ESRF Workshop

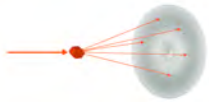
Has my experiment worked? MAD/SAD

Thomas R. Schneider

thomas.schneider@embl-hamburg.de

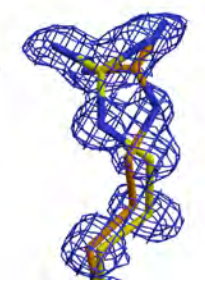
03/02/2009

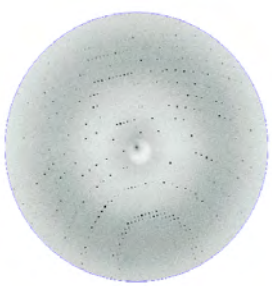
The Phase problem



- to calculate an electron density, we need structure factor amplitudes and 'phases' for every reflection:

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



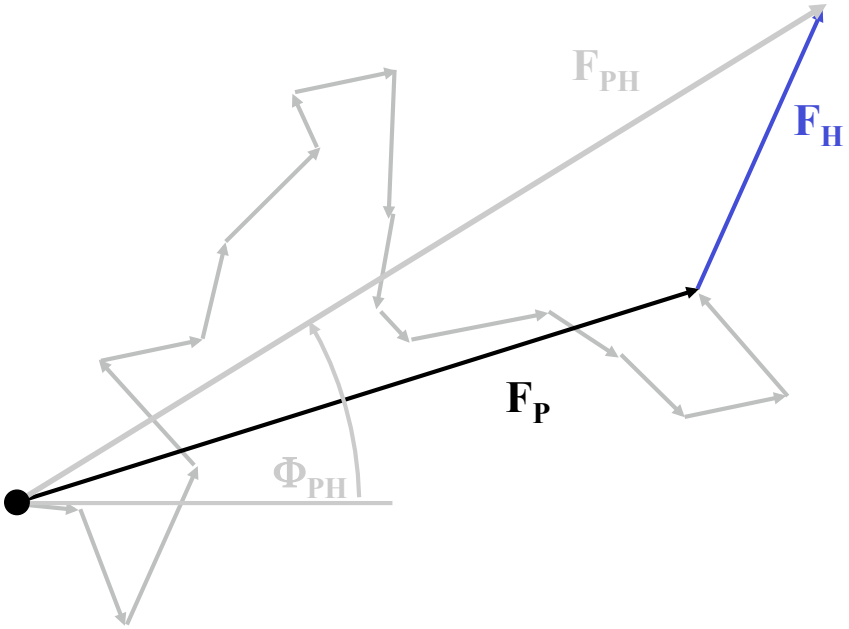


usually, we use phases calculated from the model ...

Connects hkl and xyz

$N_{XYZ} = 10^6$
$N_{HKL} = 50k$
FFT !

A protein and an extra scatterer H



Normal Scattering

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

Anomalous Scattering

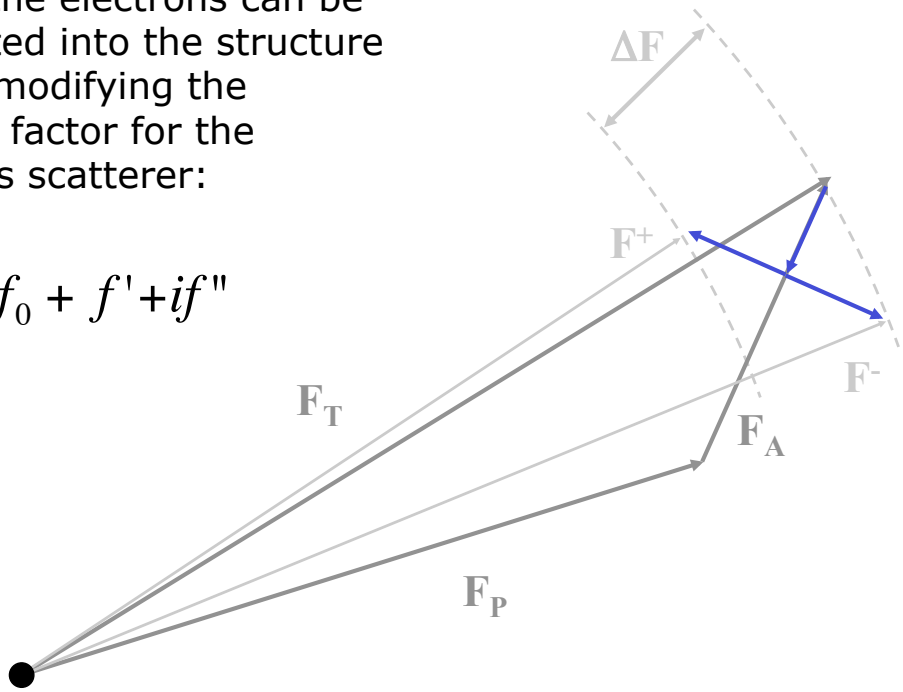
Friedel's law is broken $F_{hkl} \neq F_{\overline{hkl}}$

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

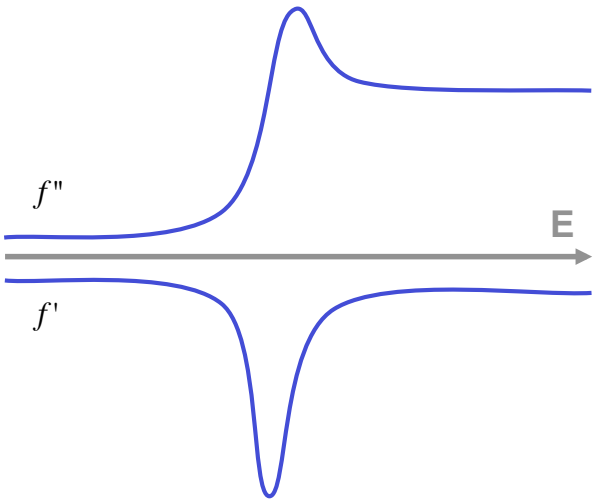
A protein and an anomalous scatterer A

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

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



Anomalous Dispersion



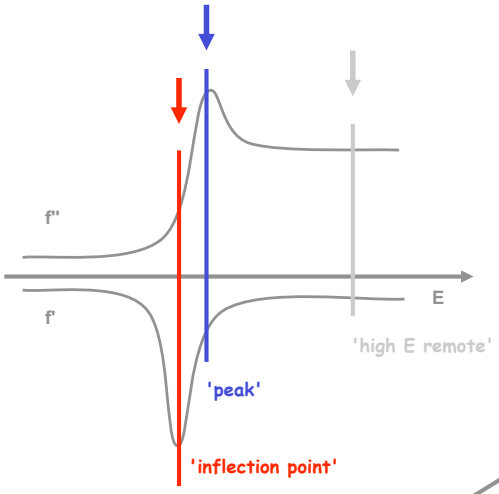
f"(E) can be measured by X-ray absorption spectroscopy:

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

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

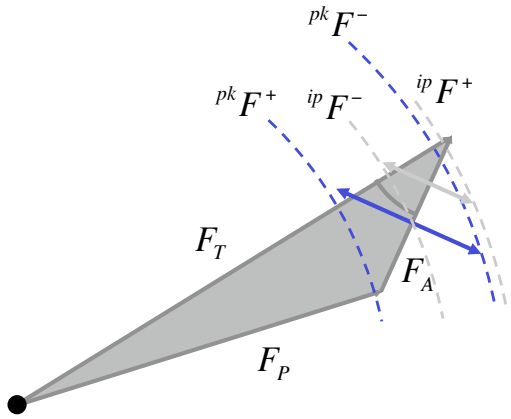


for one pair of refl:

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

from the same crystal

The MAD Observational Equations



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

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

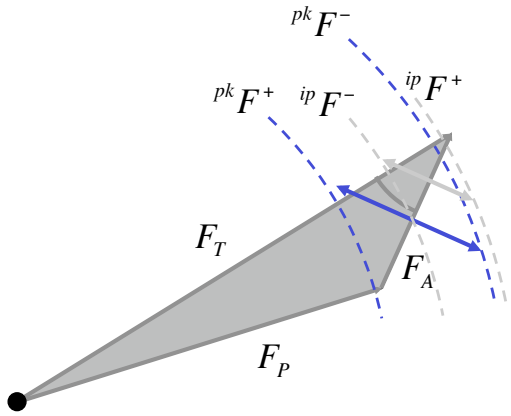
$$\left|{}^{pk}F^+_{obs}\right|^2 = \left|F_T\right|^2 + a(pk)\left|F_A\right|^2 + b(pk)\left|F_T\right|\left|F_A\right|\cos\alpha + c(pk)\left|F_T\right|\left|F_A\right|\sin\alpha$$

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

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

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

The MAD Observational Equations



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

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

$$^{pk}F^+ \quad ^{pk}F^- \quad ^{ip}F^- \quad ^{ip}F^+ \rightarrow |F_A| \quad |F_T| \quad \alpha$$

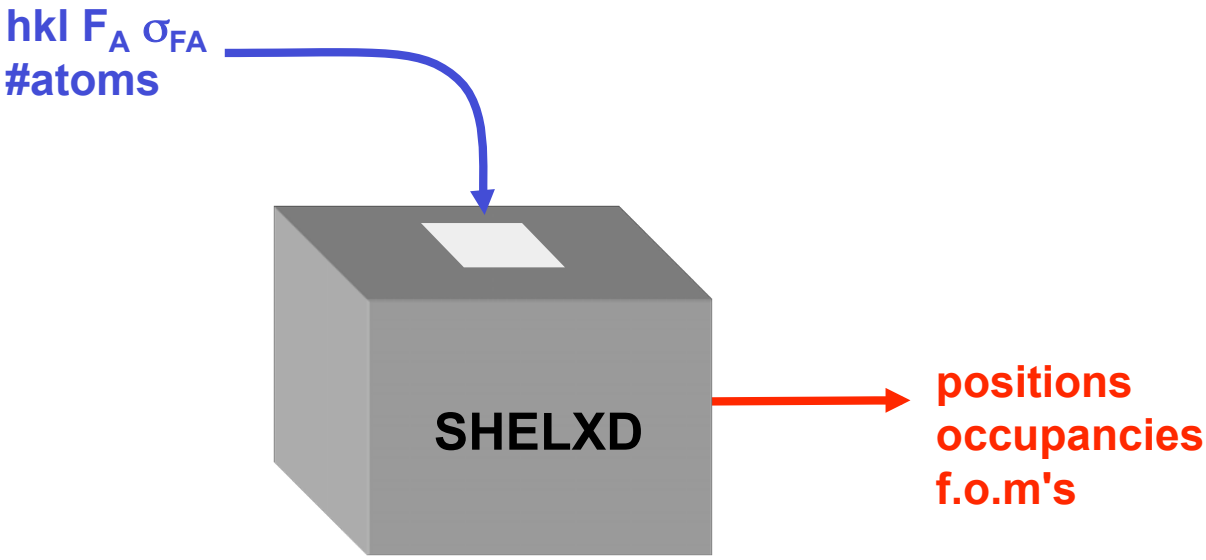
The meaning of $F_{H/A}$-values

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.

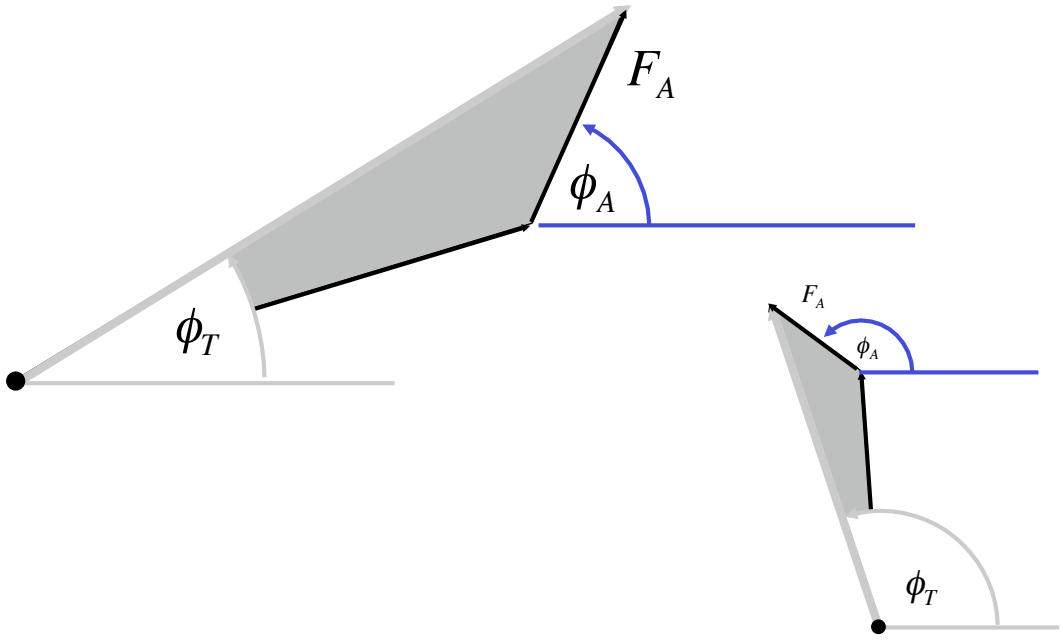
Ab initio methods can be used for finding the anomalous / heavy scatterers.

ΔF-values can be considered as lower limit estimates of F_A-values.

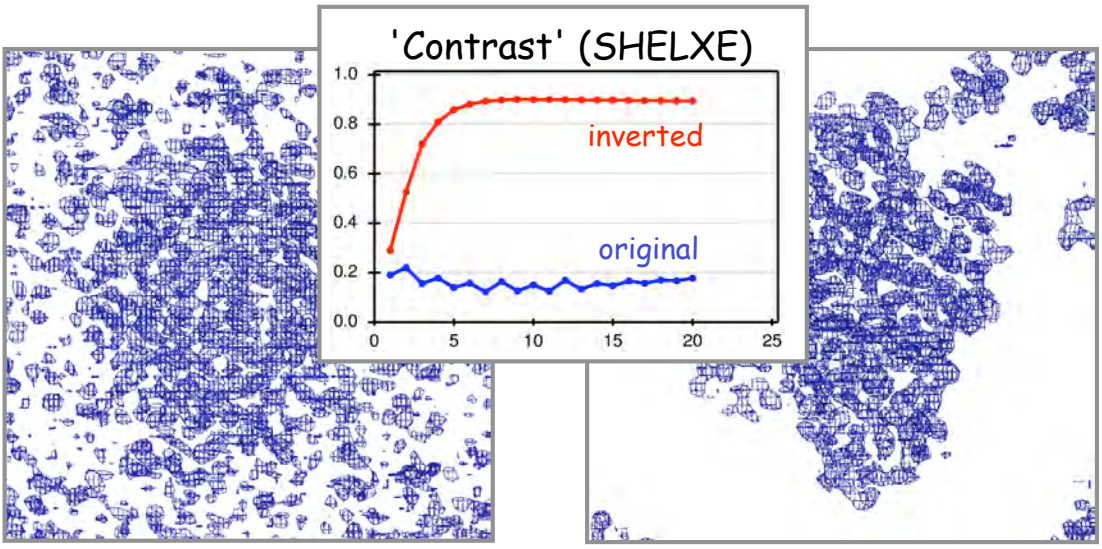
Solving the substructure against $|F_A|$



Knowing the substructure allows to calculate ϕ_T and ϕ_A



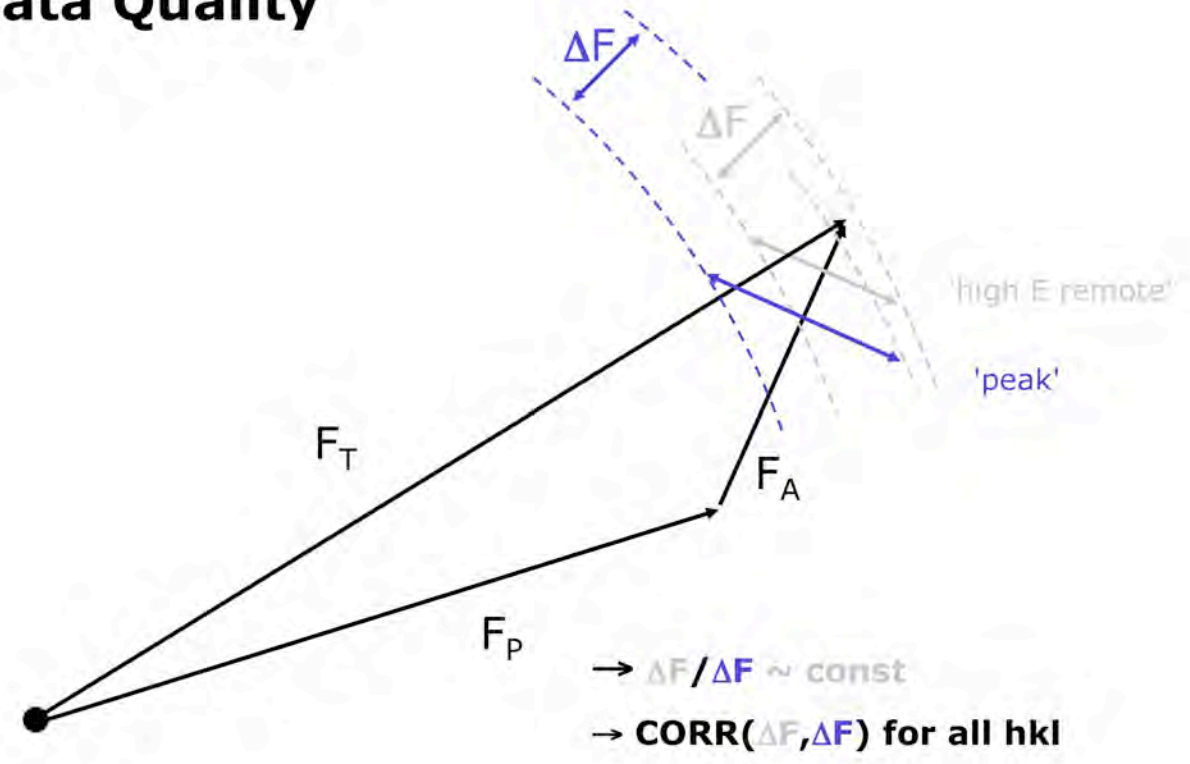
Choosing the 'right hand'



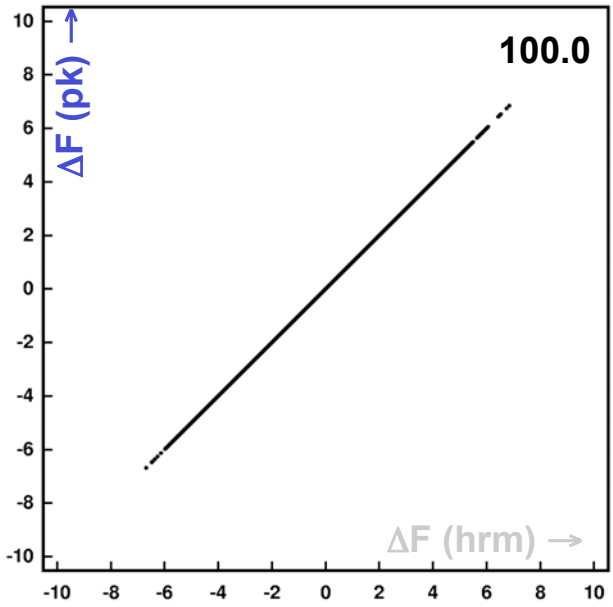
(x, y, z) $(-x, -y, -z)$

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

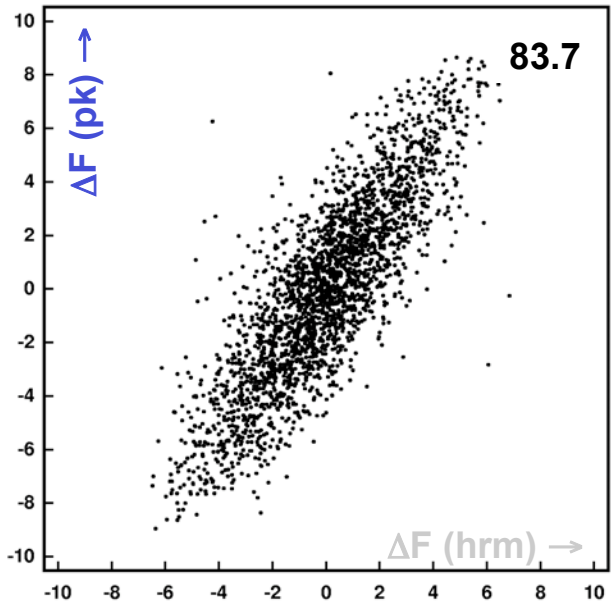
Data Quality



CORR(ΔF,ΔF)

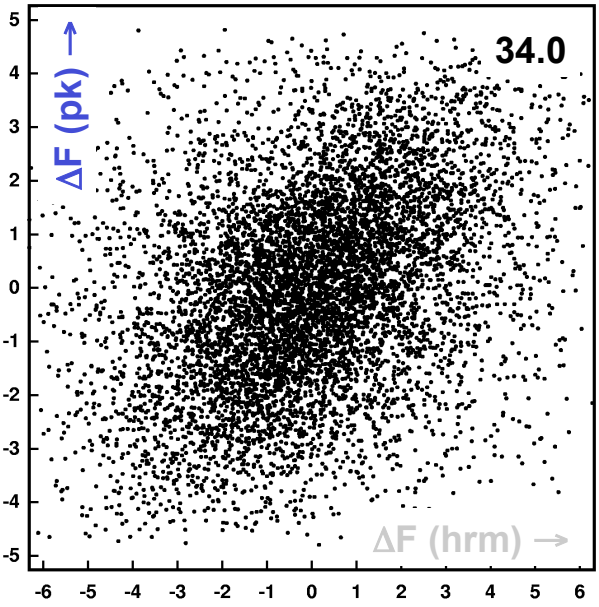


ideal case

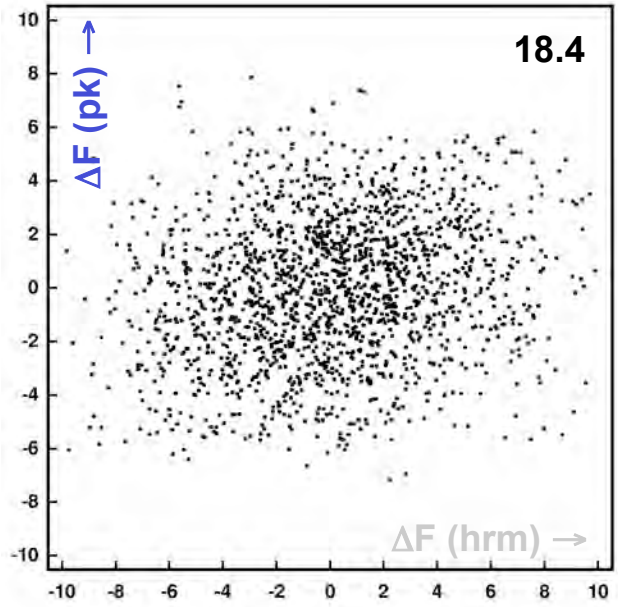


good real case

CORR(ΔF,ΔF)



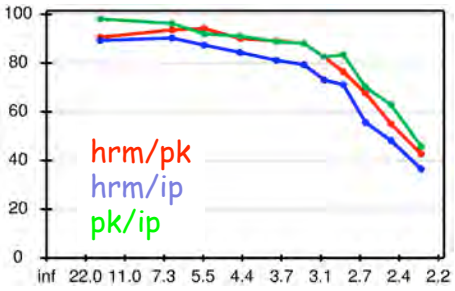
acceptable real case



unacceptable real case

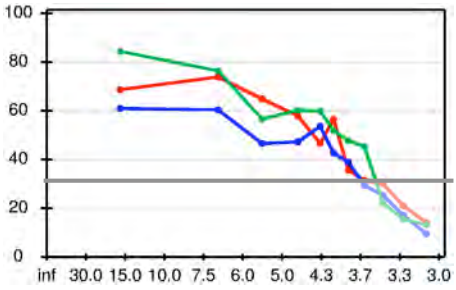
Apical Domain: $P2_12_12$
1x3Se/224 res = 1/74
SC = 42%

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



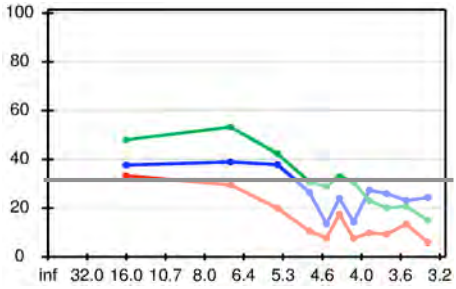
RRF: $P4_32_12$
1x3Se/185 res = 1/63
SC = 65%

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



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

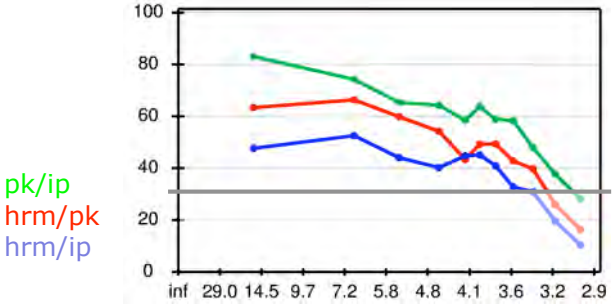
not published ...



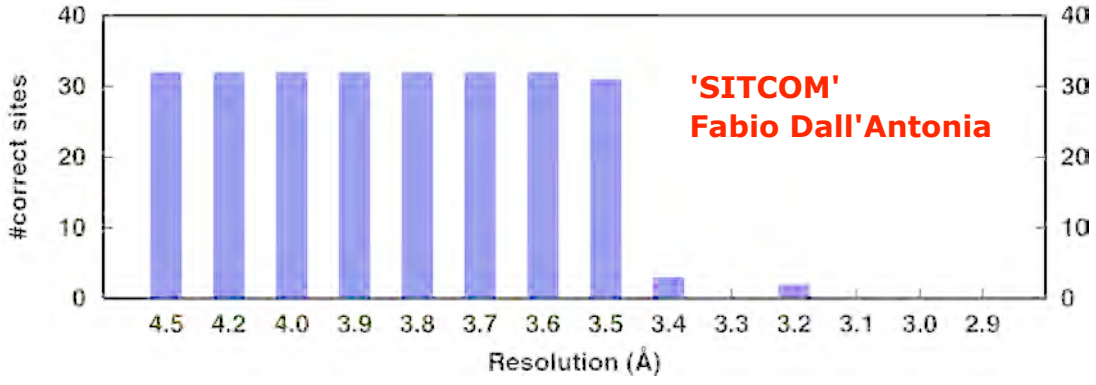
Useful / useless / dangerous data

X1: C2
4x8Se/1600 res = 1/50
SC = 65%

to be published



Number of correct sites for different resolution cutoffs:



Statistically independent data ?

Table I. Crystallographic data

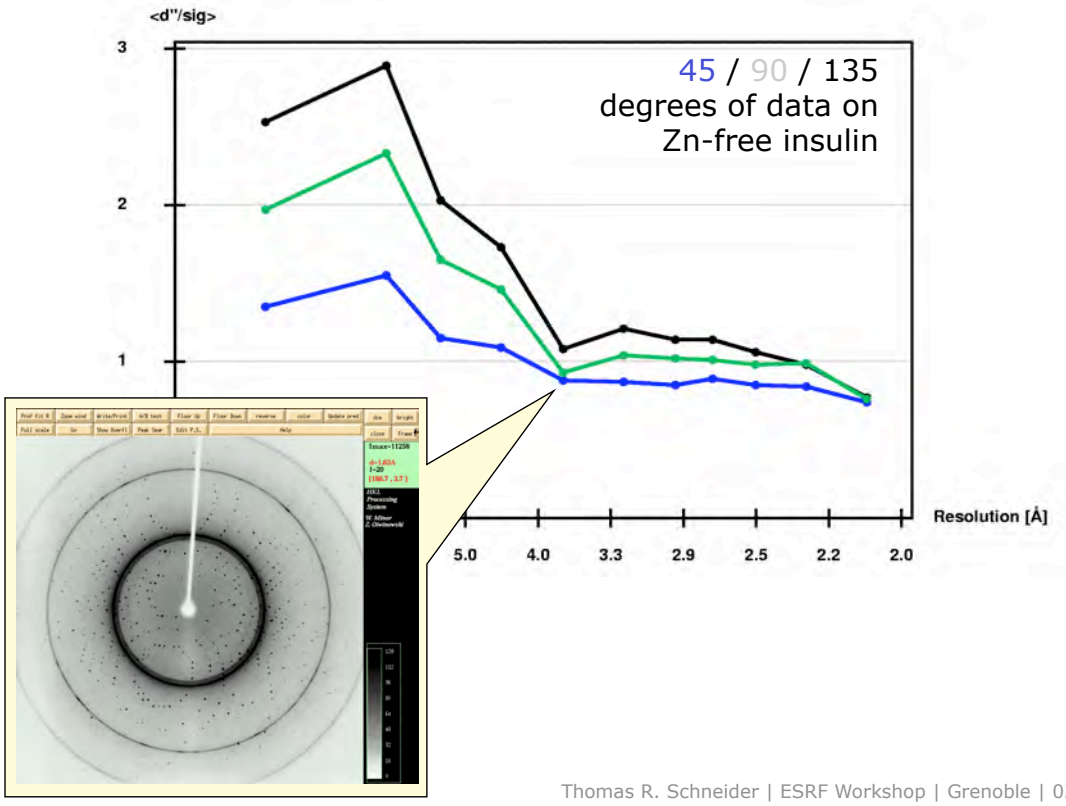
Crystal	λ(Å)	Resolution (Å)	Multiplicity[c]	Completeness (%)[d]	f'/f''[e]
PEAK1	0.9790	3.0	13.8 (6.9)	98.8	–5/7
PEAK2	0.9790	3.2	13.2 (6.6)	97.8	–5/7
INFL1	0.9792	3.0	13.8 (6.9)	98.8	–9/3
INFL2	0.9792	3.2	13.2 (6.6)	98.4	–9/3
HREM1	0.9393	3.0	13.8 (6.9)	98.8	–3/3
HREM2	0.9393	3.2	13.2 (6.6)	97.6	–3/3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

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

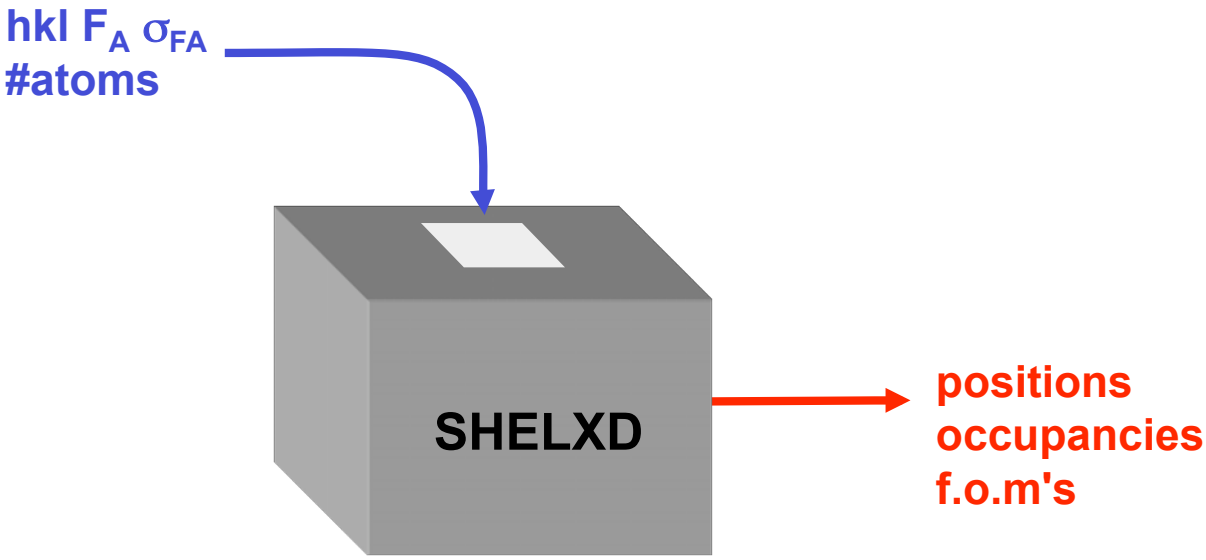
[d]Completeness for unique reflections; anomalous completeness is identical because inverse beam geometry was used.
[e]f'/f'' ratio, as determined from a fluorescence scan of the crystal.
Correlation coefficients of anomalous differences at different wavelengths for MAD experiment 1: PEAK1 versus INFL1,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

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

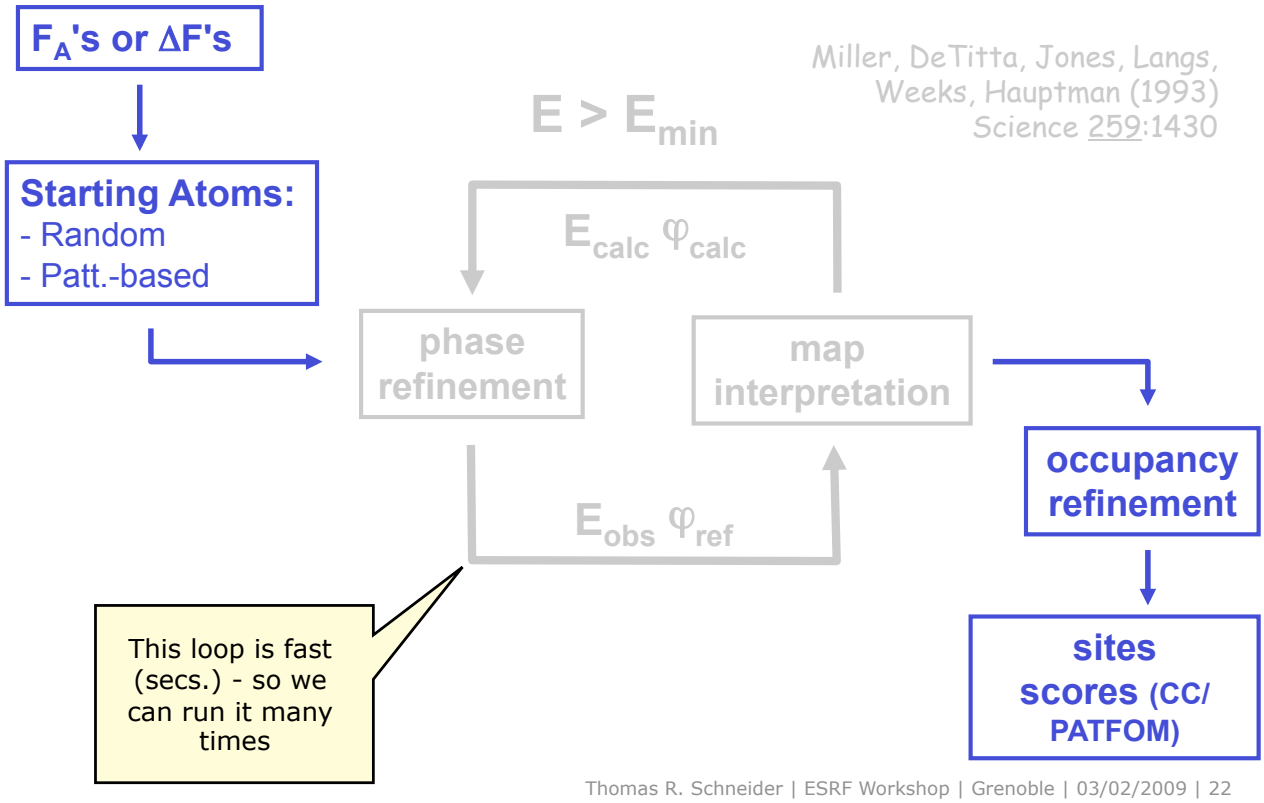
Signal-to-noise and redundancy



Solving the substructure against $|F_A|$



SHELXD: Dual Space Recycling with Patterson Seeding

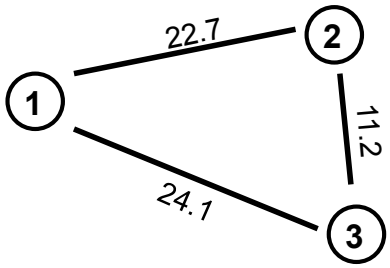


Substructure Validation

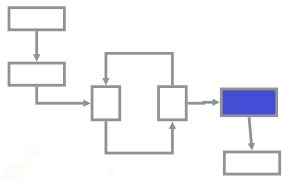
- CC(E_o,E_c) M. Fujinaga & R. Read J.Appl.Cryst. 20:517 (1987)
- PATFOM - Crossword table

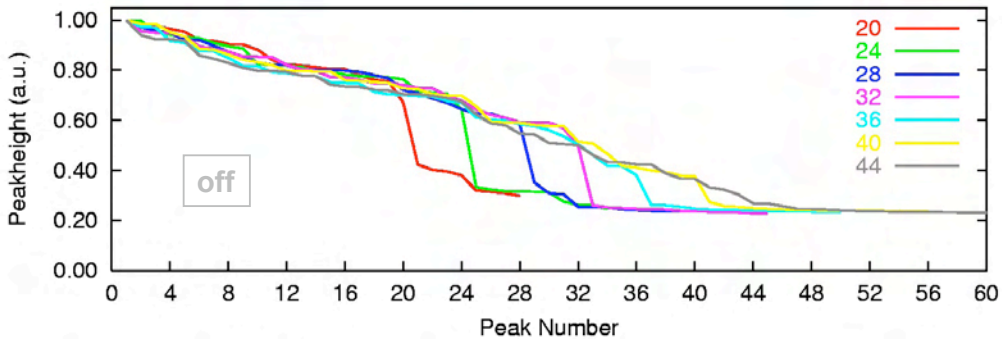
Se2			
22.7	Se1		
6.6			
11.2	24.1	Se3	
2.4	4.1		
18.6	16.8	11.8	Se4
1.5	0.1	0.9	

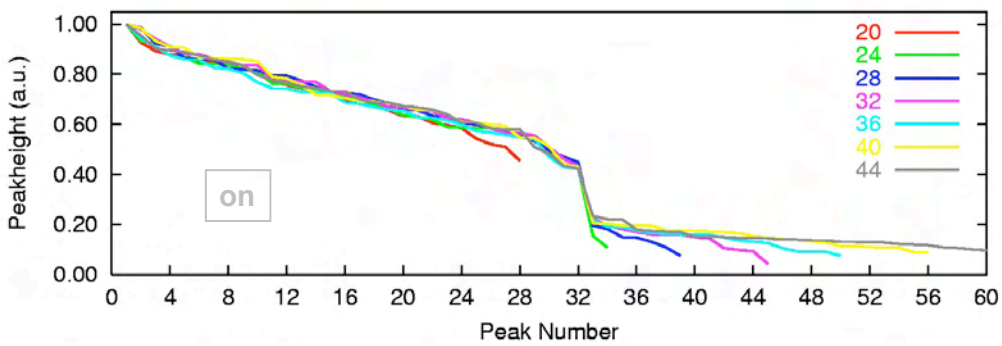
X



Occupancy Refinement





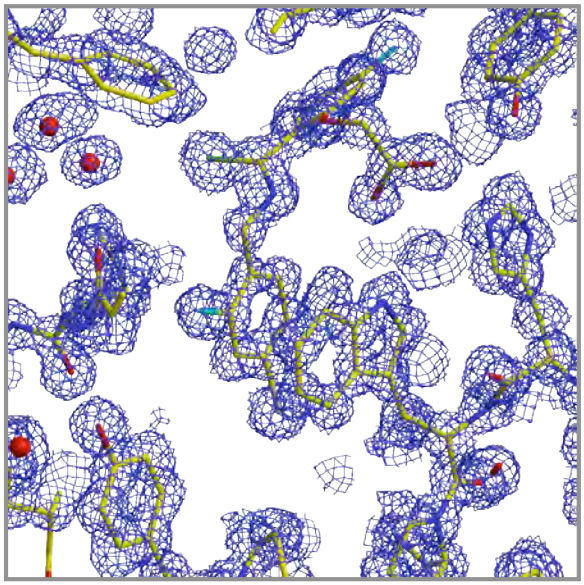


And now ?

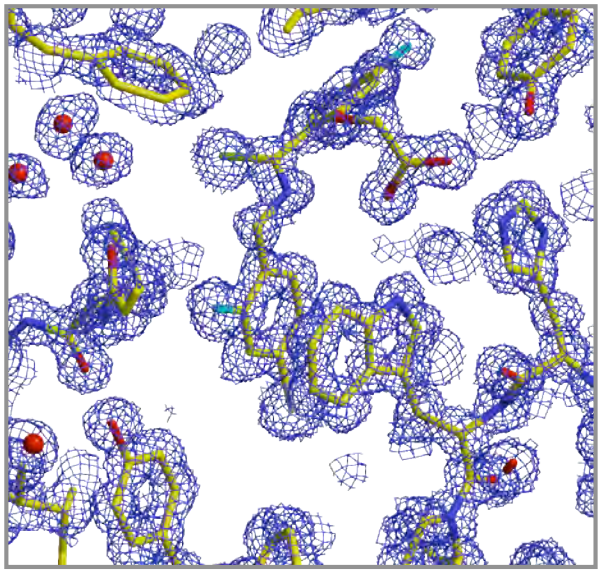
- Use site positions and anomalous/dispersive differences to derive phases and 'figures of merit'
- Use the phases to calculate an electron density map:

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

Aldose Reductase – MAD phases at 0.93 Å



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

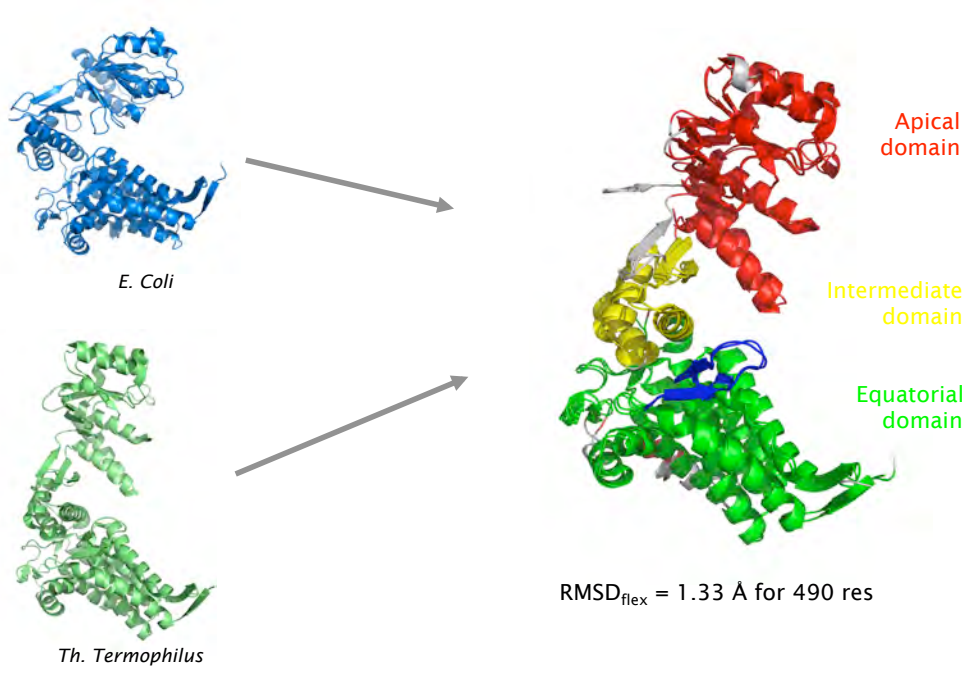


refined (remote)
σ_A weighted ($2F_{obs}$-$1F_{calc}$,Φ_{calc})-map
0.9 Å, contoured at 1σ

Commercials

RAPIDO –a web-server for 3D alignment

http://webapps.embl-hamburg.de/rapido



$RMSD_{flex}$ = 1.33 Å for 490 res

Mosca & Schneider (2008) Nucleic Acids Res 36:42
Mosca, Brannetti, Schneider (2008) BMC Bioinf. 9:352

Acknowledgements

- George Sheldrick
- Donators of data
- The TC of phasing
- EU for funding

Has my experiment worked?

- **(Multi-λ) anomalous data → sites → phases → map → model**
- Necessary conditions (*cum grano salis!*):
  - Anomalous signal can be measured from the the crystal
  - The original data are complete (>90%, better > 95%)
  - The original data are sufficiently strong (I/sigI(hires) > 5, better >10)
  - The orginal data have no/few 'problems' (ice-rings, beam stops etc.)
  - The anomalous differences have been measured with sufficient accuracy (CC_{anom} > 30% @ 4A) and acceptable signal-to-noise ratio (d"/sig > 1). N.B.: Use scaled but unmerged data for statistics
  - The substructure can be solved (CC > 39/16, PATFOM, occ-drop, (NCS))
  - The 'correct hand' can be identified (contrast)
  - The combination of the substructure and the original data allows to calculate an interpretable electron density map (helices, sheets)
  - The anomalous sites are in sensible places (not in vacuum, inside the protein)
- Sufficient condition:
  - The map is traceable
  - ... and the initial model can be refined

http://shelx.uni-ac.gwdg.de/SHELX
http://webapps.embl-hamburg.de/hkl2map

References

- Taylor G. (2003) The phase problem. Acta Cryst. D59:1881-90
  The grand map of phasing for the novice
- Hendrickson WA et al. (1985) Meth.Enz. 115:41-55
  Comprehensive overview, algebraic approach
- Scharff AJ et al. (2000) Acta Cryst. D 56:785-788
  Modification of f"(λ) as result of radiation damage
- Smith JL (1991) Structure 1:1002-1011
  Short overview, estimation of the expected signal
- Schneider TR & Sheldrick GM (2002) Acta Cryst. D58:1772-9
  Description of SHELXD
- Grosse-Kunstleve RW & Schneider TR (2006) Meth. Mol. Biol. 364:197-217.
  Text on substructure determination
- Podjarny A, Schneider TR, et al. (2003) Meth Enzymol 374:321-41
  MAD phases for Aldose Reductase at atomic resolution