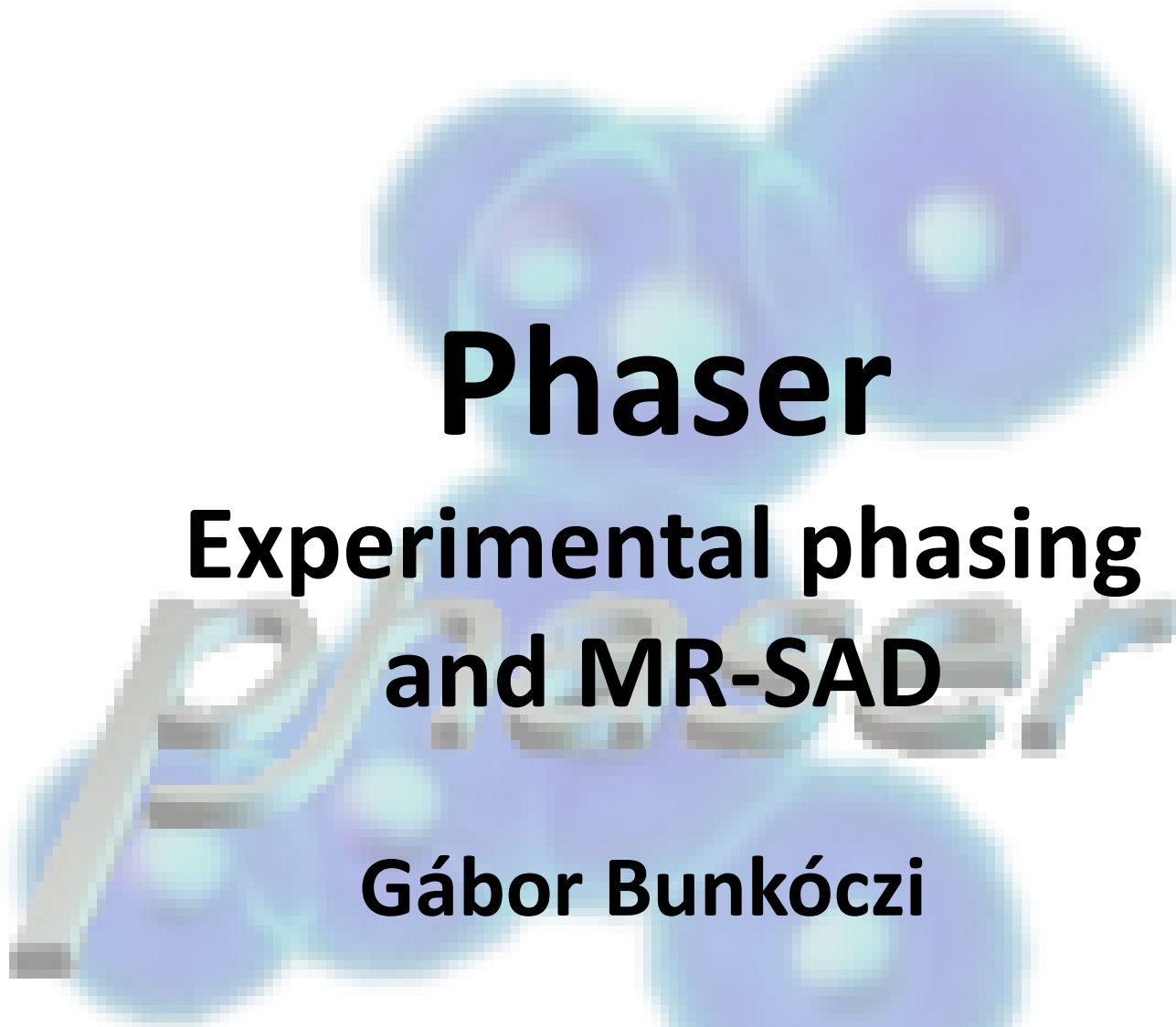




Phaser

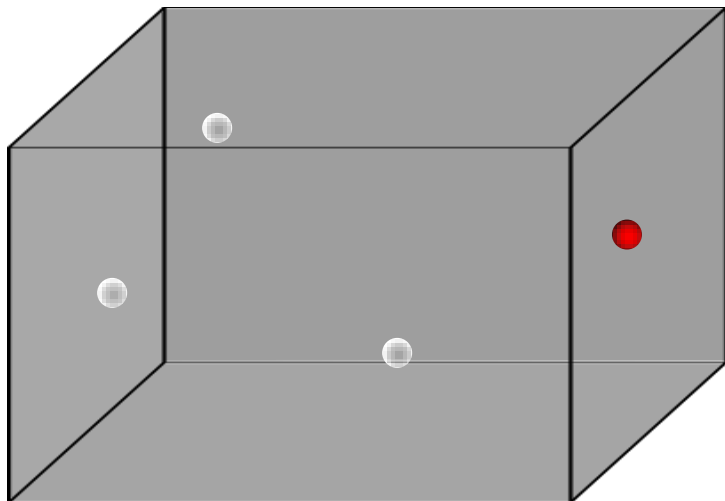
**Experimental phasing
and MR-SAD**

Gábor Bunkóczi

***CCP4 Fukuoka School,
31 October 2012***

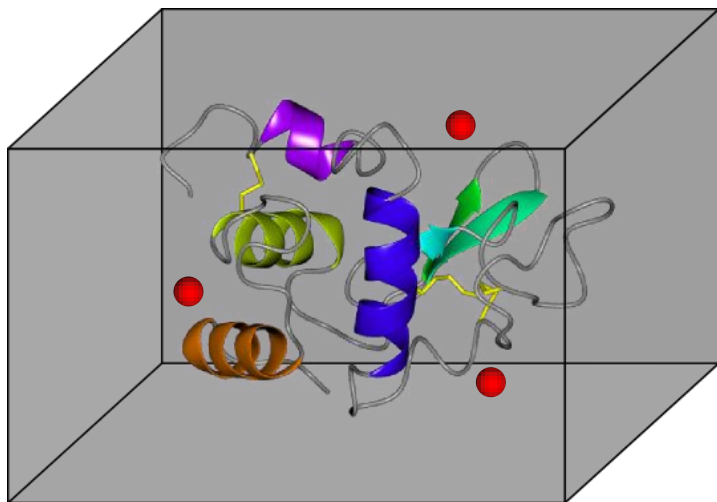


Phasing scenarios



Known anomalous substructure

- missing atoms
- model errors (xyz , occ , B)



Known partial structure

- missing (anomalous) atoms
- (partial) model errors



SAD function – normal scattering

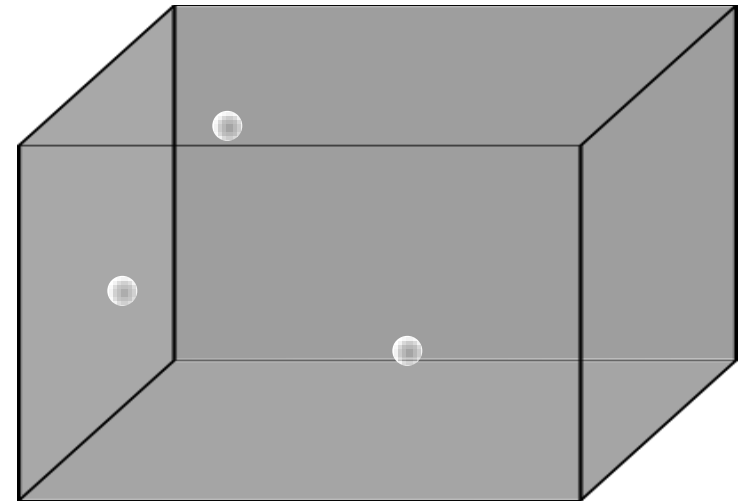
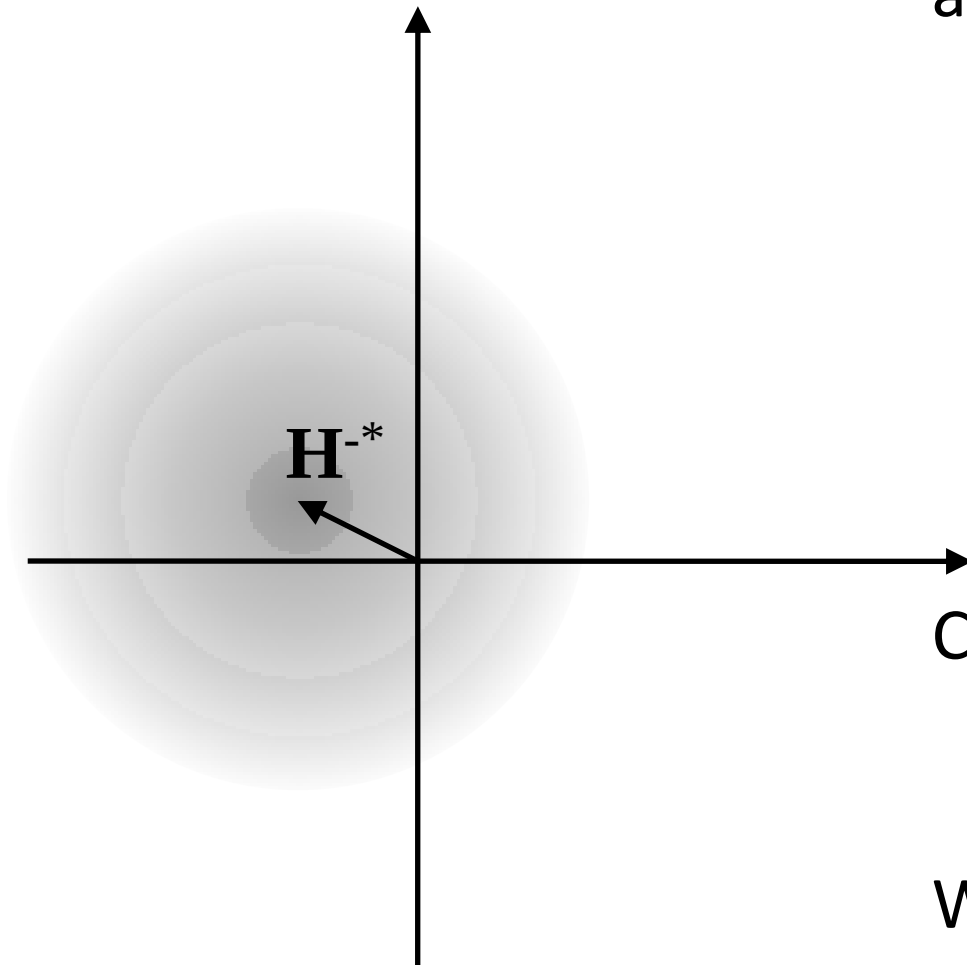
$$P(F_o^-, \alpha^-; \mathbf{H}^{-*}) = \frac{F_o^-}{\pi \Sigma^-} \exp \left(- \frac{|F_o^- e^{i\alpha^-} - D \mathbf{H}^{-*}|^2}{\Sigma^-} \right)$$

- F_o^- – measured amplitude
- α^- – phase angle
- $\mathbf{H}^-$ – structure factor from known substructure
- D – substructure error
- Σ^- – uncertainty



SAD function – normal scattering

Scenario:
anomalous substructure



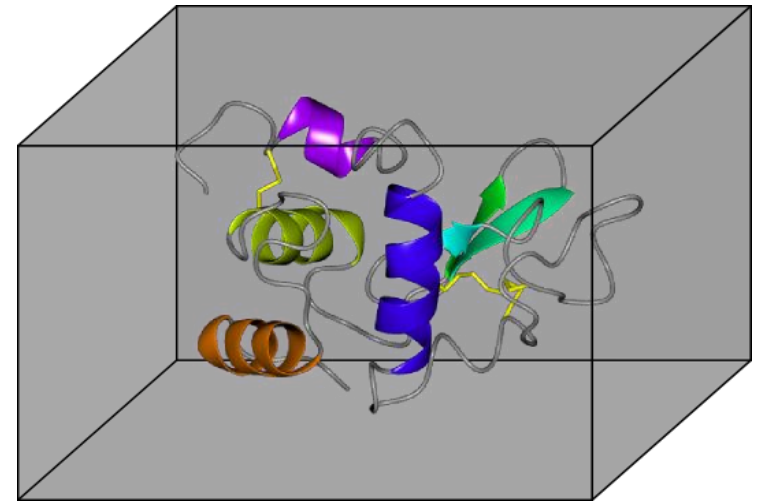
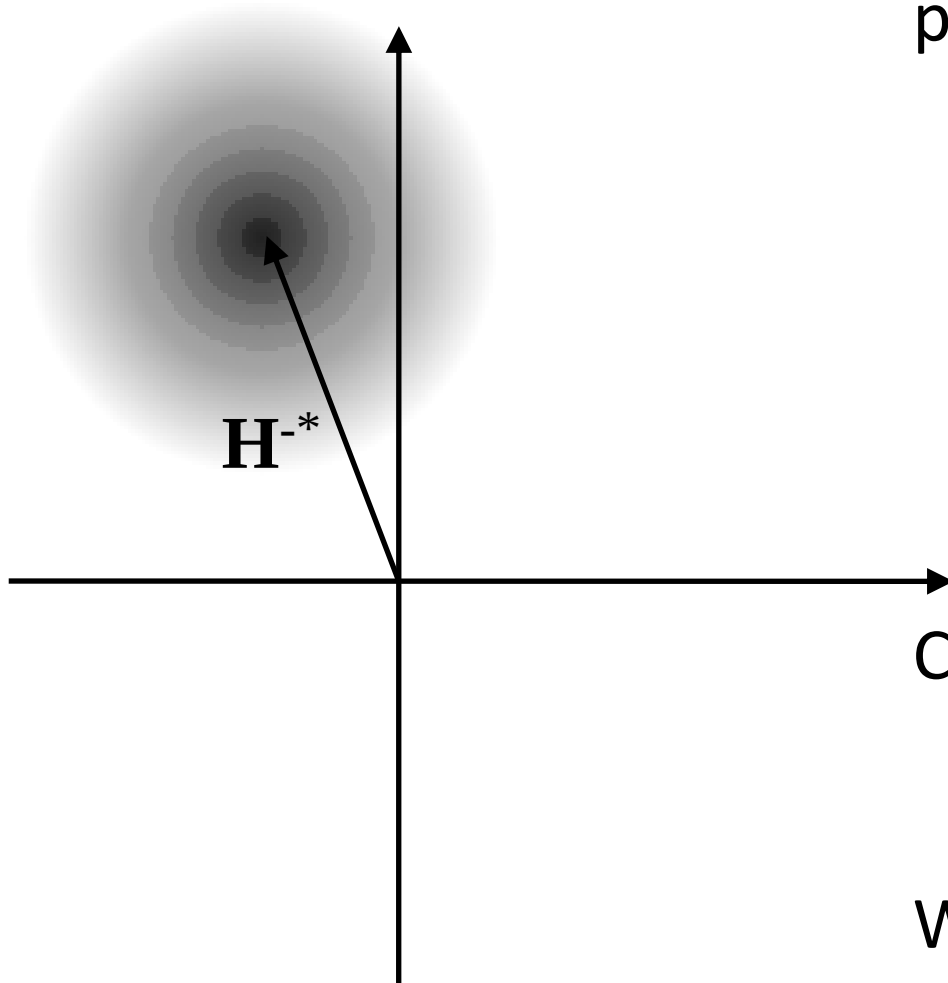
Centered on:
**structure factor of known
substructure**

Width determined by:
model incompleteness



SAD function – normal scattering

Scenario:
partial MR model



Centered on:
**structure factor of known
substructure**
Width determined by:
model errors



SAD function – anomalous scattering

$$P(F_o^+; \mathbf{F}^-, \mathbf{H}^+, \mathbf{H}^{-*}) = \frac{2F_o^+}{\Sigma^+} \exp\left(-\frac{(F_o^+ - F_c^+)^2}{\Sigma^+}\right) eI_0\left(\frac{2F_o^+ F_c^+}{\Sigma^+}\right)$$

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

$\mathbf{H}^\pm$ – structure factor of
known substructure

F_c^+ – estimate of F_o^+ from
 F_o^- , $\mathbf{H}^+$ and $\mathbf{H}^-$

Σ^+ – error in F_c^+



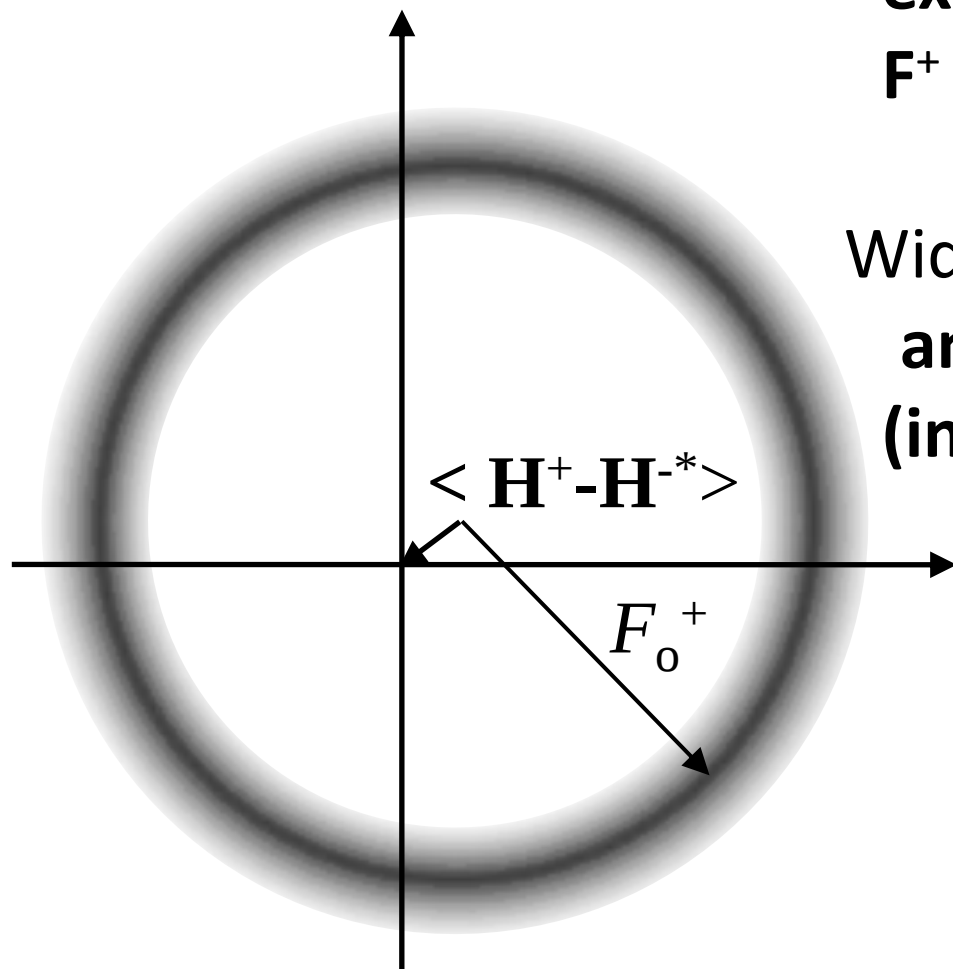
SAD function – anomalous scattering

Centered on:

expected difference between F^+ and F^-

Width determined by:

**anomalous substructure error
(including measurement error)**

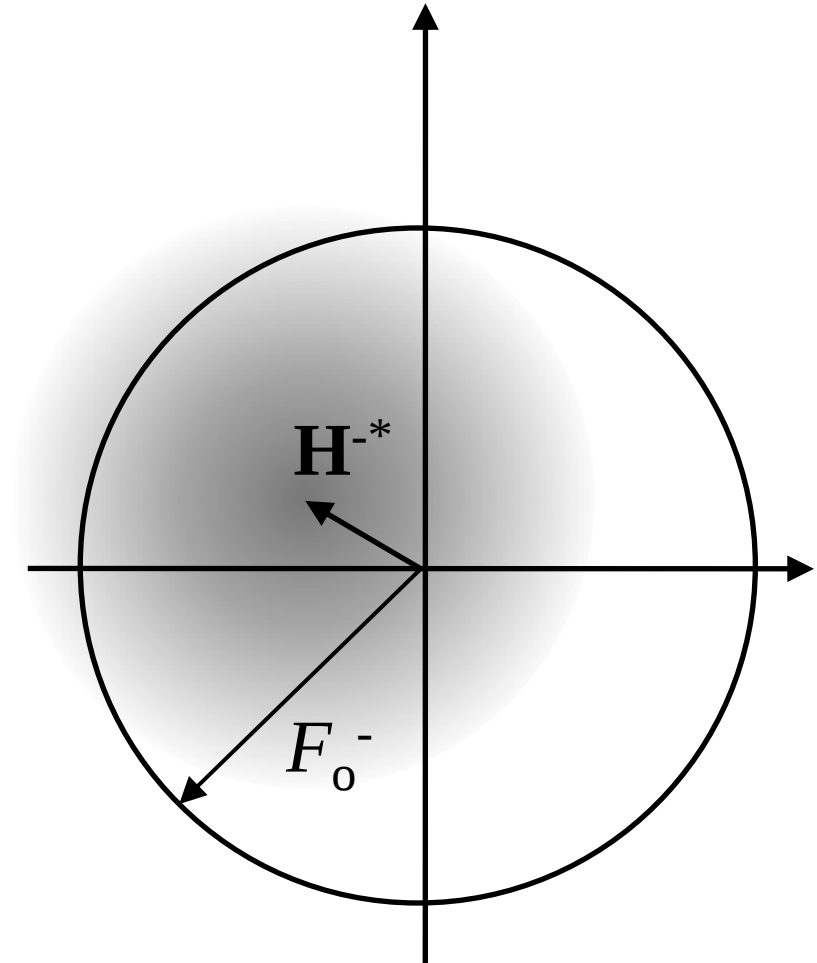
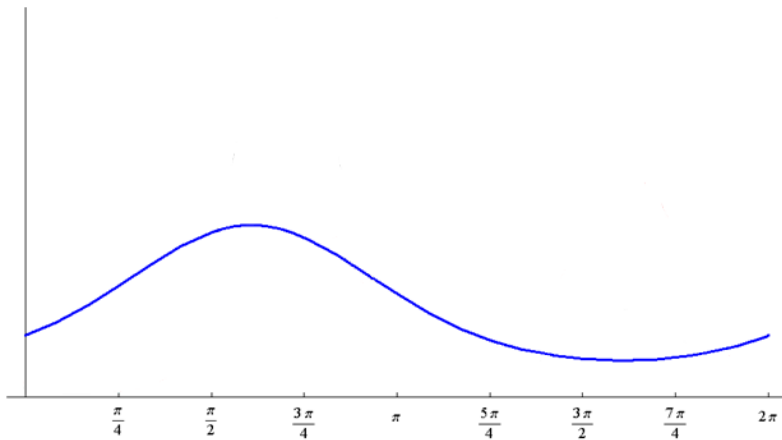




Phase probability

Normal scattering

Phase probability

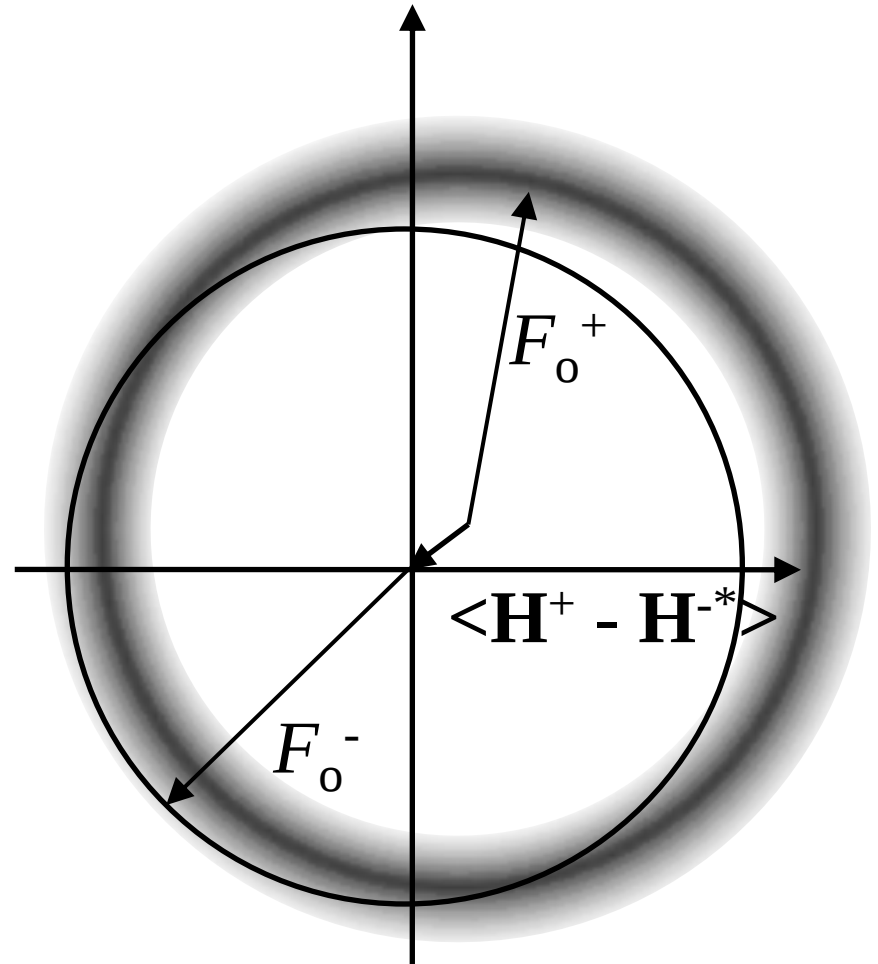
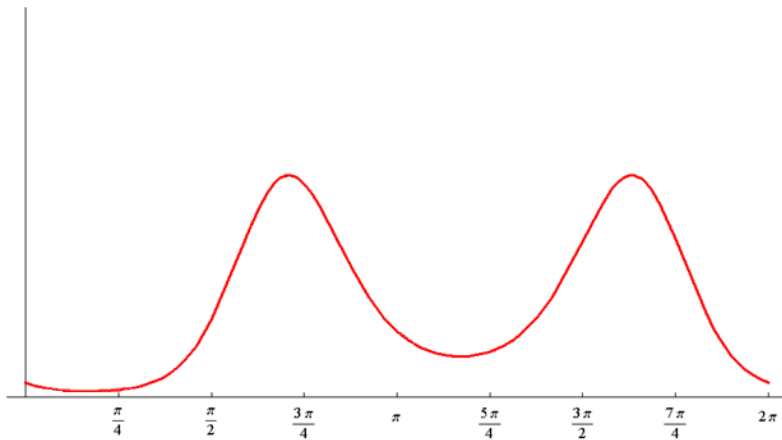




Phase probability

Anomalous scattering

Phase probability





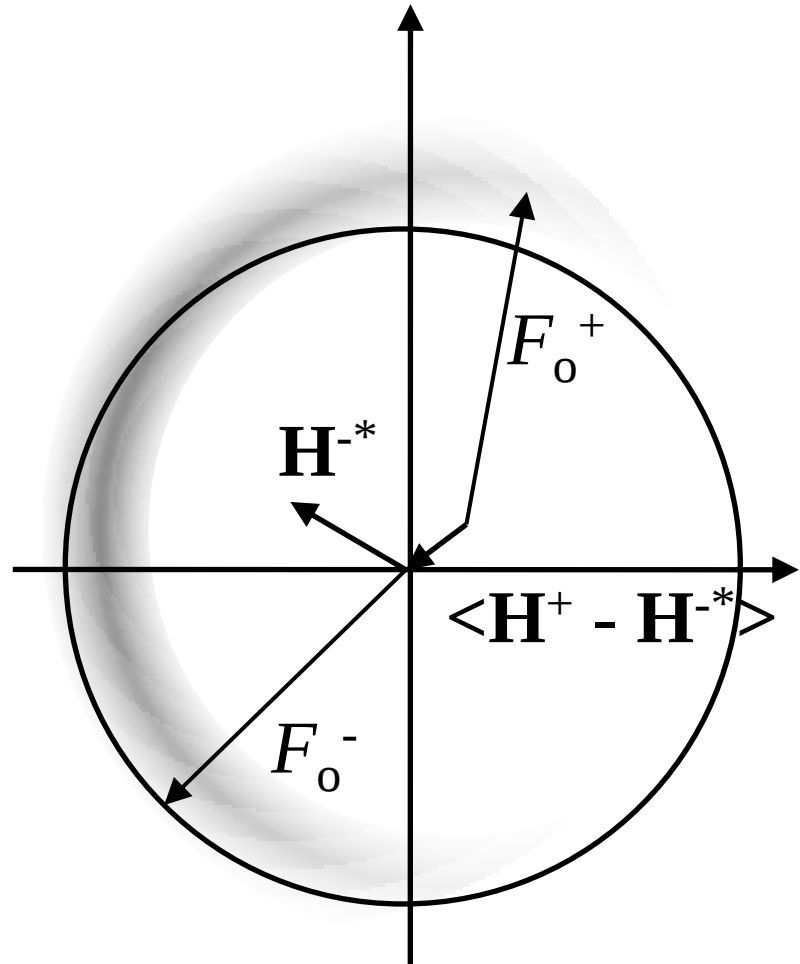
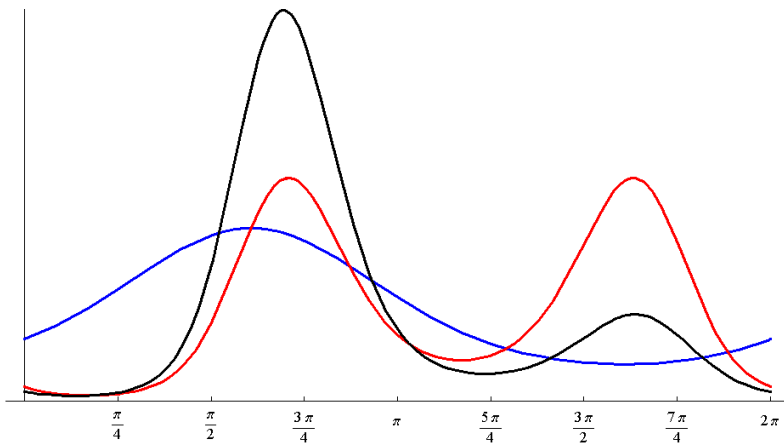
Phase probability

Total

Normal scattering

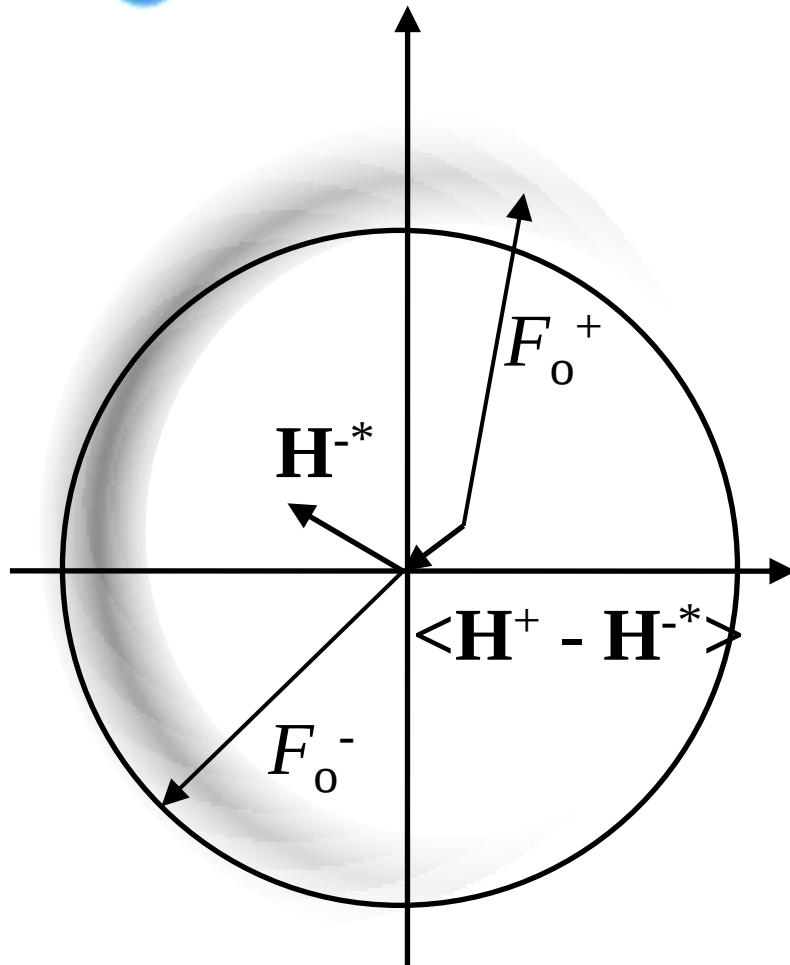
Anomalous scattering

Phase probability

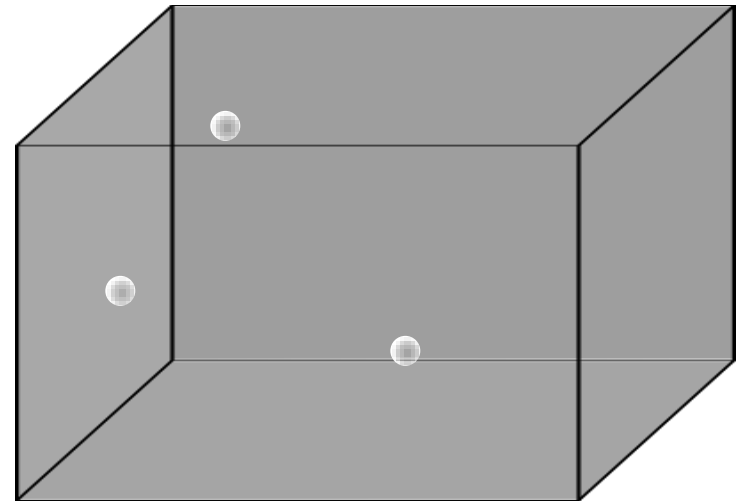




Phase probability – SAD phasing

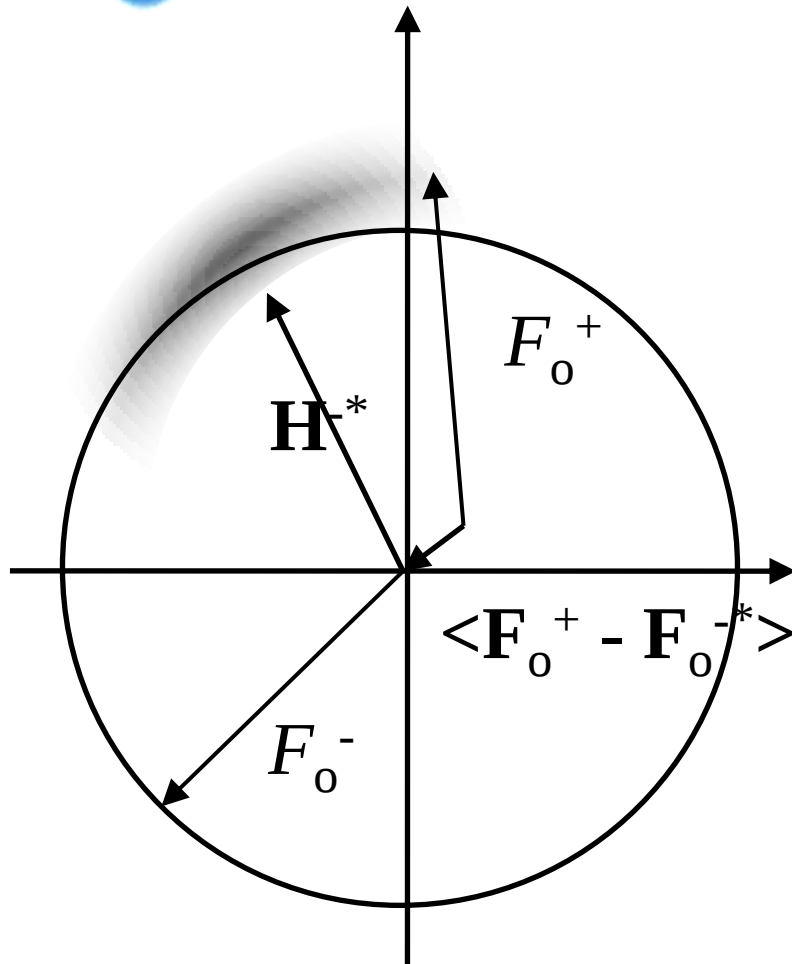


Scenario:
anomalous substructure

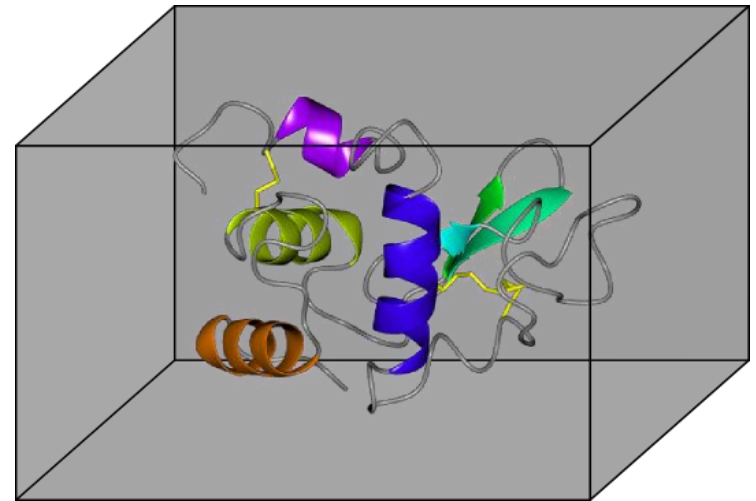




Phase probability – MR-SAD phasing



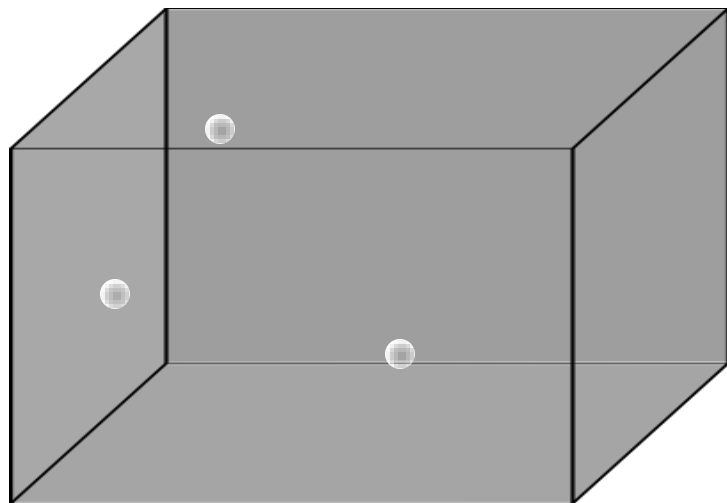
Scenario:
partial MR model



Phases contain **contribution from** known partial (MR) **model** that are weighted with **expected model error**

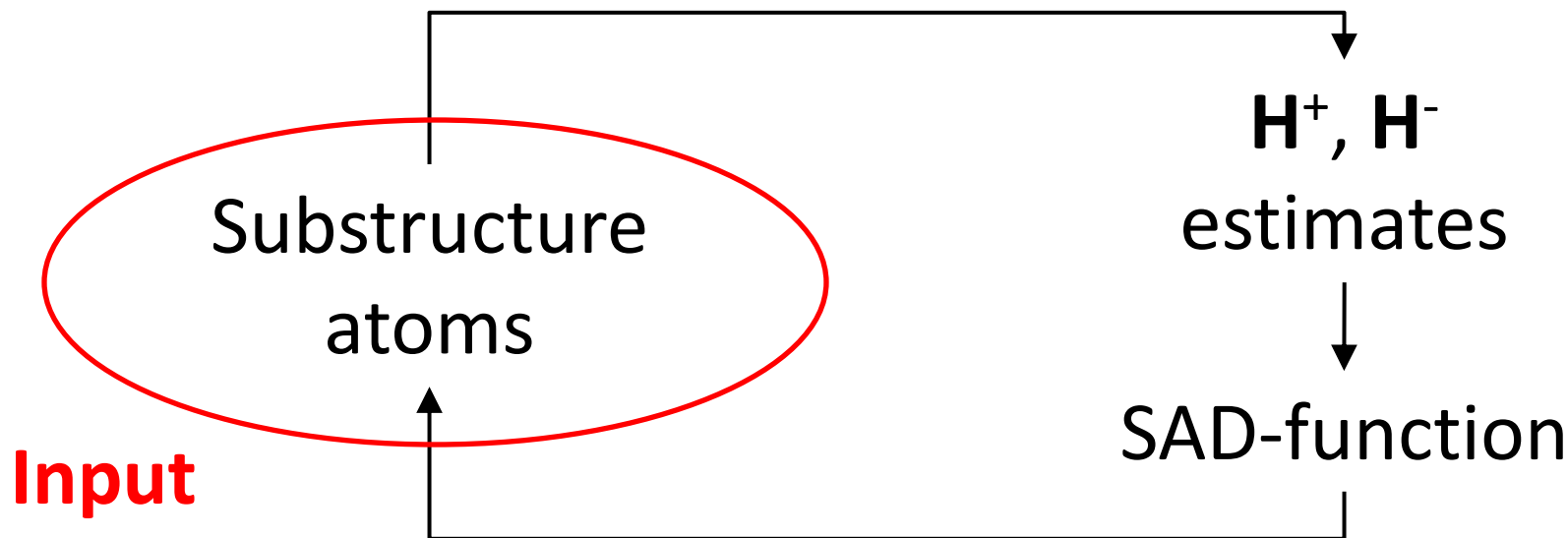


SAD phasing workflow



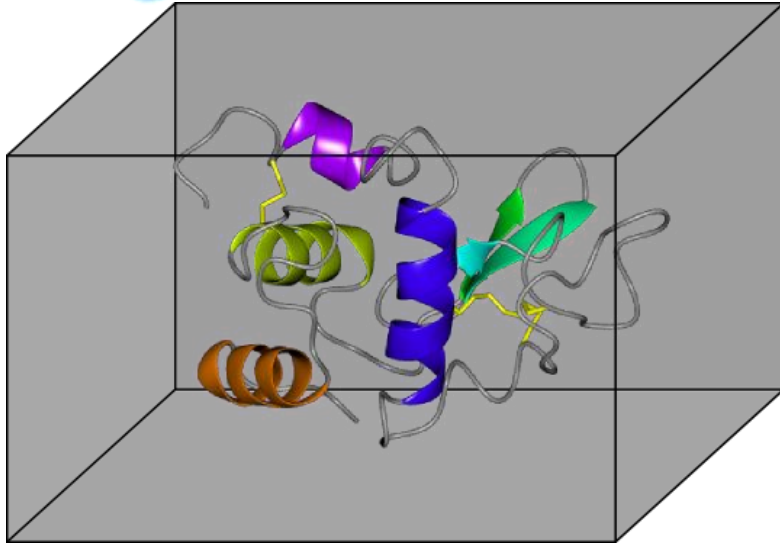
Parameters:

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





MR-SAD phasing workflow



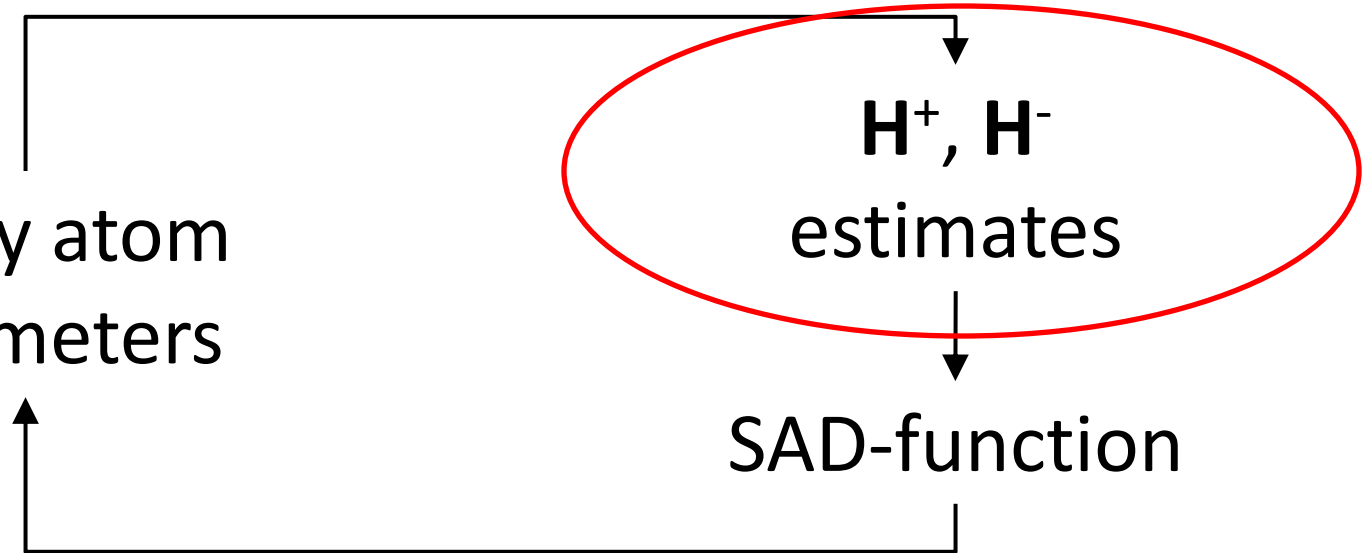
Input:

- partial structure
- possible atom types

Heavy atom
parameters

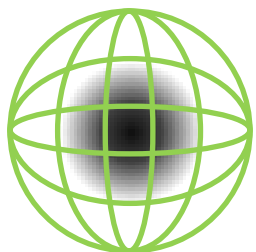
H^+ , H^-
estimates

SAD-function



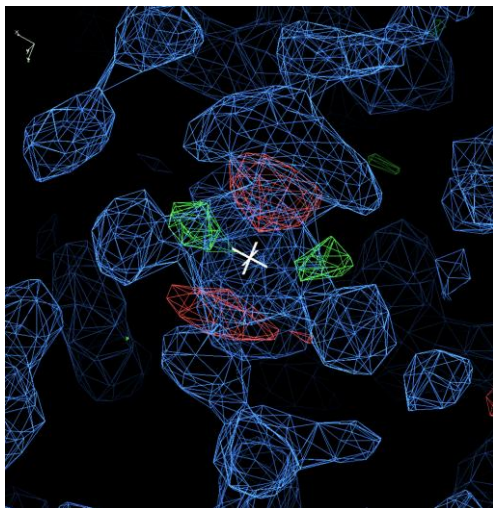


Log-likelihood gradient maps



Peak height dependent on f'/f'' ratio

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



Very sensitive to detail

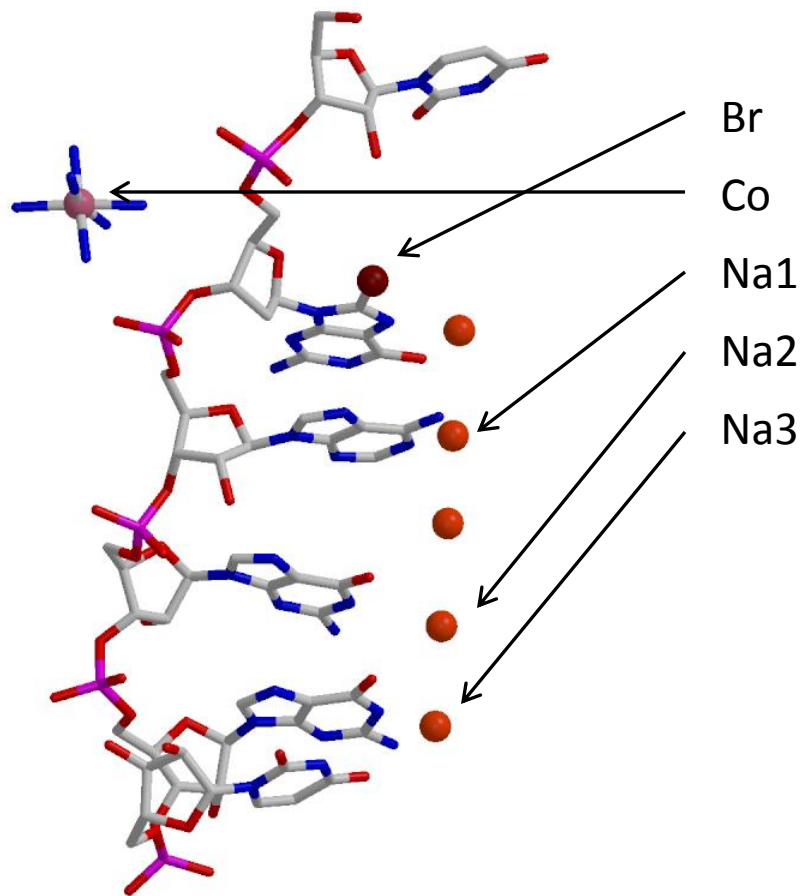
- e.g. anisotropic temperature motion

Iterative

- improvements in substructure description increases sensitivity

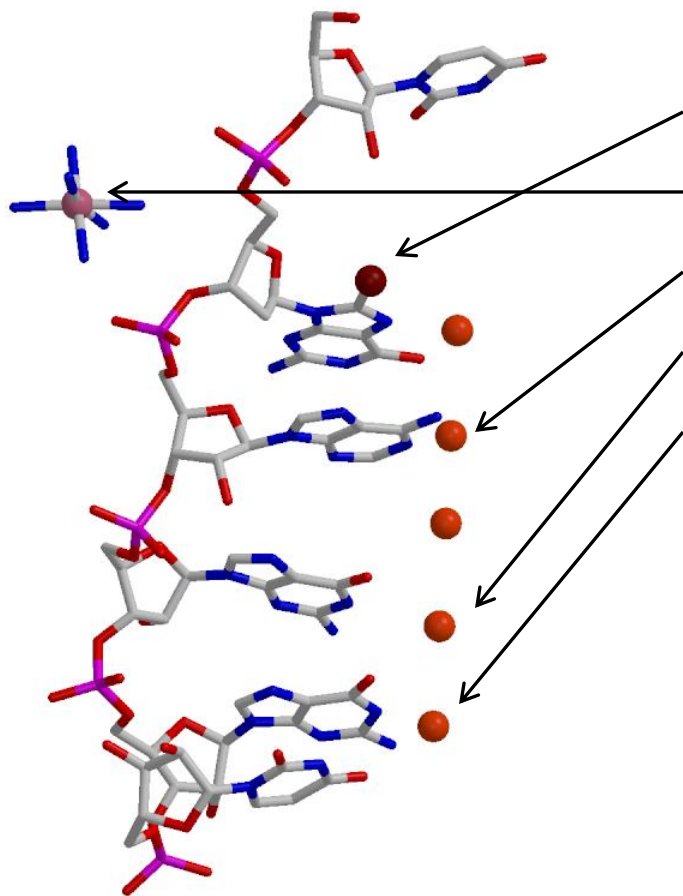


Log-likelihood gradient maps





Log-likelihood gradient maps



	Δ_{ano}	LLG1	LLG2	LLG3
Br	26.5	28.5	-	-
Co	5.0	5.0	18.5	-
Na1	-	-	7.6	-
Na2	-	-	-	6.6
Na3	-	-	-	6.0

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



Common uses of MR-SAD

- Model refinement
identify/validate anomalous scatterers
- Model building
Marker atoms to help sequence docking
- Escape model bias
add weak but independent phase information
- Improve anomalous substructure/phasing



Iterative substructure improvement

- Nitrate reductase (Natalie Strynadka)
 - integral membrane protein, 1976 residues
 - solved by combination of Fe-MAD and MIRAS
 - 21 Fe, 1 Mo and 113 S atoms
- Solve using Fe peak SAD data



Iterative substructure improvement

- Find heavy atoms sites
 - 11 “Fe” sites from phenix.hyss
 - several of super-sites of Fe_4S_4 clusters
- Complete and phase using Phaser
 - 38 Fe-sites, some of which are S atoms
 - still ghosts of super-sites



Iterative substructure improvement

- Improve phases with density modification
- Build structure with ARP/wARP
798 residues, 18 docked in the sequence
- Completion in Phaser
 - using partial poly-Ala model from ARP/wARP
 - complete using Fe, Mo and S atoms
 - 21 Fe, 1 Mo and 84 of 113 S found
- Rebuild with ARP/wARP
1392 of 1976 build, 736 docked in the sequence



F1-ATPase testcase

3500 residue protein complex
81 S atoms, 3.50 Å resolution
Wavelength: 1.77 Å ($f_s''=0.72$ e⁻)

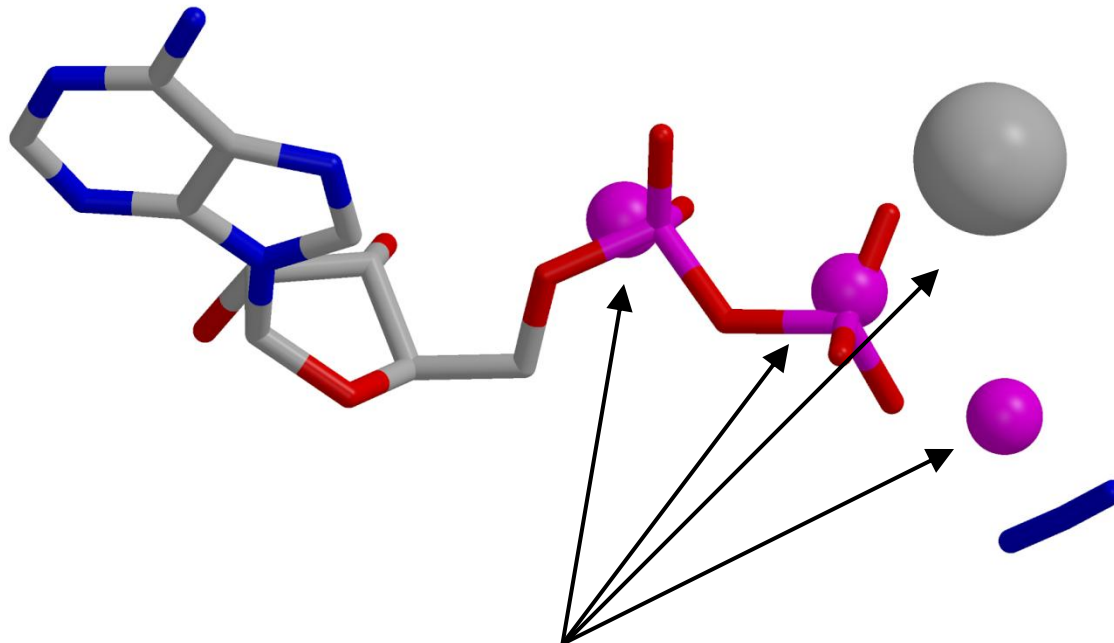
phenix.xtriage:

The anomalous signal seems to extend to about **21.5 Å** (or to **12.7 Å**, from a more **optimistic point of view**). The quoted resolution limits can be used as a guideline to decide where to cut the resolution for phenix.hyss. As the anomalous signal is not very strong in this dataset substructure solution via **SAD might prove to be a challenge**. Especially if only low resolution reflections are used, the resulting substructures could contain a significant amount of false positives.



F1-ATPase testcase

ADP in active site
(not included in model)



Located sites



Phaser

Macrocycle

- Outlier rejection
- Substructure refinement

convergence?

Substructure editing

- Find new atoms in LLG map
- Reject atoms with low occupancy

Phase calculation

Automated mode



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!)



Phaser

Mode

- SAD (start from heavy atoms)

Reflection data

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

Scatterer information

- anomalous substructure
- atom type and wavelength
- measured f'/f'' values

Absolute scale

macromolecule weight

The screenshot shows the 'Maximum Likelihood Experimental Phasing' window. The 'Job title' is 'find iodides with 2fbd_clustalw MR solution'. The 'Mode for experimental phasing' is 'Single-wavelength anomalous dispersion (SAD)'. Under 'Define data', 'MTZ in' is 'TUTORIAL' and 'hewl_iod.mtz'. 'Crystal Name' is 'New' and 'Wavelength Name' is 'New'. 'F(+)' is 'F_New(+)' and 'SIGF(+)' is 'SIGF_New(+)'. 'F(-)' is 'F_New(-)' and 'SIGF(-)' is 'SIGF_New(-)'. 'Resolution' is '55.216' A to '1.861' A. 'Space group read from mtz file' is 'P43212'. 'Enantiomorph choice' is 'Original enantiomorph'. 'Scattering' is 'at CuK-alpha wavelength' and 'fix FDP' is unchecked. 'LLG-map completion' is 'on' and 'Maximum number of cycles of completion' is '50'. 'LLG-map sigma cut-off for adding new atom sites' is '6.0'. 'LLG-map atomic separation distance cut-off' is 'by optical resolution'. 'LLG-map calculation atom type' is 'I'. There are 'Edit list' and 'Add another atomtype' buttons. Under 'Define atoms', 'Anomalous atom sites' is 'in PDB file' and 'Set B-factors to' is 'Wilson B'. 'PDB file' is 'TUTORIAL' and 'TUTORIAL_19.pdb'. There are 'Browse' and 'View' buttons. Under 'Composition of the asymmetric unit', 'Total scattering determined by' is 'components in asymmetric unit'. 'Component #1' is 'protein' and 'sequence file' is 'hewl.seq'. 'Number in asymmetric unit' is '1'. There are 'Edit list' and 'Define another component' buttons. Under 'Define refinement parameters' and 'Additional parameters', there are checkboxes. At the bottom, there are 'Run', 'Save or Restore', and 'Close' buttons.



Phaser

Mode

- MRSAD (partial MR solution)

Reflection data

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

Scatterer information

- partial MR solution
- (anomalous substructure)
- atom type and wavelength
- measured f'/f'' values

Absolute scale

macromolecule weight

The screenshot shows the 'Maximum Likelihood Experimental Phasing' window. The 'Job title' is 'Redetermine anomalous scatterers from ARP/wARP model, rephase'. The 'Mode for experimental phasing' is 'SAD with molecular replacement partial structure'. Under 'Define data', 'MTZ in' is 'ESRF2007' and 'fae_rm_P212121.mtz'. 'Crystal Name' is 'fae' and 'Wavelength Name' is 'remote'. 'F(+)' and 'F(-)' are linked to 'F_rm(+)' and 'F_rm(-)' respectively, with 'SIGF(+)' and 'SIGF(-)' also linked to 'SIGF_rm(+)' and 'SIGF_rm(-)'. 'Resolution' is '78.326 A to 2.000 A'. 'Space group read from mtz file' is 'P212121'. 'Enantiomorph choice' is 'Original enantiomorph'. 'Scattering' is 'at other wavelength' with '0.9756 A'. 'LLG-map completion' is 'on' with 'Maximum number of cycles of completion' set to '50'. 'LLG-map sigma cut-off for adding new atom sites' is '6.0'. 'LLG-map atomic separation distance cut-off' is 'by optical resolution'. 'LLG-map calculation atom type' is 'Se' and 'Cd'. Under 'Define atoms', 'PDB' is 'ESRF2007' and '7_warpNtrace.pdb'. 'Similarity of PDB of partial structure to the target structure' is 'rms difference' with '1.0'. 'Anomalous atom sites' is 'none' and 'Set B-factors to' is 'Wilson B'. Under 'Composition of the asymmetric unit', 'Total scattering determined by' is 'components in asymmetric unit'. 'Component #1' is 'protein' with 'Number in asymmetric unit' set to '2'. 'SEQ file' is 'ESRF2007' and '1GKK.seq'. At the bottom, there are buttons for 'Run', 'Save or Restore', and 'Close'.



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



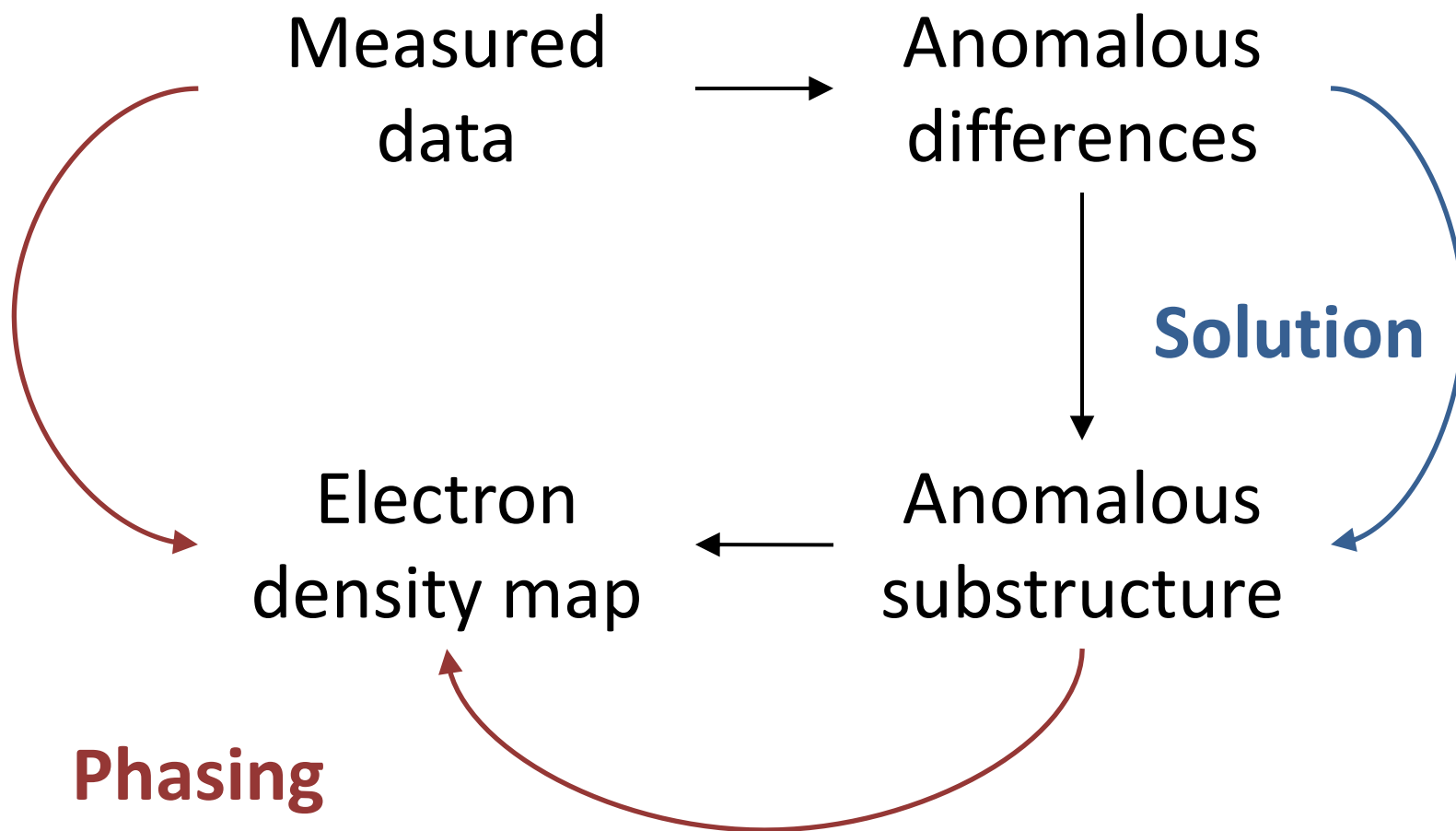
DUKE University

Duke University

Jane and Dave Richardson, Vincent Chen

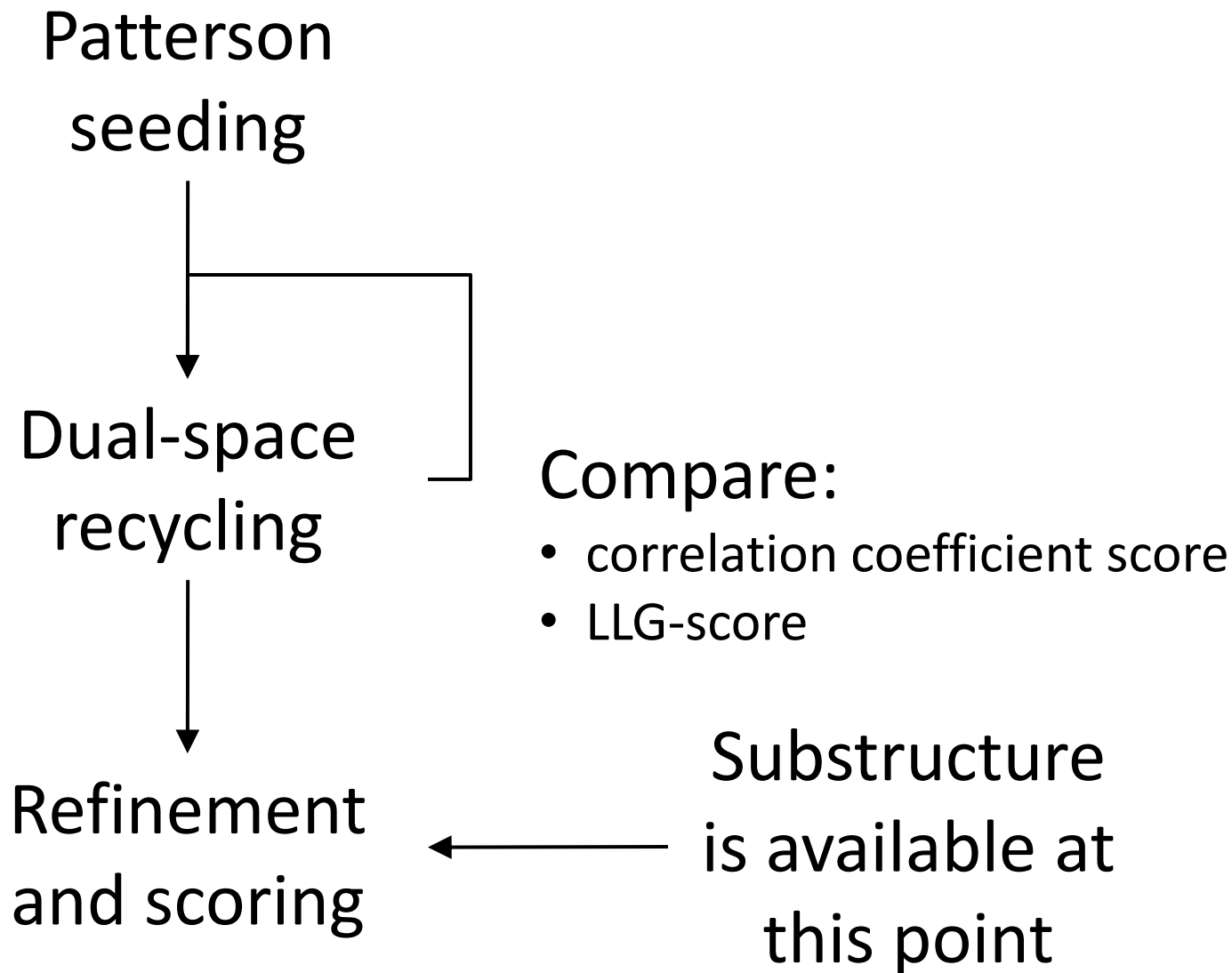


Substructure solution





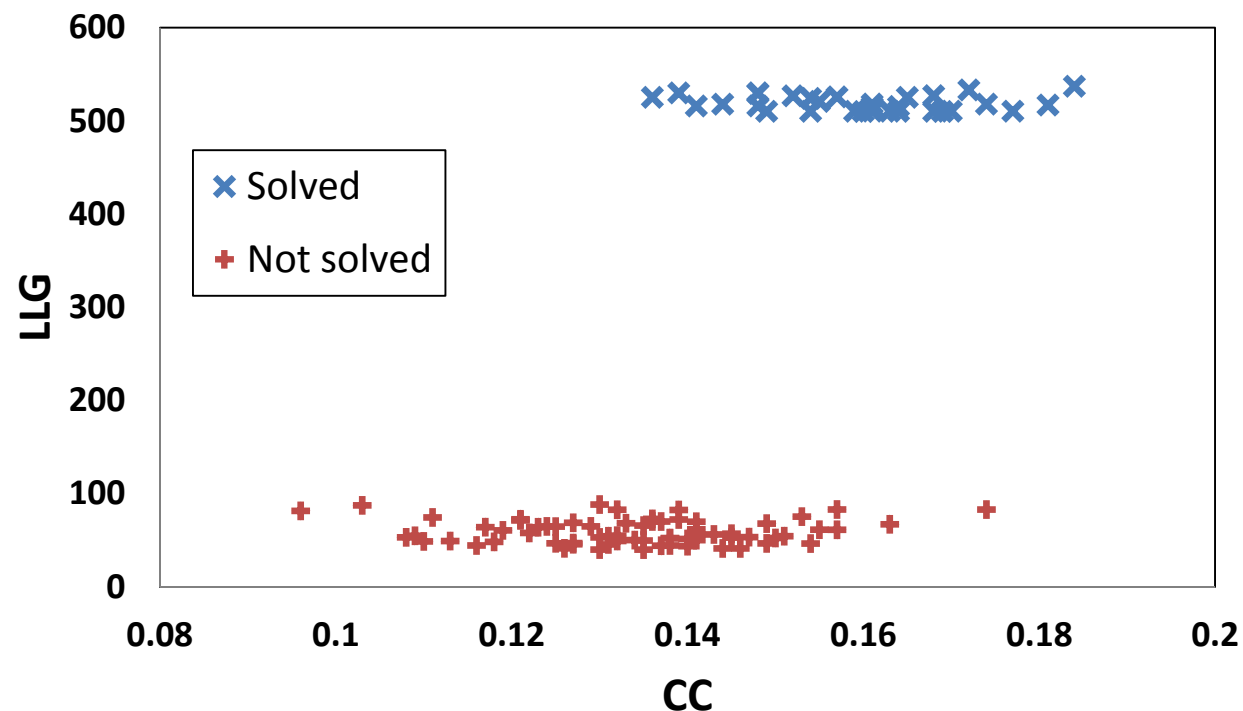
Substructure solution workflow





Rescore trials with phaser

Insulin full search against 1.7 Å data



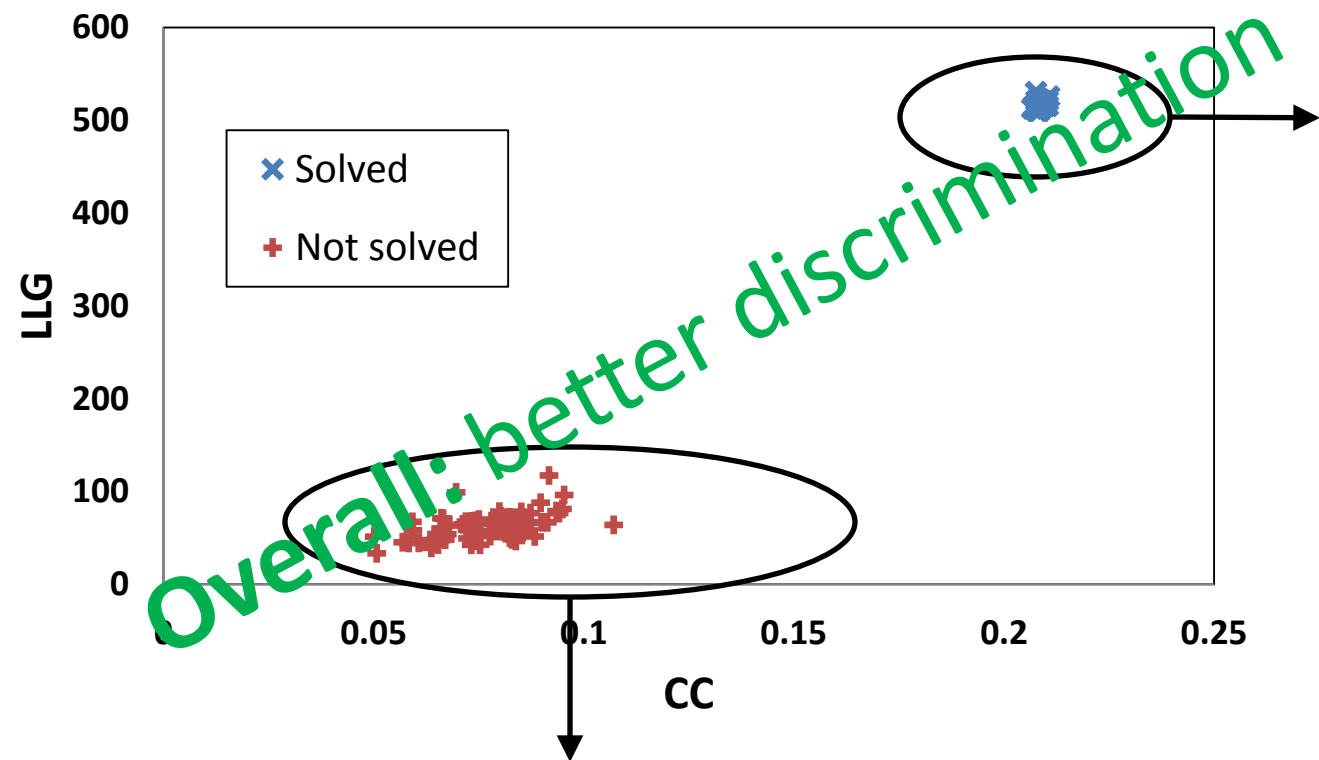
Clearly
separated
solution and
non-solution
groups

Impossible to identify solutions



CC for phaser-refined substructures

Insulin full search against 1.7 Å data



Increased CC
for solutions

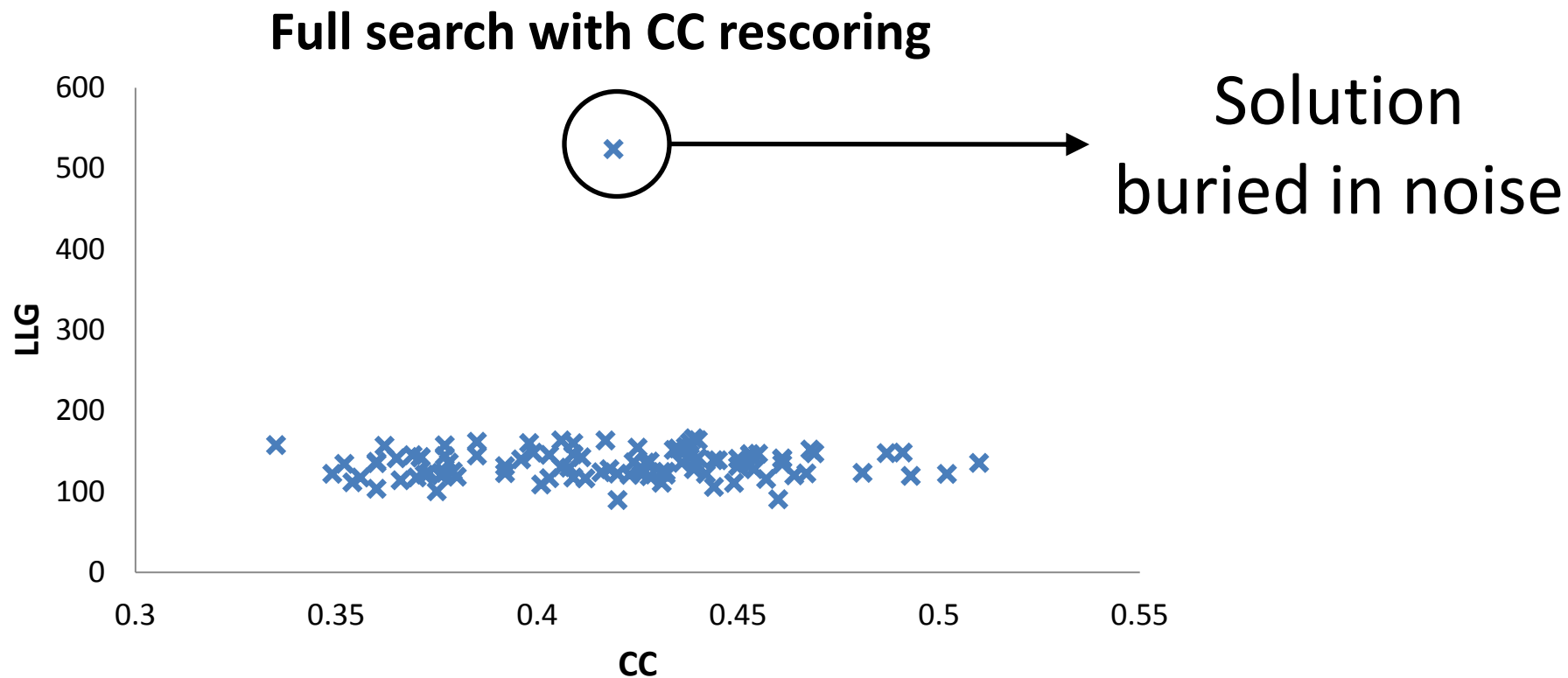
Completion:
improved
substructure

Lower CC for non-solutions

Refinement: reduced overfitting



Fibronectin test case

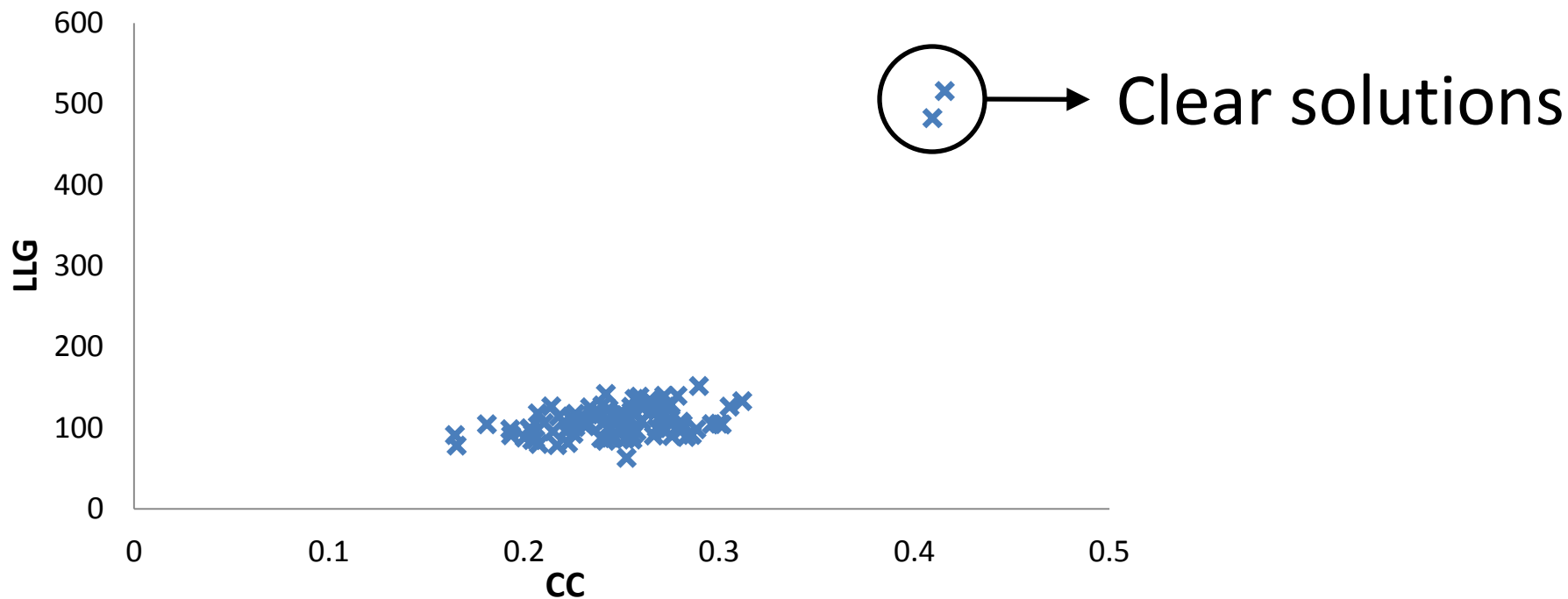


Rudiño-Piñera E, Ravelli RB, Sheldrick GM, Nanao MH, Korostelev VV,
Werner JM, Schwarz-Linek U, Potts JR, Garman EF
J Mol Biol. **368**, 833-44



Fibronectin test case

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