Experimental phasing

Charles Ballard

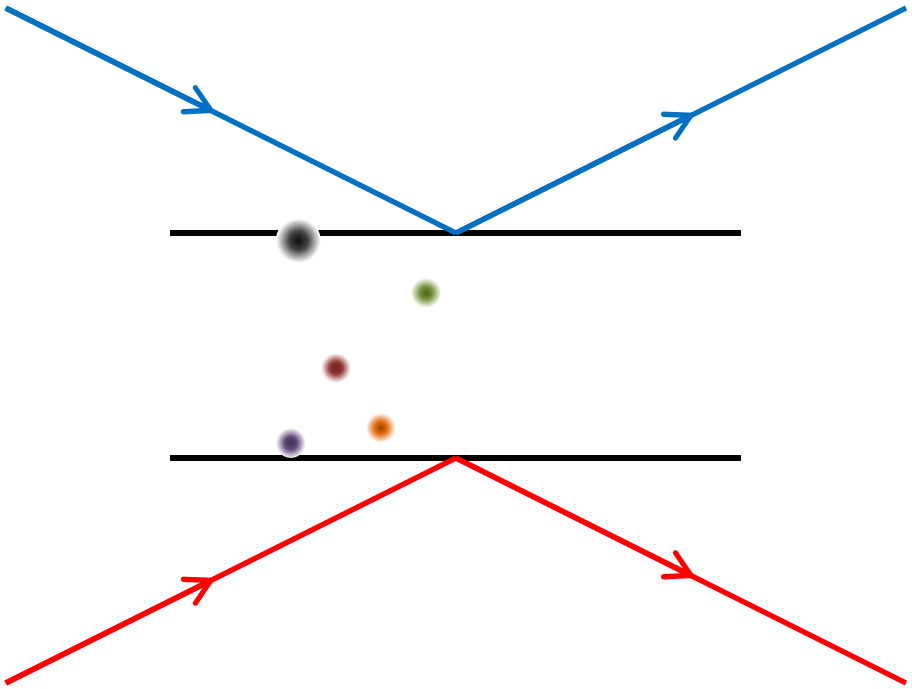
(original Gábor Bunkóczi)

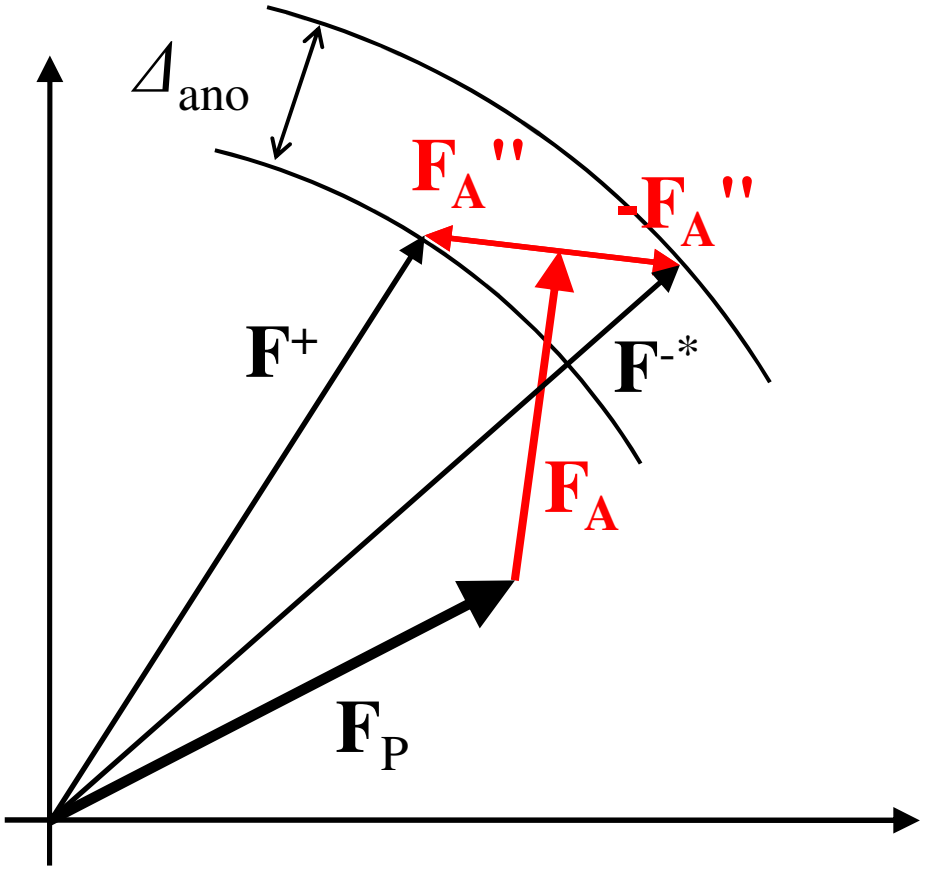
CCP4 Workshop

7 December 2011

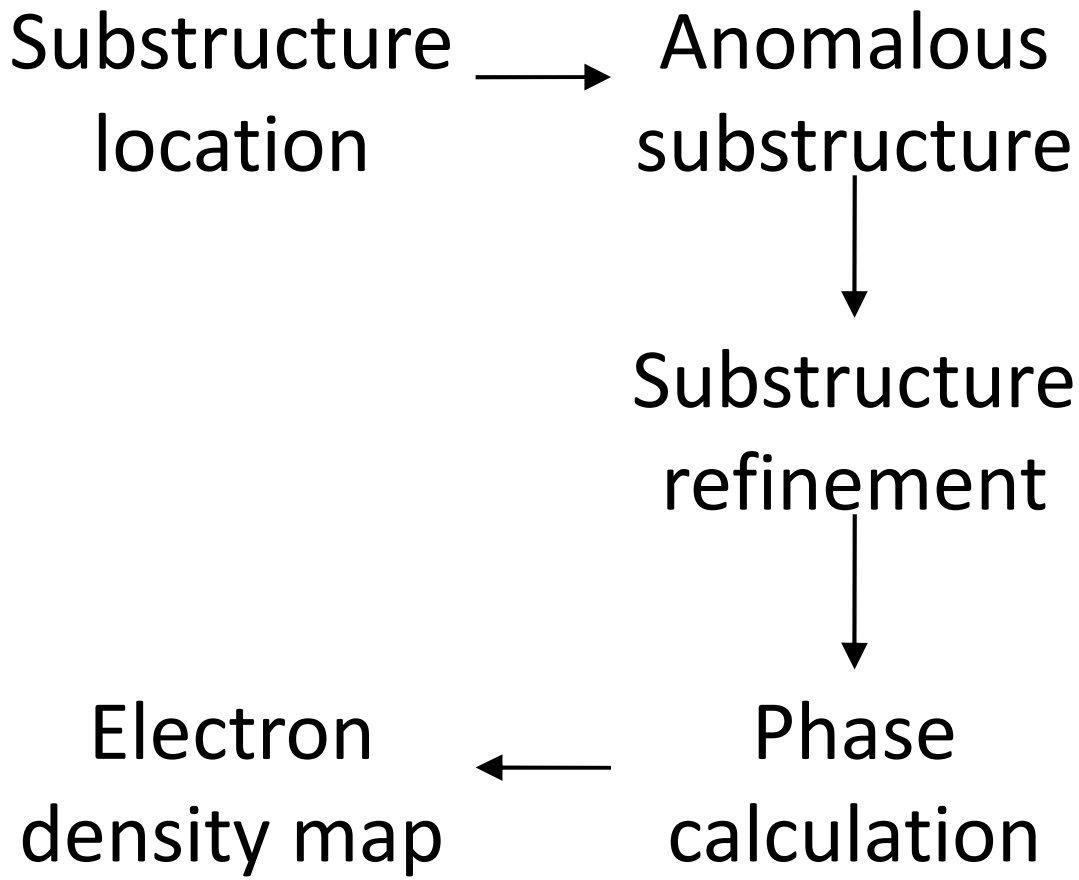
Anomalous diffraction

$\mathbf{F}_P$ – protein $\mathbf{F}_A$ – anomalous substructure





Phasing


Substructure location
Anomalous substructure
Substructure refinement
Phase calculation
Electron density map

Phase calculation in SAD

F_o^+, F_o^- – measurements

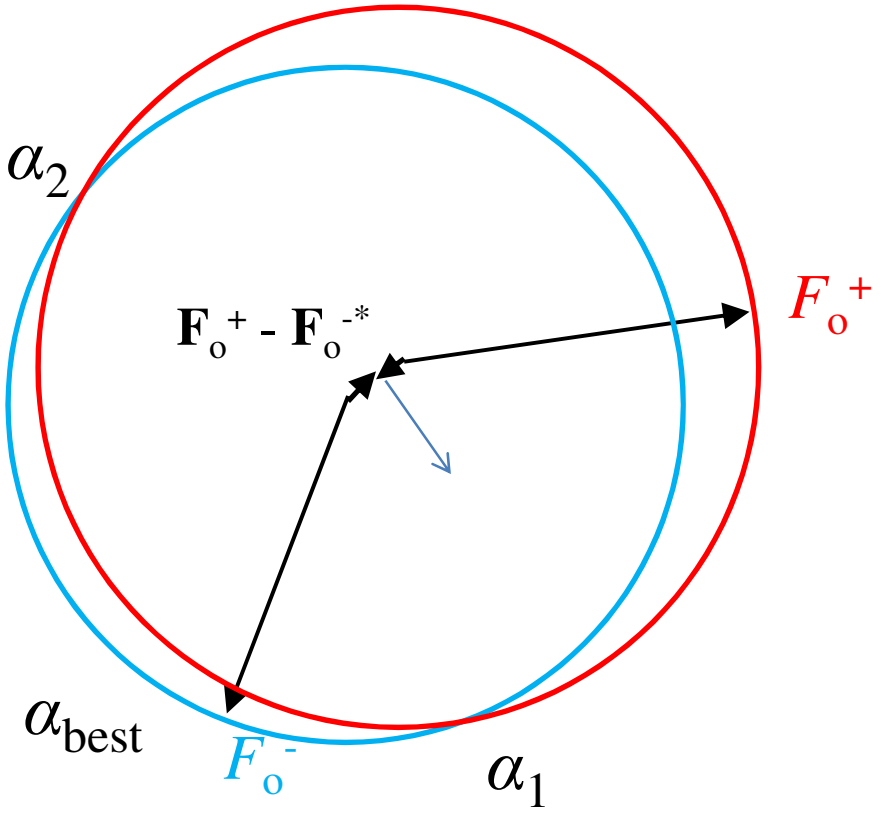
$\mathbf{F}_o^+ - \mathbf{F}_o^{-*}$ – calculated from substructure ($-2\mathbf{F}_A''$)

$\alpha_1, \alpha_2, \boldsymbol{\alpha}_{\mathbf{best}}$ – phase angles

Approximate map given by

$$mF_{best}e^{i\alpha_{best}}$$

With $F_{best} = \frac{|F^+| + |F^-|}{2.0}$, $m = \cos\frac{1}{2}(\alpha_1 - \alpha_2)$



SAD likelihood function

Probability of experimental observations. Observations correlated, multivariate distribution.

$$P(\mathrm{F}_O^+, \mathrm{F}_O^{-*}, \mathrm{H}^+, \mathrm{H}^{-*}) \rightarrow P(\mathrm{F}_O^+, \mathrm{F}_O^{-*}; \mathrm{H}^+, \mathrm{H}^{-*})$$
$$\rightarrow P(F_O^+, \alpha^+, F_O^-, \alpha^-; \mathrm{H}^+, \mathrm{H}^{-*})$$

Factor joint probability into two parts

$$P(\mathrm{F}_O^+, \mathrm{F}_O^-; \mathrm{H}^+, \mathrm{H}^{-*}) = P(\mathrm{F}_O^+; \mathrm{F}_O^-, \mathrm{H}^+, \mathrm{H}^{-*}) P(\mathrm{F}_O^-, \mathrm{H}^{-*})$$

Marginalise phases α^+ and α^-

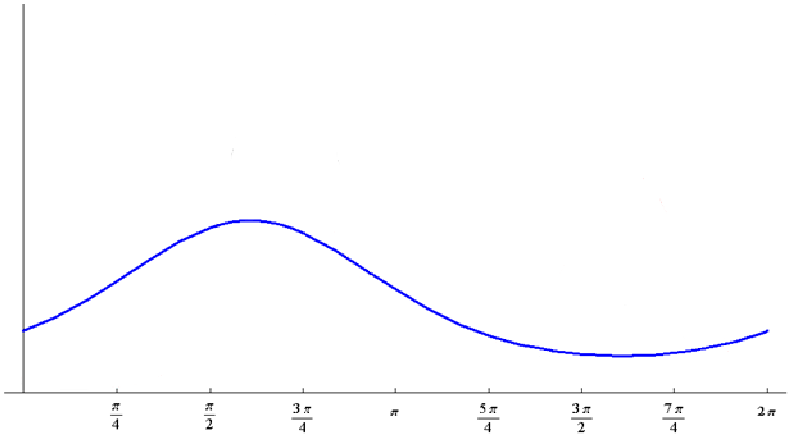
SAD function – normal scattering

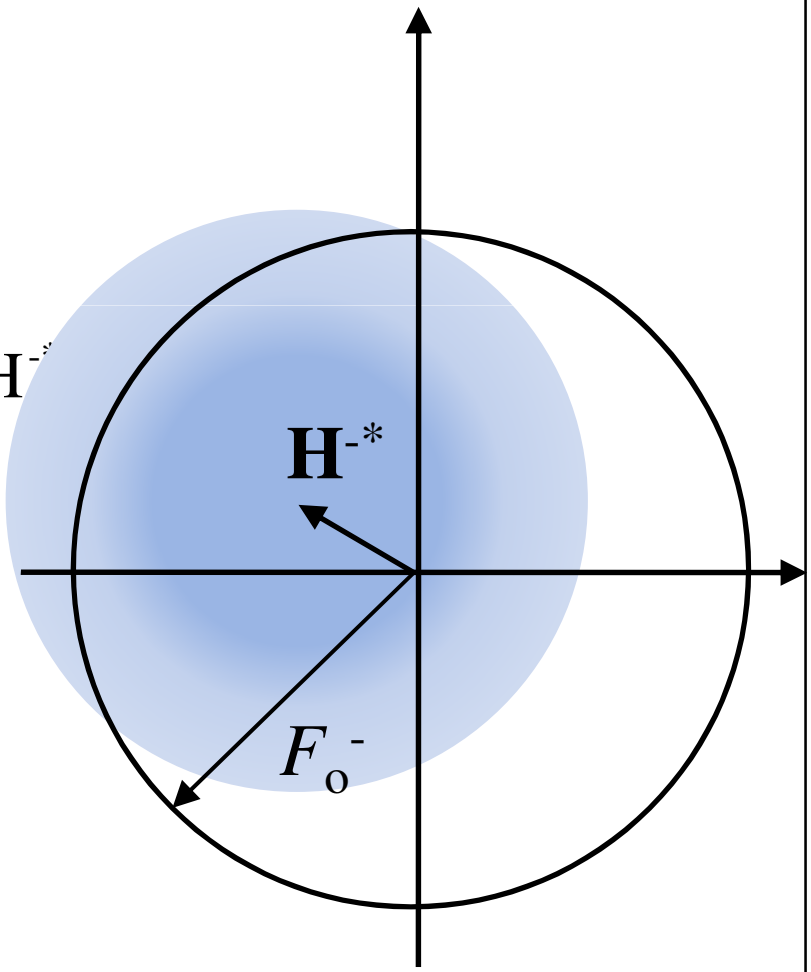
$$P\left(F_o^-, \alpha^-; \mathbf{H}^-\right) =$$

$$\frac{F_o^-}{\pi \Sigma^-} \exp\left(-\frac{\left| F_o^- e^{i\alpha^-} - D\mathbf{H}^{-*} \right|^2}{\Sigma^-} \right)$$

Σ^- accounts for errors between F_o^- and H^{-*}

Primarily normal scattering.





SAD function - anomalous

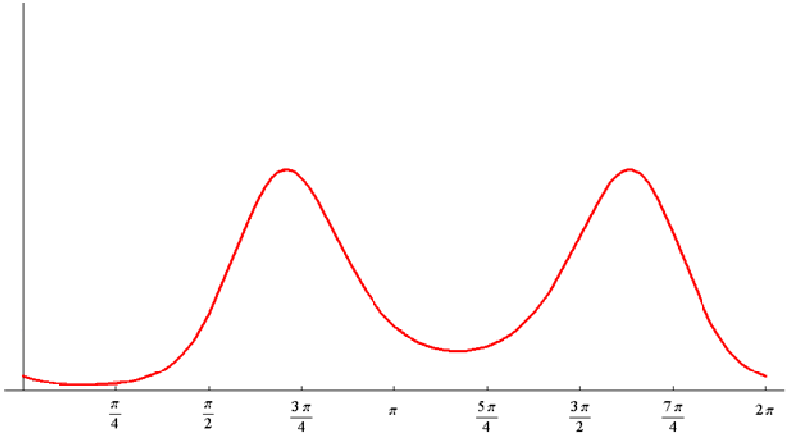
$$P\left(F_o^+;\mathbf{F}_O^-,\mathbf{H}^+,\mathbf{H}^{-*}\right)=$$

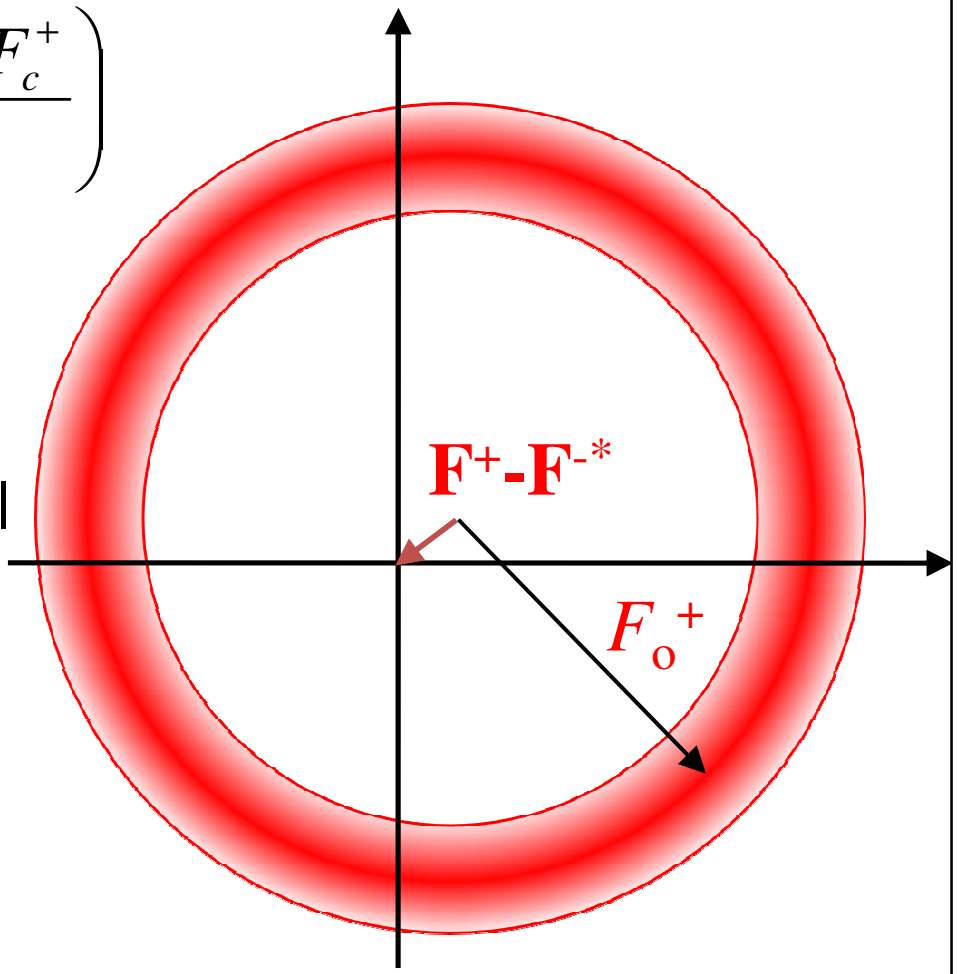
$$\frac{2F_o^+}{\Sigma^+}\exp\left(-\frac{\left(F_o^+-F_c^+\right)^2}{\Sigma^+}\right)\mathrm{eI}_0\left(\frac{2F_o^+F_c^+}{\Sigma^+}\right)$$

Σ^+ accounts for errors between

F_O^+ and F_C^+, where

$$F_C^+ = \left| H^+ + D_\Phi e^{i\alpha_\Phi}\left(F^- e^{i\alpha_-} - H^{-*}\right)\right|$$





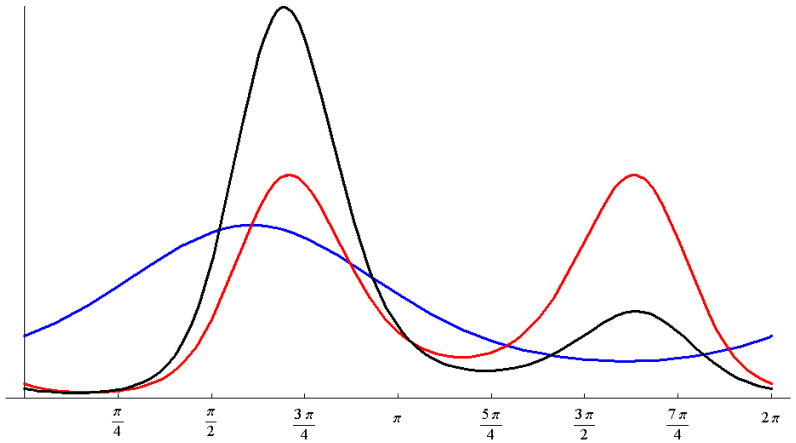
SAD function

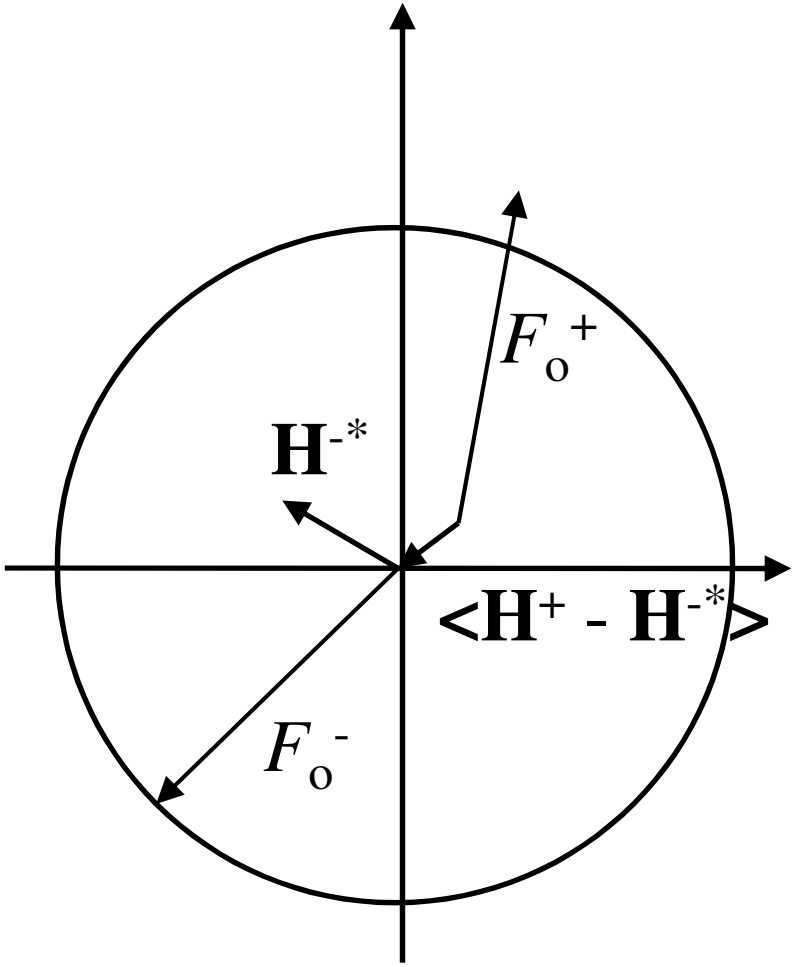
Total likelihood is integral product of the distributions under the black circle

Normal scattering

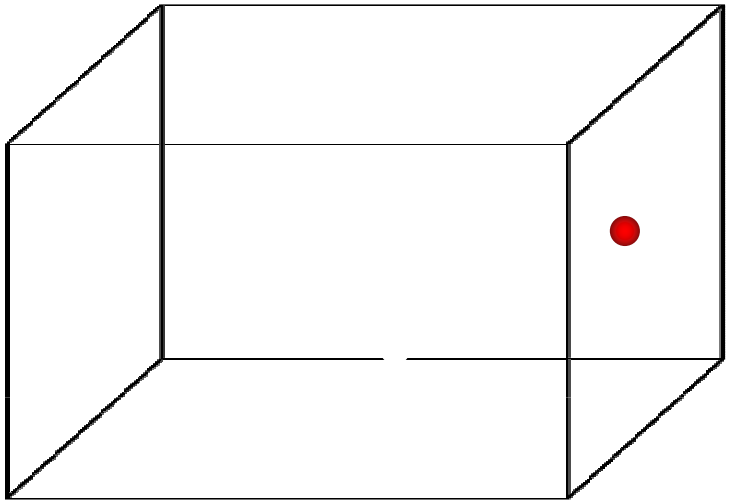
Anomalous scattering

Phase probability





Phaser phasing scenarios

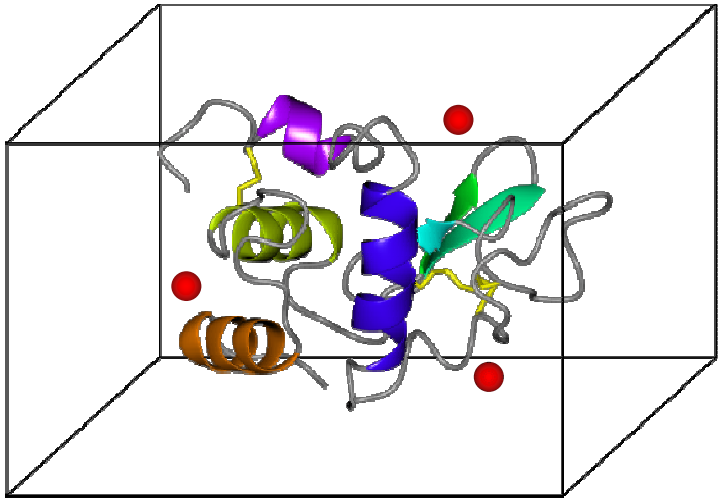


Known anomalous substructure

— missing atoms

— model errors (*xyz*, *occ*, *B*)

— weak normal component



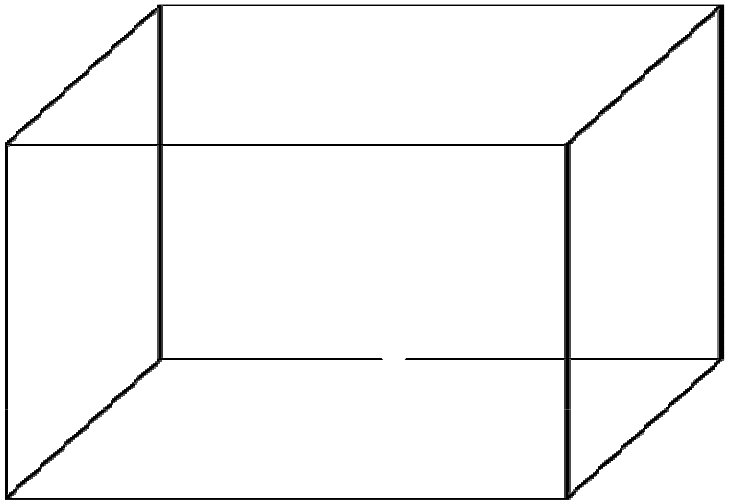
Known partial structure

— missing (anomalous) atoms

— (partial) model errors

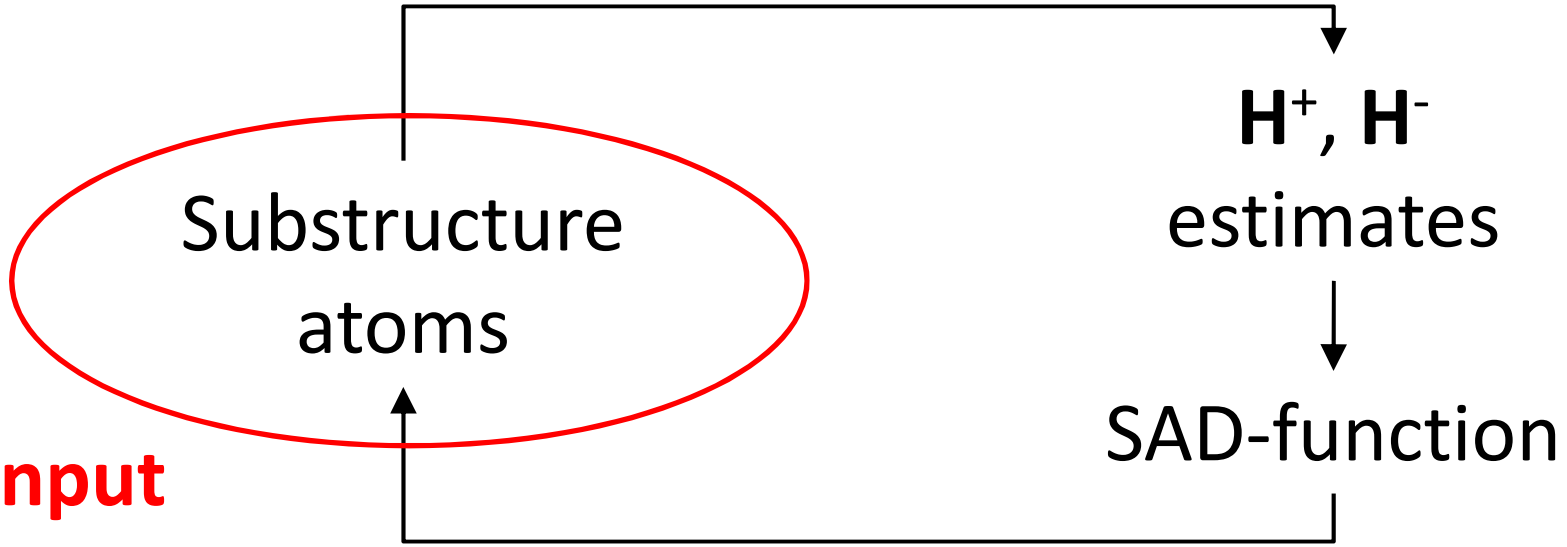
— weak anomalous component

SAD phasing

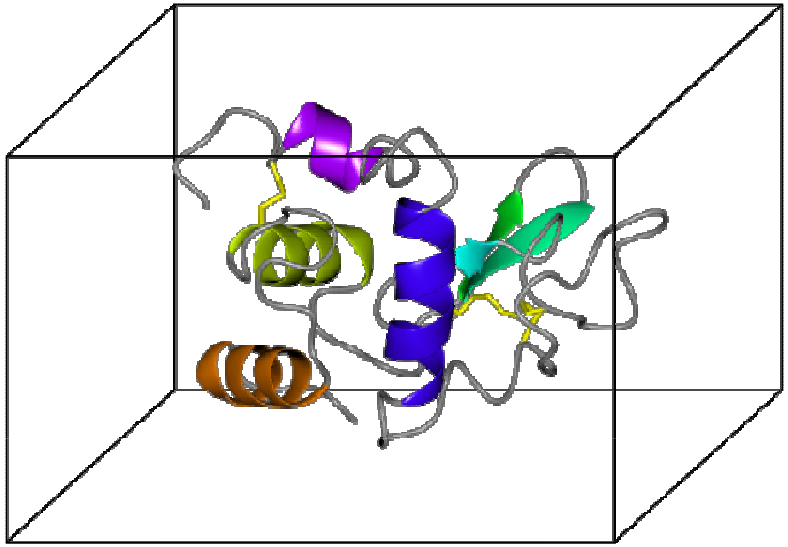


Parameters:

- atom type
- position
- occupancy
- B-factor
- f”

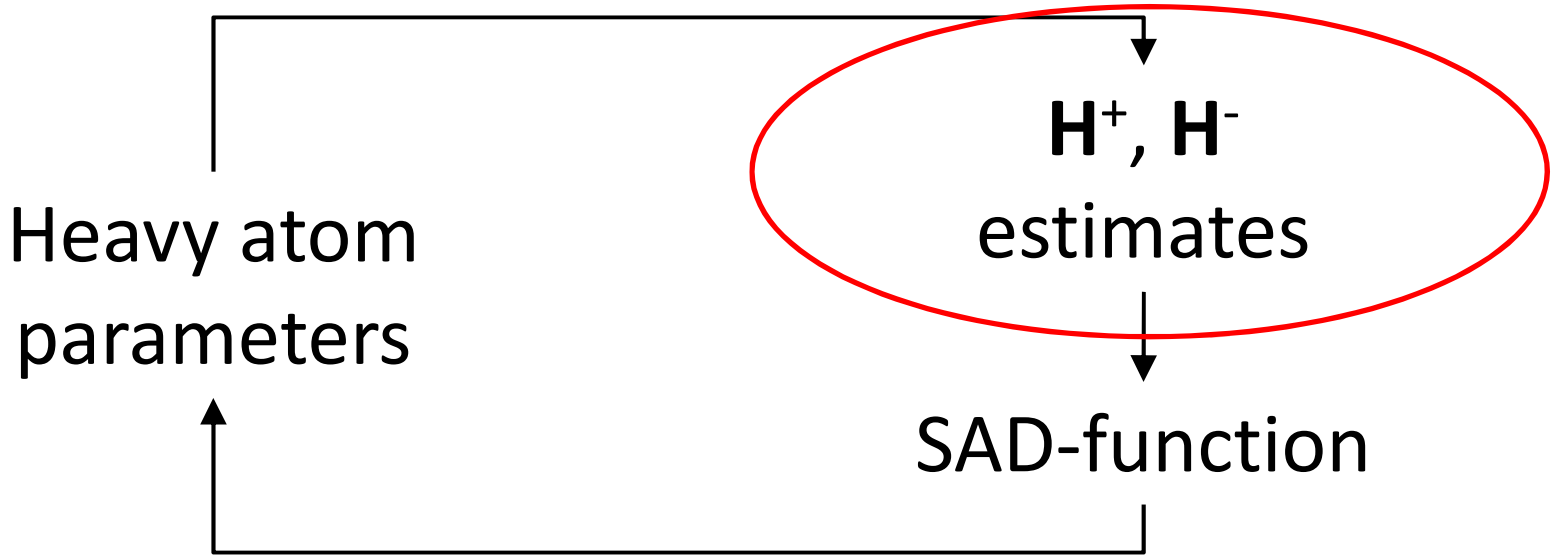


MR-SAD phasing

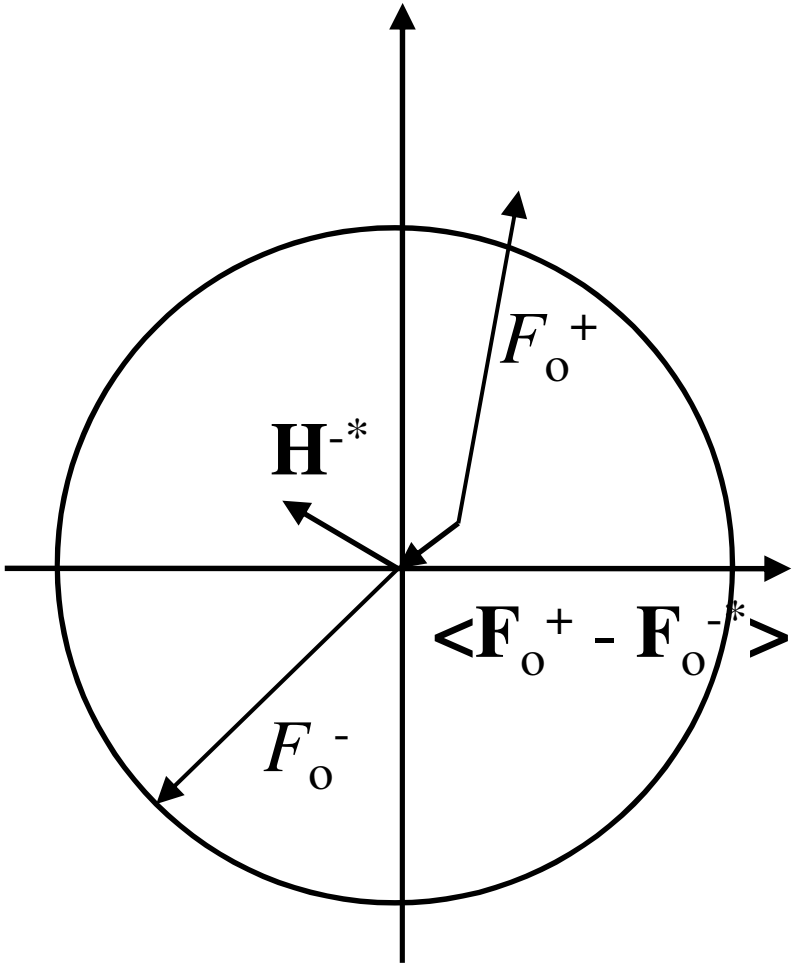


Input:

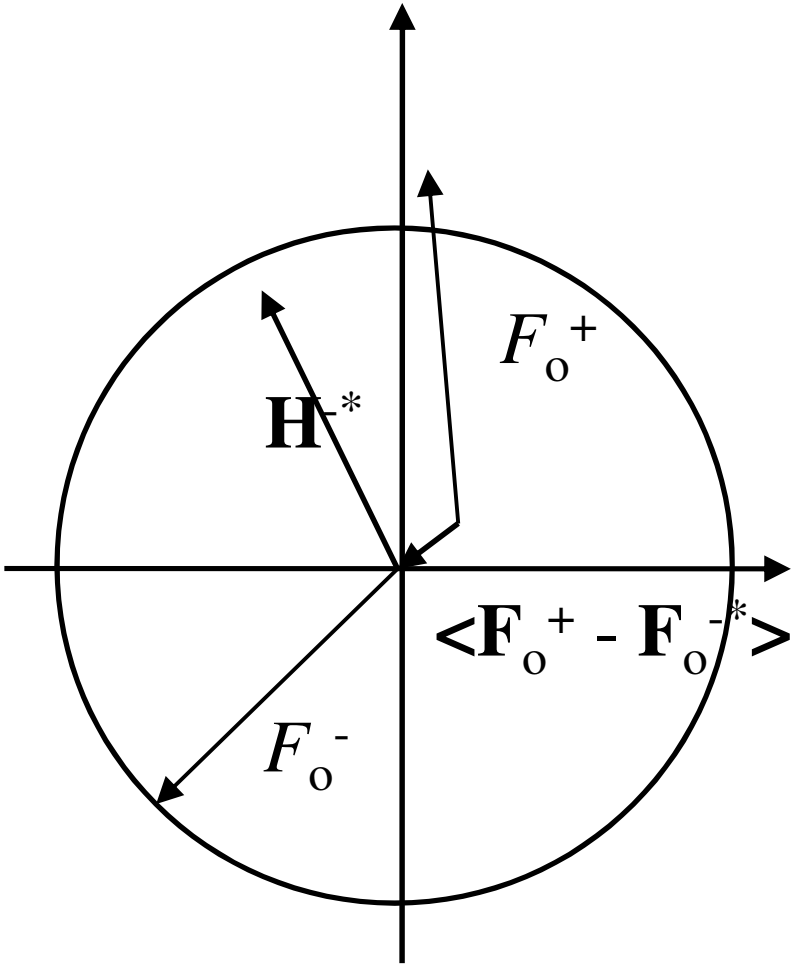
- partial structure
- possible atom types



SAD function

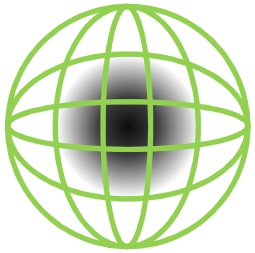


Scenario:
anomalous substructure



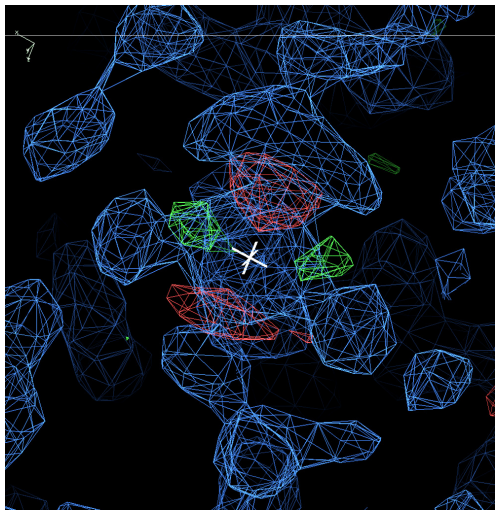
Scenario:
partial MR model

Log-likelihood gradient maps



Peak height dependent on f'/f" ratio

- distinguish atom type
- (fooled if normal scattering is present at site)



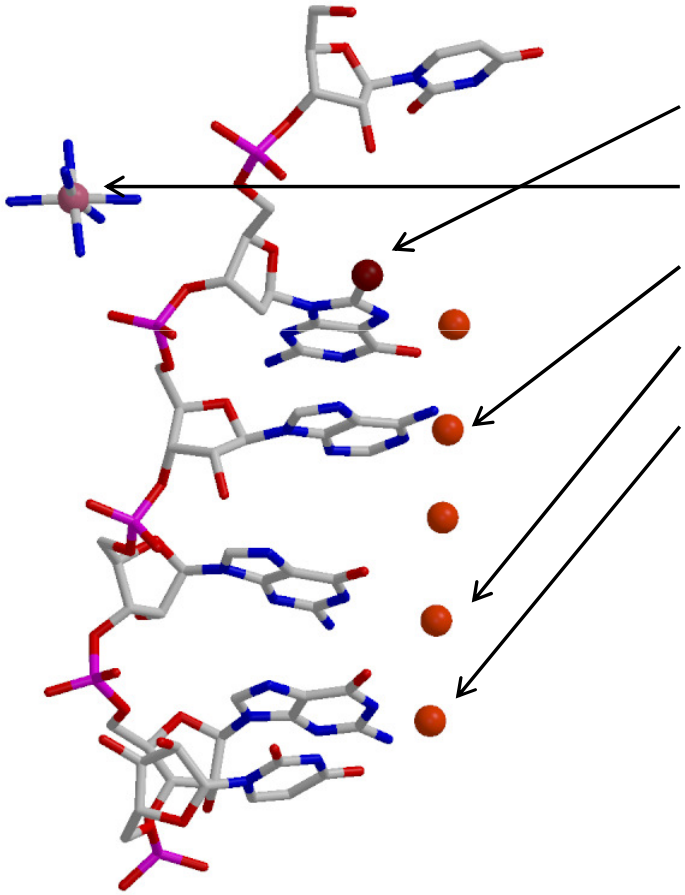
Very sensitive to detail

- e.g. anisotropic temperature motion

Iterative

- improvements in substructure description increases sensitivity

Log-likelihood gradient maps

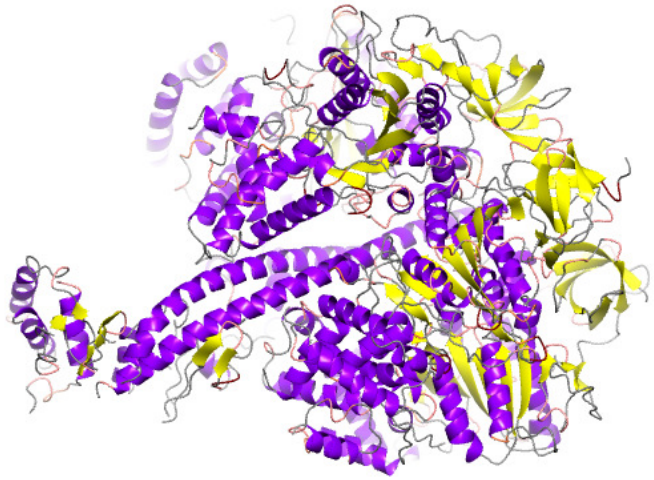


Δ_{ano}	LLG1	LLG2
26.5	28.5	-
5.0	5.0	18.5
-	-	7.6
-	-	-
-	-	-

Element	f’	f”
Na	0.05	0.04
P	0.14	0.16
K	0.25	0.41
Ca	0.28	0.50

F1-ATPase testcase

3500 residue protein complex
81 S atoms, 3.50 Å resolution
Wavelength: 1.77 Å (f_S"=0.72 e^-)
Anomalous signal extends to 21.5 Å, or to 12.7 Å very optimistically.



1bmf -
Abrahams J.P., Leslie A.G.W., Lutter R., Walker J.E.
Nature (1994) ***370****:621*
2ck3 - high resolution (1.95 Å) equivalent
Bowler M.D., Montgomery M.G. Leslie, A.G Walker J.E.
Proc. Natl. Acad. Sci. USA (2006) ***103****:8646*

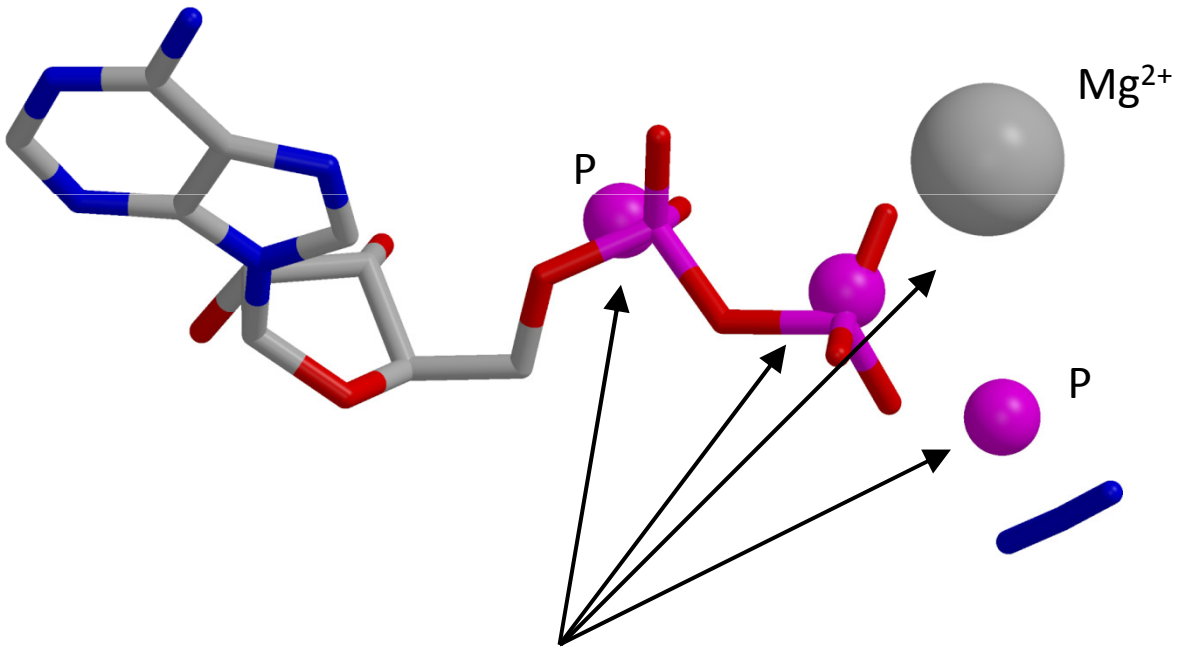
F1 ATPase testcase

	Present sites			
	Met	**Cys**	**PO_4**	**Ion**
pdb 2CK3	73	8	15	?

	Found sites			
Model	**Met**	**Cys**	**PO_4**	**Ion**
F1-ATPase	53	8	12	8
F1 ATPase 3α chains	10	3	7	5
Yeast ATPase	24	5	8	3

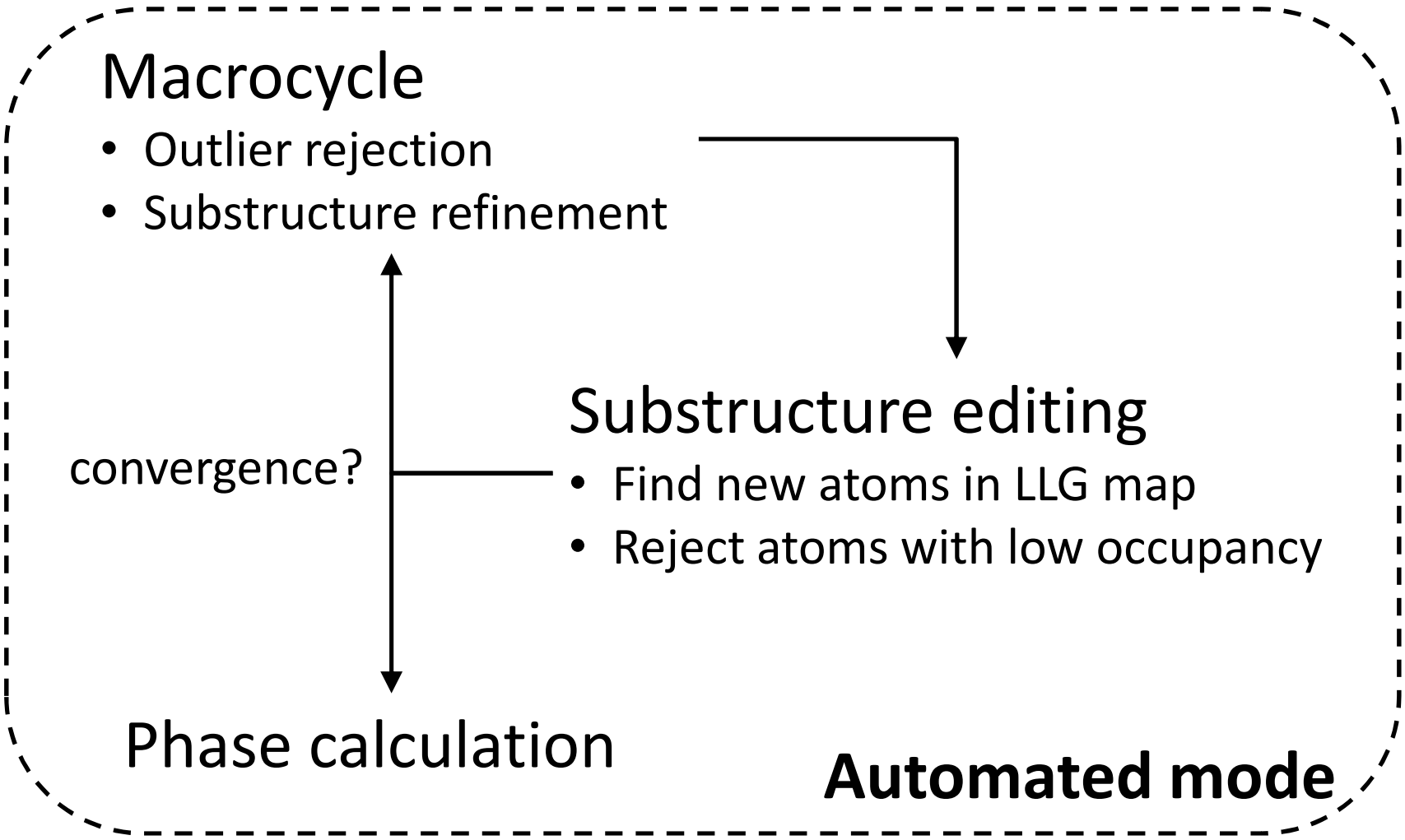
F1 ATPase testcase

ADP in active site
(not included in model)

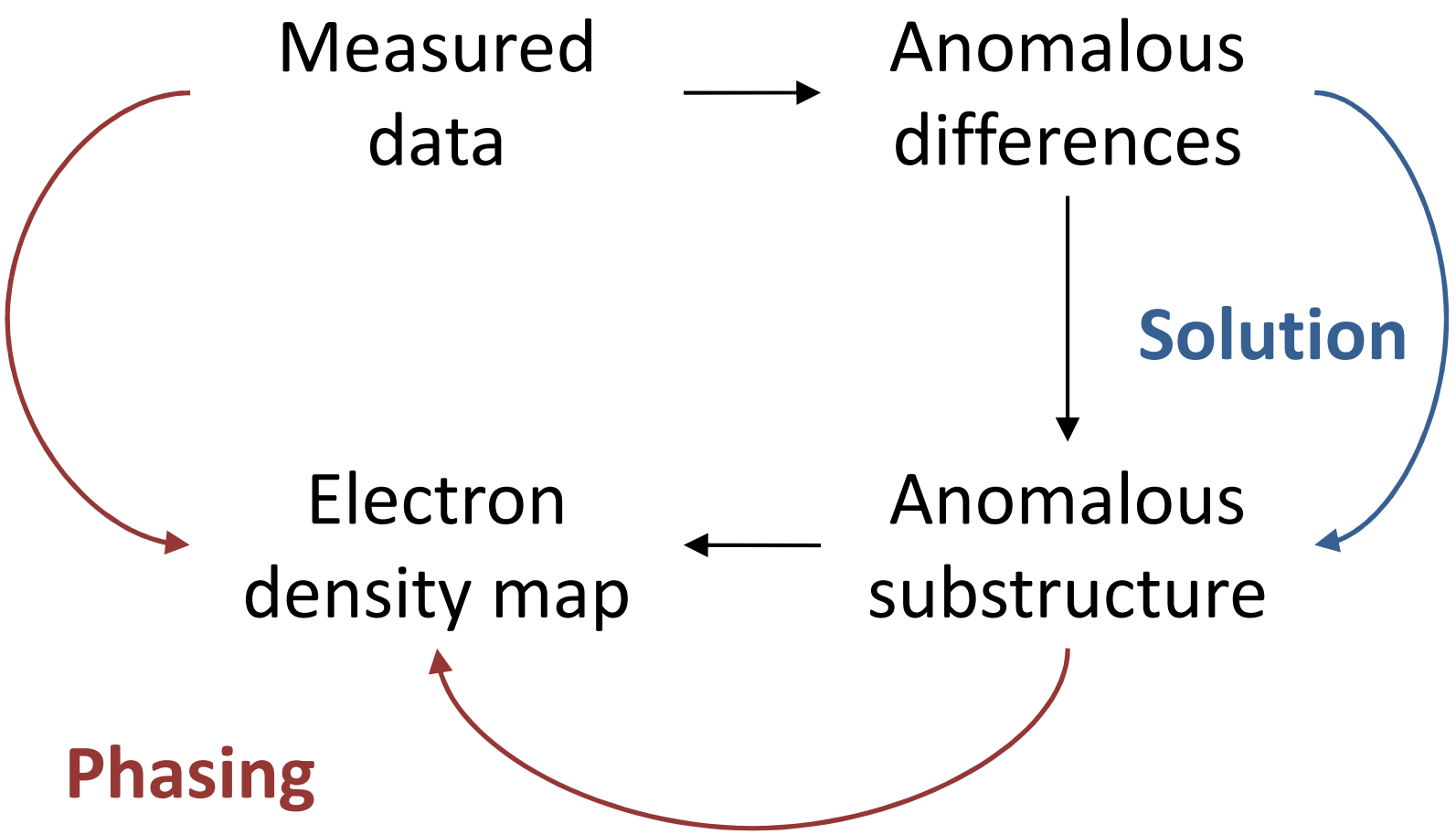


Located sites

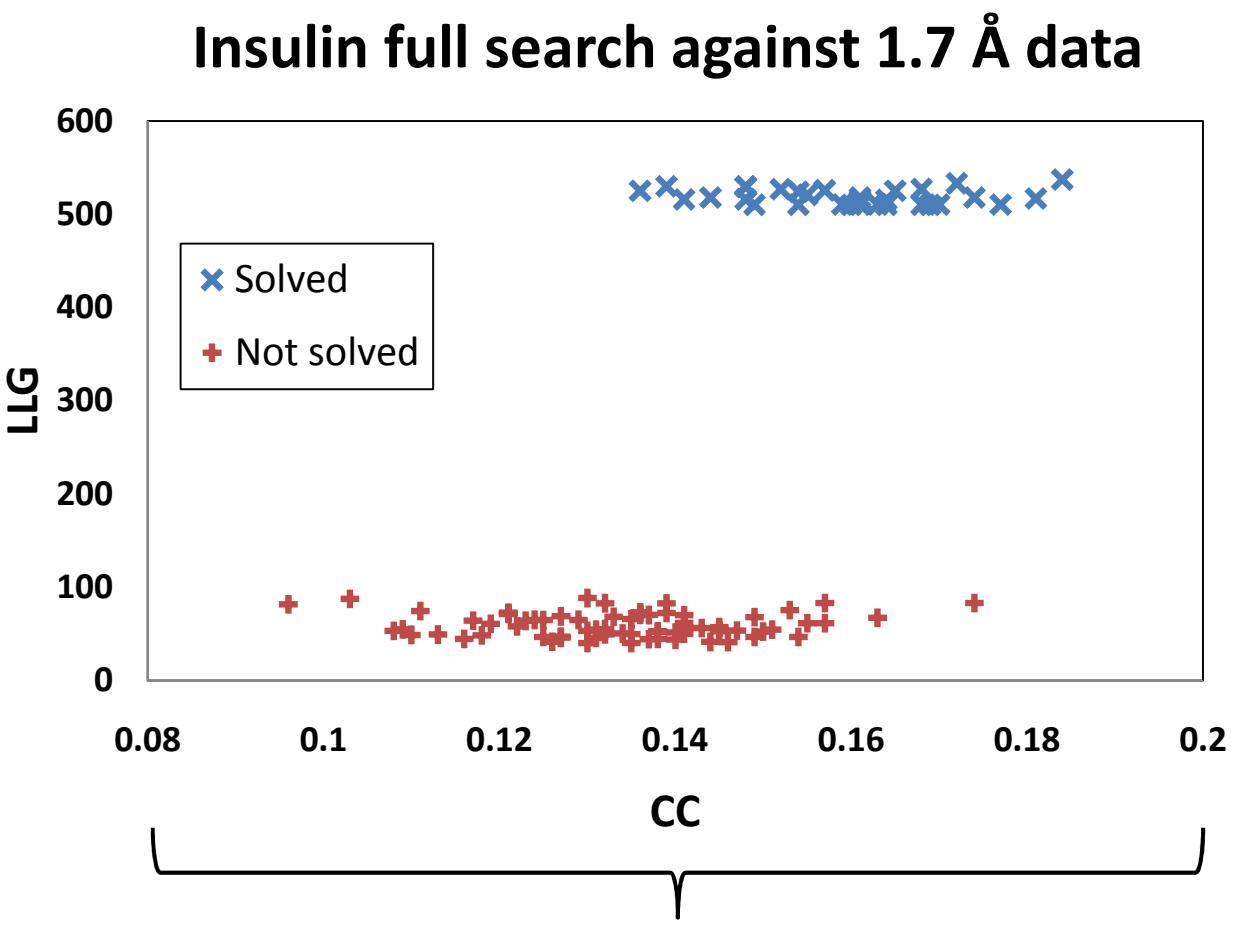
Phaser


Macrocycle
• Outlier rejection
• Substructure refinement
Substructure editing
• Find new atoms in LLG map
• Reject atoms with low occupancy
convergence?
Phase calculation
Automated mode

Data flow in substructure solution



Rescore trials with phaser

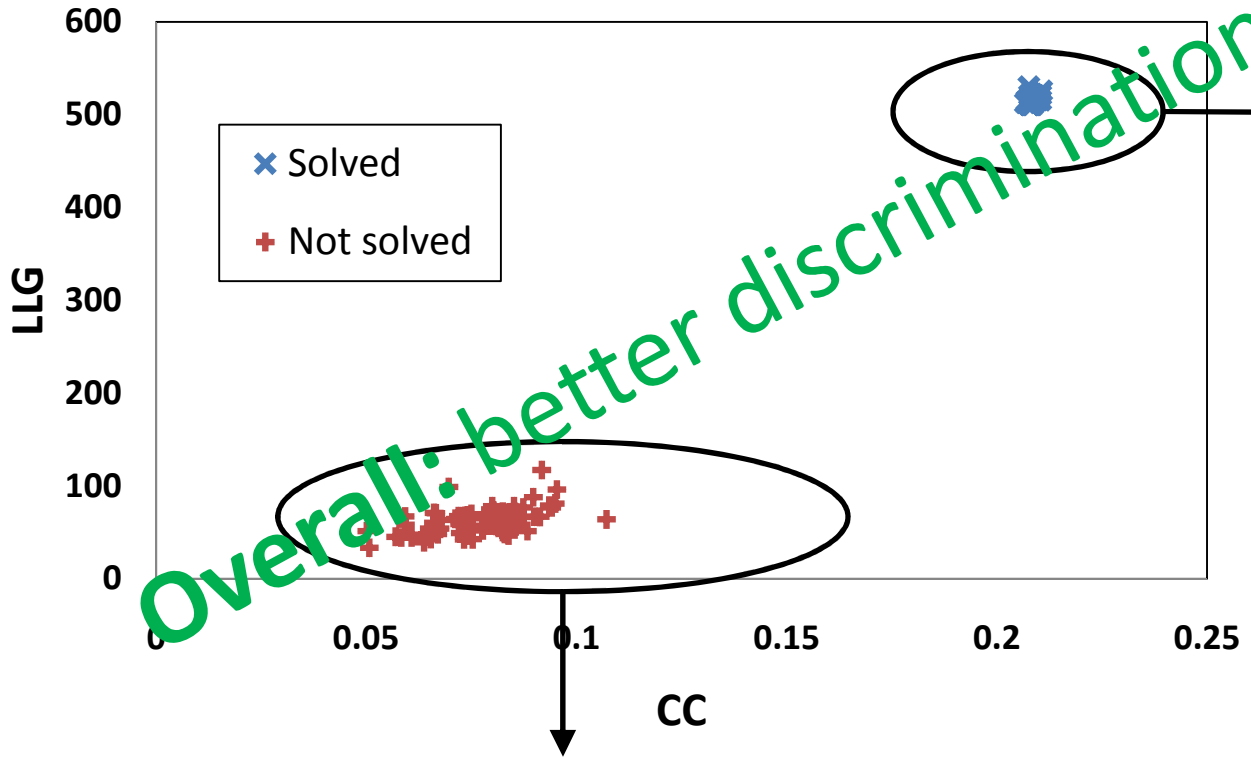


Clearly separated solution and non-solution groups

Impossible to identify solutions

CC for phaser-refined substructures

Insulin full search against 1.7 Å data



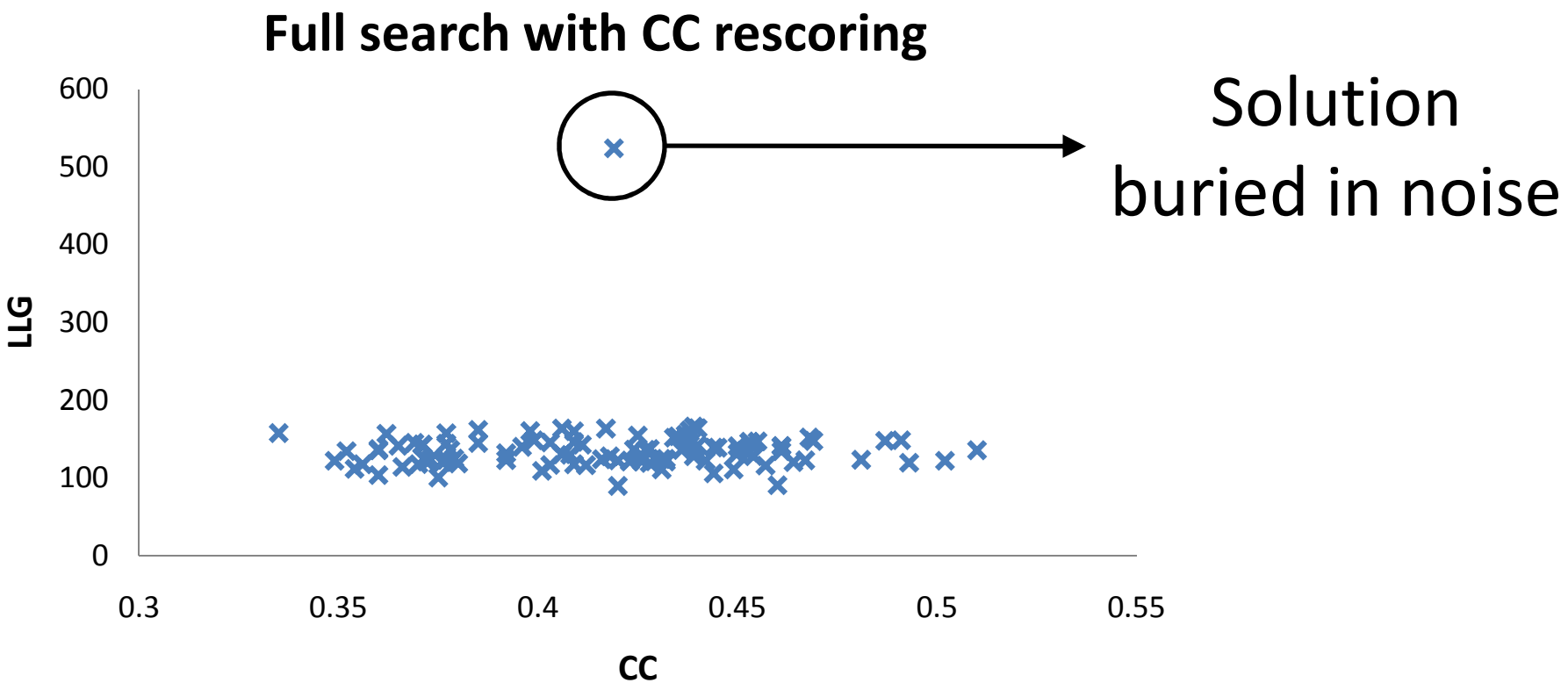
Increased CC for solutions

Completion: improved substructure

Lower CC for non-solutions

Refinement: reduced overfitting

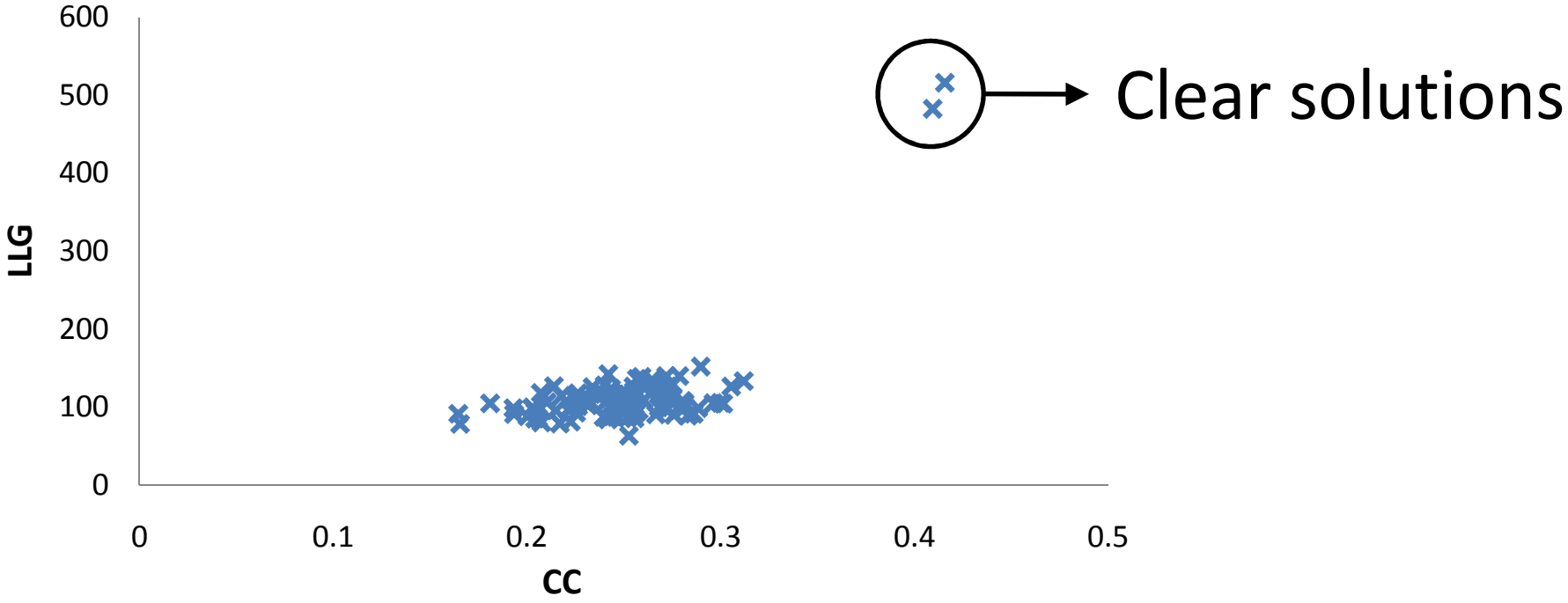
Fibronectin testcase



Rudiño-Piñera E, Ravelli RB, Sheldrick GM, Nanao MH, Korostelev VV, Werner JM, Schwarz-Linek U, Potts JR, Garman EF
J Mol Biol. **368**, 833-44

Fibronectin testcase

Full search with phaser rescoring



Rudiño-Piñera E, Ravelli RB, Sheldrick GM, Nanao MH, Korostelev VV, Werner JM, Schwarz-Linek U, Potts JR, Garman EF
J Mol Biol. **368**, 833-44

Substructure seeding

Substructure is available at this point

Patterson seeding

Dual-space recycling

Refinement and scoring

Compare:

- atoms compatible with the difference Patterson
- atoms compatible with the SAD LLG map

Comparison

Transaminase, 60 Se sites

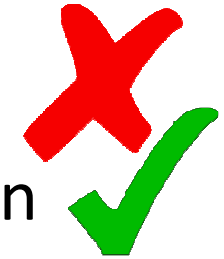
	Patterson seeding	LLG-map seeding
CC scoring	17	25
Phaser scoring	20	25

KSM42, 160 Se sites

	Patterson seeding	LLG-map seeding
CC scoring	23	28
Phaser scoring	28	37

When to use

- Phaser scoring
  - default run
  - default give no solution



- LLG-map seeding
  - default run
  - large substructures
  - with phaser scoring

Phaser

Mode

- SAD (start from heavy atoms)
- MRSAD (start with partial MR model)

Reflection data

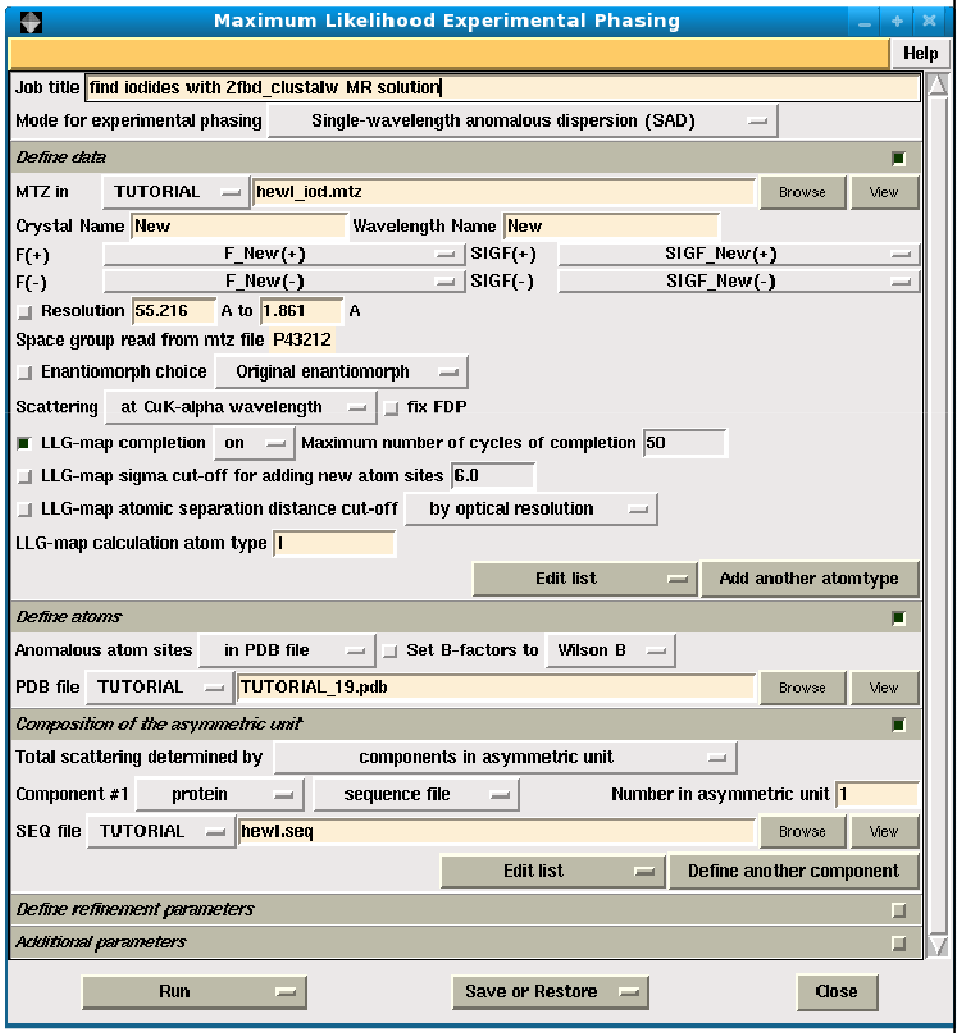
- F^+/F^-
- possible space groups

Scatterer information

- atom type and wavelength
- measured f'/f" values

Absolute scale

macromolecule weight



Results

Phase information

- FWT, PHWT
- HLA, HLB, HLC, HLD
- PHIB
- FLLG, PHILLG

PDB file

anomalous substructure

Logfile

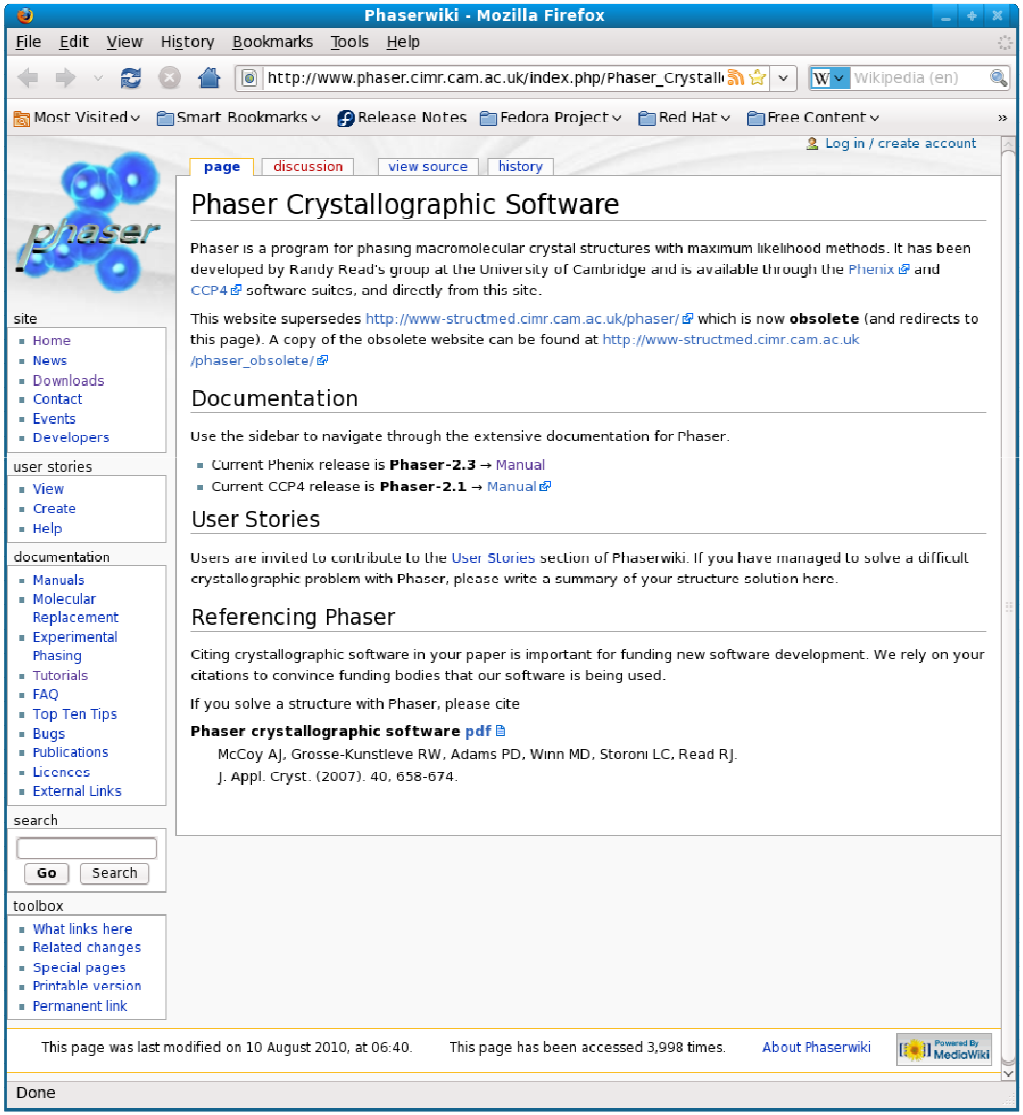
Best map or initial map for density modification

Phase probability for density modification or refinement

Centroid phase for density modification/refinement (use HL-coefficients!)

Log-likelihood gradient map (flat after automated mode!)

www.phaser.cimr.cam.ac.uk



Acknowledgements

Python-based Hierarchical Environment for Integrated Xtallography

Computational Crystallography Initiative / LBNL
Paul Adams, Ralf Grosse-Kunstleve, Pavel Afonine, Nathaniel Echols, Jeffrey Headd, Nigel Moriarty, Nicholas Sauter, Peter Zwart

Los Alamos National Lab
Tom Terwilliger, Li-Wei Hung

UNIVERSITY OF CAMBRIDGE

University of Cambridge
Randy Read, Airlie McCoy, Robert Oeffner, Gabor Bunkoczi

DUKE University

Duke University
Jane and Dave Richardson, Vincent Chen