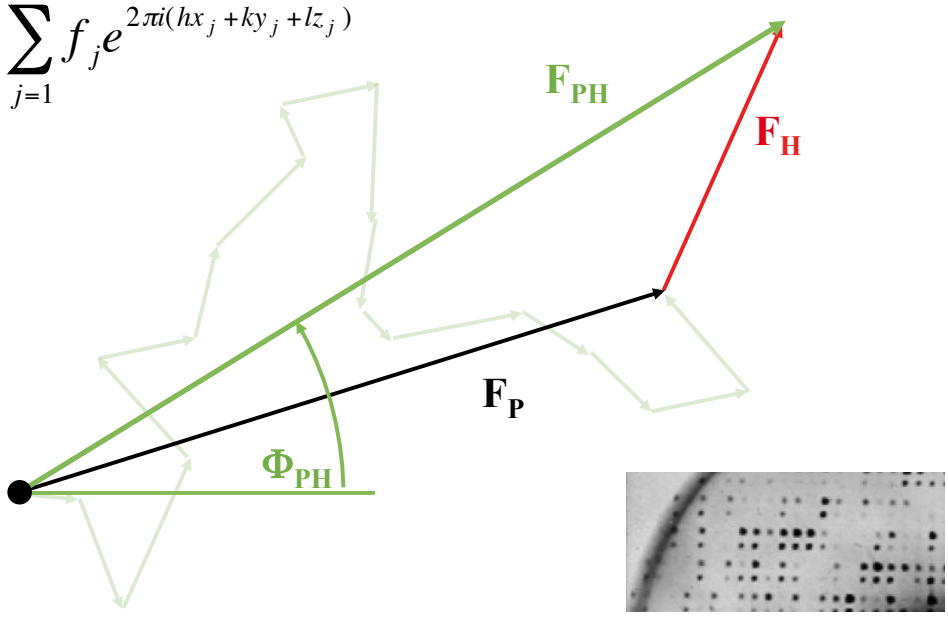
Isomorphous Replacement

Taylor (2010) Acta Cryst D66:325

- By introducing extra-atoms into a crystal, its diffraction properties become altered.
- Determining phases experimentally is better than inventing phases based on insecure assumptions.

A protein and an extra scatterer

$$F_{hkl} = \sum_{j=1}^{N} f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$



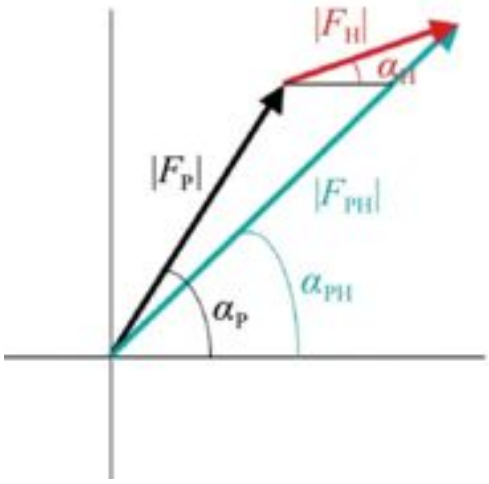
Assumption. Apart from adding the extra scatterer, 'nothing' changes -> isomorphism

Isomorphous replacement

- Addition of heavy atoms leads to changes in diffracted intensities. E.g. a single Hg-atom (80 e^-) in 1000 CNO-atoms (6-8 e^-) changes the intensities on average by 25%.
- The *isomorphous difference*:

$$|F_H| = |F_{PH}| - |F_P|$$

can be used as an estimate of the heavy atom structure factor amplitude



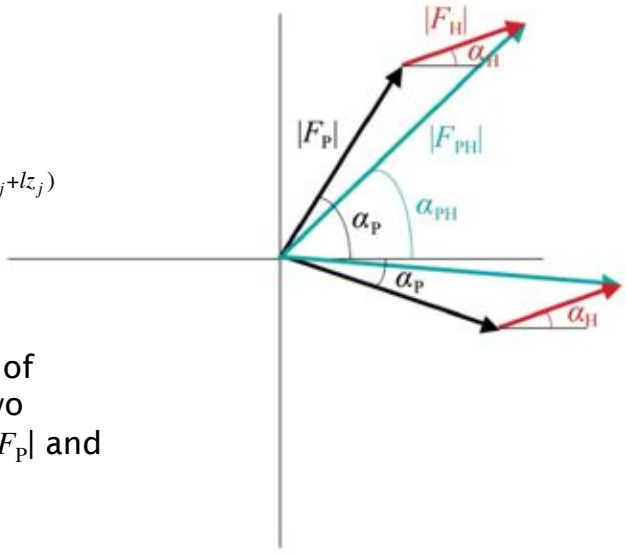
Taylor (2010) Acta Cryst D66:325

Single Isomorphous Replacement

- After finding the heavy atoms, complex structure factors for the heavy atoms can be calculated.

$$F_{H,hkl,calc} = \sum_{j=1}^{N} f_j e^{2\pi i(hx_j+ky_j+lz_j)}$$

- However, for a given set of heavy atom positions, two phases are possible for $|F_P|$ and $|F_{PH}|$

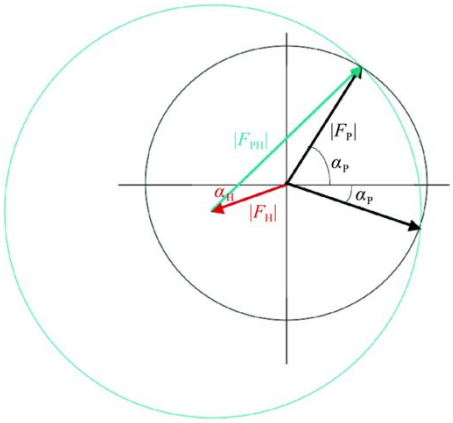


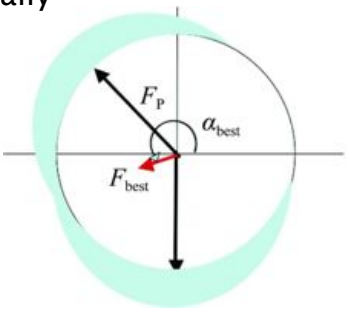
Tafel – Harker construction

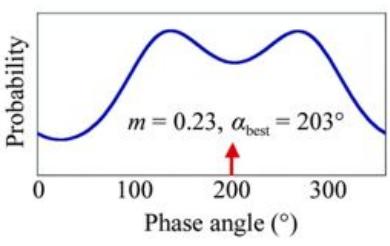
Isomorphous Replacement

Taylor (2010) Acta Cryst D66:325

- From the differences observed between the measurements on the native and the derivate crystal, the positions of the extra atoms can be found.
- By clever combination of the knowledge about the positions of the extra atoms and the differences in the measurements, phase information can be derived.
- Unfortunately, usually two phase angles are consistent with the data. 'Phase ambiguity'

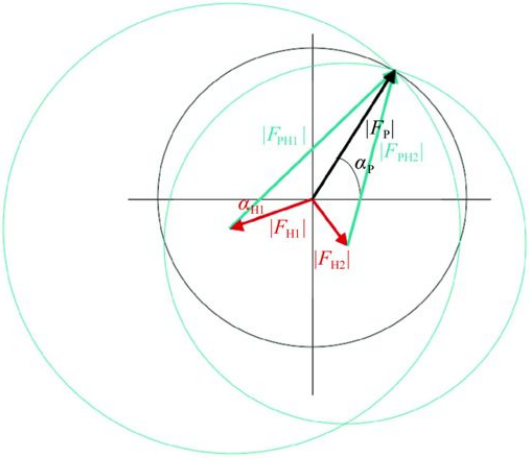




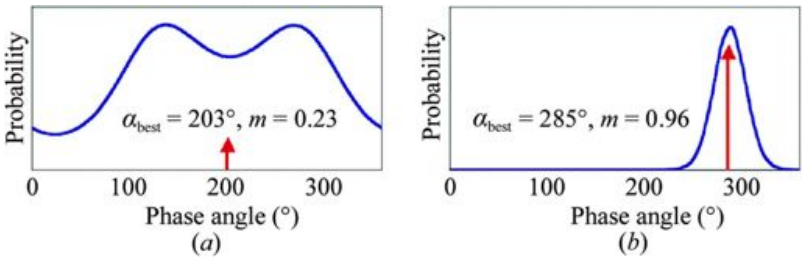


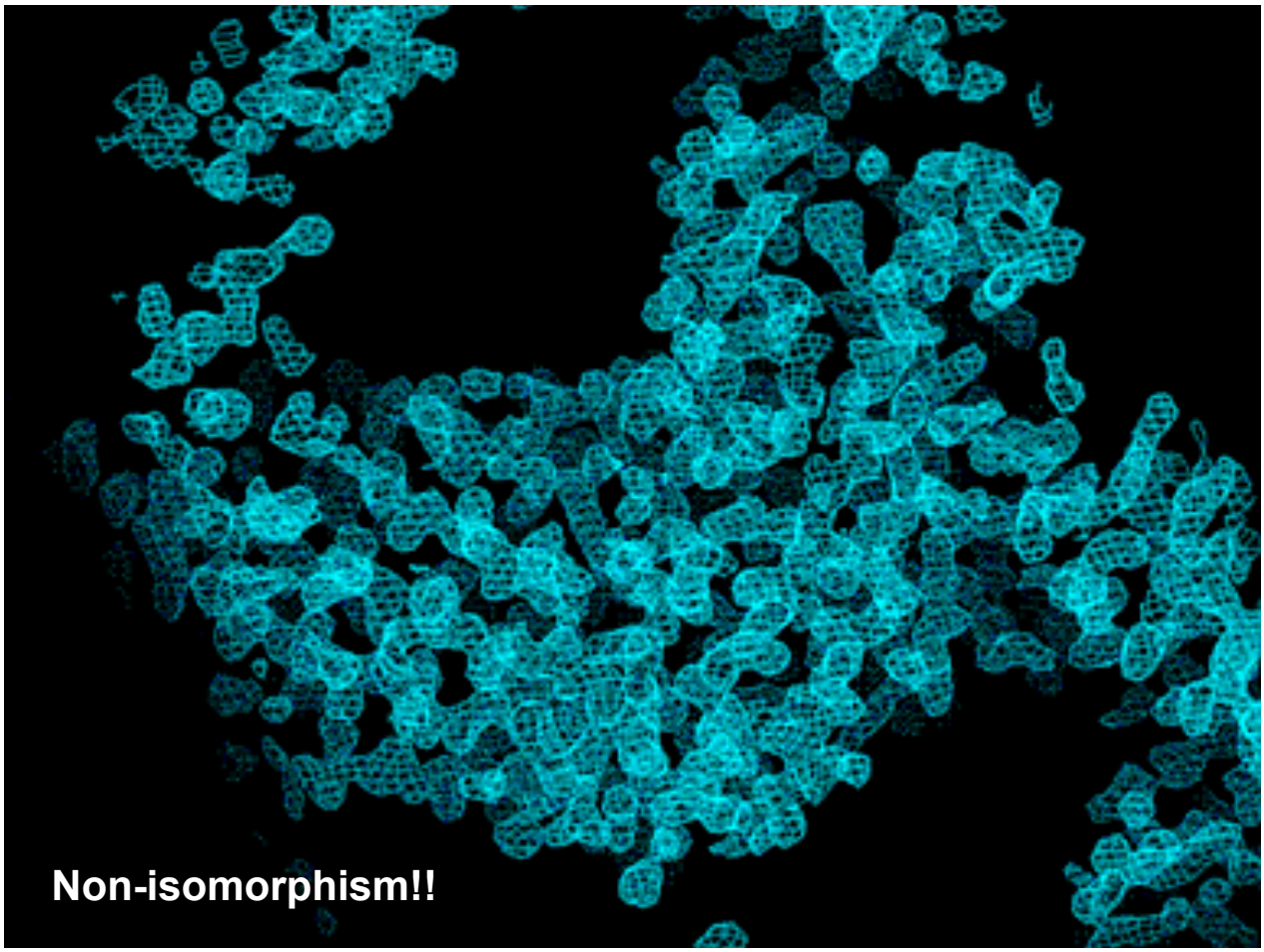
Multiple Isomorphous Replacement

- When two 'derivatives' are used, the phase ambiguity can be resolved and a uni-model phase distribution can be constructed.
- An electron density map can be calculated.



Taylor (2010) Acta Cryst D66:325





Anomalous Scattering

Normal Scattering (centrosymmetry)

Friedel's law $\left|F_{hkl}\right| = \left|F_{\overline{hkl}}\right|$

Anomalous Scattering

Friedel's law is broken $\left|F_{hkl}\right| \neq \left|F_{\overline{hkl}}\right|$

$F_{hkl} \neq F_{\overline{hkl}}$ can be measured experimentally

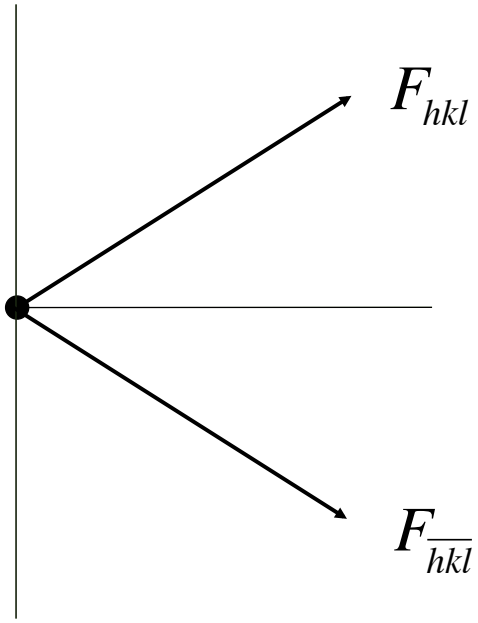
Physics of Anomalous Scattering

- Normal scattering
  - elastic
  - direction-dependent

- Anomalous scattering
  - resonant / inelastic
  - direction independent

Pictures courtesy of Zbyszek Dauter

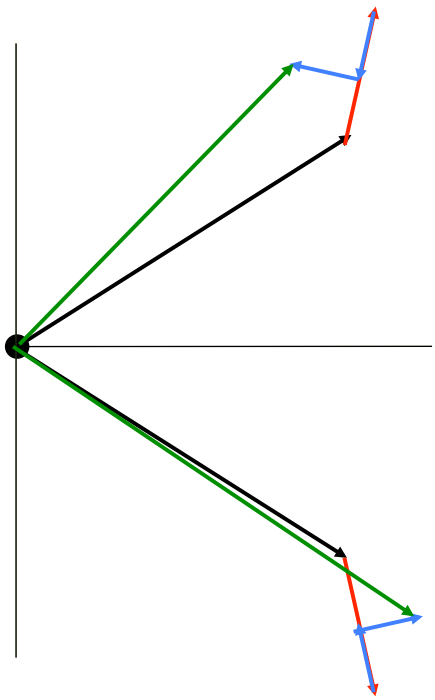
Normal diffraction into hkl and -h-k-l



$$\left|F_{hkl}\right| = \left|F_{\overline{hkl}}\right|$$

$$\varphi_{hkl} = -\varphi_{\overline{hkl}}$$

Anomalous diffraction into hkl and -h-k-l

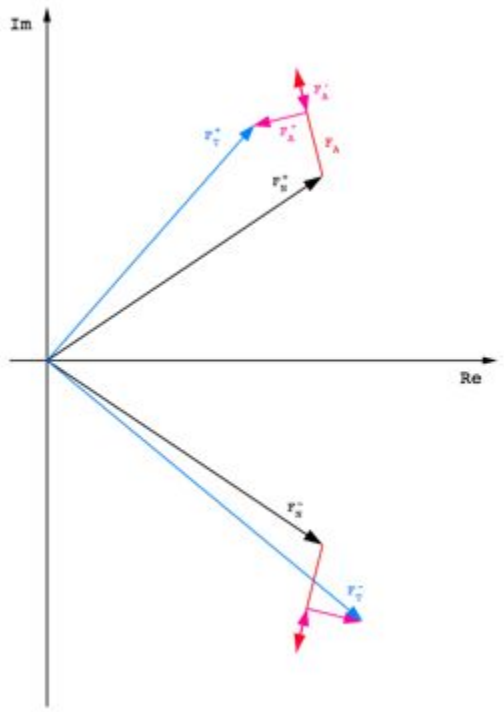


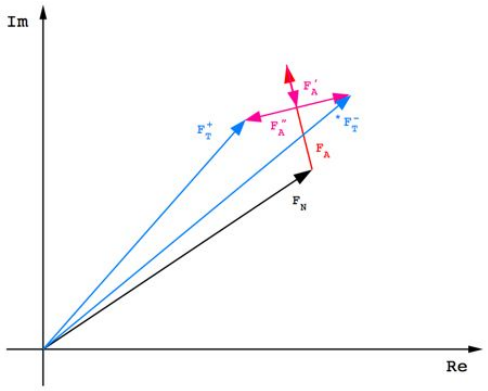
$$f = f_0 + f' + if''$$

$$\left|F_{hkl}\right| \neq \left|F_{\overline{hkl}}\right|$$

$$\left|F^+\right| \neq \left|F^-\right|$$

Anomalous Diffraction into hkl and -h-k-l



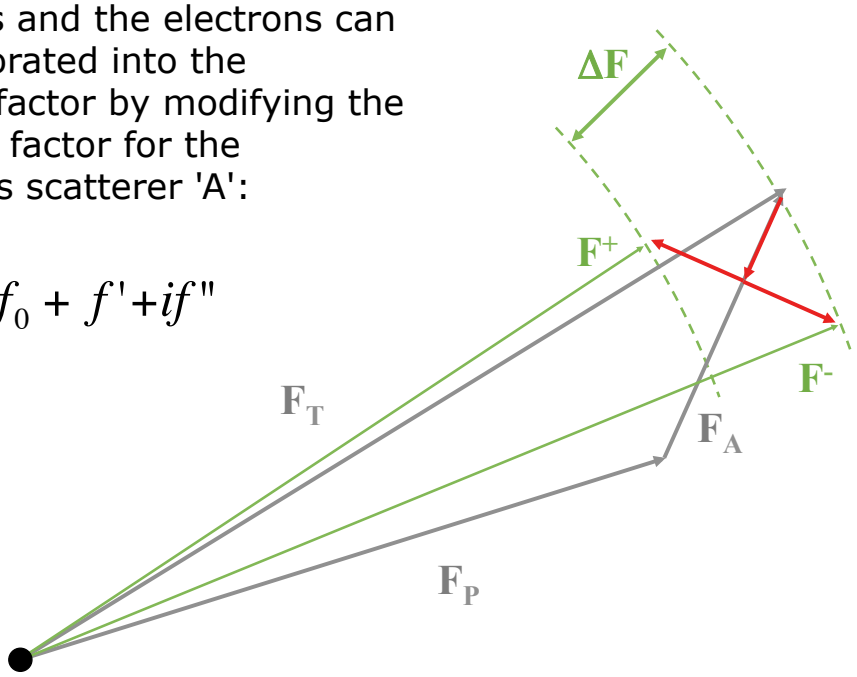


Pictures courtesy of Zbyszek Dauter

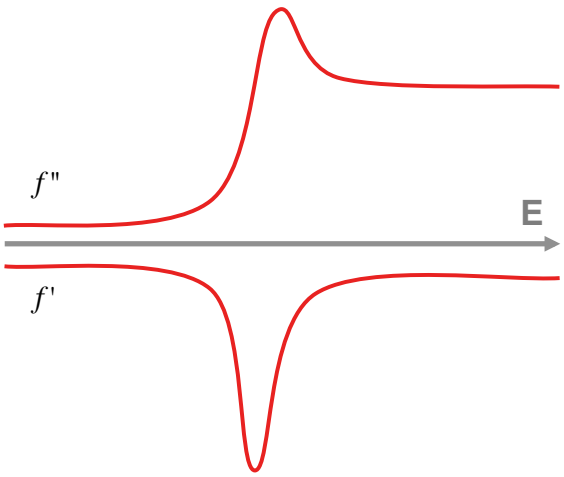
A protein and an anomalous scatterer A

A non-elastic interaction between the X-rays and the electrons can be incorporated into the structure factor by modifying the scattering factor for the anomalous scatterer 'A':

$$f = f_0 + f' + if''$$



Anomalous Dispersion



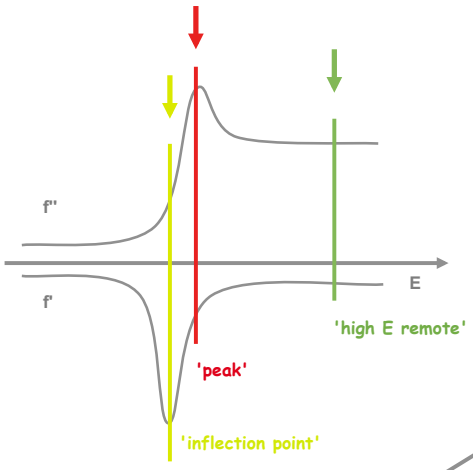
f"(E) can be measured by X-ray absorption spectroscopy:

$$f''(E) = \frac{mc}{2he^2} E\mu(E)$$

f'(E) can be calculated from f"(E) by a Kramers-Kronig transformation

$$f'(E) = \frac{2}{\pi} \int_0^\infty \frac{E' f''(E')}{(E^2 - E'^2)} dE'$$

A MAD Experiment

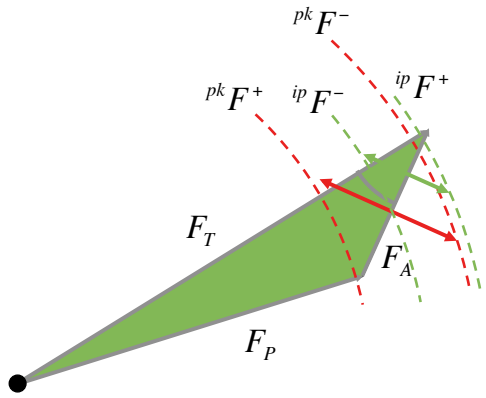


for one pair of refl:

F^+ F^- F^+ F^- F^+ F^-

from the same crystal

The MAD Observational Equations



Hendrickson WA et al. (1985)
Meth.Enz. 115:41-55

$$\left|{}^{\lambda}F^{\pm}_{obs}\right|^2 = \left|F_T\right|^2 + a(\lambda)\left|F_A\right|^2 + b(\lambda)\left|F_T\right|\left|F_A\right|\cos\alpha \pm c(\lambda)\left|F_T\right|\left|F_A\right|\sin\alpha$$

$$\left|{}^{pk}F^+_{obs}\right|^2 = \left|F_T\right|^2 + a(pk)\left|F_A\right|^2 + b(pk)\left|F_T\right|\left|F_A\right|\cos\alpha + c(pk)\left|F_T\right|\left|F_A\right|\sin\alpha$$

$$\left|{}^{pk}F^-_{obs}\right|^2 = \left|F_T\right|^2 + a(pk)\left|F_A\right|^2 + b(pk)\left|F_T\right|\left|F_A\right|\cos\alpha - c(pk)\left|F_T\right|\left|F_A\right|\sin\alpha$$

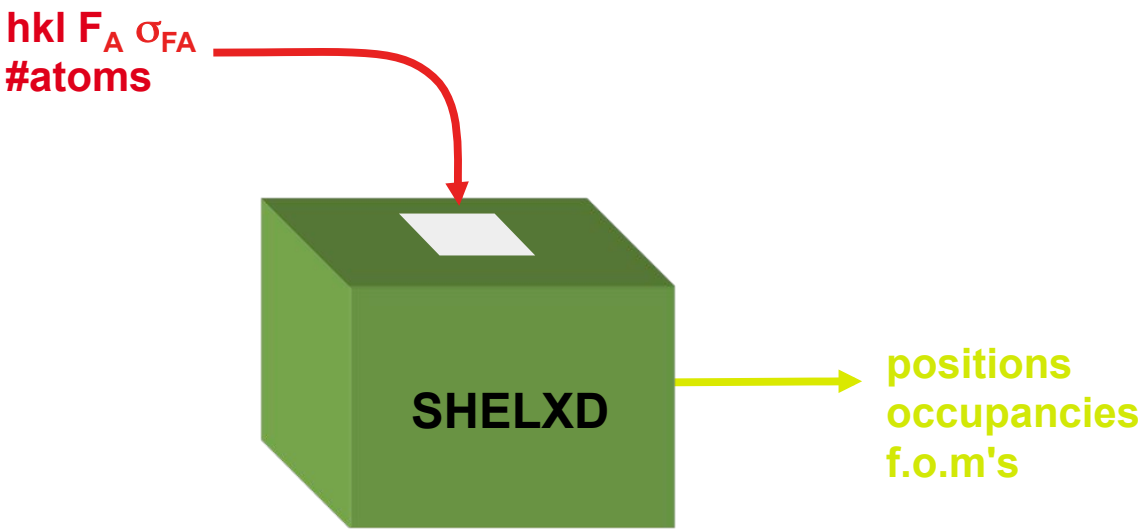
$$\left|{}^{ip}F^+_{obs}\right|^2 = \left|F_T\right|^2 + a(ip)\left|F_A\right|^2 + b(ip)\left|F_T\right|\left|F_A\right|\cos\alpha + c(ip)\left|F_T\right|\left|F_A\right|\sin\alpha$$

$$\left|{}^{ip}F^-_{obs}\right|^2 = \left|F_T\right|^2 + a(ip)\left|F_A\right|^2 + b(ip)\left|F_T\right|\left|F_A\right|\cos\alpha - c(ip)\left|F_T\right|\left|F_A\right|\sin\alpha$$

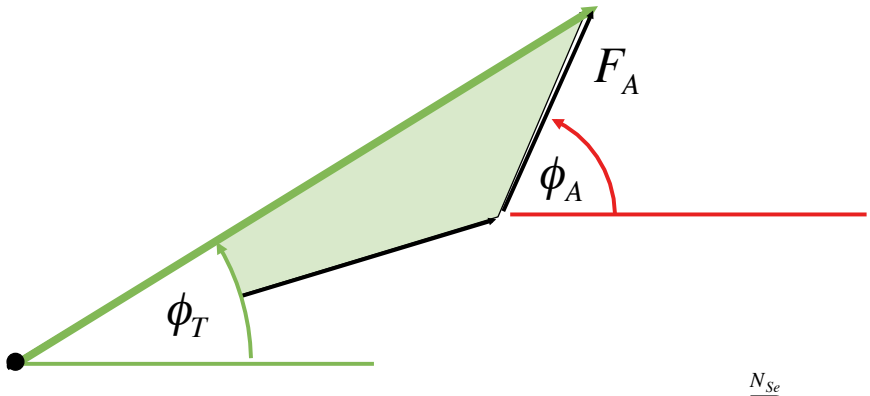
The meaning of FH/A-values

- In principle, $F_{H/A}$-values represent the substructure of anomalous / heavy scatterers only. This reduces the initial step of phasing to finding a few atoms in a large unit cell.
- *Ab initio* methods can be used for finding the anomalous / heavy scatterers.
- ΔF-values can be considered as lower limit estimates of F_A-values.

Solving the substructure against $|F_A|$



Knowing the substructure allows to calculate ϕ_A and ϕ_T

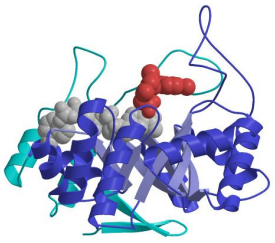


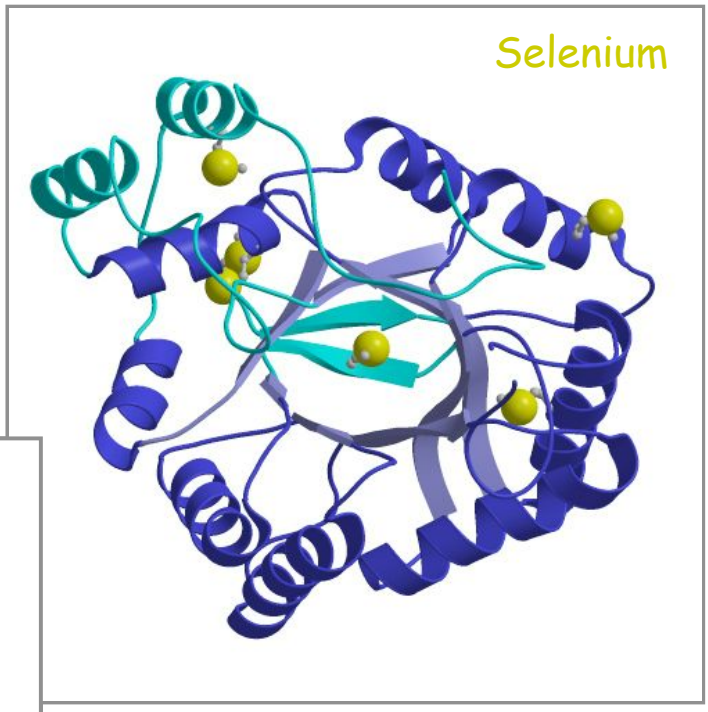
$$F_{A,hkl} = \sum_{j=1}^{N_{Se}} f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

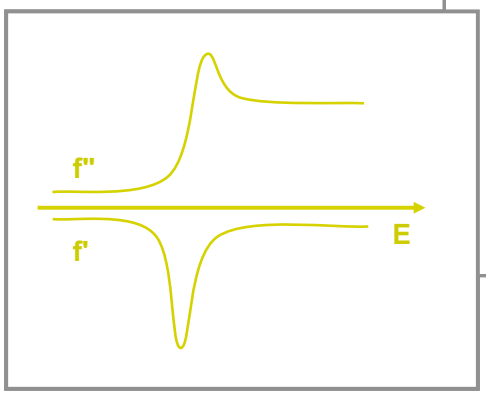
$$\Rightarrow |F_{A,hkl}|, e^{i\varphi_{A,hkl}}$$

Real life is not that simple - ML as implemented e.g. in SHARP is the way to go.

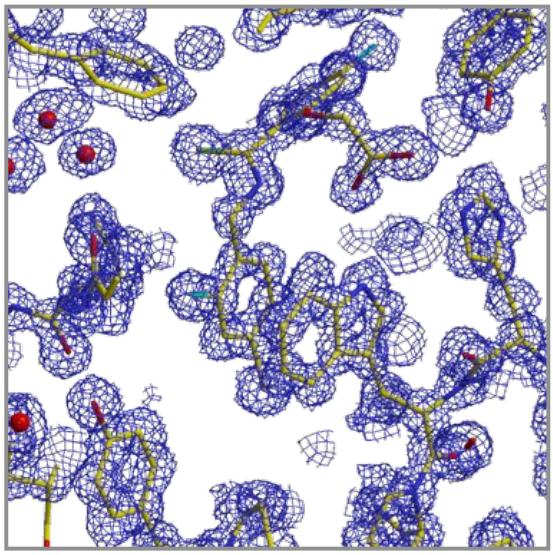
... in proteins





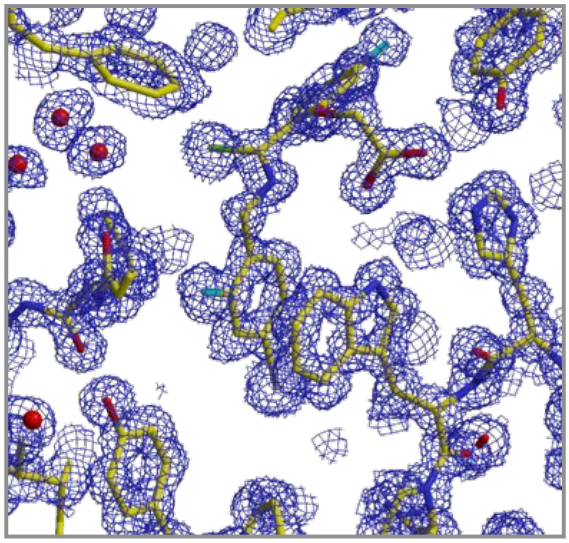


Aldose Reductase - experimental vs. refined



experimental

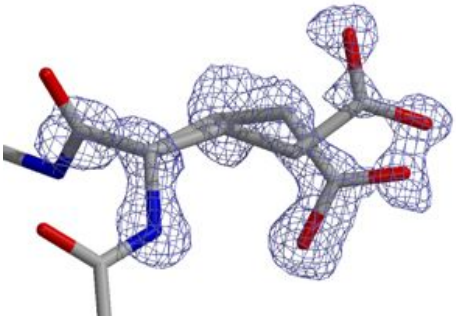
(F_{obs},Φ_{MAD})- map
0.9 Å, contoured at 1σ

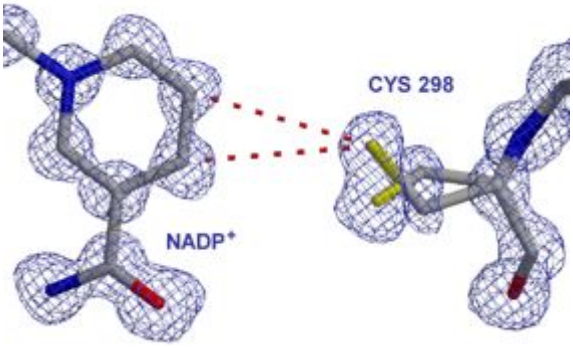


refined (remote)

σ_A weighted $(2F_{obs}-1F_{calc},\Phi_{calc})$-map
0.9 Å, contoured at 1σ

Aldose Reductase - Level of Detail

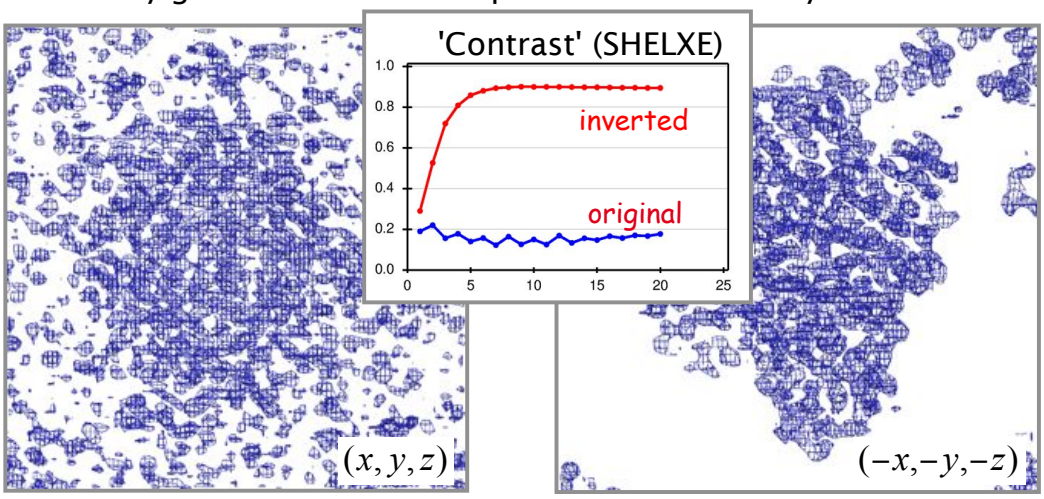




- Multiple conformations
- O > N > C

Choosing the right hand

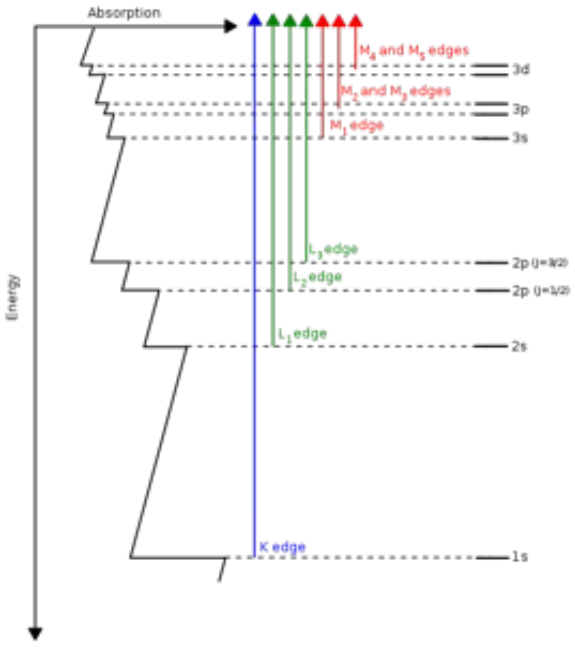
- The correct solution for the substructure and its enantiomorph have the same agreement with the experimental data.
- But they give rise to different phase sets for the crystal structure



- local r.m.s. density Terwilliger & Berendsen (1999) Acta Cryst. D55:1872
- size of the PNG-file: 591451 vs. 399487 Bytes

Back to the experiment

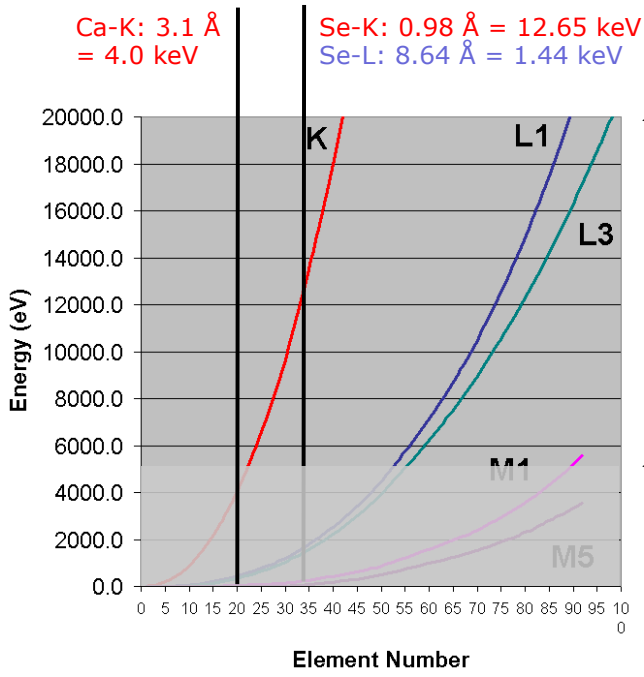
- Anomalous diffraction occurs when when the energy / wavelength used is close to an absorption edge where transitions of electrons can be excited.



http://en.wikipedia.org/wiki/X-ray_absorption_spectroscopy

Accessible Edges

http://ruppweb.dyndns.org/Xray/101index.html

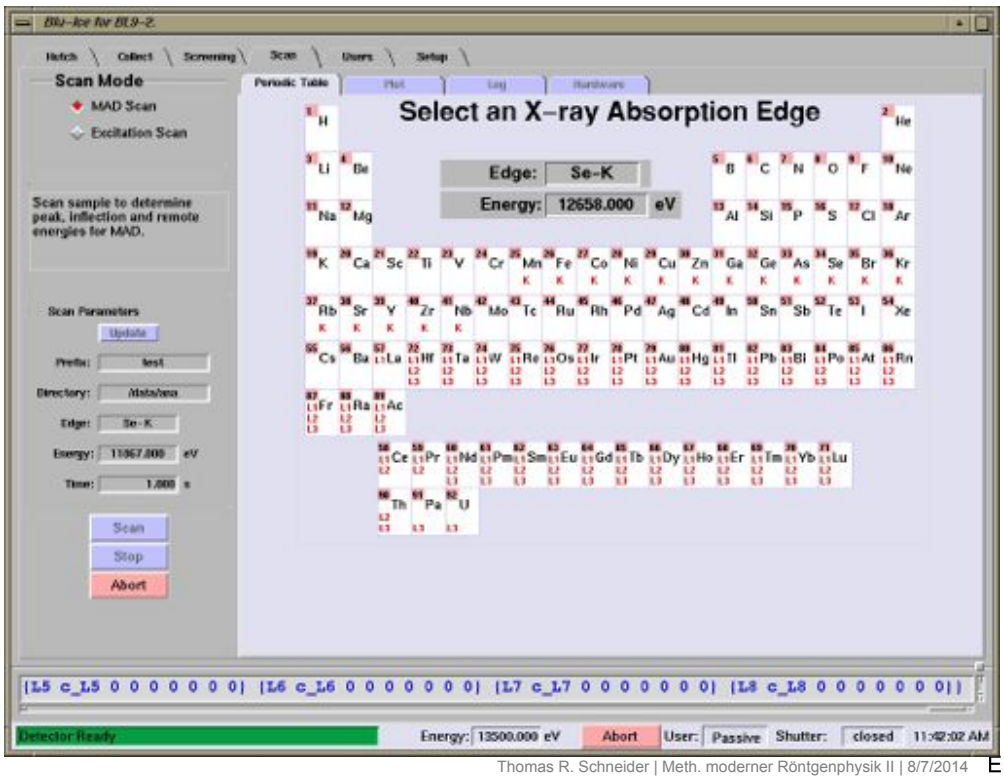


12.4 keV / **20 keV** = 0.62 Å

12.4 keV / **5 keV** = 2.48 Å

Accessible Edges

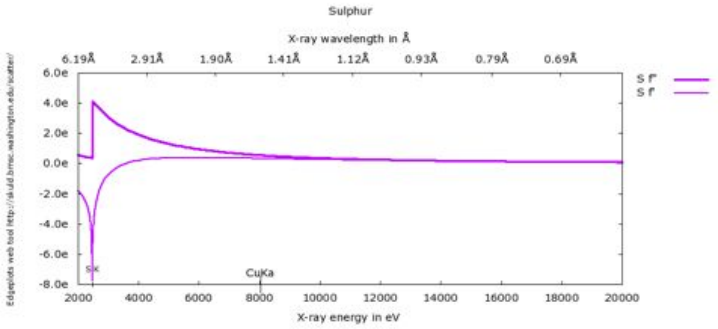
BL9-2 @ SSRL: 0.85-2.06 Å / 6.0-14.5 keV



Derivatization Laboratory, (11 m^2)

SAD Phasing – Only one wavelength

- A good option when:
  - the absorption edge is not reachable, e.g. for sulphur, the K edge is at 2.4 keV, 5.04 A (http://skuld.bmsc.washington.edu/scatter)/AS_form.html)



  - when crystals can not stand 3 successive data collections
  - when the beamline / home source is not tunable
  - when the beamline is not sufficiently stable
- Best explanation: Dauter (2002) Acta Cryst D58:1958

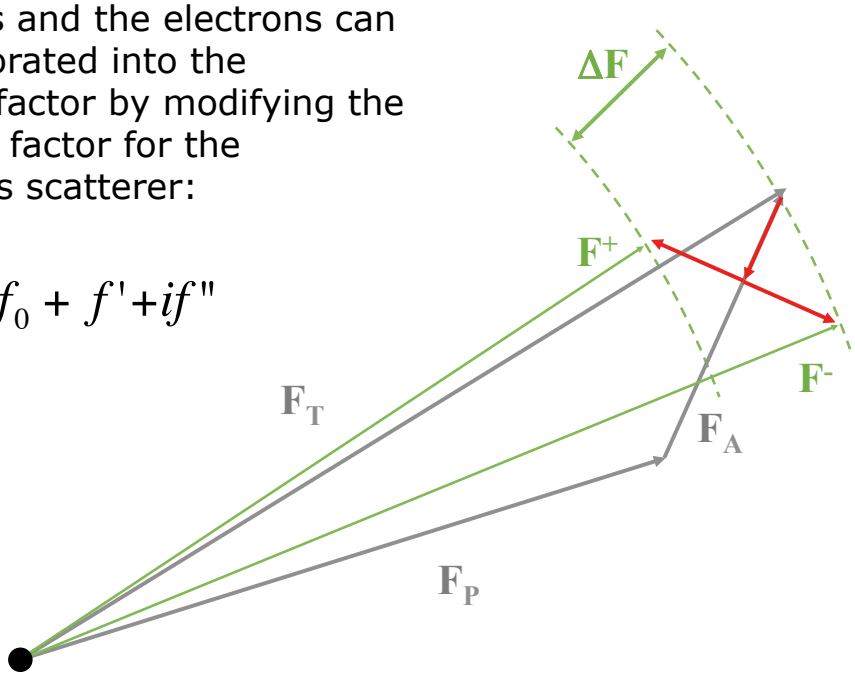
SAD

- How can we derive phase information if we have only one wavelength data?
  1. extract substructure structure factors
  2. solve substructure
  3. derive phase angles

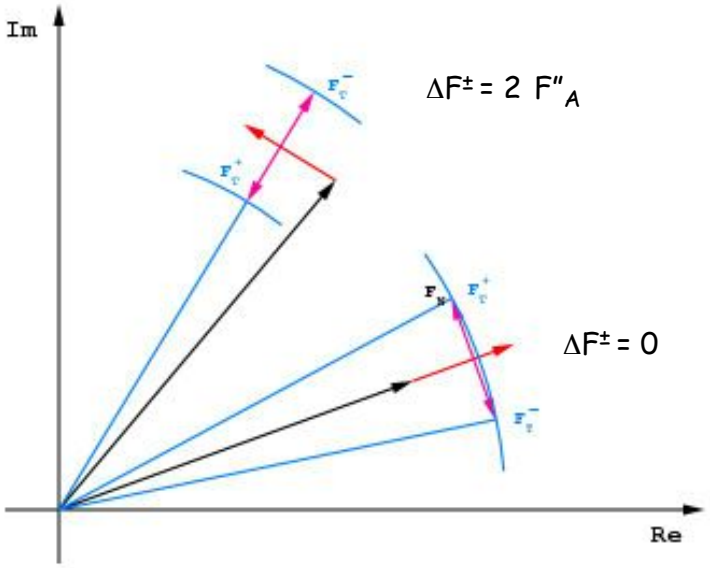
A protein and an anomalous scatterer A

A non-elastic interaction between the X-rays and the electrons can be incorporated into the structure factor by modifying the scattering factor for the anomalous scatterer:

$$f = f_0 + f' + if''$$



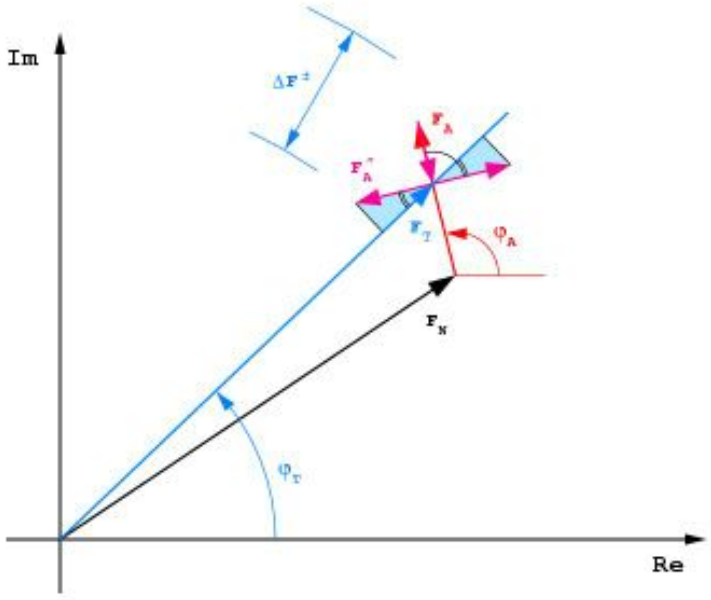
Relation between F_A and ΔF



- The magnitude of ΔF depends on the relative orientation between FP and F_A

Picture courtesy of Zbyszek Dauter

Sinusoidal dependence of ΔF and F_A



$$\Delta F^{\pm} = F^{+} - F^{-}$$
$$= 2F''_{A}\sin(\varphi_T - \varphi_A)$$

Hendrickson (1979)
Acta Cryst. A35:245

For large anomalous differences:

$$\Delta F^{\pm} \cong 2F''_A = 2F_A(f'' / f^0)$$

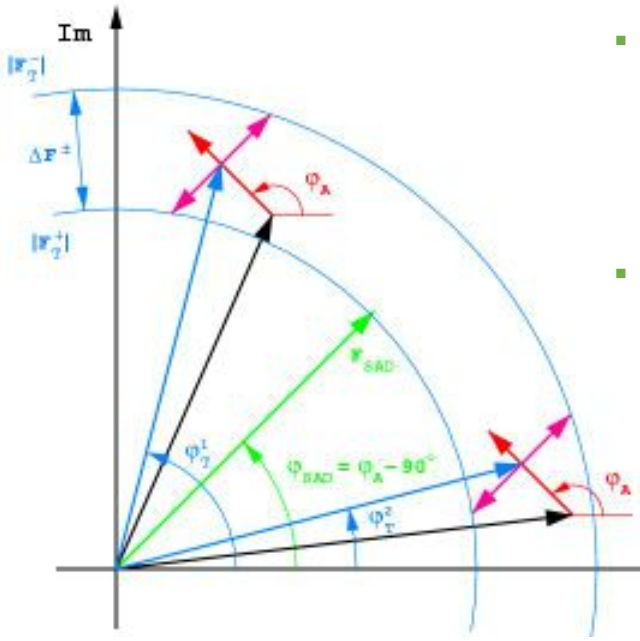
Rossmann (1961)
Acta Cryst.14:383

Picture courtesy of Zbyszek Dauter

SAD Phasing – Substructure Solution

- Anomalous differences can be considered as lower limit estimates to the F_A values for the anomalously scattering substructure.
- If we only use the large anomalous differences, these are actually proportional to the F_A values for the anomalously scattering substructure.
- As *ab initio* methods rely only on the 10-20% strongest reflections anyway, these methods can solve substructures based on anomalous differences.
- When we know the anomalous substructure, we can calculate the ϕ_A and then orient the ΔF pair of vectors.

SAD Phasing – Phase Ambiguity

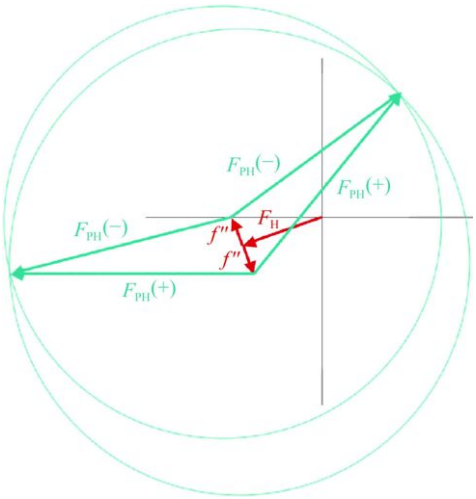


Picture courtesy of Zbyszek Dauter

- If the anomalous sites are known ($\Delta F\pm$, F_A , F'_A , F''_A , φ_A), there are two possible solutions for the phase φ_T of the crystal structure.
- There are various tricks to resolve the phase ambiguity
  - resolved anomalous phasing
  - density modification
  - direct methods
  - NCS

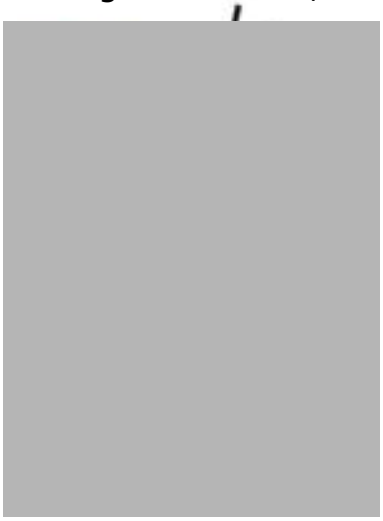
Anomalous Phasing

- By using X-rays with a wavelength/energy close to an absorption edge of atoms inside the crystal, Friedel's law $F^+=F^-$ is broken.
- 'The crystal becomes a derivative of itself'
- From the knowledge of the positions of the anomalous scatterers and the differences between the F^+ and the F^- reflections, phase information can be derived.
- This phase information is bimodal.
- The bimodality can be resolved
  - experimentally by collecting data at different energies on the same crystal
  - computationally by using various tricks, including density modification (e.g. solvent flattening).



Sulphur SAD phasing of a 66.3 kDa protein

- 1120 images of 1.0° for space group C2 (redundancy = 23) at a wavelength of 1.9 Å at BESSY (Berlin).
- Expected Bijvoet ratio with f''_S=0.82 at 1.9 Å was 1.8%.
- 21 S-sites were found with SHELXD.
- Phasing with SHARP / Density modification with SOLOMON



Lakomek et al. (2009) Acta Cryst D65:220

Electron-density map after density modification. One N-acetylglucosamine moiety of the glycan attached to Asn115 (for example) is already clearly visible in the experimental map at a level of 1.5 Å before the inclusion of any model phases and manual intervention.

Data Quality

Pictures from Wikipedia

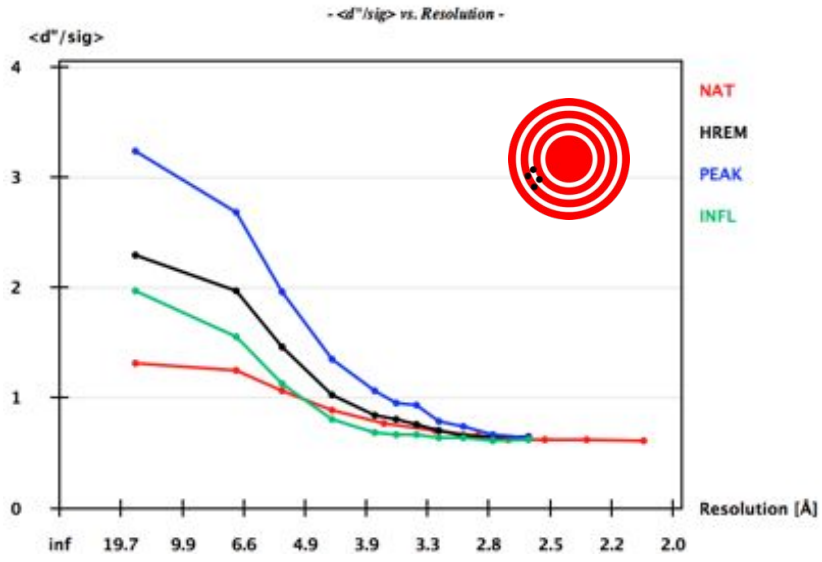
High precision, low accuracy

High accuracy, low precision

Of course we would like to achieve high accuracy and high precision. We need accurate measures of accuracy and precision to make sensible decisions during difficult structure determinations

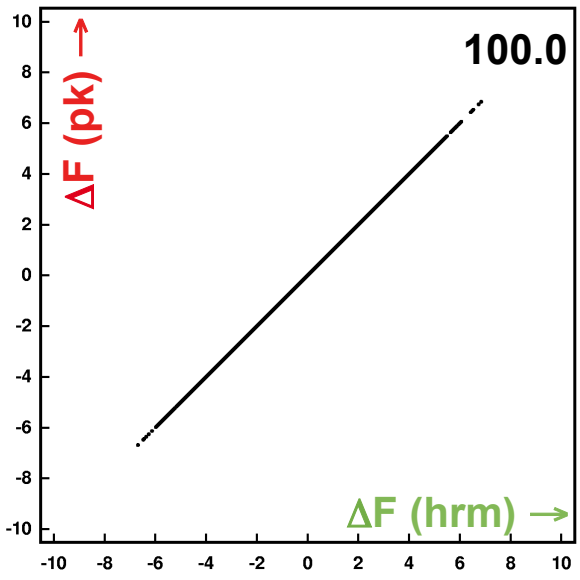
Anomalous signal-to-noise-ratio

- Plot $\langle \Delta F \rangle / \langle \sigma(\Delta F) \rangle$ as a function of resolution
- It is better to use scaled but unmerged data to do this (in order to not depend on the accuracy of the error estimates from data processing ...)

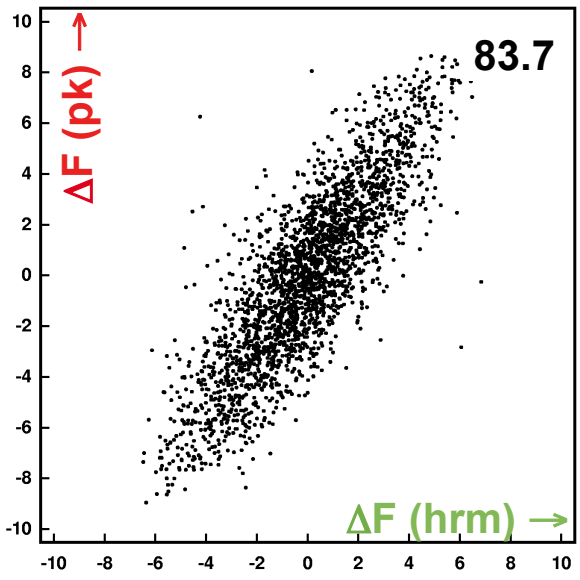


Plot is for unpublished MAD data (Schneider et al.) on Rabex-5 * Ub complex

CORR(ΔF,ΔF)

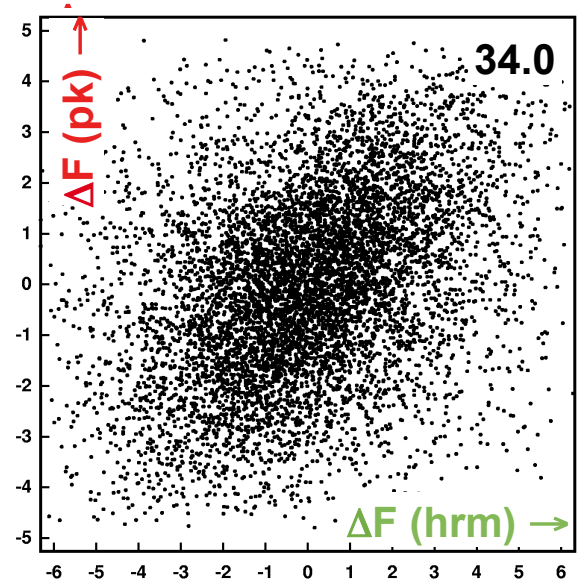


ideal case

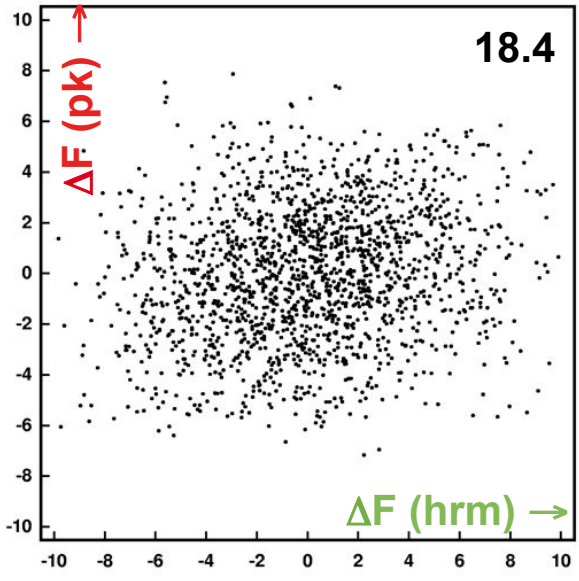


good real case

CORR(ΔF,ΔF)



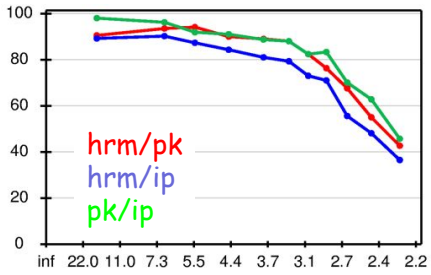
acceptable real case



unacceptable real case

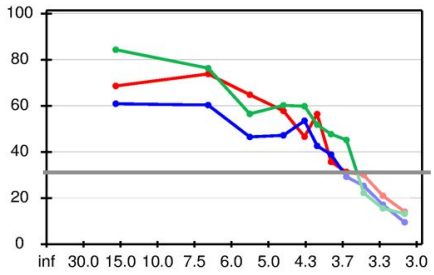
Apical Domain: $P2_12_12$
1x3Se/224 res = 1/74
SC = 42%

Martin Walsh et al. (1999)
Acta Cryst. D55:1168-1173



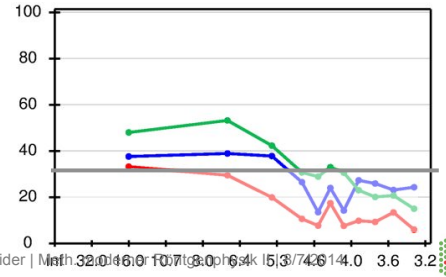
RRF: $P4_32_12$
1x3Se/185 res = 1/63
SC = 65%

Maria Selmer et al. (1999)
Science 286: 2349



PFH: $P2_1$
4x4Se/350 res
SC = 40%

not published ...



Statistically independent data ?

Table I. Crystallographic data

Crystal	λ(Å)	Resolution (Å)	Multiplicity[c]	Completeness (%)[d]	f'/f''[e]
PEAK1	0.9790	3.0	13.8 (6.9)	98.8	–5/7
PEAK2	0.9790	3.2	13.2 (6.6)	97.8	–5/7
INFL1	0.9792	3.0	13.8 (6.9)	98.8	–9/3
INFL2	0.9792	3.2	13.2 (6.6)	98.4	–9/3
HREM1	0.9393	3.0	13.8 (6.9)	98.8	–3/3
HREM2	0.9393	3.2	13.2 (6.6)	97.6	–3/3
NATI	0.9393	2.2	3.2	95.0	

is the mean intensity of the reflection with unique index h.

[d]Completeness for unique reflections; anomalous completeness is identical because inverse beam geometry was used.
[e]f'/f'' ratio, as determined from a fluorescence scan of the crystal.
Correlation coefficients of anomalous differences at different wavelengths for MAD experiment 1: PEAK1 versus INFL1, 0.54; PEAK1 versus HREM1, 0.46; INFL1 versus HREM1, 0.39.

Cordell et al. EMBO J. (2001). 20:2454