Phaser: Experimental phasing

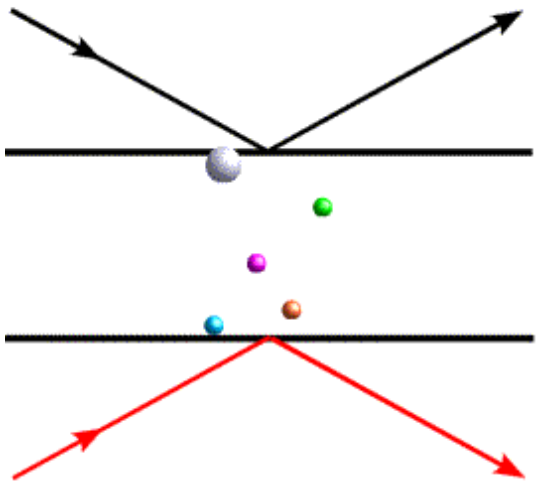
Using SAD data in *Phaser*

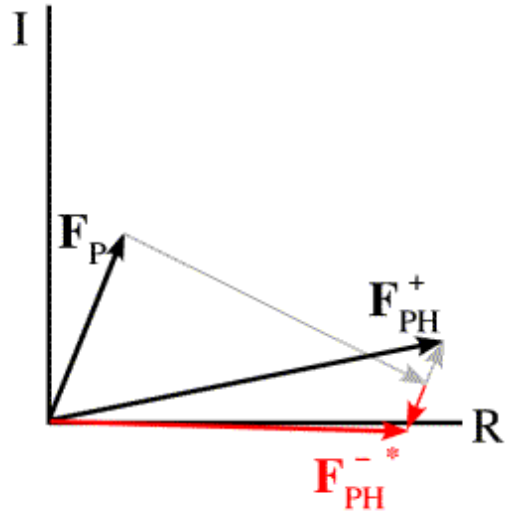
UNIVERSITY OF CAMBRIDGE

R J Read, Department of Haematology
Cambridge Institute for Medical Research

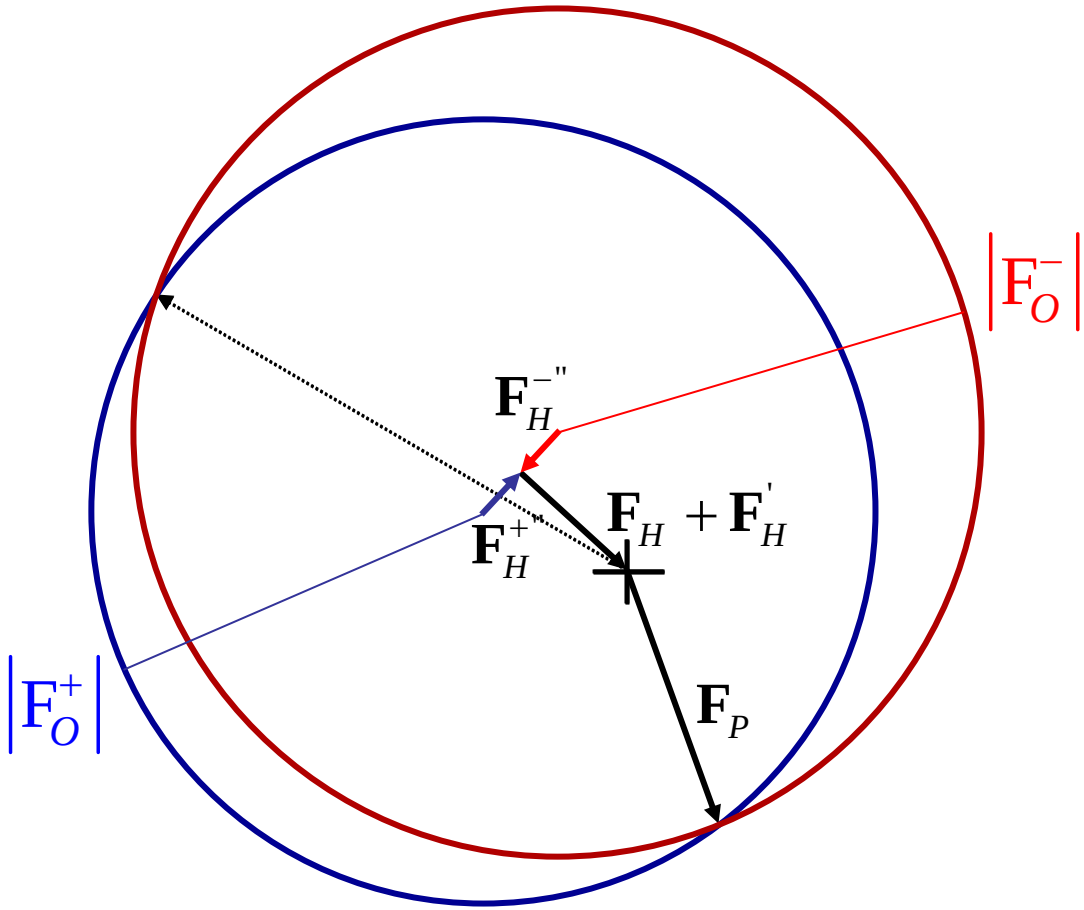
Diffraction with anomalous scatterers

- SAD: single-wavelength anomalous diffraction

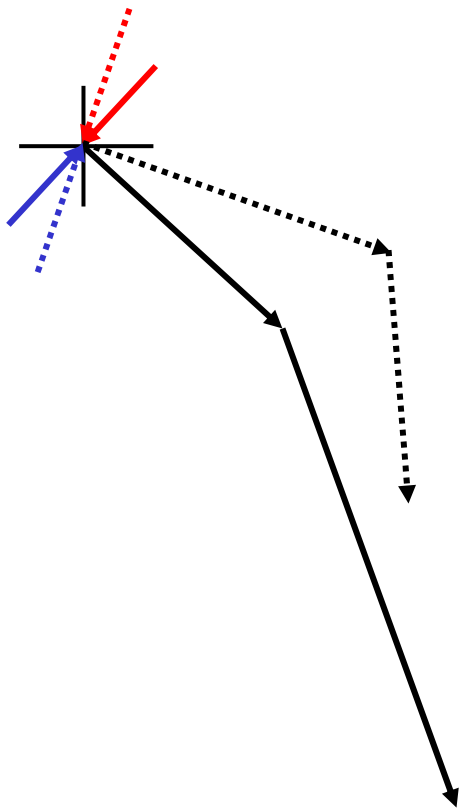


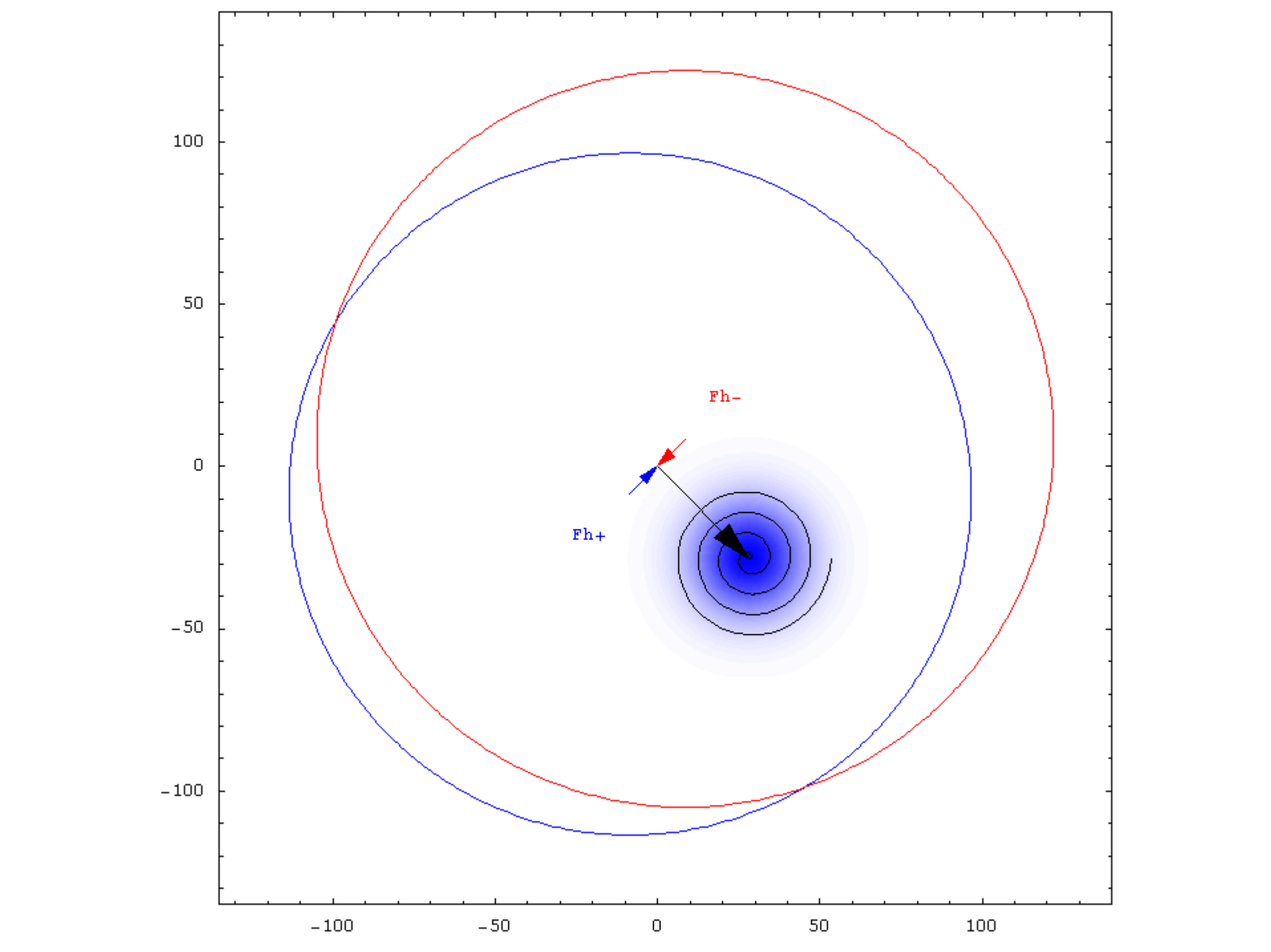


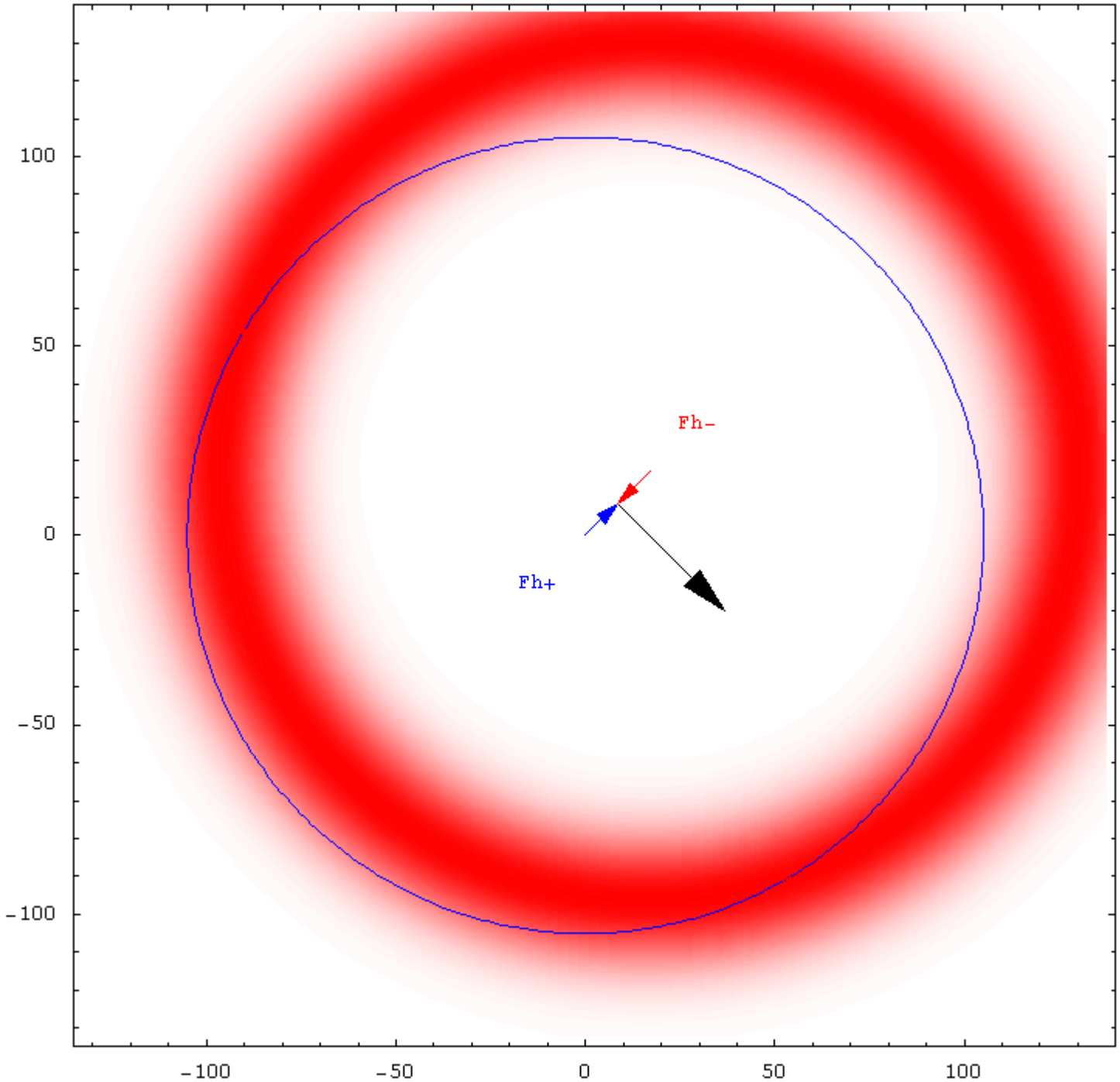
Harker construction for SAD phasing



Real *vs.* calculated SAD contributions




100
50
0
-50
-100
-100
-50
0
50
100
Fh-
Fh+


100
50
0
-50
-100
-100
-50
0
50
100
Fh-
Fh+

Principle of maximum likelihood

- How consistent is the model with the data?
- *What is the probability that the data would be measured if the model were correct?*

$$L = \mathrm{p}(data; model)$$

- Optimise model by adjusting parameters in probability distribution

SAD likelihood function

- Based on probability of F^+ and F^- given model

$$p\left(\mathbf{F}_O^+, \mathbf{F}_O^-, \mathbf{H}^+, \mathbf{H}^-\right) \rightarrow p\left(\mathbf{F}_O^+, \mathbf{F}_O^-; \mathbf{H}^+, \mathbf{H}^-\right)$$

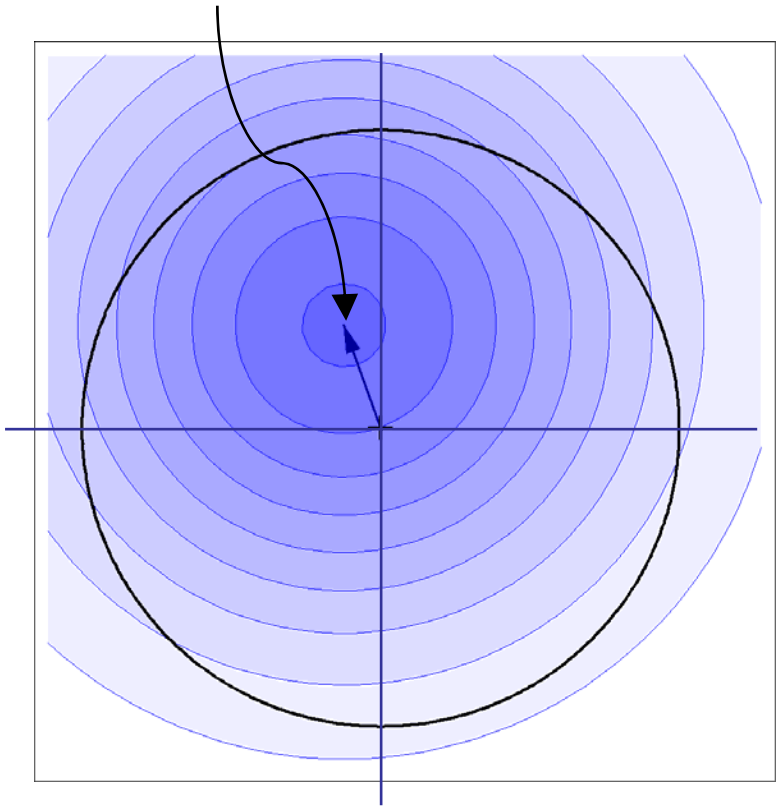
- Factor joint probability into two parts

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

- Integrate out unknown phases, α^+ and α^-

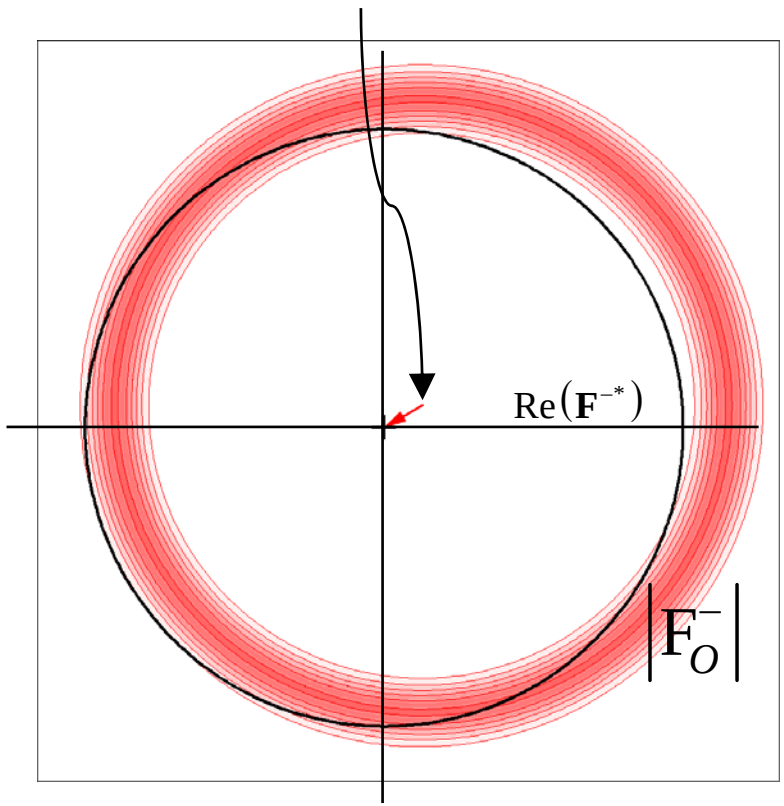
Intuitive understanding of SAD phasing

Expected value of $\mathbf{F}^{-*}$ $(\mathbf{H}^{-*})$



$P\left(\mathrm{F}_O^-, \alpha_O^-; \mathbf{H}^{-*}\right)$

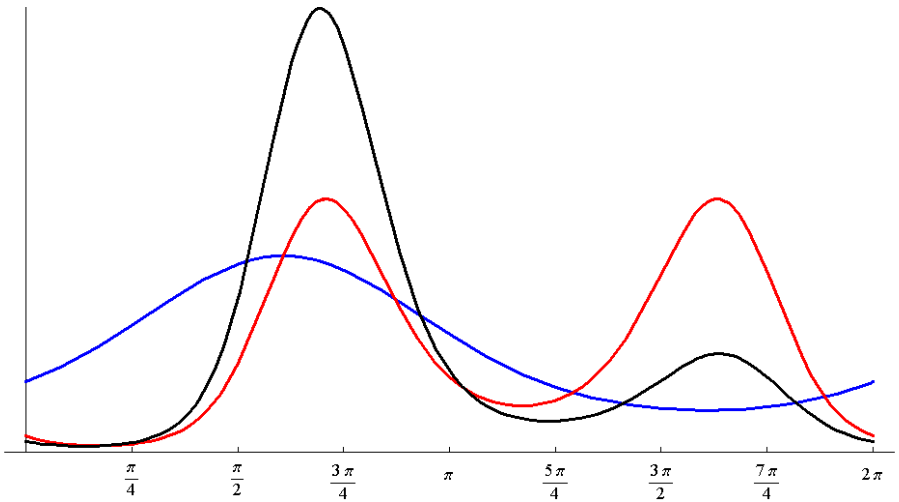
Expected difference between $\mathbf{F}^{+}$ and $\mathbf{F}^{-*}$



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

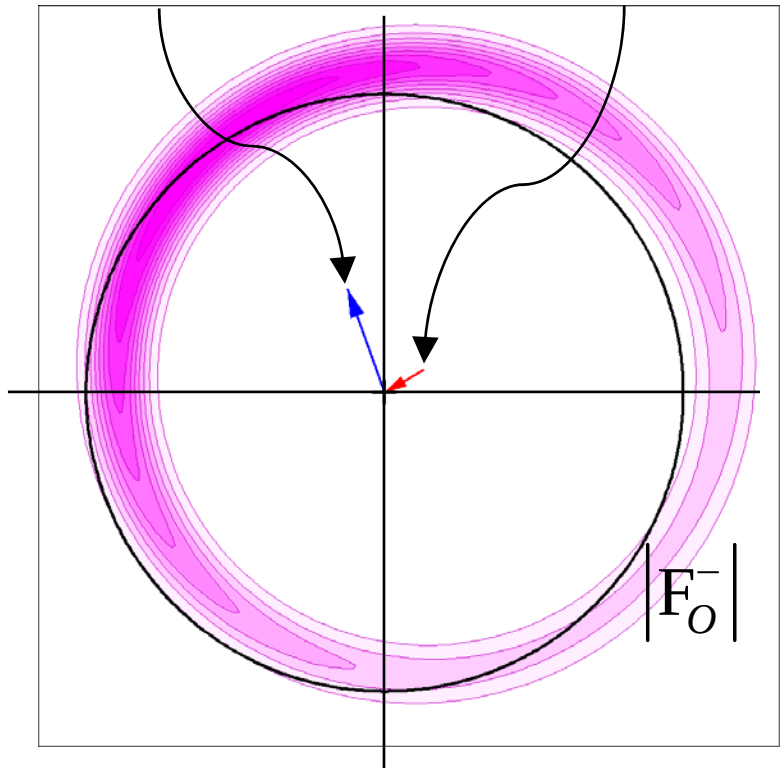
Intuitive understanding of SAD phasing

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



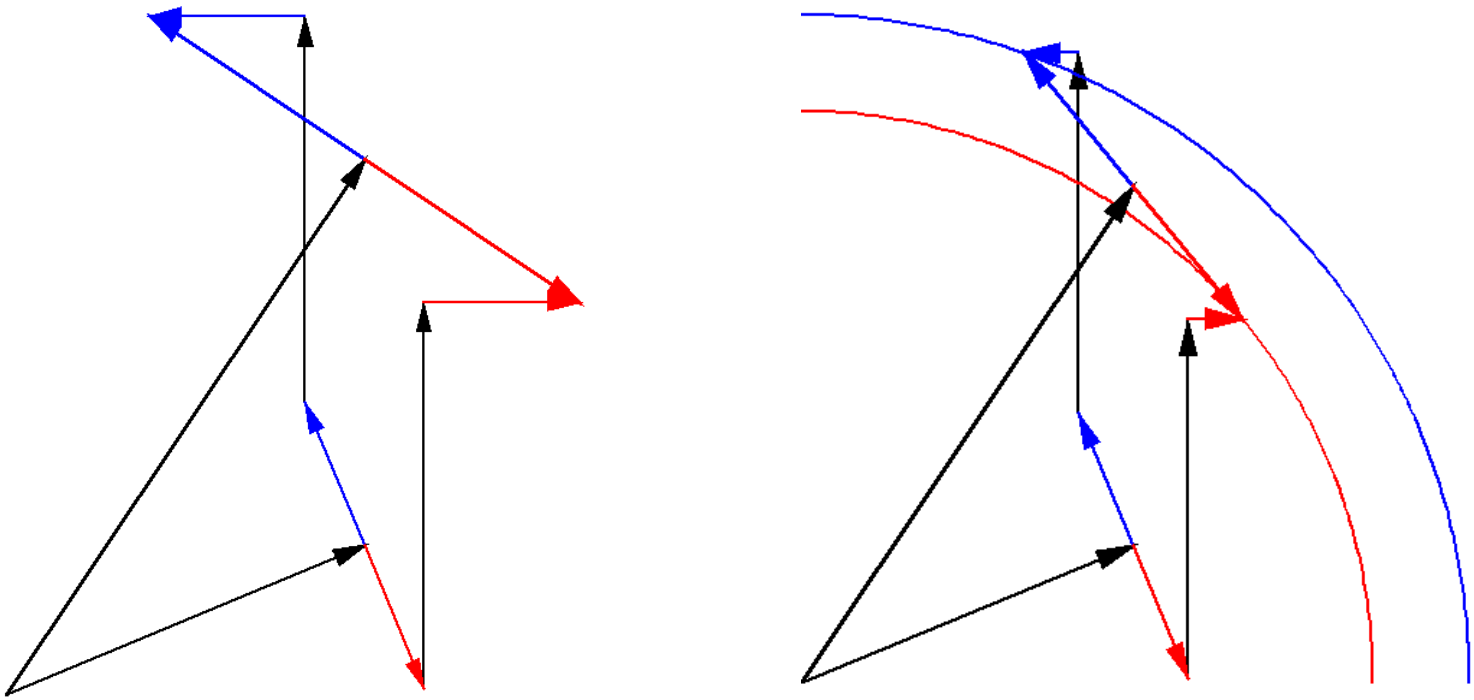
Expected value of $\mathbf{F}^{-*}$ $(\mathbf{H}^{-*})$

Expected difference between $\mathbf{F}^{+}$ and $\mathbf{F}^{-*}$



Breakdown of Friedel's law

- Friedel's law breaks down for mixture of scatterers differing in real:anomalous ratio

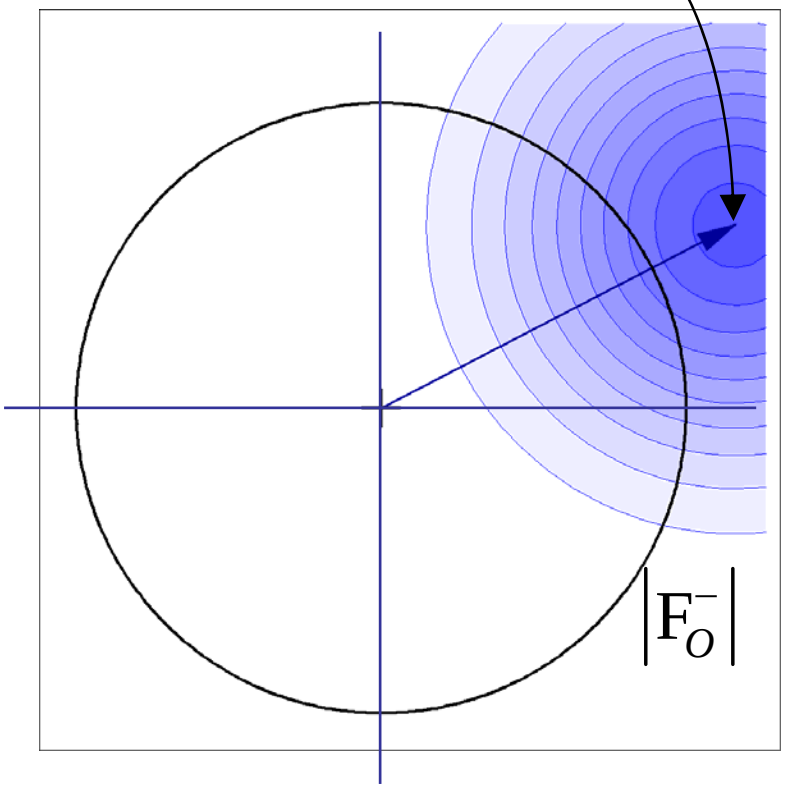


SAD log-likelihood gradient (LLG) map

- Compute derivative of log-likelihood with respect to heavy atom structure factor
- Fourier transform gives map of where likelihood target would like to see changes in anomalous scatterer model
- Very sensitive to minor sites
  - picks up sites identified as water molecules in refined structures determined by halide soaks
- http://www-structmed.cimr.cam.ac.uk/phaser/tutorial
  - tutorial with data for lysozyme iodide soak

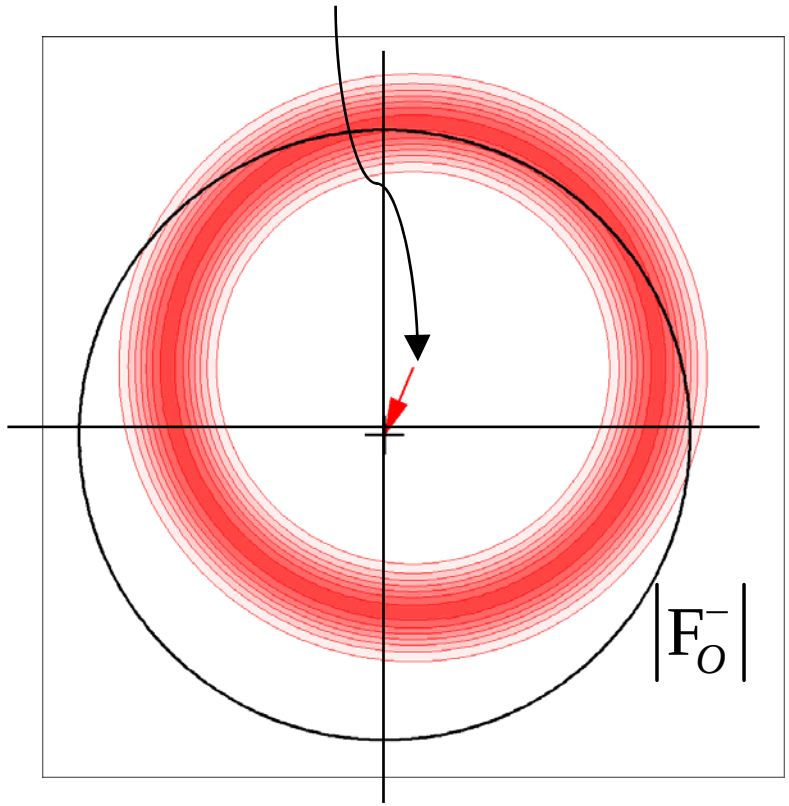
SAD with partial model

Expected value of $\mathbf{F}^{-*}\left(D\mathbf{F}_C^{-*}\right)$



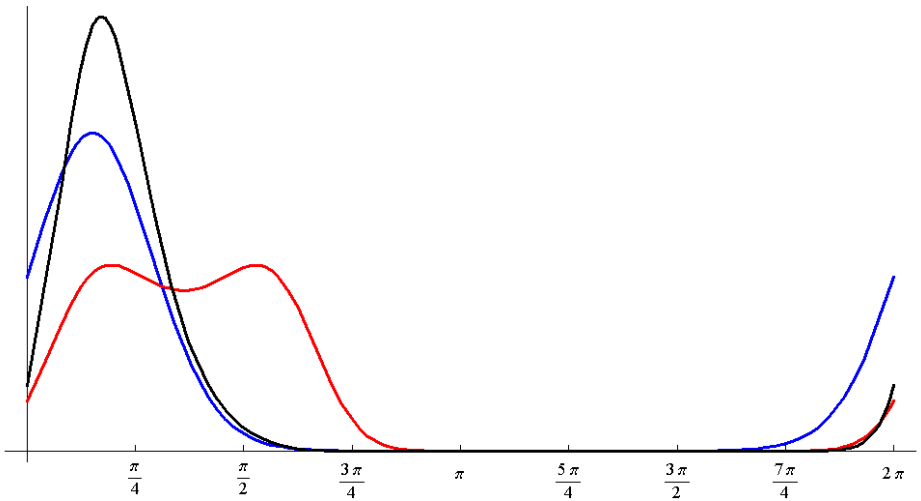
$P\left(F_O^-, \alpha_O^- \mid \mathbf{F}_C^{-*}\right)$

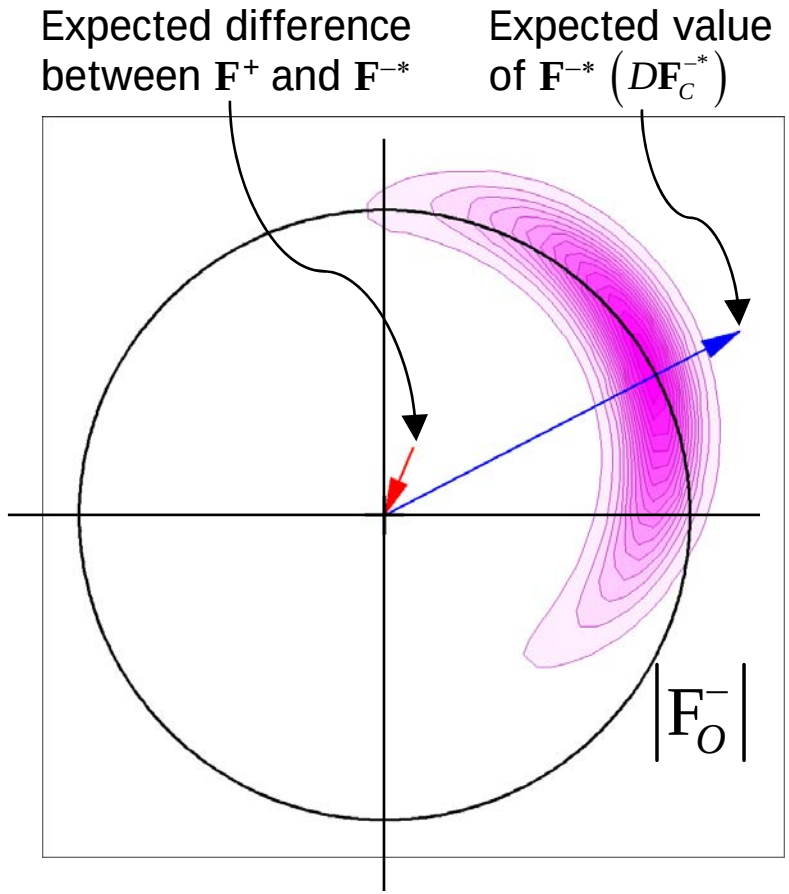
Expected difference between $\mathbf{F}^{+}$ and $\mathbf{F}^{-*}$



$P\left(F_O^+ \mid F_O^-, \mathbf{F}_C^+, \mathbf{F}_C^{-*}\right)$

SAD with partial model





Combining MR and SAD

- CuKα data to 1.9Å on hen egg-white lysozyme
  - can't find sulfurs with HySS or SHELXD
- Solve by MR with goat alpha-lactalbumin (40% identical)
- Use MR model as "substructure" for SAD
  - look for S atoms in LLG map (finds all 10 S, 5-9 Cl^-)
  - phases automatically combine MR and SAD
- Automated fitting with density-modified map
- http://www-structmed.cimr.cam.ac.uk/phaser/tutorial
  - tutorial with these data

Automation of SAD phasing

- Functions are all available from Python
  - used for SAD in AutoSolve wizard
  - can run from HAPPy (CCP4)
- Log-likelihood-gradient completion
  - look for one or several types of scatterer
    - start from MR model or partial substructure
  - analyse map to add sites, make atoms anisotropic
  - delete atoms that fade away
  - change atom type if occupancy far from one
  - repeat to convergence

Iterative model-building with SAD

- Nitrate reductase structure (Natalie Strynadka)
  - integral membrane protein, 1976 residues
  - contains 21 Fe atoms, 1 Mo, 118 S, 5 P (146 total)
  - solved using combination of Fe-MAD, MIRAS
- Fe peak SAD data only
  - find 11 “Fe” sites with phenix.hyss
    - several are super-sites of Fe_4S_4 clusters
  - phase and complete adding Fe, Mo, S with *Phaser*
    - total of 57 sites: 20 Fe, 6 Mo, 31 S
    - superatoms are resolved, 51 of 57 are identified correctly
    - correct hand indicated by number of sites, LLG score

Iterative model-building and phasing

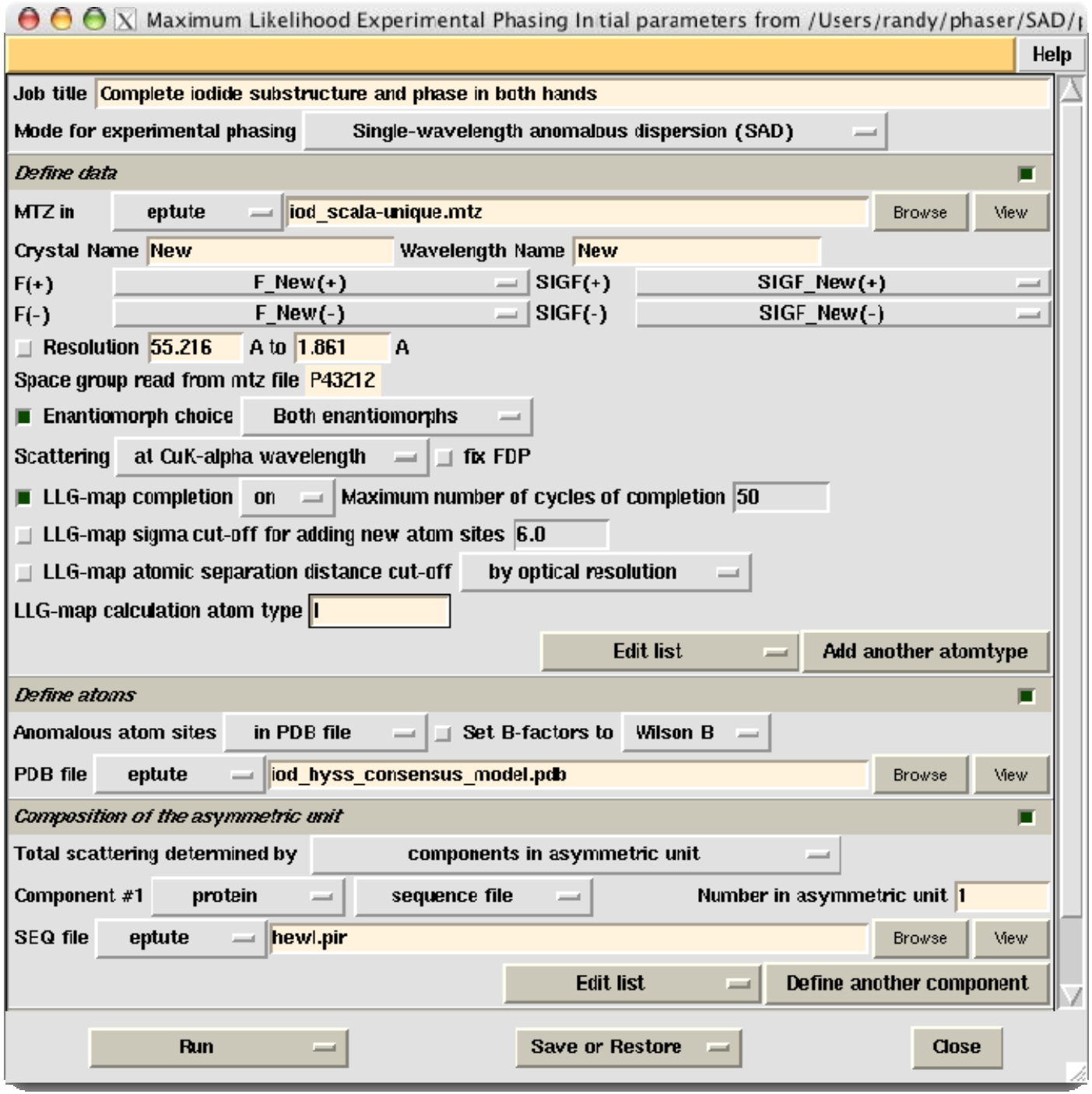
- Improve phases by density modification
- Build with ARP/wARP (or Resolve)
  - 1607 residues, 1368 docked in sequence
- LLG completion from ARP/wARP model
  - 105 sites, 92 correctly identified
- Repeat DM and ARP/wARP
  - 1813 residues, 1775 docked in sequence

Practical aspects of SAD phasing in *Phaser*

- Start with substructure or protein model
- Provide information about cell content
  - sequence, molecular weight, percent solvent...
  - used to put data on absolute scale
    - occupancies are reasonably accurate
- Provide information about f” values
  - wavelength (table lookup) or measured
  - refined by default if close to peak wavelength
- Try both hands if uncertain
  - separate completion if mixture of atom types

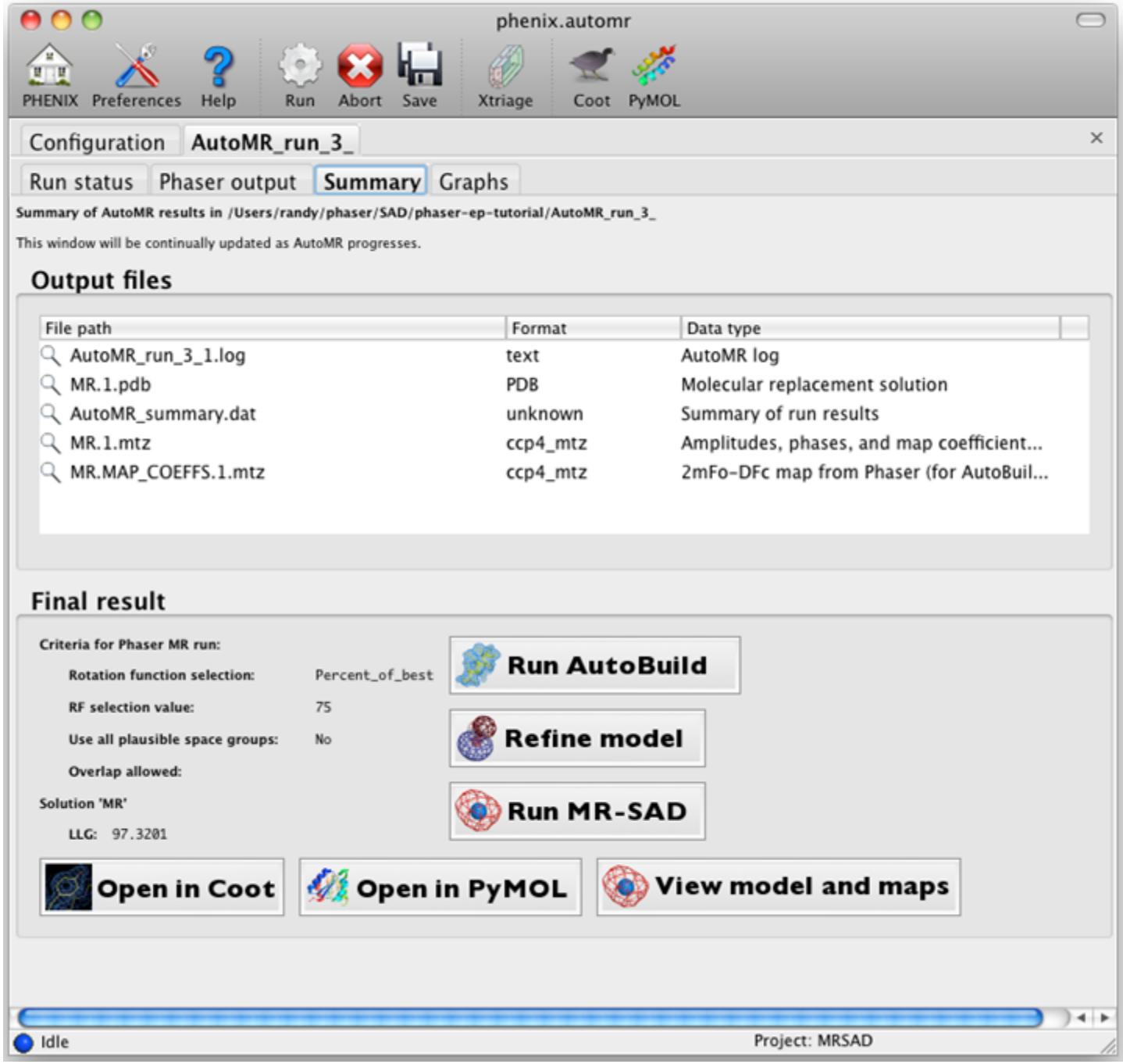
SAD phasing in CCP4

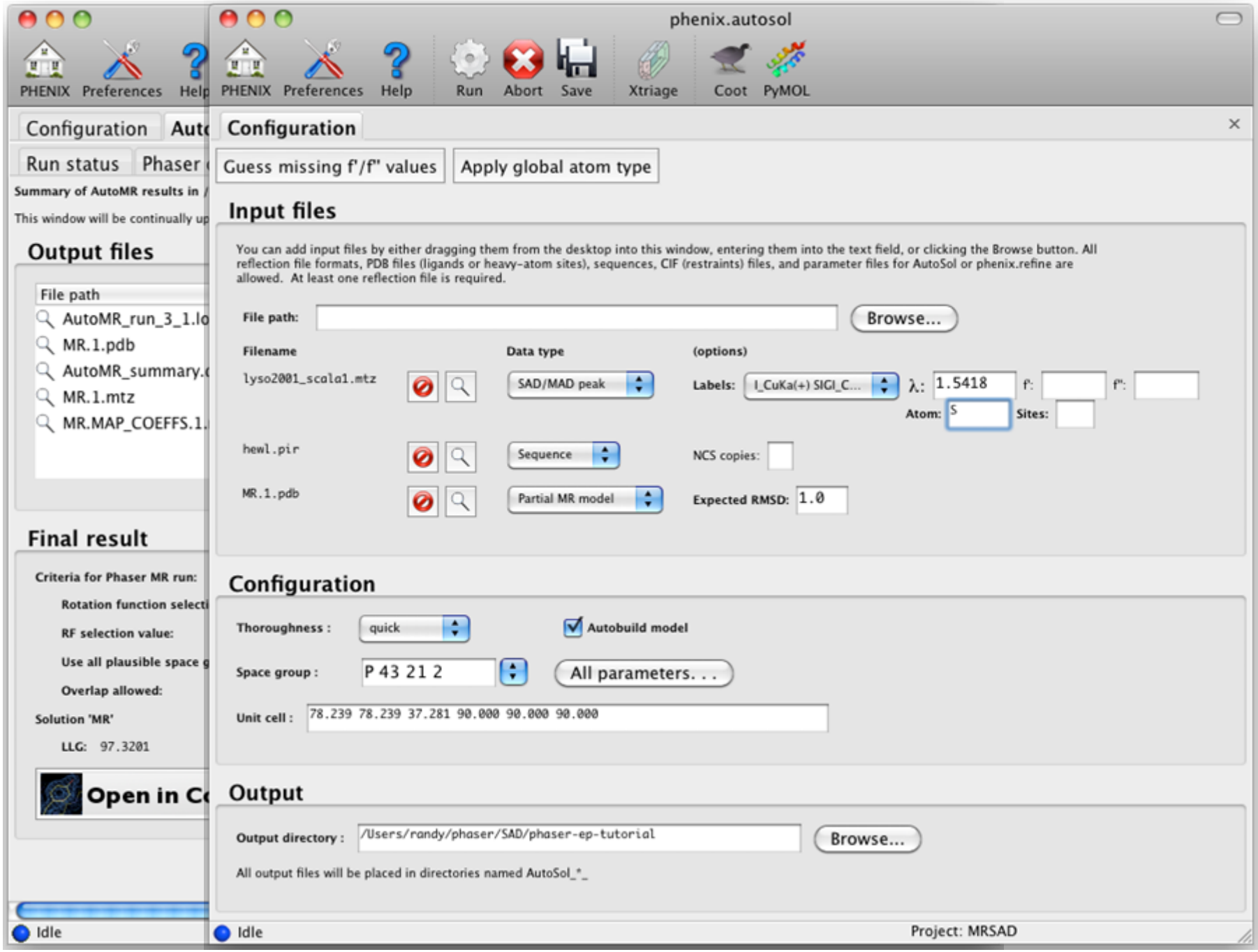
- ccp4i interface has *Phaser* SAD phasing module
- Two modes:
  - “Single-wavelength anomalous dispersion (SAD)”
    - start from substructure of anomalous scatterers
    - can test both hands, complete with multiple scatterers
  - “SAD with molecular replacement partial structure”
    - start from substructure of non-anomalous scatterers
    - optionally include known anomalous scatterers


Maximum Likelihood Experimental Phasing Initial parameters from /Users/randy/phaser/SAD/
Help
Job title
Complete iodide substructure and phase in both hands
Mode for experimental phasing
Single-wavelength anomalous dispersion (SAD)
Define data
MTZ in
eptute
iod_scala-unique.mtz
Browse
View
Crystal Name
New
Wavelength Name
New
F(+)
F_New(+)
SIGF(+)
SIGF_New(+)
F(-)
F_New(-)
SIGF(-)
SIGF_New(-)
Resolution
55.216
A to
1.861
A
Space group read from mtz file
P43212
Enantiomorph choice
Both enantiomorphs
Scattering
at CuK-alpha wavelength
fix FDP
LLG-map completion
on
Maximum number of cycles of completion
50
LLG-map sigma cut-off for adding new atom sites
6.0
LLG-map atomic separation distance cut-off
by optical resolution
LLG-map calculation atom type
I
Edit list
Add another atomtype
Define atoms
Anomalous atom sites
in PDB file
Set B-factors to
Wilson B
PDB file
eptute
iod_hyss_consensus_model.pdb
Browse
View
Composition of the asymmetric unit
Total scattering determined by
components in asymmetric unit
Component #1
protein
sequence file
Number in asymmetric unit
1
SEQ file
eptute
hewl.pir
Browse
View
Edit list
Define another component
Run
Save or Restore
Close

SAD phasing in Phenix

- Use AutoSol wizard
  - launch directly if no prior model
    - finds sites with Hyss
    - tests both hands, chooses best hand
    - carries out Resolve density modification and model-building
  - launch from AutoMR wizard after MR


phenix.automr
PHENIX
Preferences
Help
Run
Abort
Save
Xtriage
Coot
PyMOL
Configuration
AutoMR_run_3_
Run status
Phaser output
Summary
Graphs
Summary of AutoMR results in /Users/randy/phaser/SAD/phaser-ep-tutorial/AutoMR_run_3_
This window will be continually updated as AutoMR progresses.
Output files
File path
Format
Data type
AutoMR_run_3_1.log
text
AutoMR log
MR.1.pdb
PDB
Molecular replacement solution
AutoMR_summary.dat
unknown
Summary of run results
MR.1.mtz
ccp4_mtz
Amplitudes, phases, and map coefficient...
MR.MAP_COEFFS.1.mtz
ccp4_mtz
2mFo-DFc map from Phaser (for AutoBuil...
Final result
Criteria for Phaser MR run:
Rotation function selection:
Percent_of_best
RF selection value:
75
Use all plausible space groups:
No
Overlap allowed:
Solution 'MR'
LLG:
97.3201
Run AutoBuild
Refine model
Run MR-SAD
Open in Coot
Open in PyMOL
View model and maps
Idle
Project: MRSAD


phenix.autosol
PHENIX
Preferences
Help
Run
Abort
Save
Xtriage
Coot
PyMOL
Configuration
Run status
Summary of AutoMR results in /
This window will be continually up
Output files
File path
AutoMR_run_3_1.lo
MR.1.pdb
AutoMR_summary.
MR.1.mtz
MR.MAP_COEFFS.1.
Final result
Criteria for Phaser MR run:
Rotation function selecti
RF selection value:
Use all plausible space g
Overlap allowed:
Solution 'MR'
LLG: 97.3201
Open in C
Idle
Configuration
Guess missing f'/f" values
Apply global atom type
Input files
You can add input files by either dragging them from the desktop into this window, entering them into the text field, or clicking the Browse button. All reflection file formats, PDB files (ligands or heavy-atom sites), sequences, CIF (restraints) files, and parameter files for AutoSol or phenix.refine are allowed. At least one reflection file is required.
File path:
Browse...
Filename
Data type
(options)
lyso2001_scala1.mtz
SAD/MAD peak
Labels:
I_CuKa(+) SIGI_C...
λ: 1.5418
f':
f":
Atom: S
Sites:
hewl.pir
Sequence
NCS copies:
MR.1.pdb
Partial MR model
Expected RMSD: 1.0
Configuration
Thoroughness :
quick
Autobuild model
Space group :
P 43 21 2
All parameters. . .
Unit cell :
78.239 78.239 37.281 90.000 90.000 90.000
Output
Output directory :
/Users/randy/phaser/SAD/phaser-ep-tutorial
Browse...
All output files will be placed in directories named AutoSol_*_
Idle
Project: MRSAD

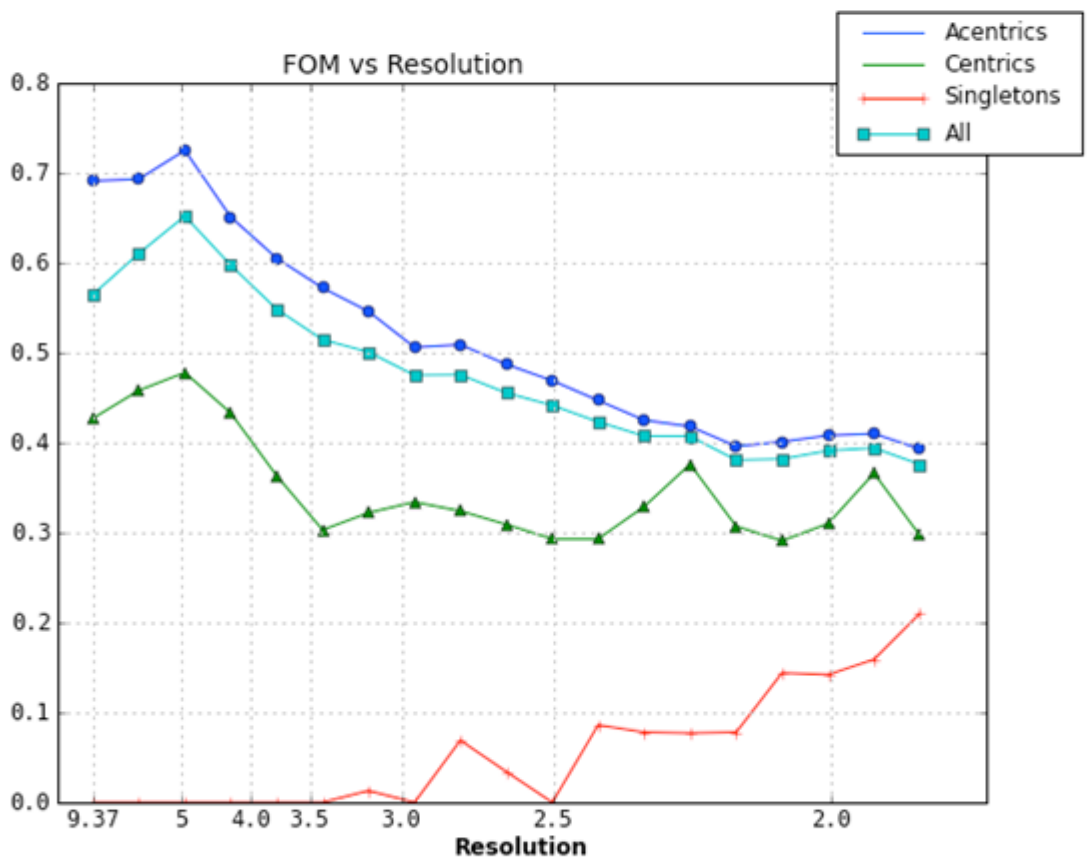
phenix.autosol

PHENIX Preferences Help Run Abort Save Xtriage Coot PyMOL

Configuration AutoSol_run_4_

Run status Phaser output Graphs

Figures of Merit



Running 1 job. . .

Project: MRSAD

Configuration Auto

Run status Phaser

Summary of AutoMR results in /

This window will be continually up

Output files

File path

AutoMR_run_3_1.lo

MR.1.pdb

AutoMR_summary.

MR.1.mtz

MR.MAP_COEFFS.1.

Final result

Criteria for Phaser MR run:

Rotation function selecti

RF selection value:

Use all plausible space g

Overlap allowed:

Solution 'MR'

LLG: 97.3201

Open in C

Idle

Background information

- "*Phaser* crystallographic software", McCoy, Grosse-Kunstleve, Adams, Winn, Storoni & Read (2007), *J. Appl. Cryst.* **40**, 658-674.
  - plus papers cited here
- "Liking likelihood", Airlie J. McCoy (2004), *Acta Cryst. D***60**, 2169-2183.

- http://www-structmed.cimr.cam.ac.uk/phaser
- http://www-structmed.cimr.cam.ac.uk/Course

Acknowledgments

- Molecular replacement
  - Airlie McCoy, Laurent Storoni, Gabor Bunkoczi, Rob Oeffner
- Experimental phasing
  - Airlie McCoy, Laurent Storoni
- PHENIX collaboration
  - Ralf Grosse-Kunstleve, Nigel Moriarty, Paul Adams
  - Nat Echols, Tom Terwilliger

wellcome trust

NATIONAL INSTITUTES OF HEALTH

CCP4