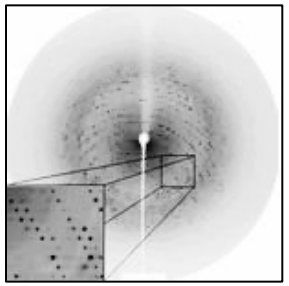
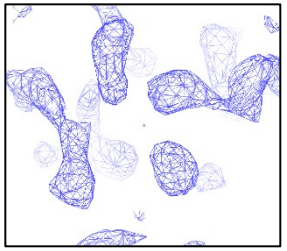
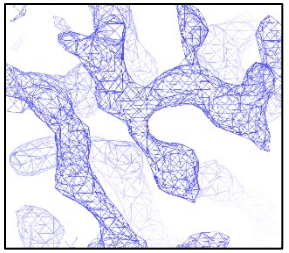
Experimental phasing in Crank2

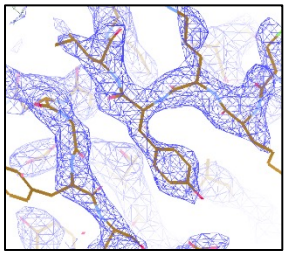
Pavol Skubak and Navraj Pannu
Leiden University Medical Center,
The Netherlands

X-ray structure solution pipeline









Data collection

Data processing

Experimental phasing

Molecular Replacement

Density Modification

Model building

Refinement

Rebuilding Validation

Crank2 for experimental phasing

- Crank2 is suitable for SAD, MR-SAD, MAD and SIRAS.

- Crank2 has been shown to be particularly powerful weak anomalous signal to low resolution SAD datasets (4.5 Angstroms)

Single isomorphous replacement

A “native” data set

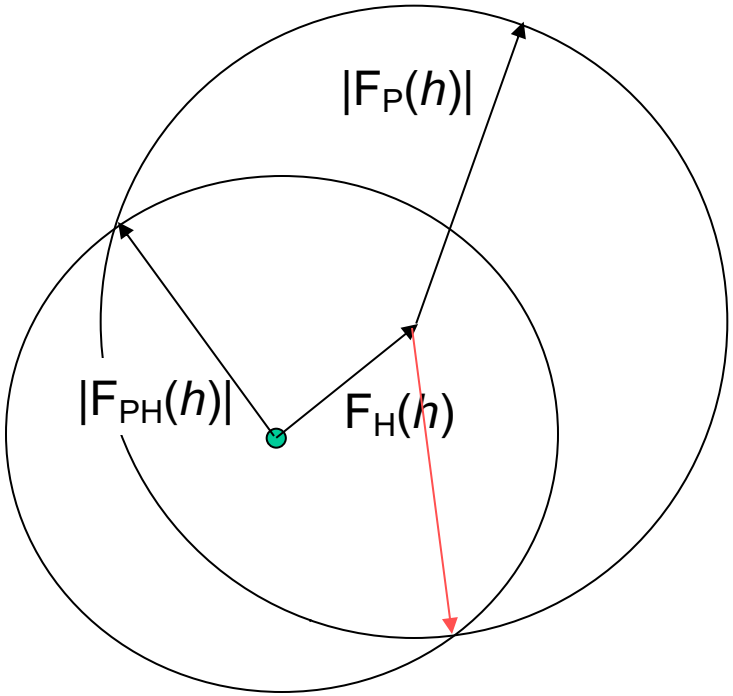
from a crystal that just contains your macromolecule.

A “derivative” data set

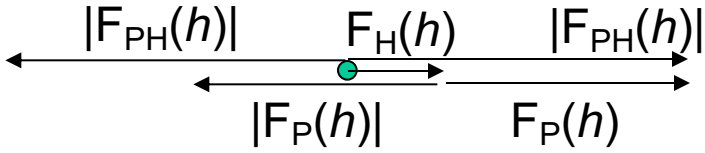
A macromolecular crystal that has been soaked with a heavy atom.

Phase information from (error-free) isomorphous replacement

Acentric
Two solutions



Centric
Single solution



$|F_{PH}(h)|$: “derivative” observed amplitude
$F_H(h)$: heavy atom structure factor
$F_{PH}(h) = F_H(h) + F_P(h)$ (vector addition)

What are some of the errors in isomorphous replacement?

Sometimes heavy atoms induce changes in the unit cell or changes in the macromolecule (non-isomorphism).

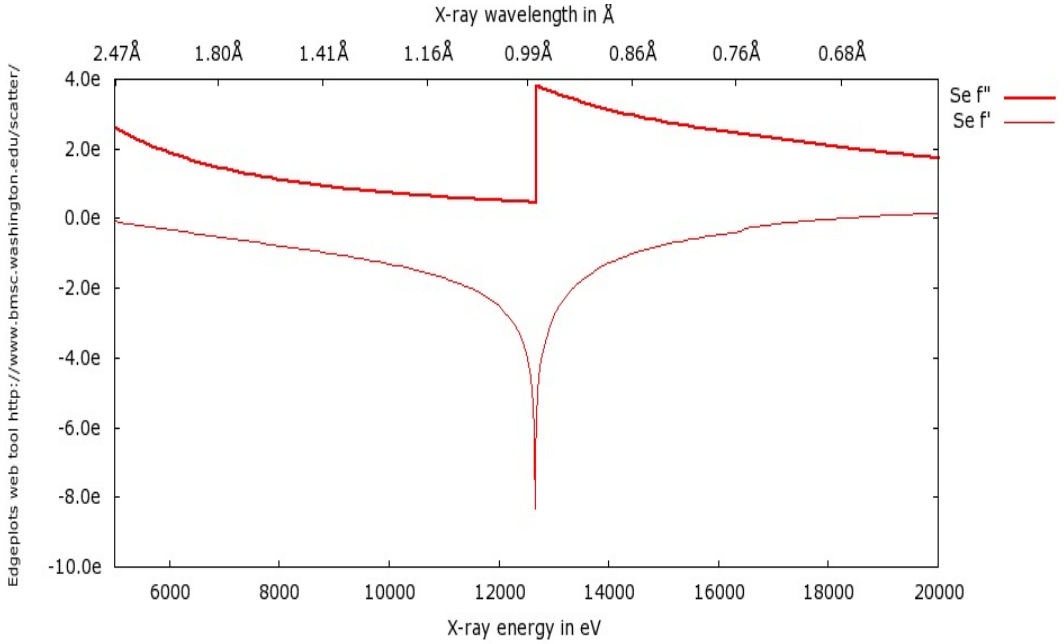
The scale between native and derivative data sets.

Errors in the heavy atom positions.

Errors in the measurement of the amplitudes.

But, heavy atoms could be easier than growing Se-met protein!

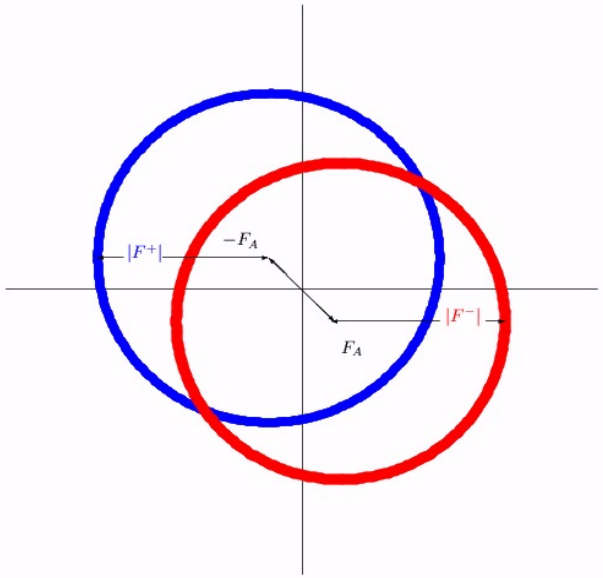
Effect of changing the wavelength: Anomalous scattering



$$\boldsymbol{F^{+}(h)} = \sum (f_j + f_j \; + i\, f_j'') \exp(2\pi i\; h \cdot x_j)$$

$$\boldsymbol{F^{-}(h)} = \sum (f_j \; f'_j - i\, f_j'') \exp(2\pi i\; h \cdot x_j)$$

Single-wavelength Anomalous Diffraction



$F_A = \sum_j f''_j \exp(2\pi i h \cdot x_j)$

$F = F^+ - F_A$, $F = F^- + F_A$

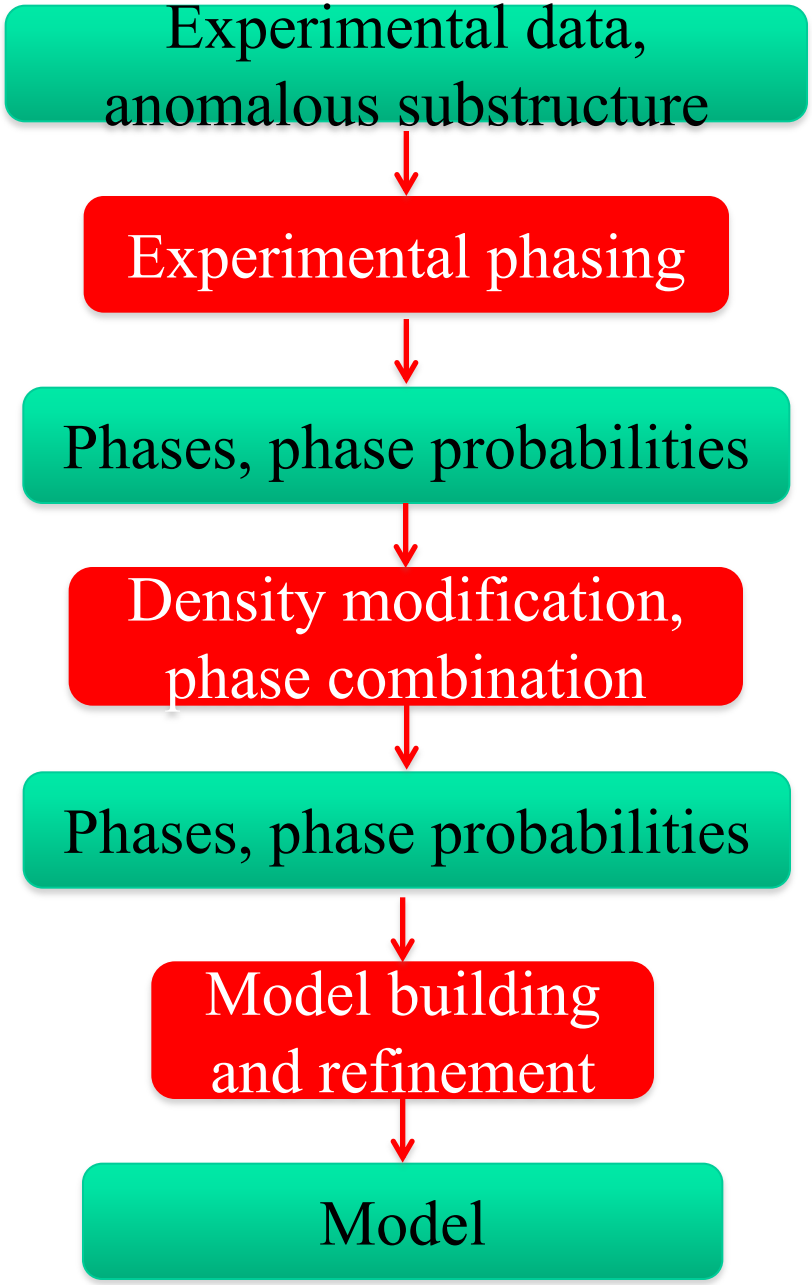
Solving structures using Friedel pairs collected at one wavelength from a crystal that contains an anomalous substructure.

Simultaneously combining experimental phasing steps to improve structure solution

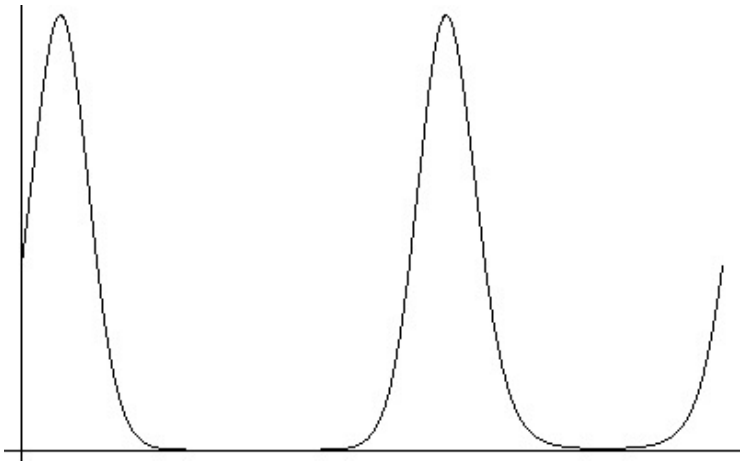
- Traditionally structure solution is divided into distinct steps:
  - Substructure detection
  - Obtain initial phases
  - (Density) modify the initial experimental map
  - Build and refine the model
- By combining these steps, we can improve the process.

Traditional structure solution

- Step-wise
- Information is propagated via 'phase probabilities'



Phase probabilities

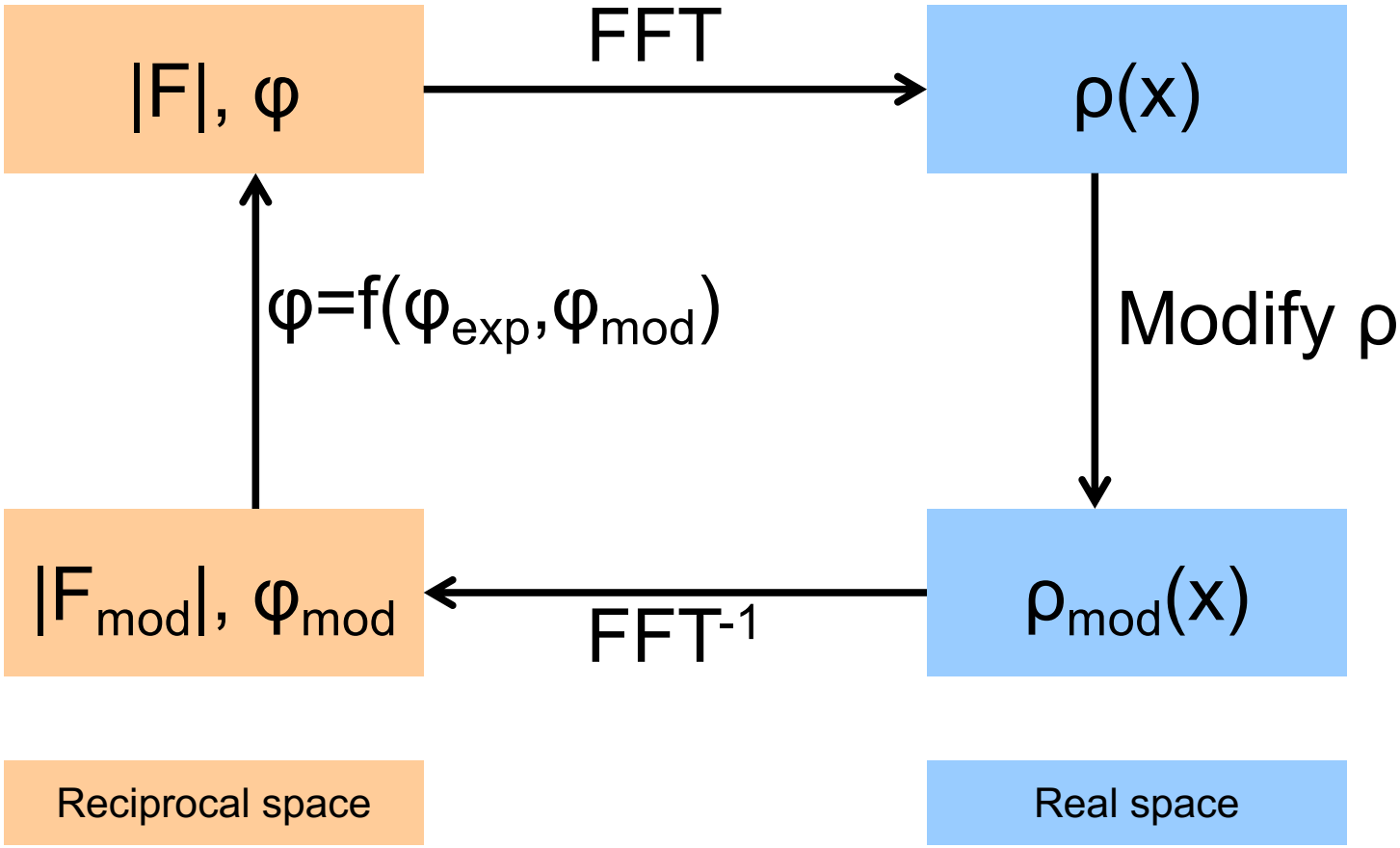


$$P(\alpha) = \exp(A\cos(\alpha) + B\sin(\alpha) + C\cos(2\alpha) + D\sin(2\alpha))$$

- The phase distribution can be approximated via 4 "Hendrickson-Lattmann" coefficients, A, B, C, D.
- We rely on programs to estimate these coefficients.
- 'Even better than the real thing?'

Density modification

2. Phase weighting:



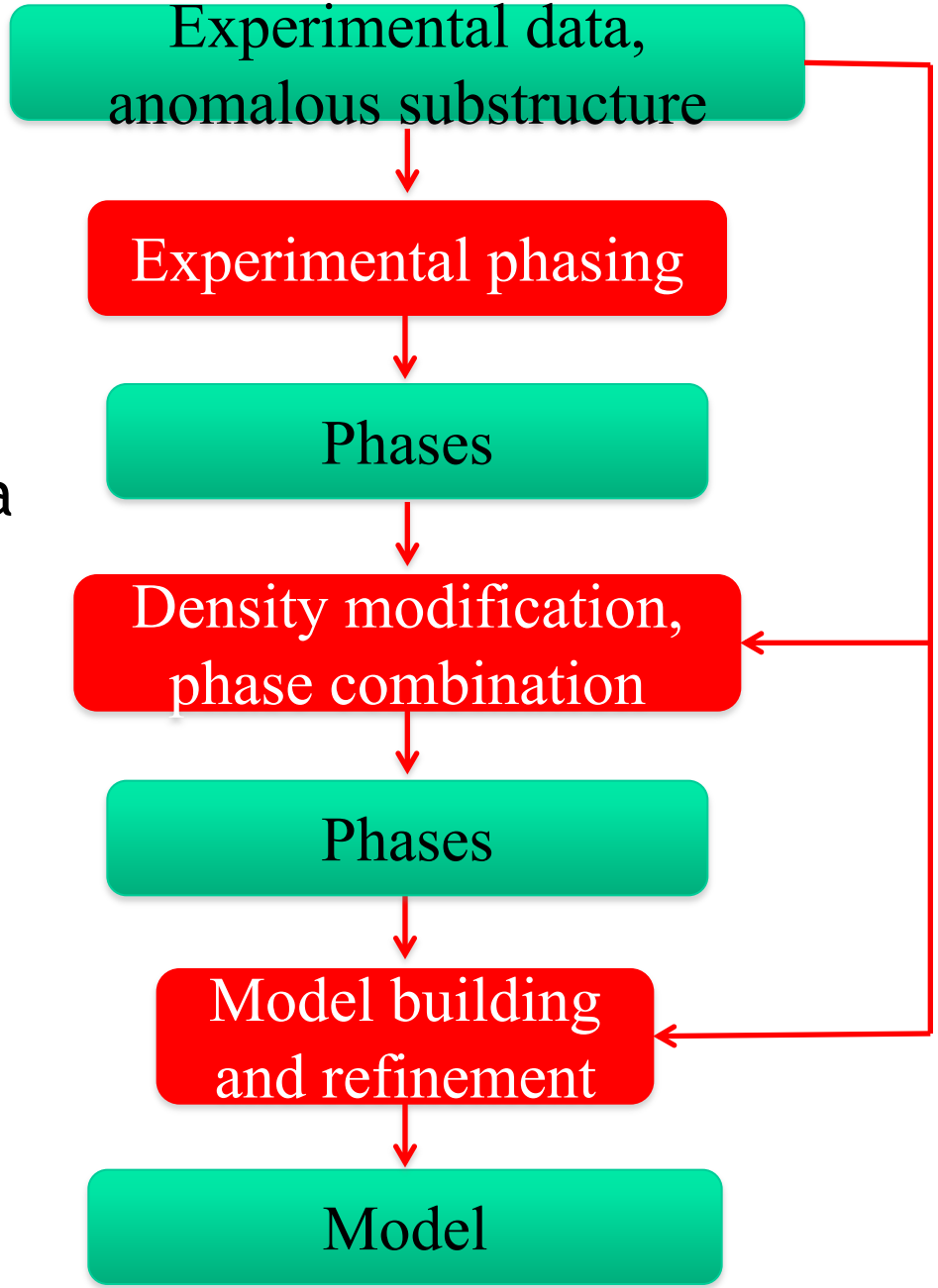
Kevin Cowtan, cowtan@ysbl.york.ac.uk

HL propagation and the independence assumption

- Use the experimental data and anomalous substructure directly!
  - Do not need to assume independence or rely on HL coefficients.
  - Need multivariate distributions at each step that take into account correlations between the model and data.

Step-wise multivariate structure solution

- Still step-wise
- Information is propagated via the data and model(s).

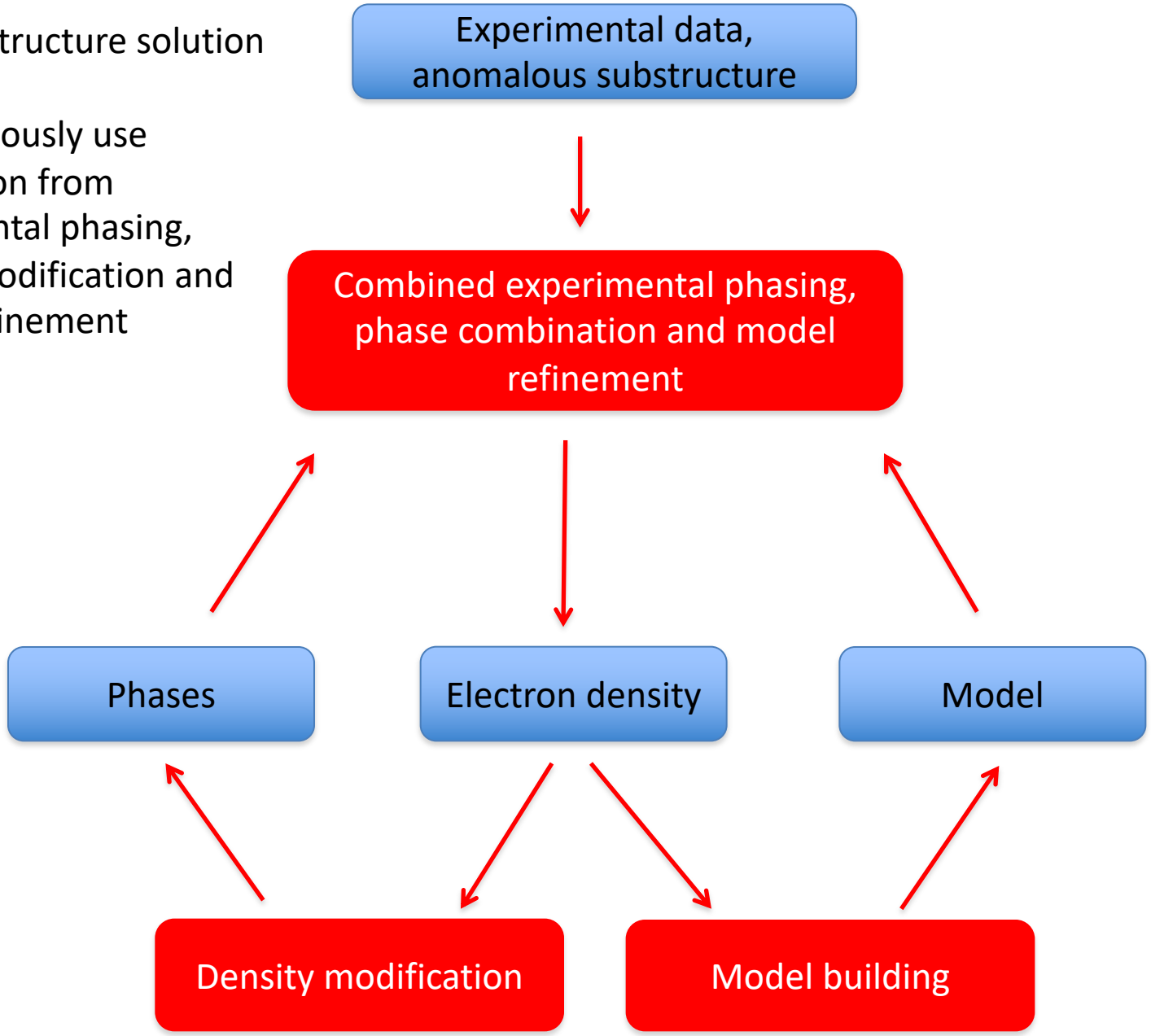


What are step-wise multivariate functions?

- The multivariate functions consider all the correlations between the data and the model together.
- Earlier methods neglected correlations and took into account, for example, “DANO” (DeltaF) instead of *F*+ and *F*-.
- By taking into account of *F*+ and *F*-, we can explicitly consider the measurement error we determine.

Combined structure solution

- Simultaneously use information from experimental phasing, density modification and model refinement



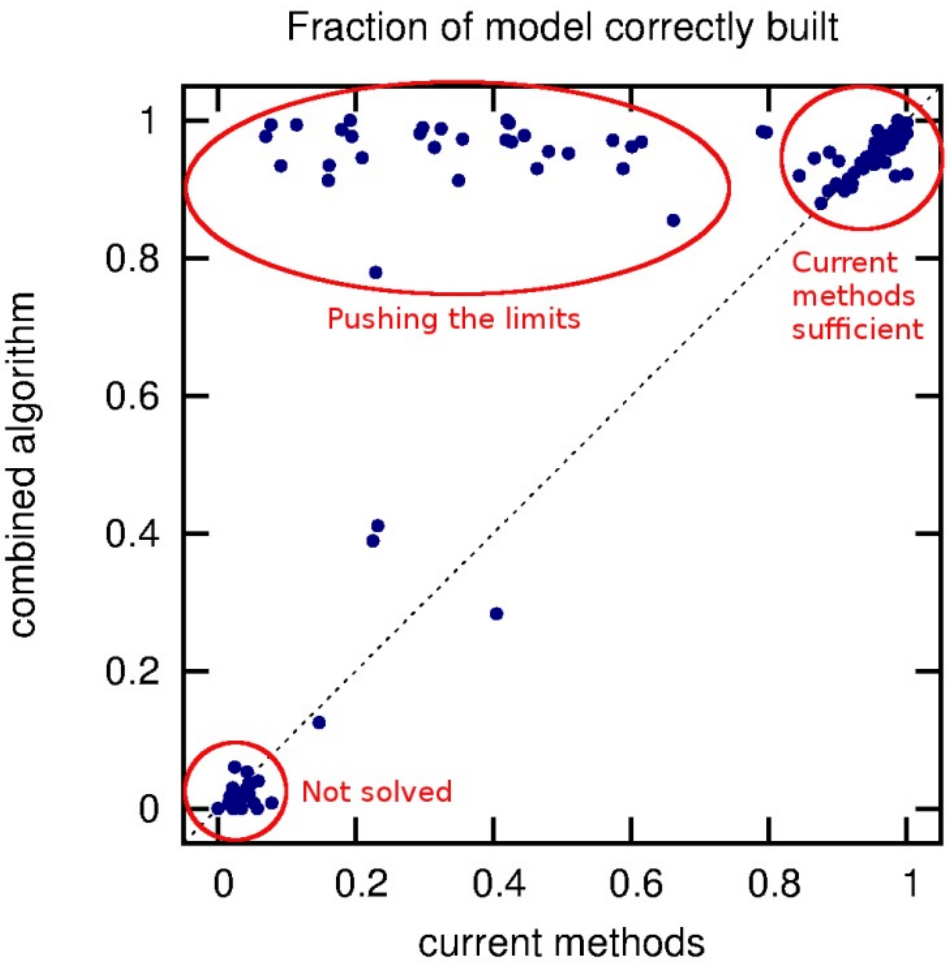
The “combined” multivariate function

- The combined multivariate function consider all the correlations between the data and the substructure, (partially-built) model and density modified phases.

Tests of > 140 real SAD data sets

- Resolution range of data sets is 0.94 to 3.88 Angstroms
- Types of anomalous scatterers: selenium, sulfur, chloride, iodide, bromide, calcium, zinc (and others).
- We compare with the step wise multivariate approach (current CRANK) versus the combined approach.

Model building results on over 140 real SAD data sets (using parrot and buccaneer)



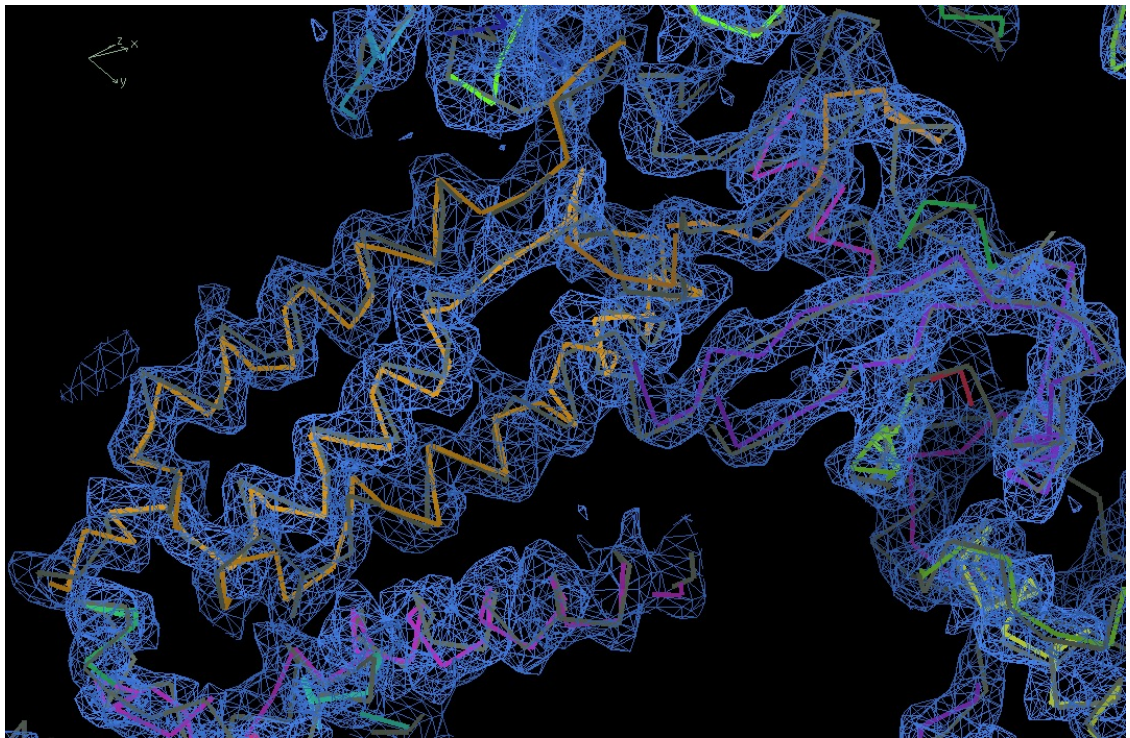
Summary of large scale test

- The average fraction of the model built increased from 60% to 74% with the new approach.

- If we exclude data sets built to 85% by the current approach or where the substructure was not found, 45 data sets remain and the average fraction of the model built increased from 28 to 77%.

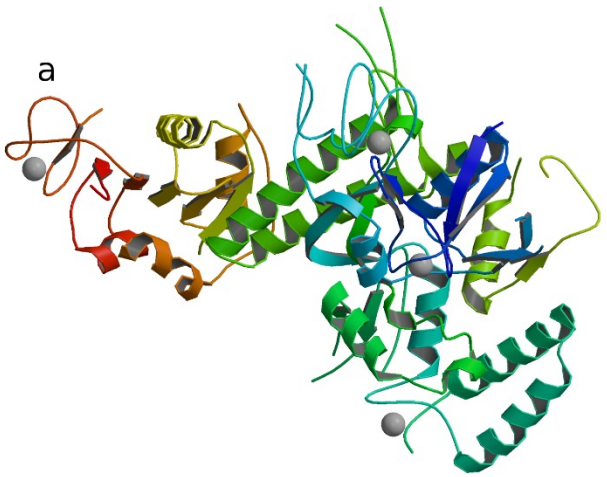
3.88 Angstrom RNA polymerase II

- 3.88 Angstrom SAD data with signal from zinc.
- Authors could not solve the structure with SAD data alone, but with a partial model, multi-crystal MAD and manual building.
- > 80% can be built with SAD data alone with the new algorithm automatically to an R-free of 37.6%



Related RNA polymerases complex from Cramer et al.

- 3.3 Angstrom data with signal from zinc.
- Could not solve the structure with anomalous data alone.
- With the new method, a majority can be built automatically in minutes.



Crank2 availability

- Crank2 is suitable for S/MAD and S/MIRAS experiments and implements multivariate functions.
- Crank2 is available in ccp4 with gui ccp4i2 and ccp4 cloud.
- Crank2 ccp4i2 doc: https://ccp4i2.gitlab.io/rstdocs/tasks/crank2/crank2.html

Important parameters in substructure detection

- The number of cycles run.
- The number of atoms to search for.
  - Should be within 10-20% of actual number
  - A first guess uses a probabilistic Matthew's coefficient
- The resolution cut-off:
  - For MAD, look at signed anomalous difference correlation.
  - For SAD, a first guess is 0.5 + high resolution limit.

FA estimation

- Substructure detection methods need to first estimate FA or substructure factor amplitudes
- Improving the estimates can improve hit rates of substructure detection.
- The simplest estimation is Delta F.

Afro: multivariate SAD equation for FA estimation

- Giacovazzo previously proposed a multivariate FA estimation assuming Friedel pairs are equal.

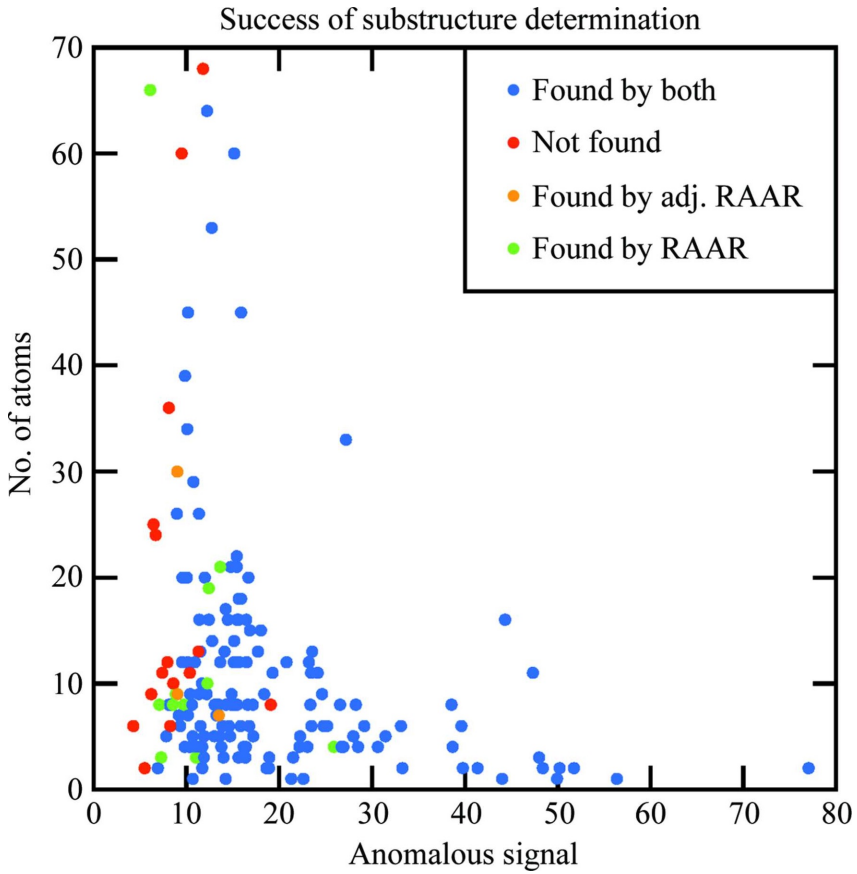
- An equation can be obtained without the equal phase assumption requiring only one numerical integration.

Prasa: A new program for substructure detection

- Phase retrieval for anomalously scattering atoms
- Charge flipping and RAAR (relaxed average alternating reflections)
- Automatic multiple high resolution cutoffs.
- Input of number of substructure atoms not needed.
- Parallelized.
- Released and stable, but not default yet.

Prasa: A new program for substructure detection

- 150 SAD data sets using charge flipping and RAAR



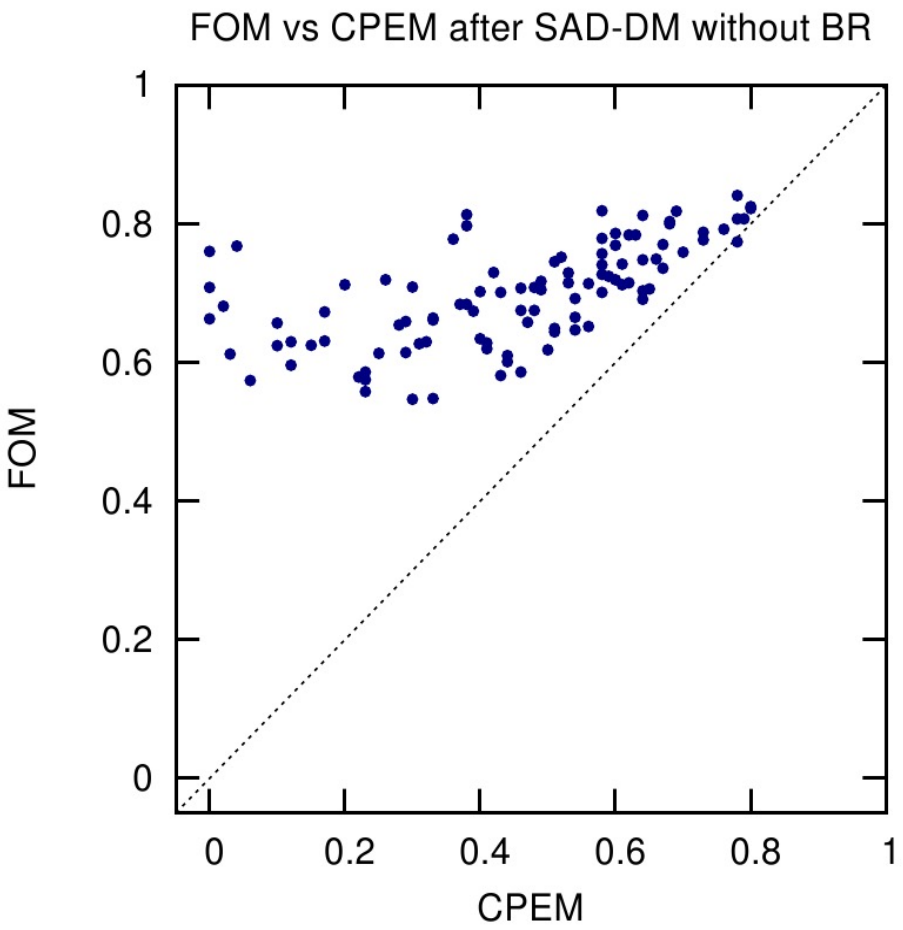
Is my map good enough?

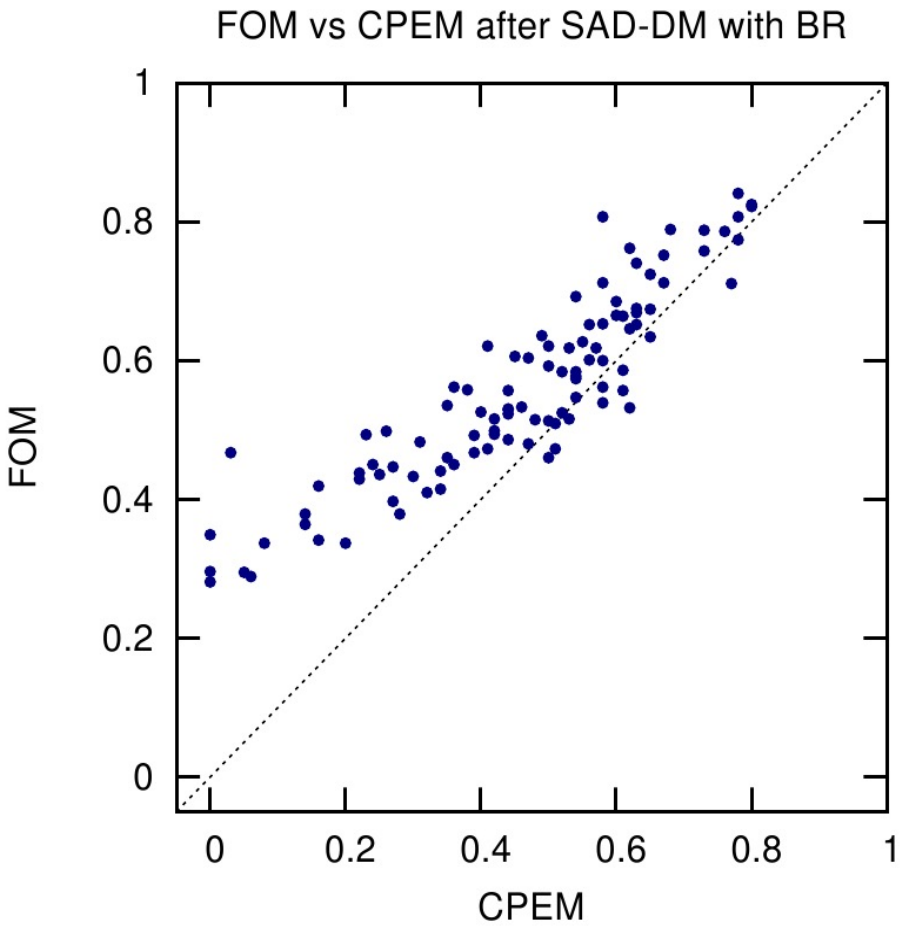
- Statistics from substructure phasing:
  - Look at FOM.
  - Refined occupancies.
- Statistics from density modification:
  - Compare the "contrast" from hand and enantiomorph (output of solomon or shelxe).
- Does it look like a protein? (model visualization)
- For Crank2, look to see if R-comb < 40%.

Bias reduction in density modification

- Density modified map is obtained from experimental map leading to artificially high correlations between the observed and modified amplitudes.
- β correction is applied to the Luzzati error parameter to reduce bias of modified data.

FOM and phase error after DM with/without bias reduction

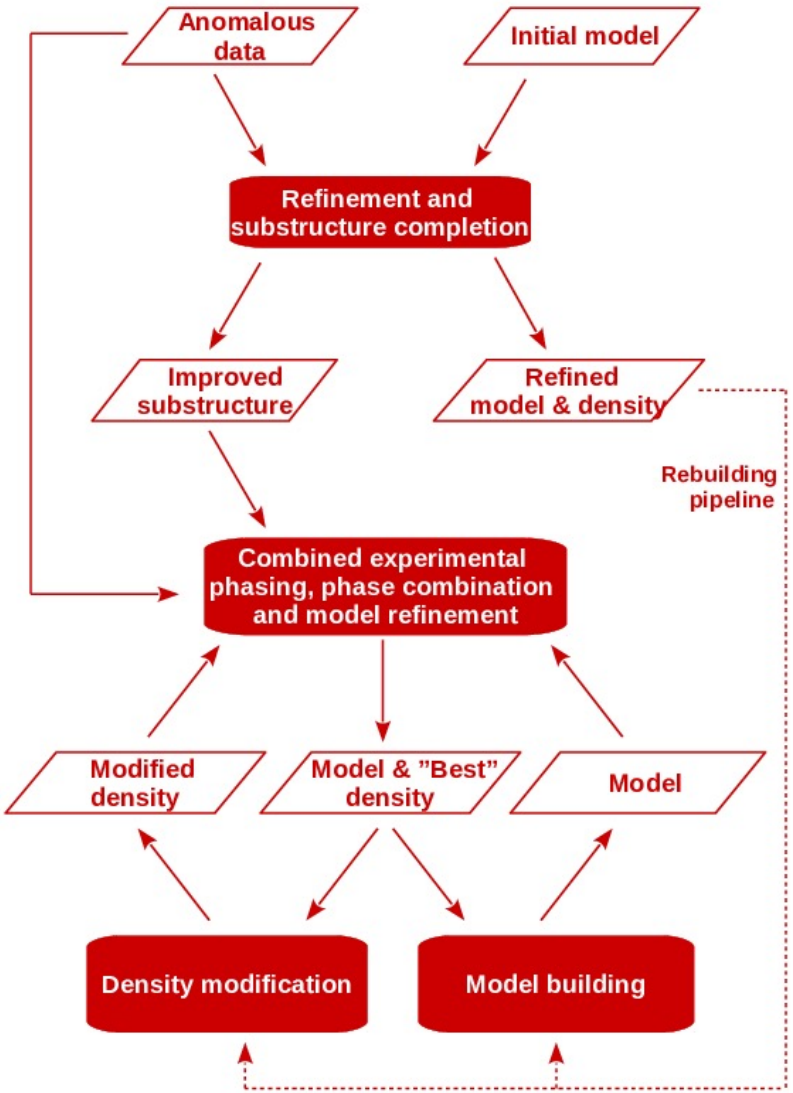




Combining molecular replacement and anomalous scattering (MR-SAD)

- When do you use “MR-SAD”?
  - If you have a partial model (usually obtained by molecular replacement) and anomalous diffraction.
- Two approaches implemented in Crank2
  - Use only the substructure model only
  - Use the substructure and partial model

The two pipelines



MR-SAD data sets

	Final PDB code	Reso-lution (Å)	Anomalous Scatterer	Number of residues	Correct MR Residues (%)	Incorrect MR Residues (%)	RMSD correct residues (Å)
Dataset 1	NA	3.6	Se	800	42.5	23.5	1.6
Dataset 2	NA	3.2	Se	378	60.8	12.9	1.7
GPCR ECR-Mb	5kvm	3.0	I	459	49.7	11.7	1.5
AAA-ATPase	4d80	3.6	Se	1776	75.2	22.1	1.7
F_1-ATPase	2w6f	3.0	S, P	3587	46.7	2.3	0.9
SecYEG-SecA	3din	4.5	Se	2886	47.9	40.7	1.7

MR-SAD results: R-free values

	Molecular replacement solution	Substructure-only pipeline	Rebuild pipeline
Dataset 1	49.8	32.6	29.8
Dataset 2	53.7	28.9	32.3
GPCR ECR-Mb	48.6	39.1	38.4
AAA-ATPase	47.5	39.0	40.9
F_1-ATPase	46.5	34.8	33.8
SecYEG-SecA	51.8	39.9	39.6

References

- Crank
  - Ness et al (2004) Structure 12, 1753-1761.
  - Pannu et al (2011) Acta Cryst D67, 331-337.
- MR-SAD
  - Skubak et al. (2018) IUCrJ 5, 166-171.
- Combined approach and Crank2
  - Skubak and Pannu (2013) Nature Communications 4: 2777.
- Using data directly in refinement
  - Skubak et al (2004) Acta Cryst D60, 2196-2201.
  - Skubak et al (2009) Acta Cryst D65, 1051-1061.
- Multivariate phase combination
  - Waterreus et al (2010) Acta Cryst D66, 783-788.

Acknowledgements

- All dataset contributors
- Garib Murshudov, Kevin Cowtan, George Sheldrick, Victor Lamzin
- http://www.bfsc.leidenuniv.nl/software/crank/

NWO Nederlandse Organisatie voor Wetenschappelijk Onderzoek

Cyttron

CCP4