Experimental phasing in Crank2

Pavol Skubak and Navraj Pannu
Biophysical Structural Chemistry,
Leiden University, The Netherlands
http://www.bfsc.leidenuniv.nl/software/crank/

Crank2 for experimental phasing

- Crank2 is suitable for SAD, MAD and SIRAS.

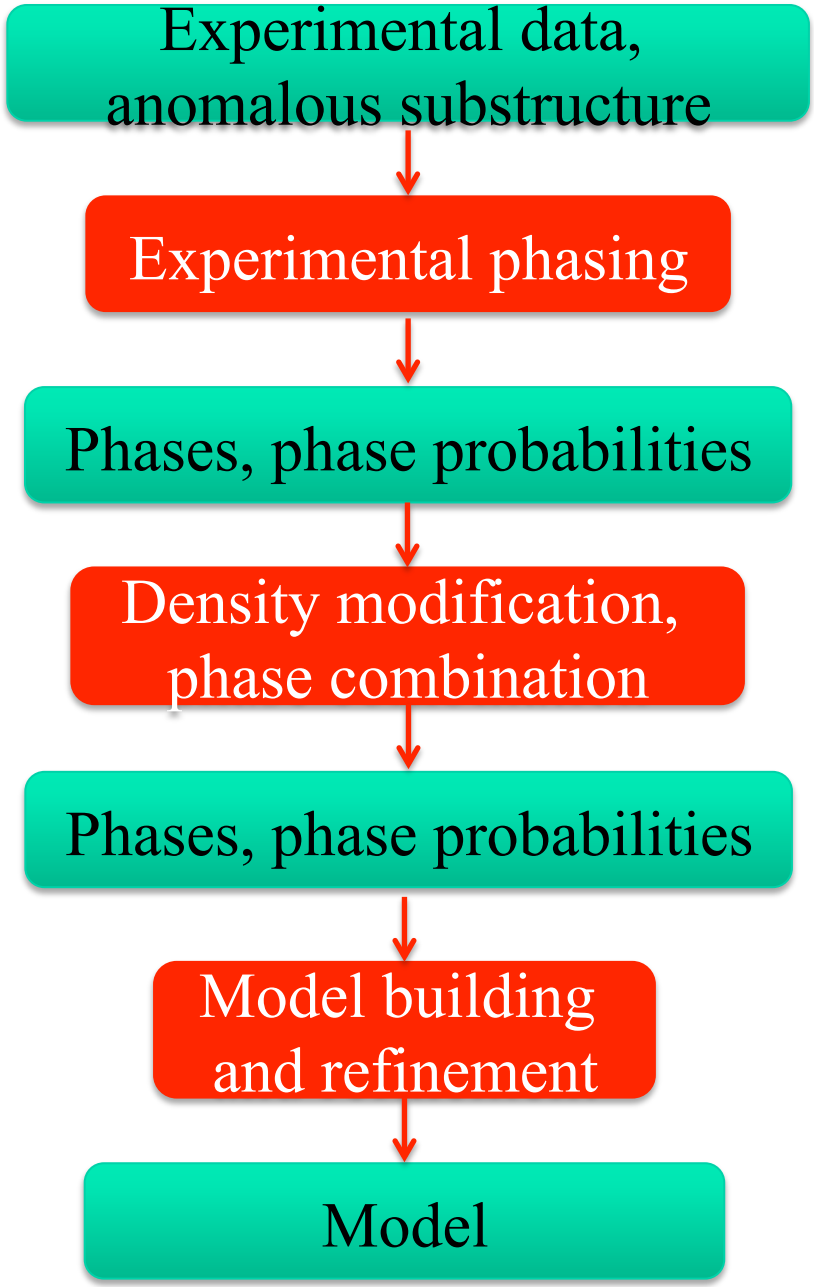
- Crank2 has been shown to be particularly powerful weak anomalous signal to low resolution SAD datasets (4.5 Angstroms)

Simultaneously combining experimental phasing steps to improve structure solution

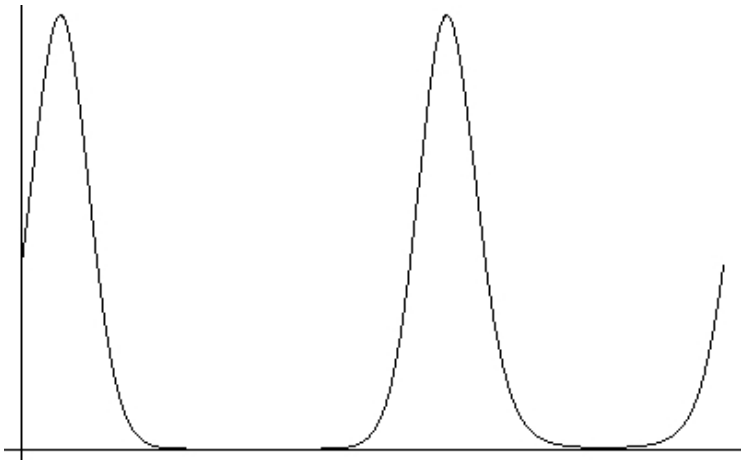
- Traditionally structure solution is divided into distinct steps:
  - Substructure detection
  - Obtain initial phases
  - (Density) modify the initial experimental map
  - Build and refine the model
- By combining these steps, we can improve the process.

Traditional structure solution

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



Phase probabilities

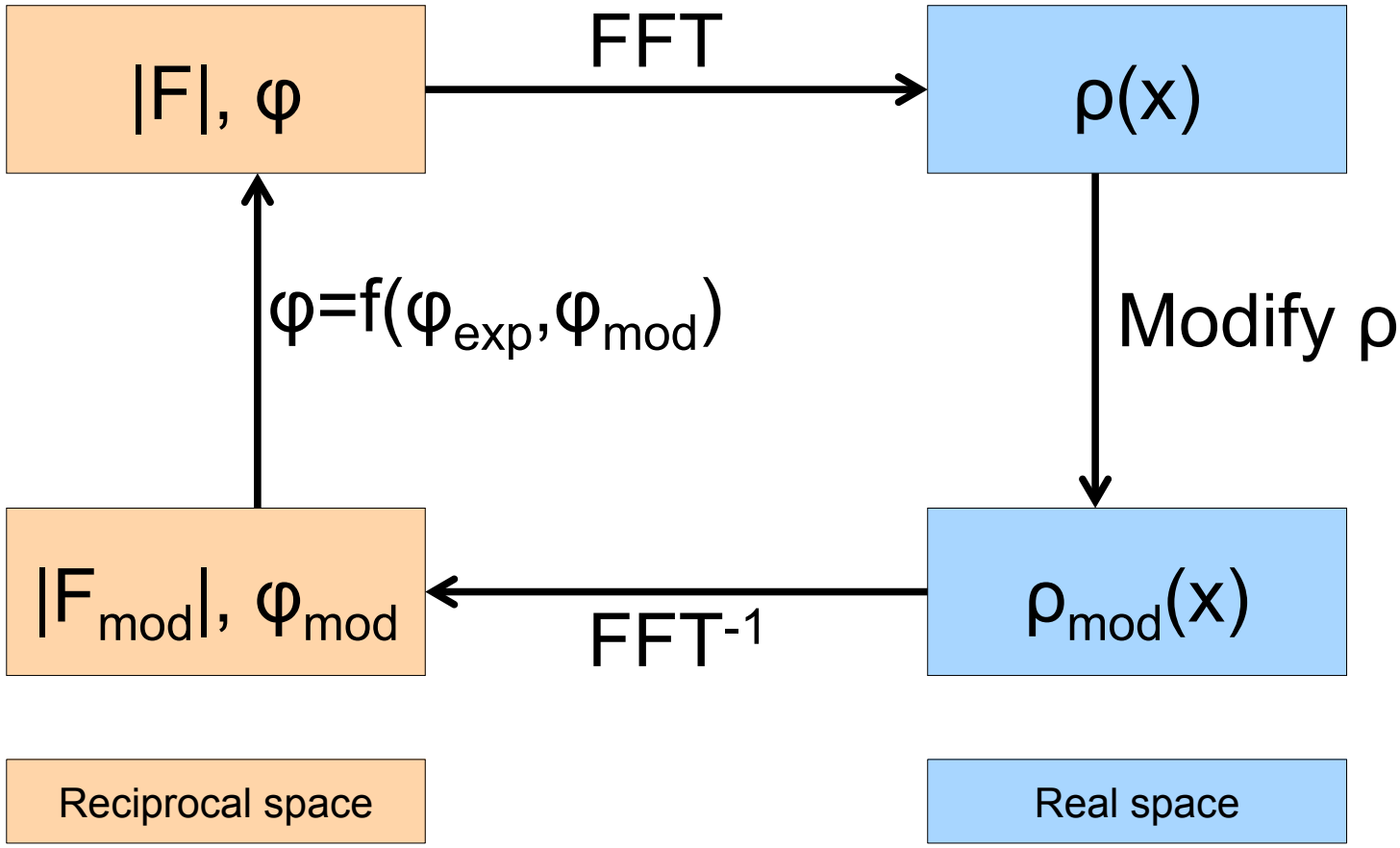


$$P(\alpha) = \exp(A\cos(\alpha) + B\sin(\alpha) + C\cos(2\alpha) + D\sin(2\alpha))$$

- 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:



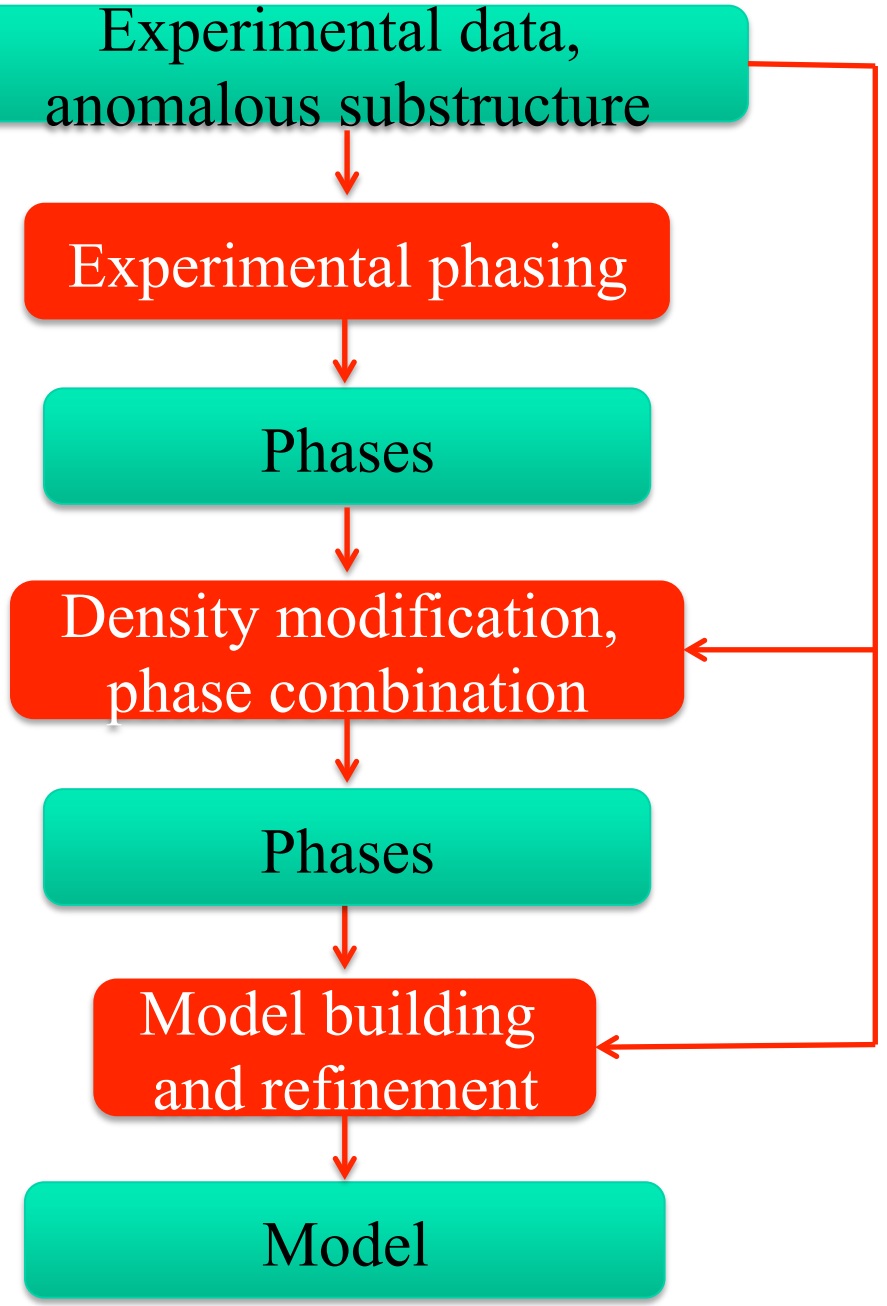
Kevin Cowtan, cowtan@ysbl.york.ac.uk

Hendrickson-Lattmann (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).

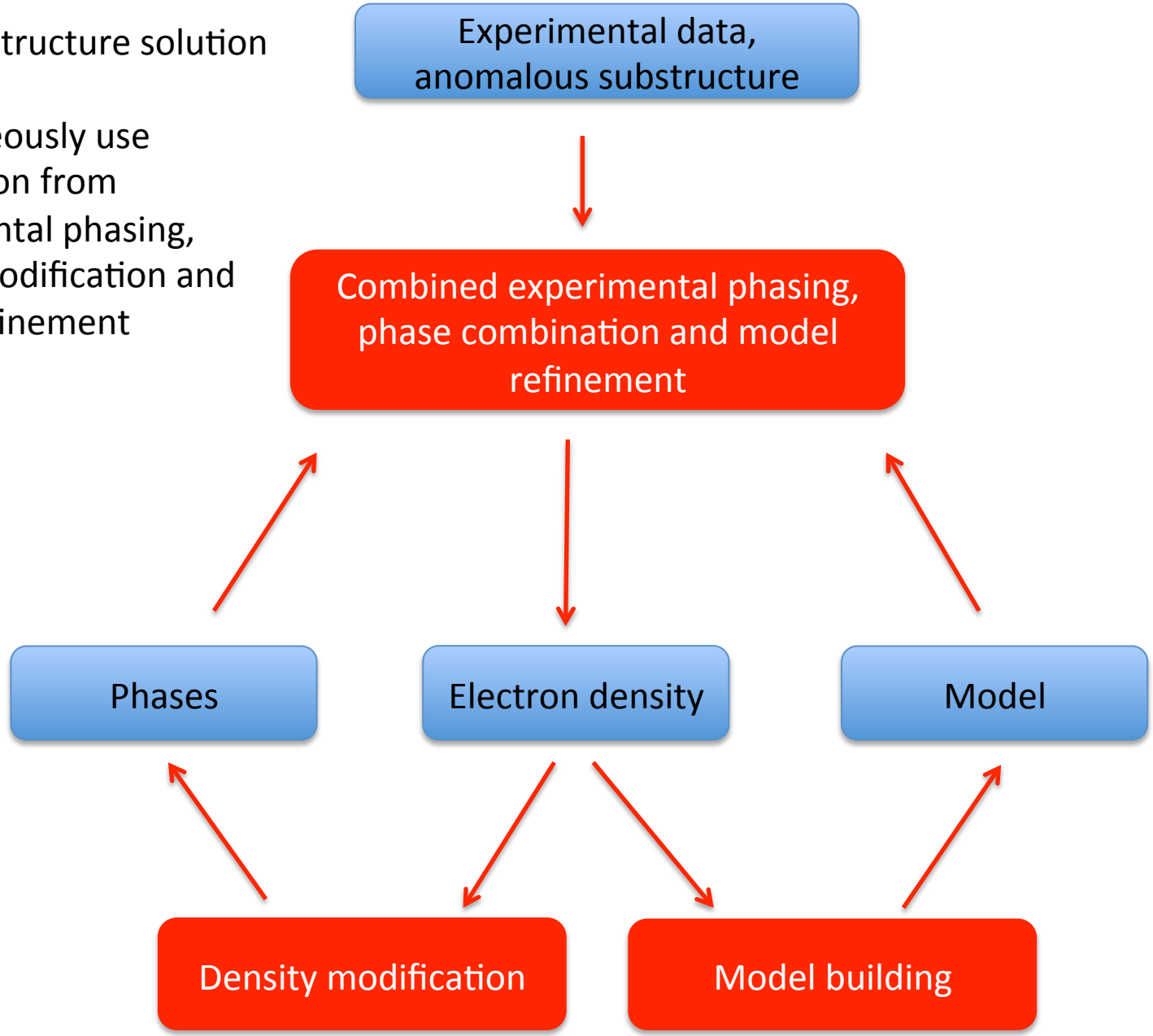


What are step-wise multivariate functions?

- 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



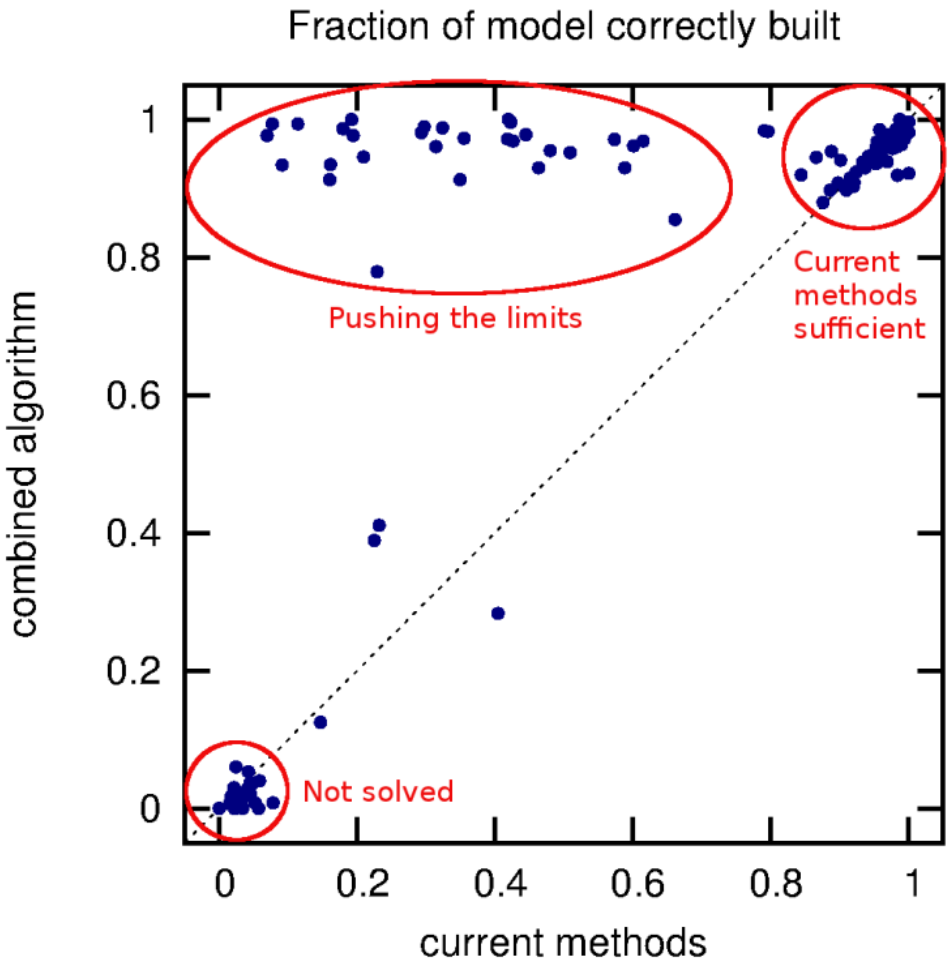
The “combined” multivariate function

- 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)



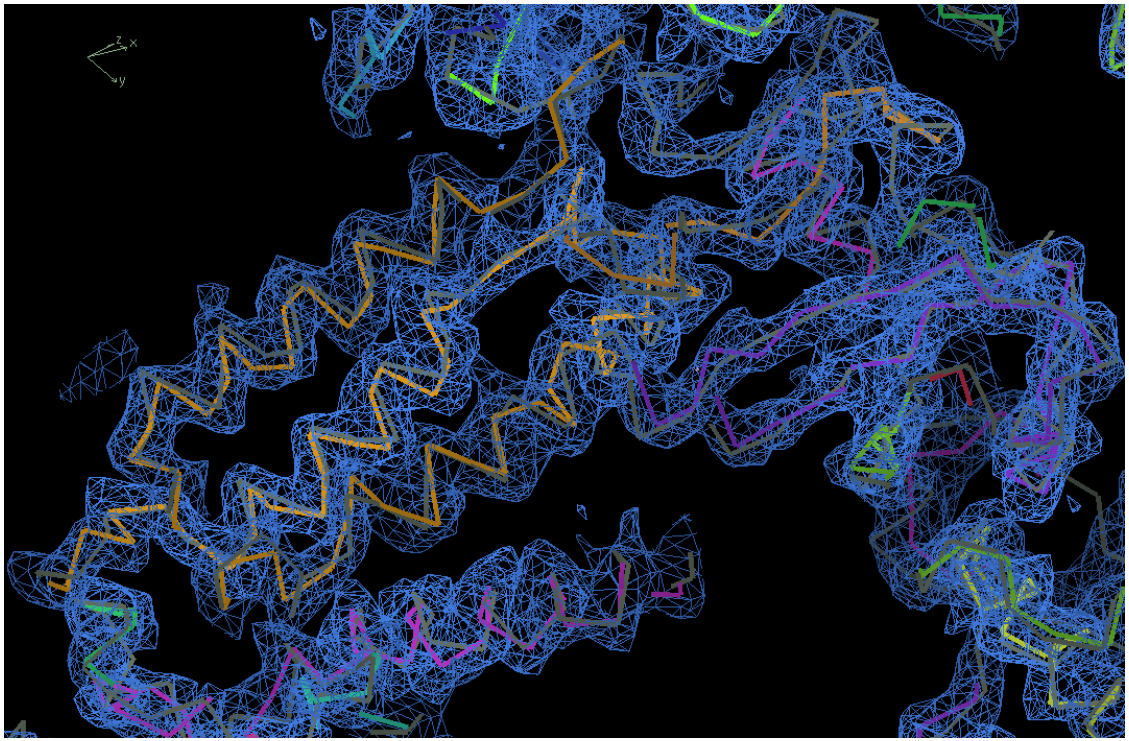
Summary of large scale test

- 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%.

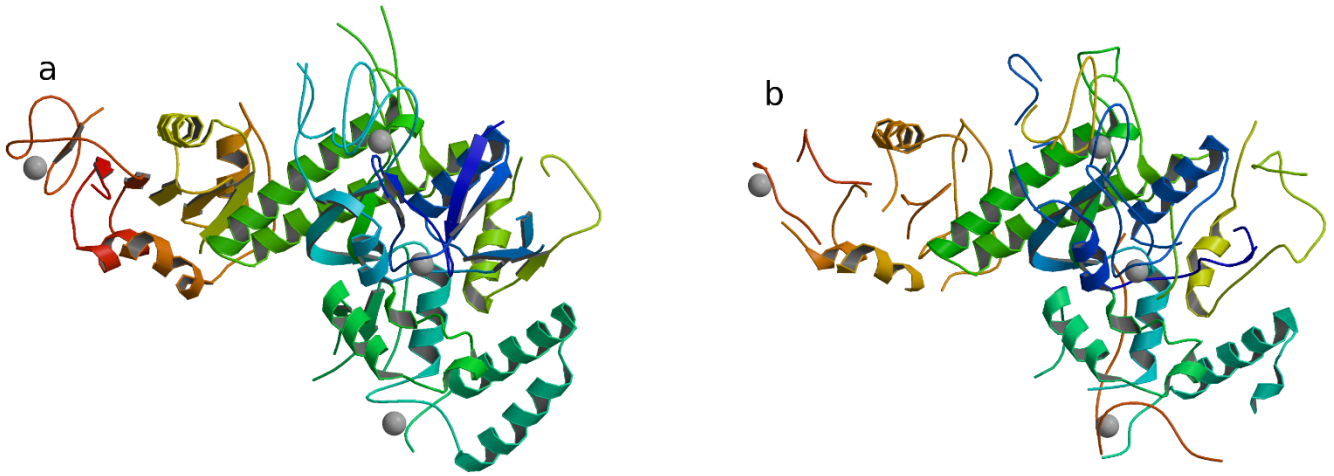
3.88 Angstrom RNA polymerase II

- 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.



Crank1 vs Crank2

- Crank1 and Crank2 are suitable for S/MAD and S/MIRAS experiments and implements multivariate functions.
- Crank1 and Crank2 are available in ccp4 6.5 and Crank2 is available in ccp4 7.0 with new gui.

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.

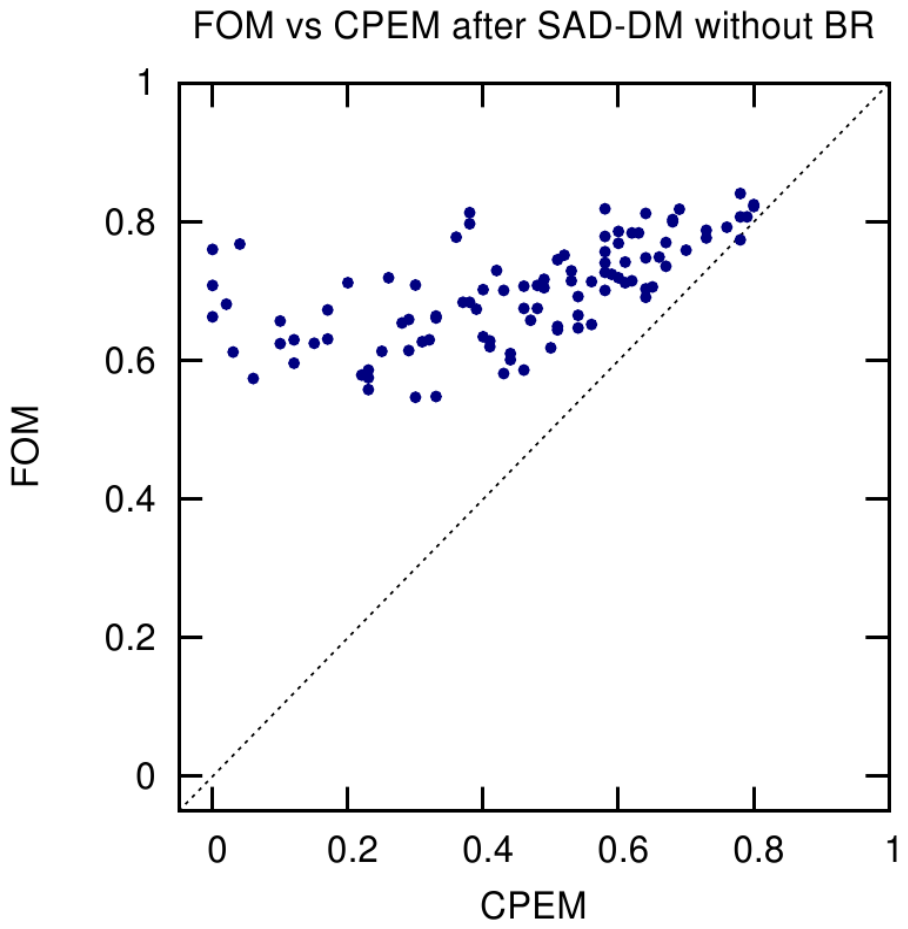
Is my map good enough?

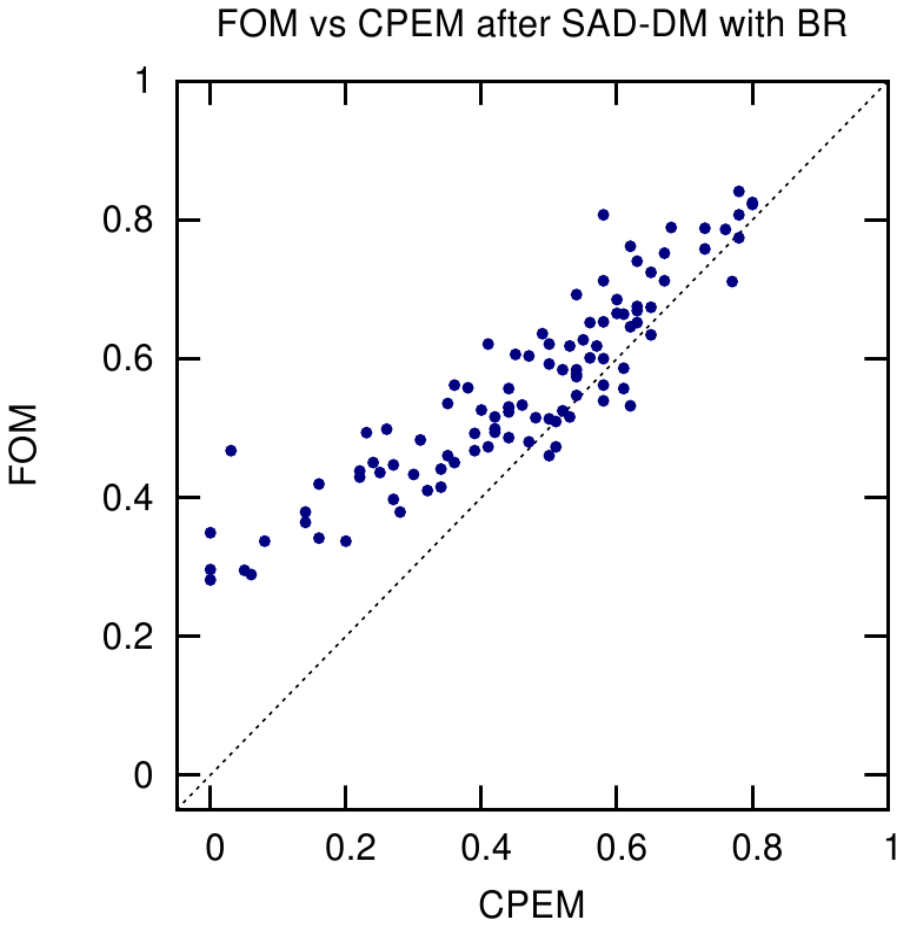
- Statistics from substructure phasing:
  - Look at FOM.
  - Refined occupancies.
- Statistics from density modification:
  - Compare the "contrast" from hand and enantiomorph (output of solomon or shelxe).
- Does it look like a protein? (model visualization)
- For Crank2, look to see if R-comb < 40%.

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





When do you use MR-SAD?

- If you have a molecular replacement solution, but you can not obtain a complete model.

- If you have an anomalous signal (look at CCano), you can use this to improve your model.

What do you need to consider?

- Your partial model may not contain all the anomalous scatterers and you will need to find the rest using anomalous maps.

- Your partial MR model may be biased and that can affect your ability to complete your model. Should you truncate your model to only what you believe is correctly modelled??

SAD and SIRAS functions in model refinement

Previous functions in REFMAC:

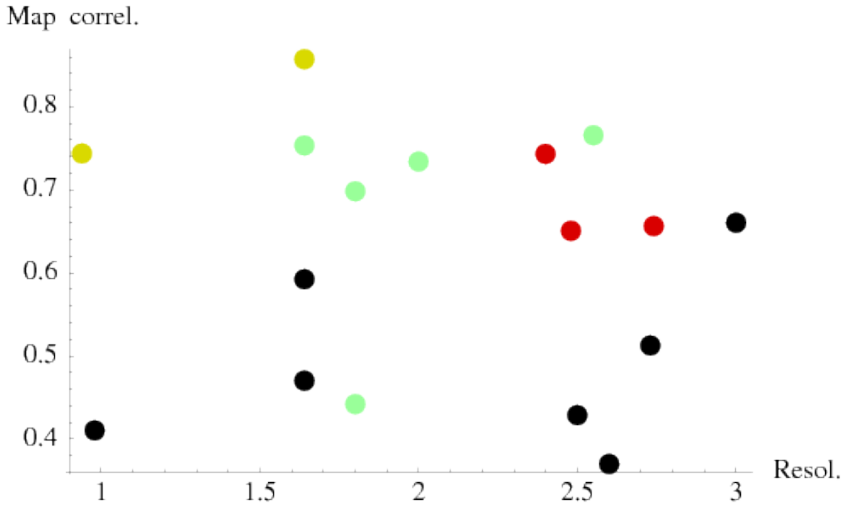
- No prior phase information (Rice function) (Murshudov *et al.*,1997), (Bricogne and Irwin, 1996), (Pannu and Read, 1996)
- Prior phase information used indirectly in the form of Hendrickson-Lattman coefficients (MLHL) (Pannu *et al.*, 1998)

Tests of SAD and SIRAS functions in refinement

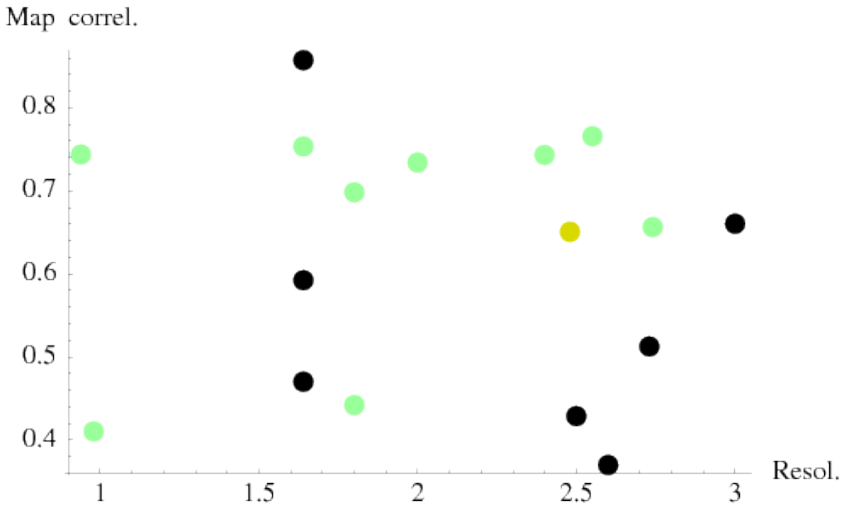
- The functions were tested on many real data sets (various phasing signals and resolution ranges) in ARP/wARP + REFMAC.

- Skubak *et al.* (2004,2005,2009) Acta D.

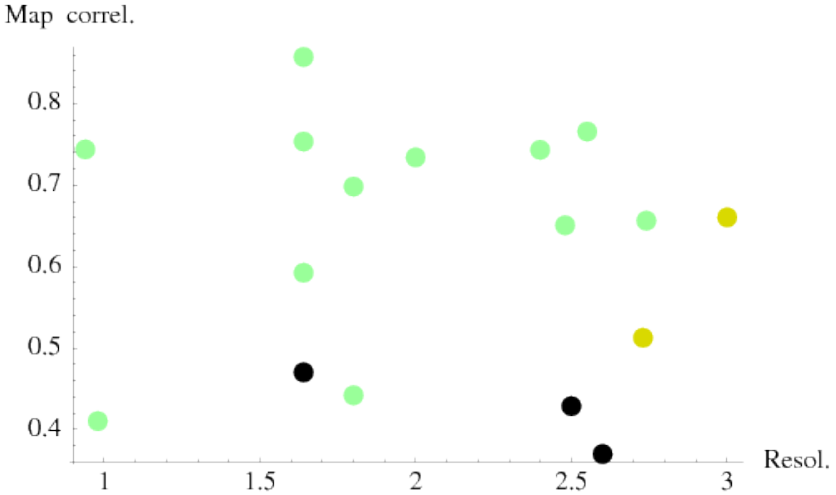
Results from SAD function



RICE



MLHL



SAD

Green circle: 100 – 70% built

Yellow circle: 70 – 50% built

Red circle: 50 – 20% built

Black circle: 20 – 0% built

MR-SAD tests
4.5 Angstrom ATPase SecA-SecY complex

- 4.5 Angstrom data set with weak anomalous signal from Se-Met SecY.
- Authors could not solve the structure with SAD data alone, but used a partial MR structure, (2-fold NCS averaging), cross-crystal averaging, and manual model building
- > 60% can be built automatically starting just from selenium positions (obtained from partial MR solution)
- R-free obtained was under 40%.

4.6 Angstrom SKI2-3-8 complex

- 4.6 Angstrom data set Se-Met data set.
- > 50% can be built automatically starting just from selenium positions.
- R-free obtained was under 40%.

Conclusions from MR-SAD

- At the moment, if a (partial) MR solution is available, it is best to run *two* crank2 jobs and manually combine runs:
  - Input the whole MR solution
  - Input just the heavy atoms

References

- Crank
  - Ness et al (2004) Structure 12, 1753-1761.
  - Pannu et al (2011) Acta Cryst D67, 331-337.
- 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 (JCSG, Z. Dauter, M.Weiss, C.Mueller-Dieckmann)
- Garib Murshudov, Kevin Cowtan, George Sheldrick, Victor Lamzin
- http://www.bfsc.leidenuniv.nl/software/crank/

Cyttron