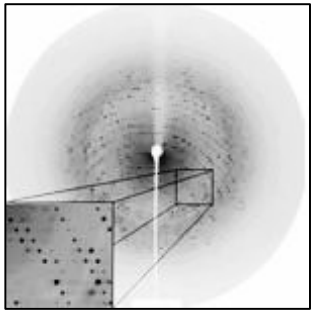
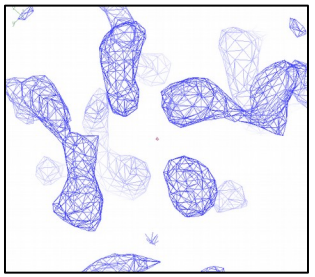
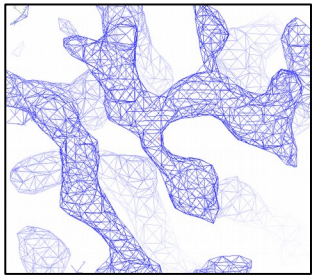
Automated phase improvement (density modification)

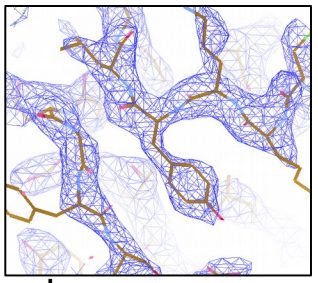
Kevin Cowtan
(they/them)
University of York

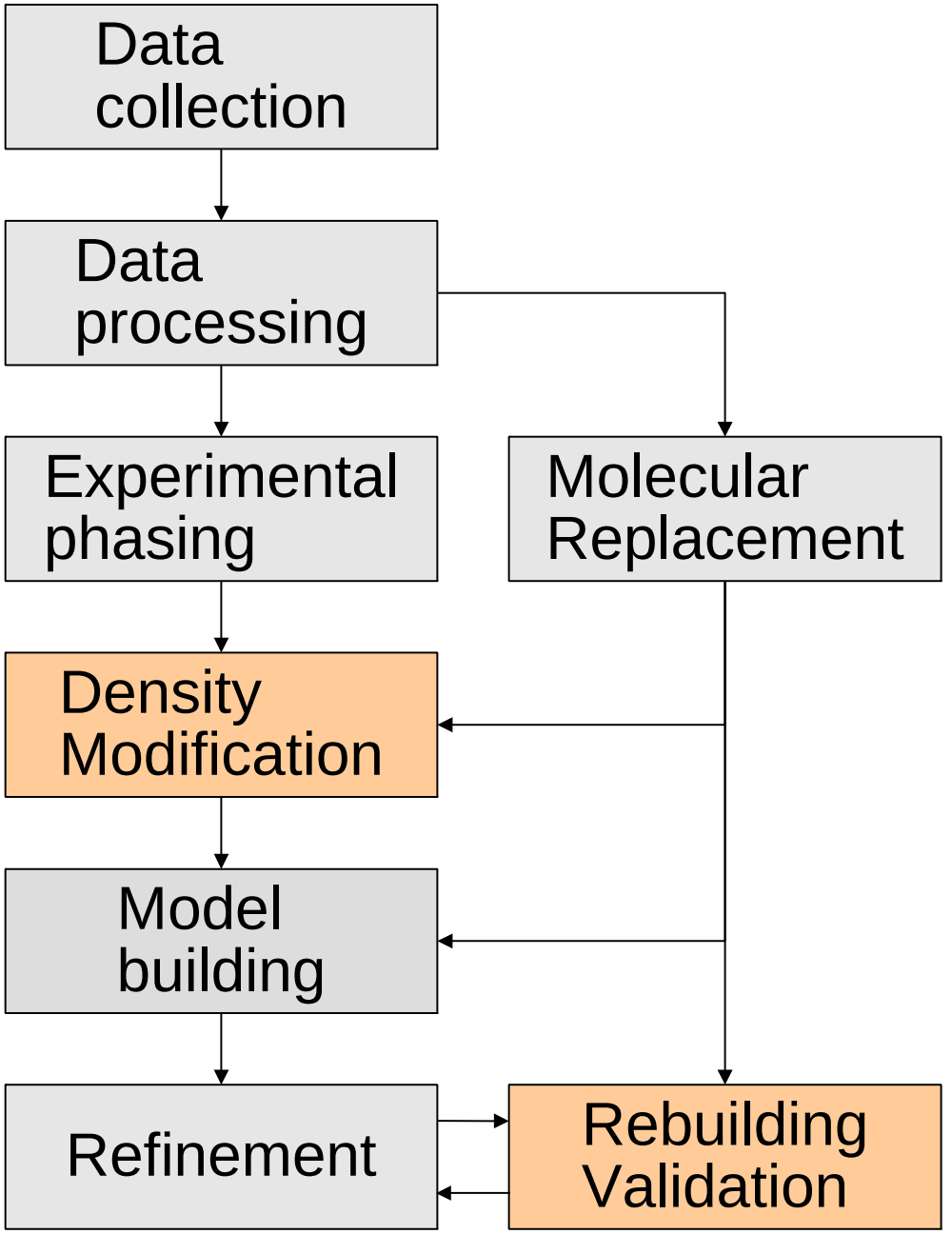
X-ray structure solution pipeline...







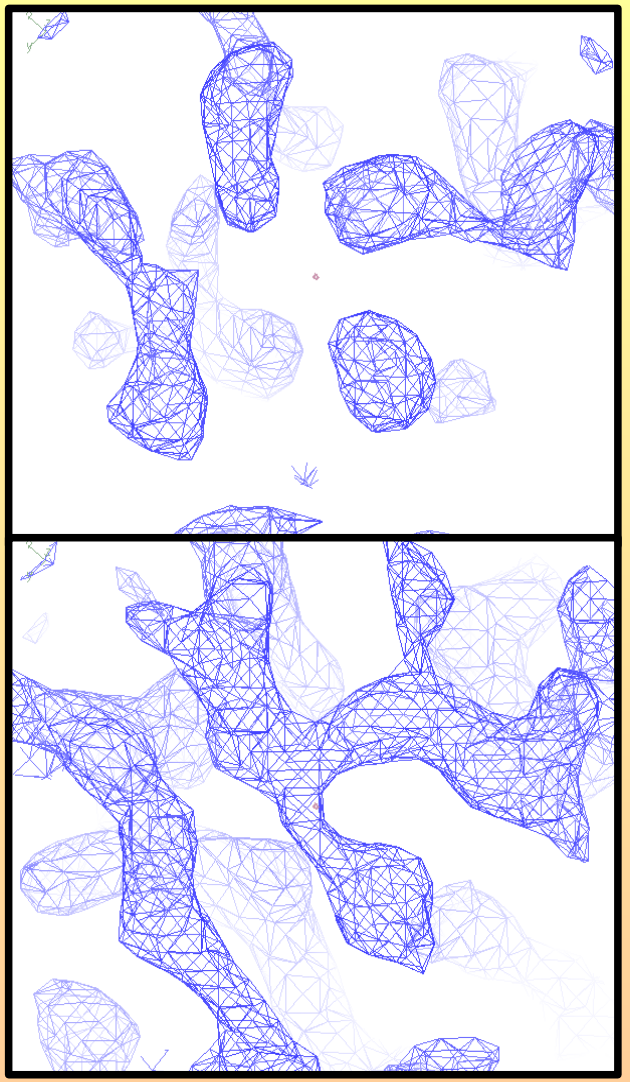




K Cowtan, kevin.cowtan@york.ac.uk

Density Modification

- Classical density modification: e.g. 'dm', 'solomon', 'parrot', CNS
- Statistical density modification: e.g. 'resolve', 'pirate'

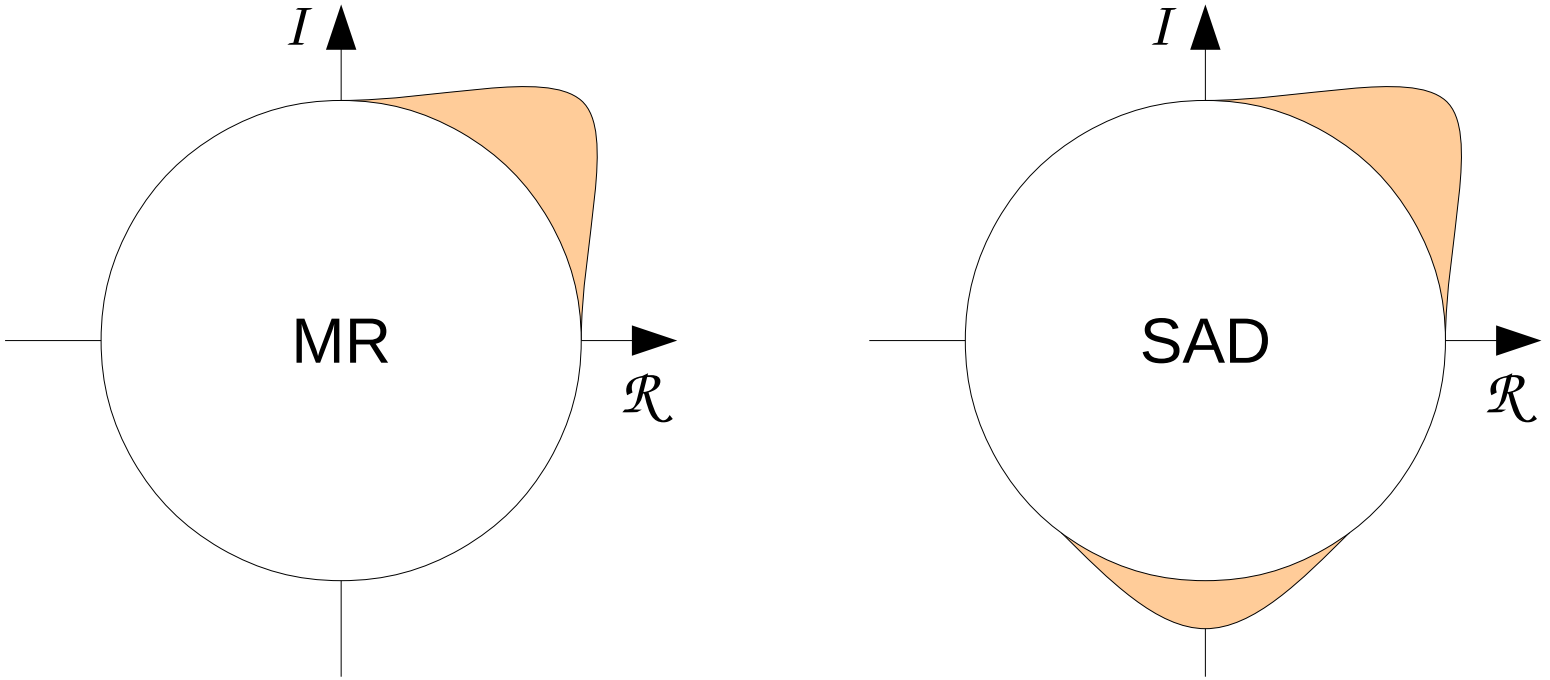


K Cowtan, kevin.cowtan@york.ac.uk

Density modification

Starting point:

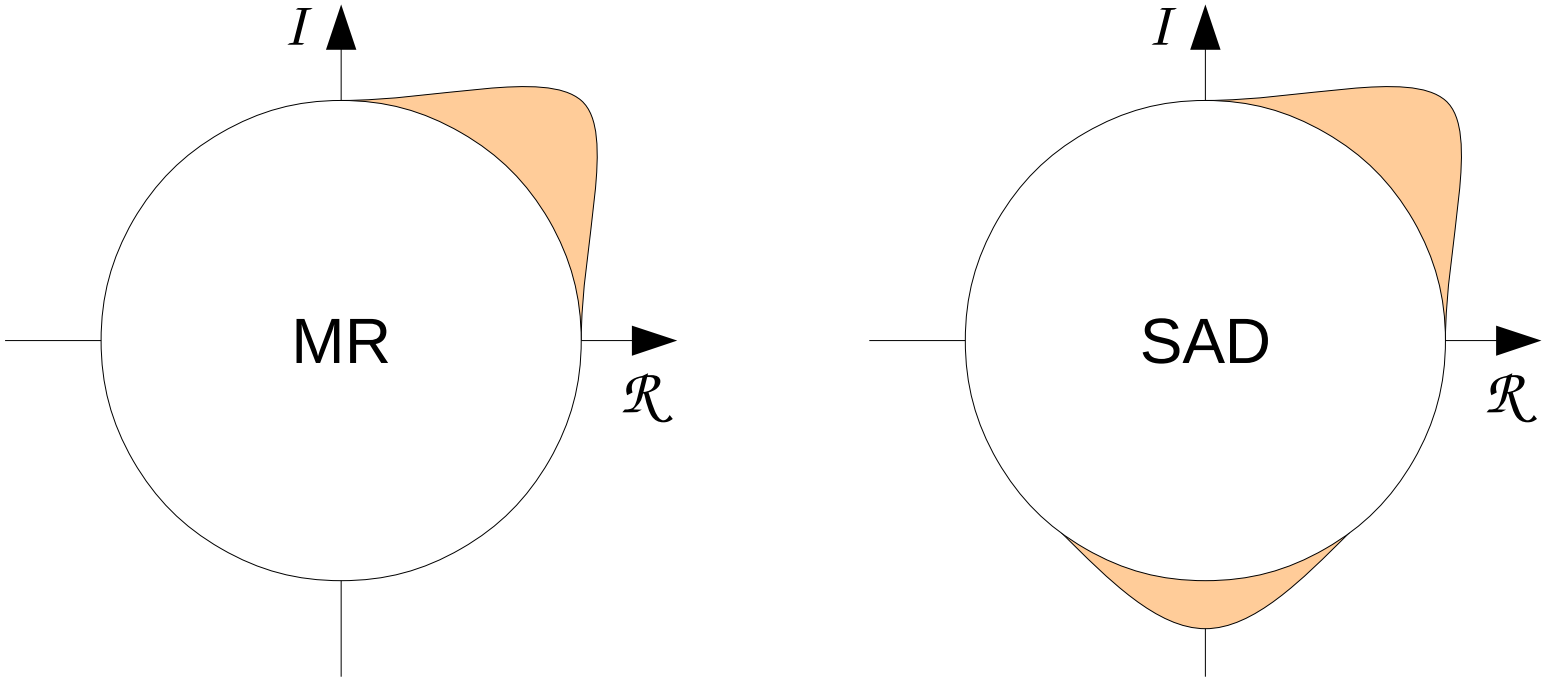
- Structure factor amplitudes
- Phase estimates:
  - MR: Unimodel distribution
  - SAD: Biomodal distribution



Density modification

How do we represent phase probability distributions?

- Phase/figure of merit - Φ, FOM
  - (unimodel, MR only)
- Henrickson-Lattman coeffs – ABCD
  - (bimodal or unimodal, general)

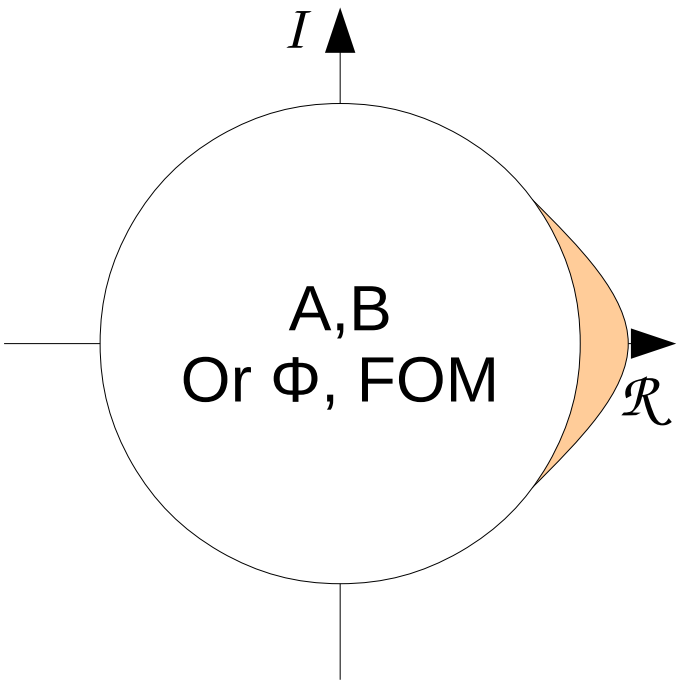


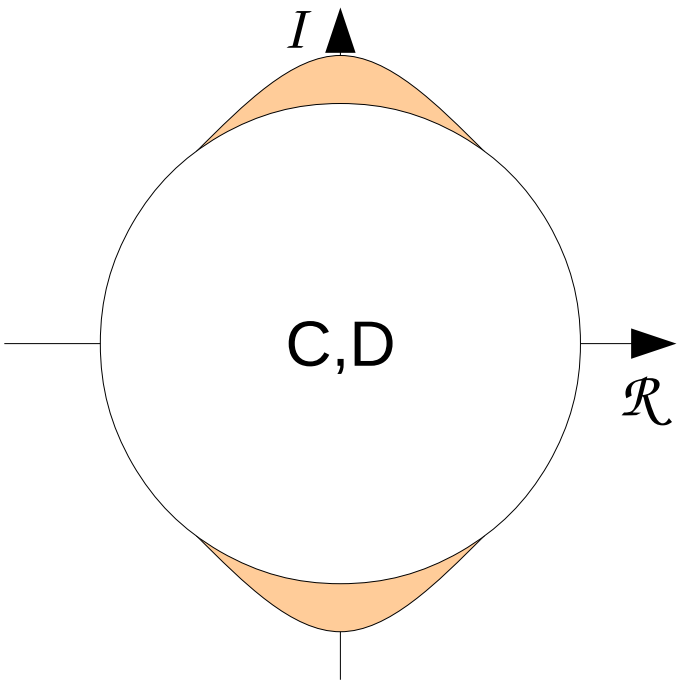
K Cowtan, kevin.cowtan@york.ac.uk

Density modification

A,B represent a unimodal distribution (equivalent to Φ, FOM)
C,D represent the superimposed biomodality.

- Relative size and sign of A,B or C,D control the direction.
- Absolute size $(A^2+B^2)^{½}$ controls the sharpness.
- For MR, we get A,B (or Φ, FOM) i.e. C=D=0.
- Together A,B,C,D can describe a bimodal distribution with any combination of peak height and direction.

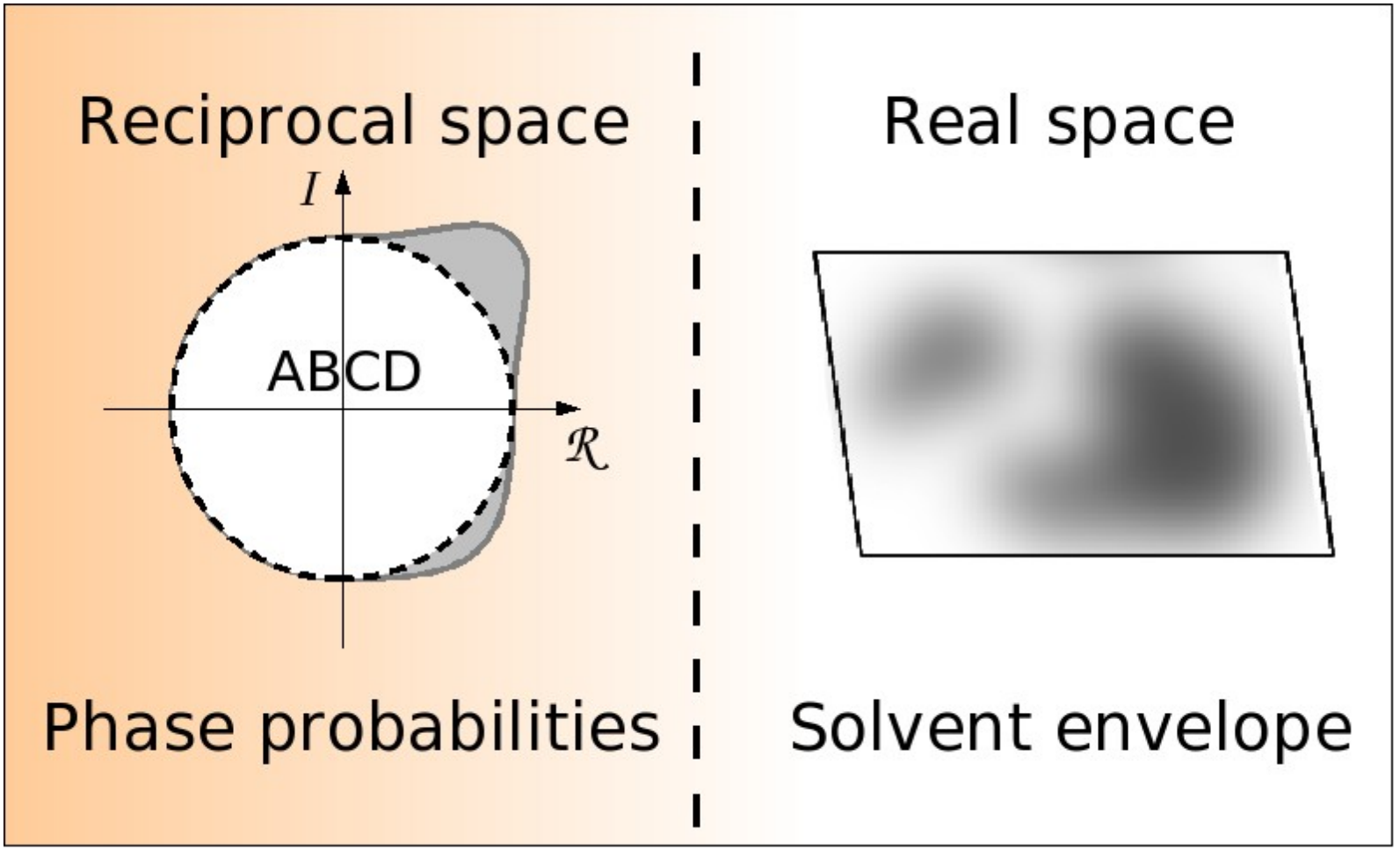




K Cowtan, kevin.cowtan@york.ac.uk

Density modification

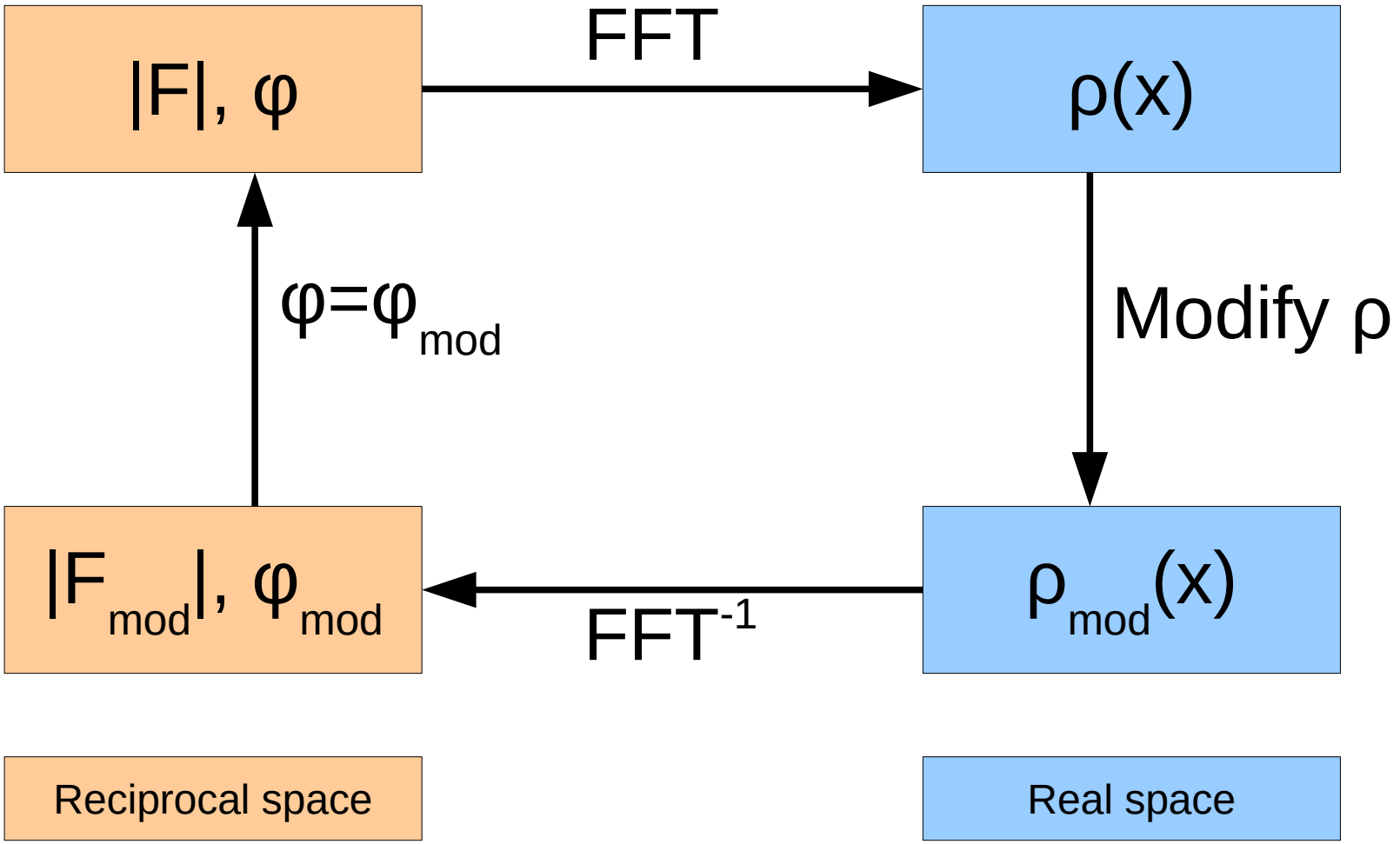
- Density modification is a problem in combining information:



K Cowtan, kevin.cowtan@york.ac.uk

Density modification

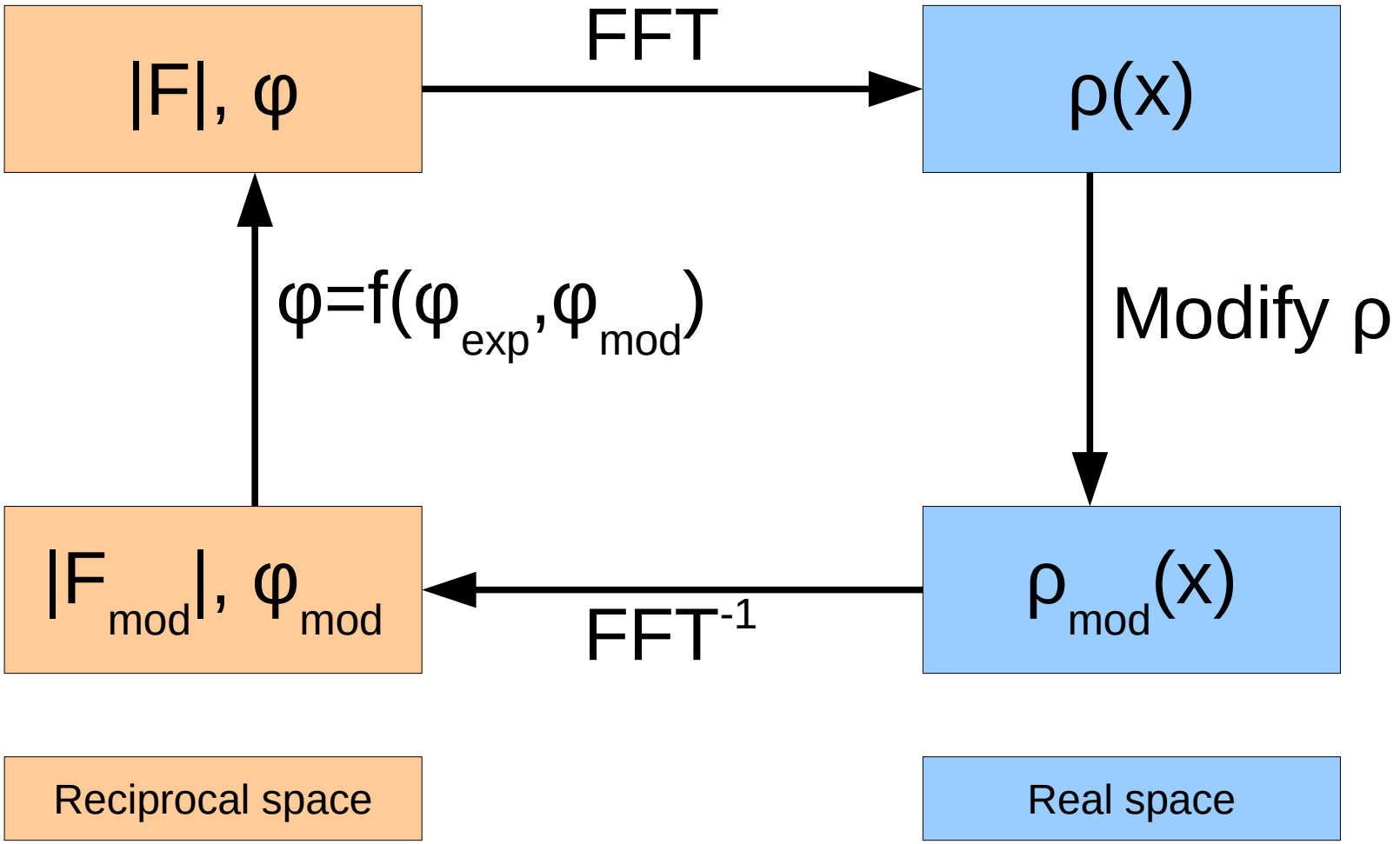
1. Rudimentary calculation:



K Cowtan, kevin.cowtan@york.ac.uk

Density modification

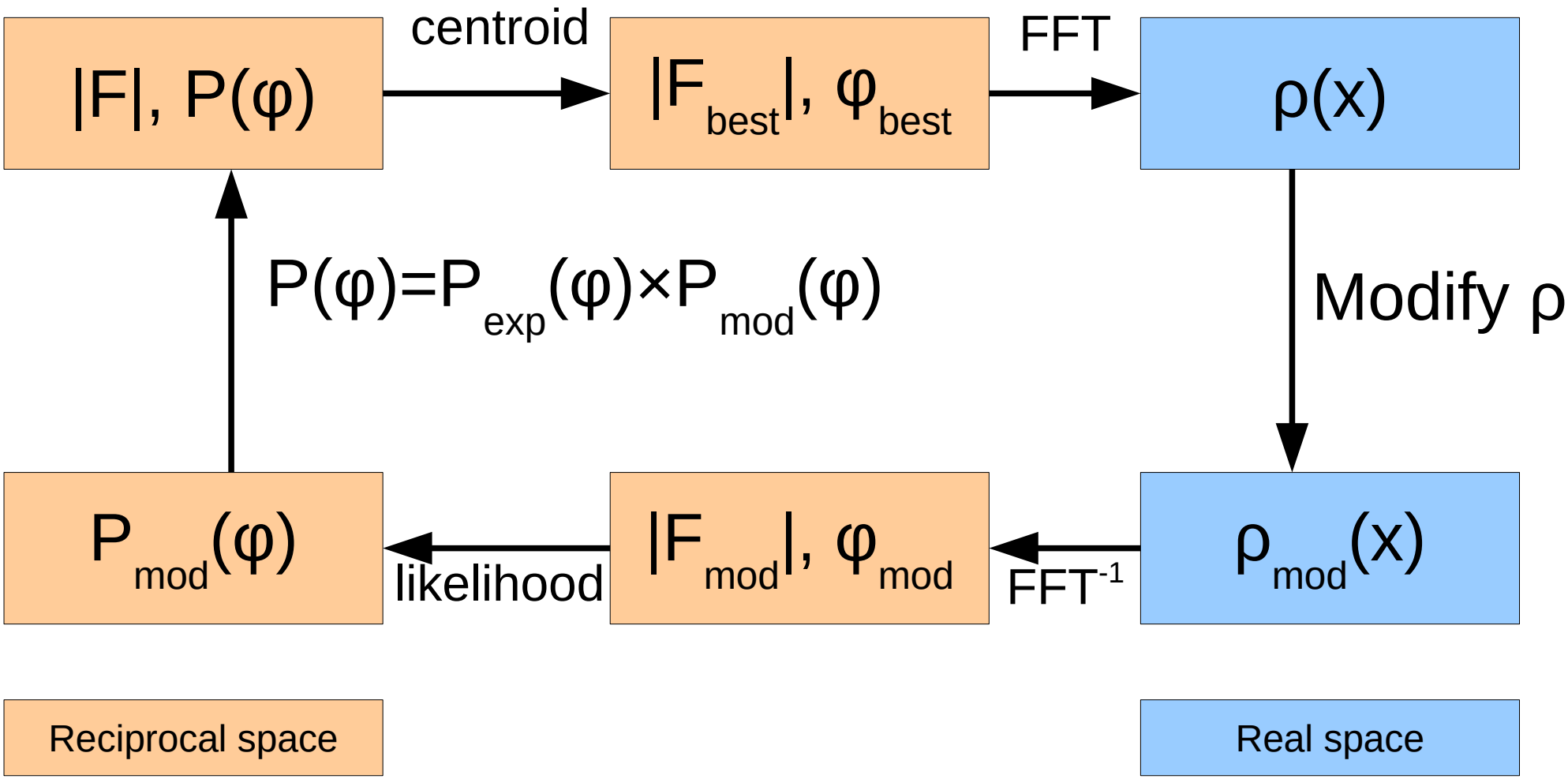
2. Phase weighting:



K Cowtan, kevin.cowtan@york.ac.uk

Density modification

3. Phase probability distributions:

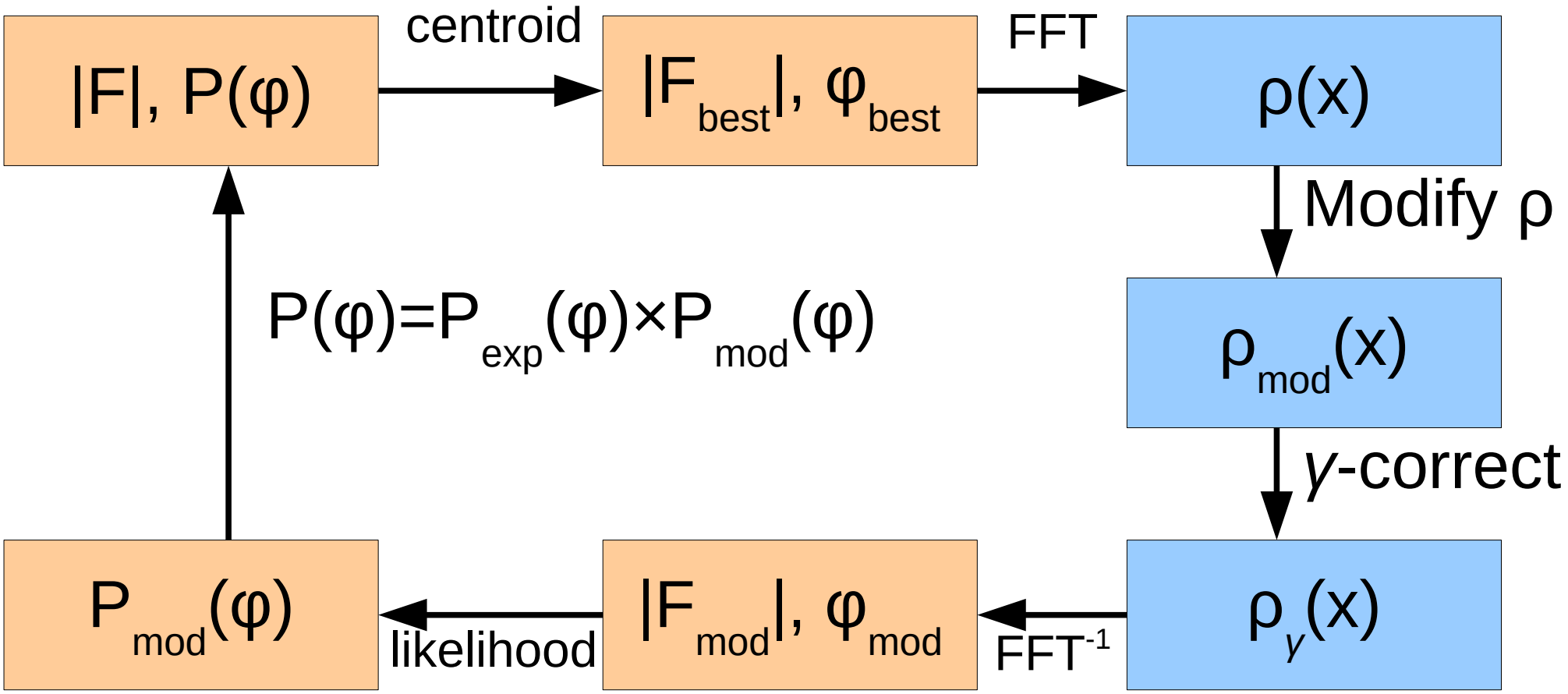


K Cowtan, kevin.cowtan@york.ac.uk

Density modification

DM, SOLOMON, (CNS)

4. Bias reduction (gamma-correction):



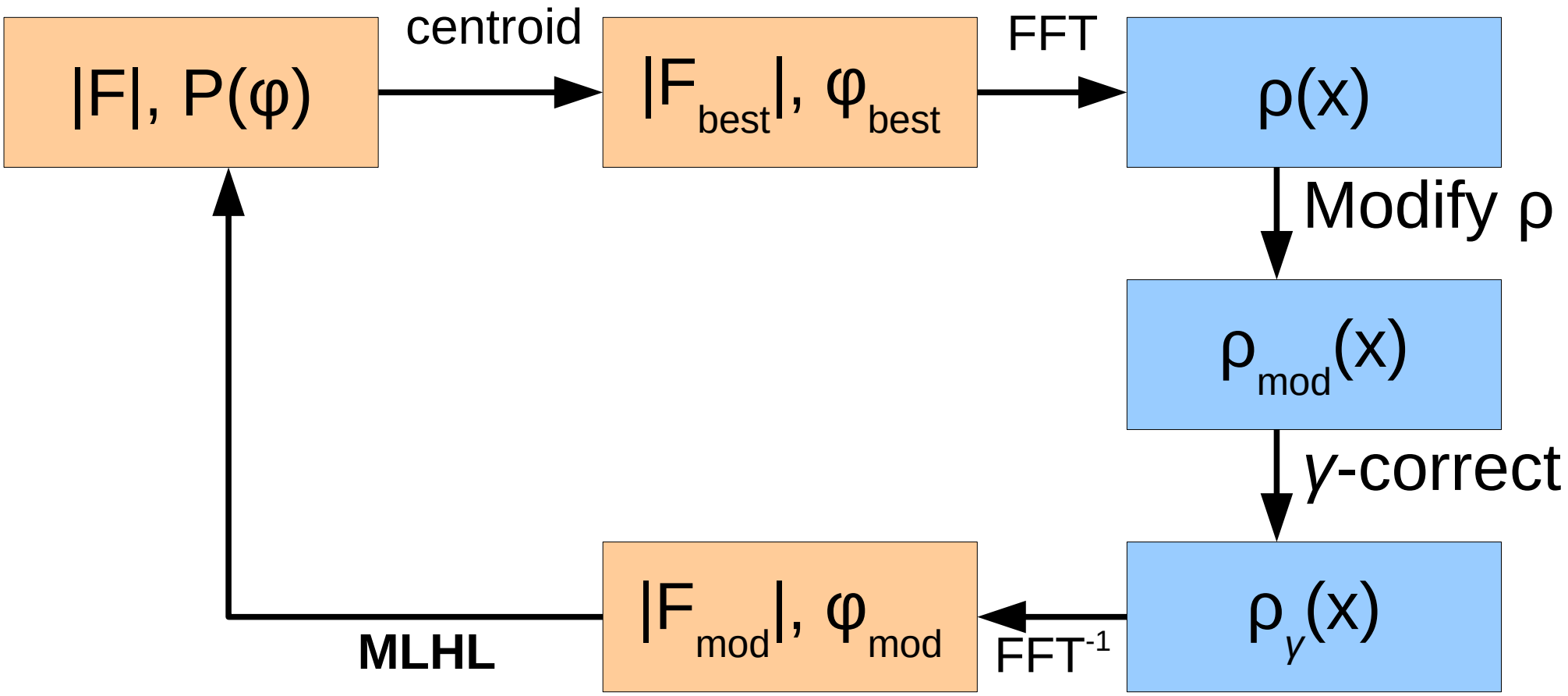
J.P.Abrahams

K Cowtan, kevin.cowtan@york.ac.uk

Density modification

PARROT

5. Maximum Likelihood H-L:

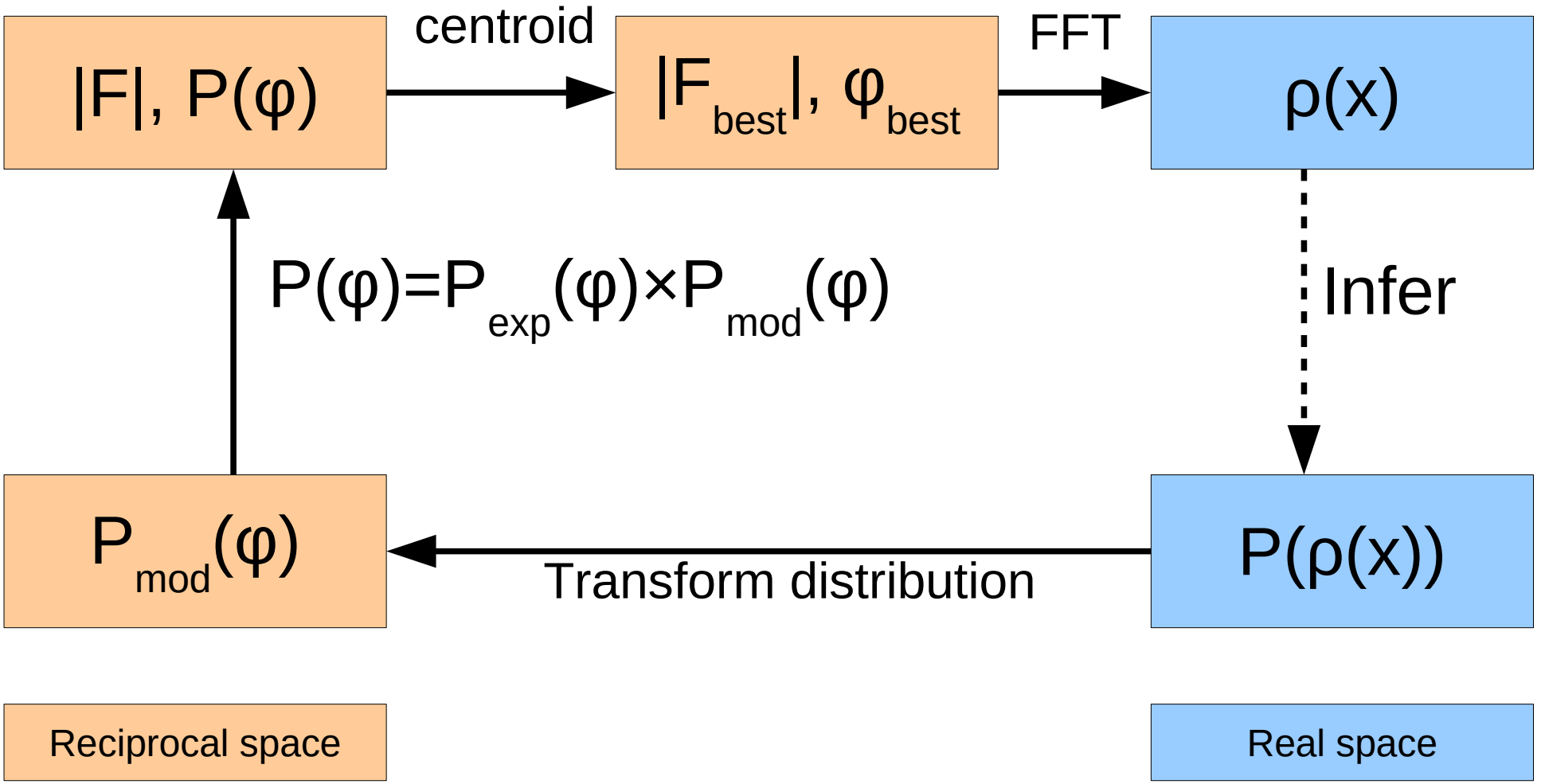


K Cowtan, kevin.cowtan@york.ac.uk

Density modification

RESOLVE, PIRATE

6. Statistical density modification:

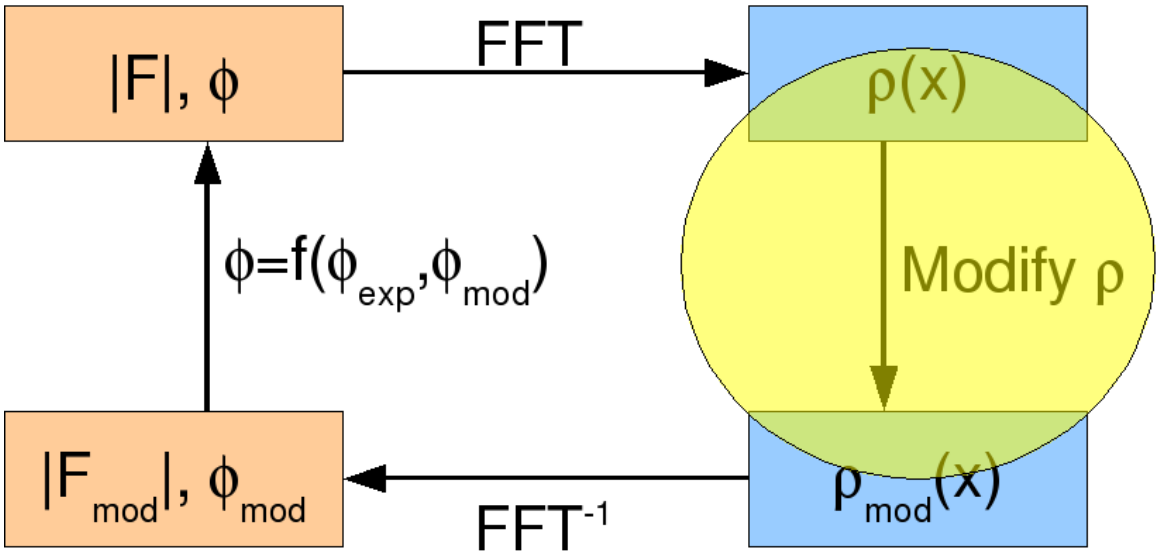


K Cowtan, kevin.cowtan@york.ac.uk

Density modification

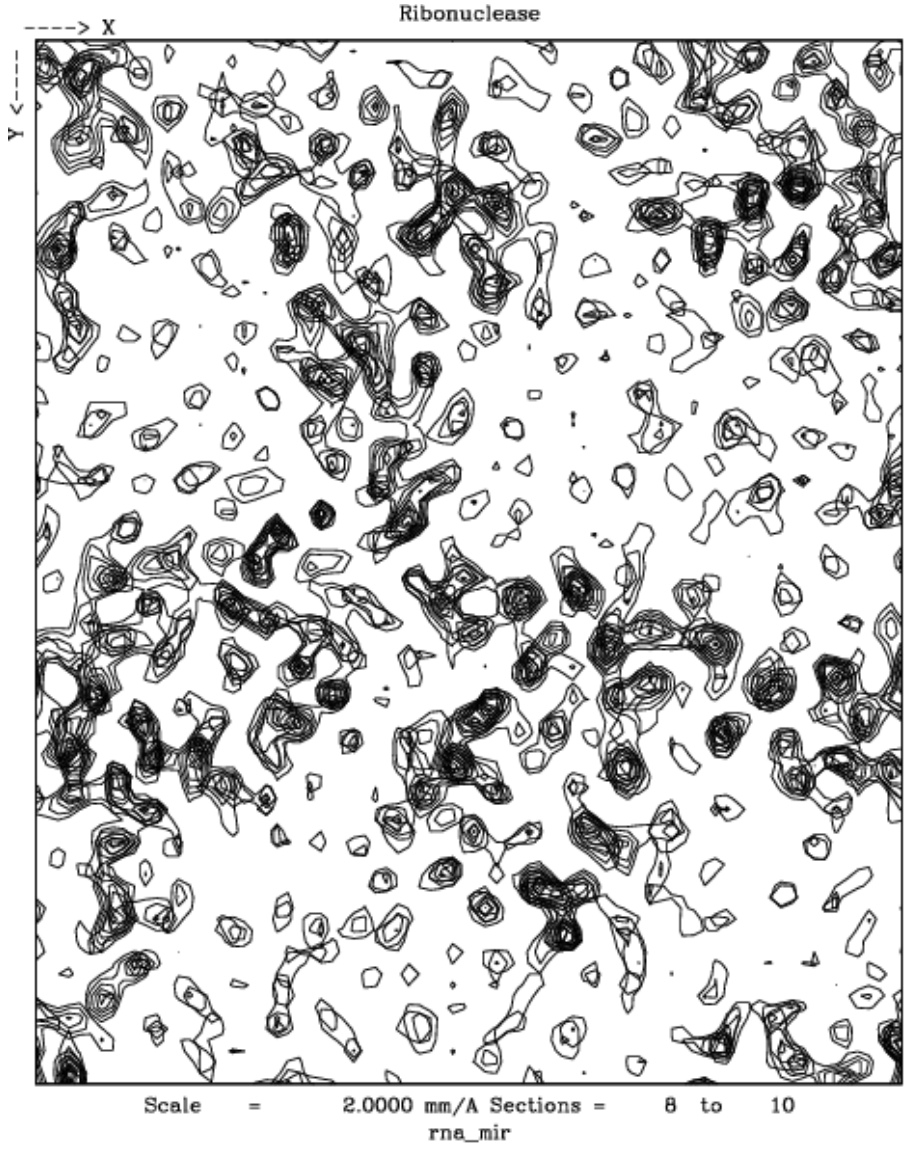
Traditional density modification techniques:

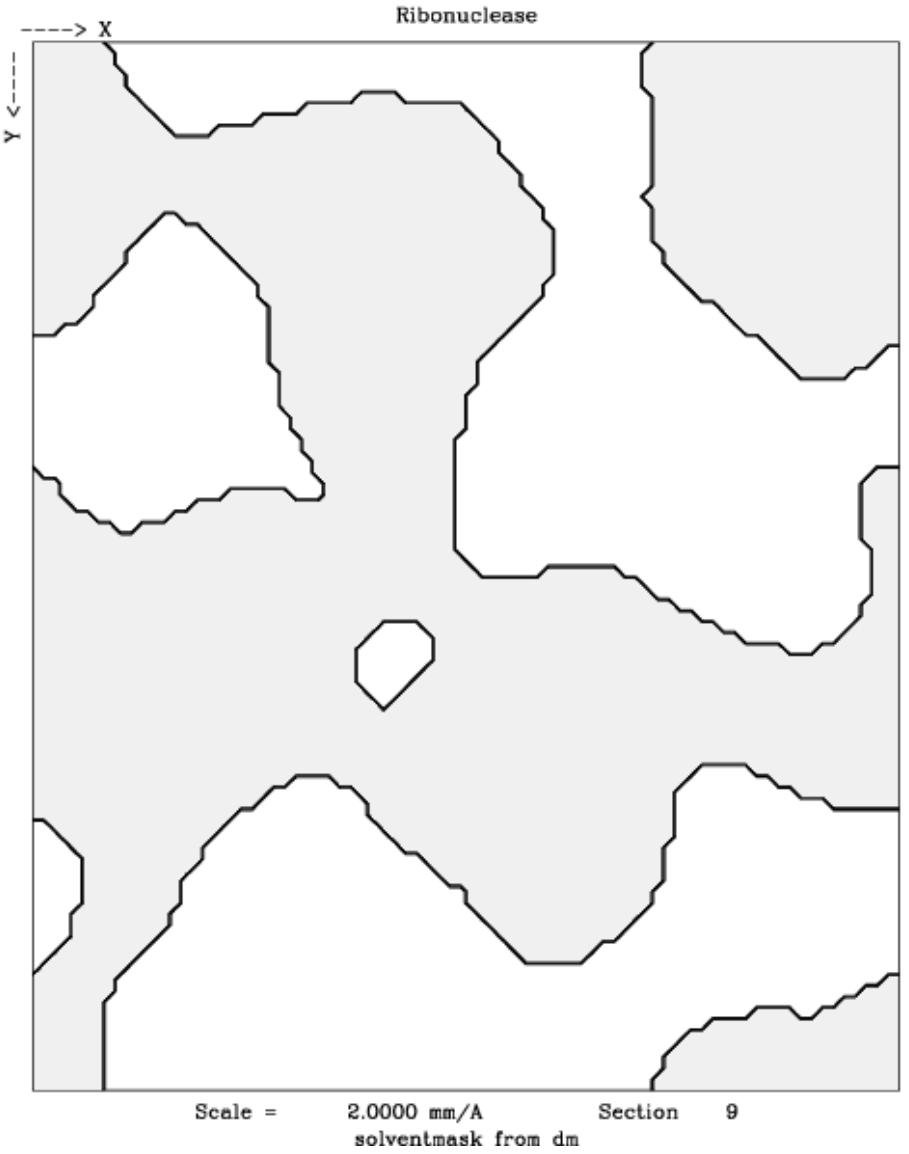
- Solvent flattening
- Histogram matching
- Non-crystallographic symmetry (NCS) averaging



K Cowtan, kevin.cowtan@york.ac.uk

Solvent flattening





K Cowtan, kevin.cowtan@york.ac.uk

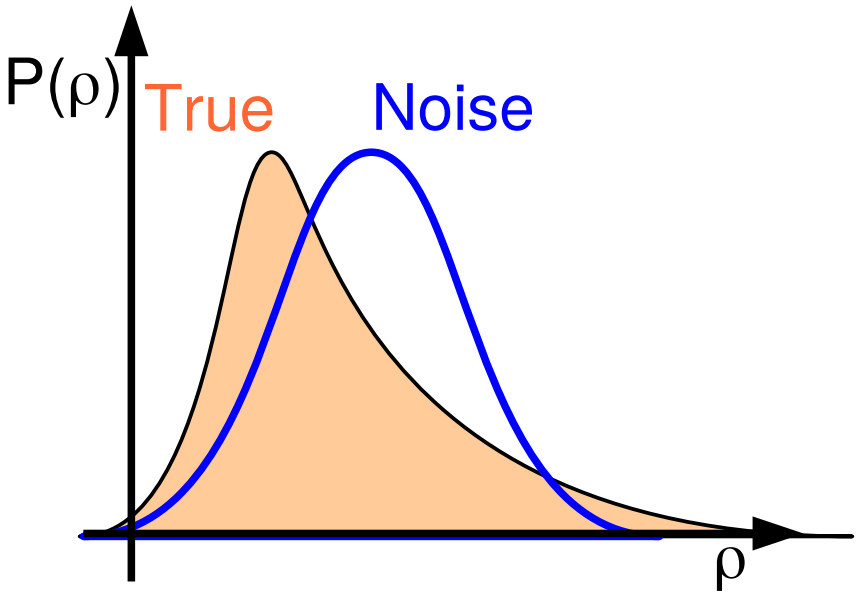
Histogram matching

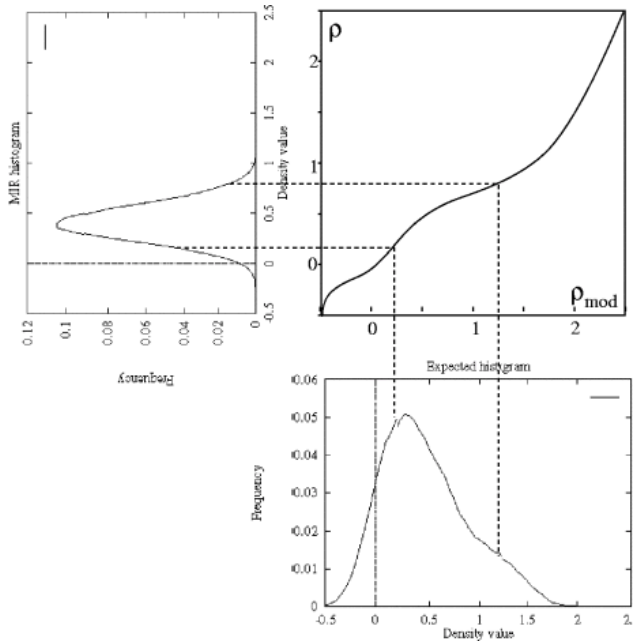
A technique from image processing for modifying the protein region.

- Noise maps have Gaussian histogram.
- Well phased maps have a skewed distribution: sharper peaks and bigger gaps.

Sharpen the protein density by a transform which matches the histogram of a well phased map.

Useful at better than 4A.

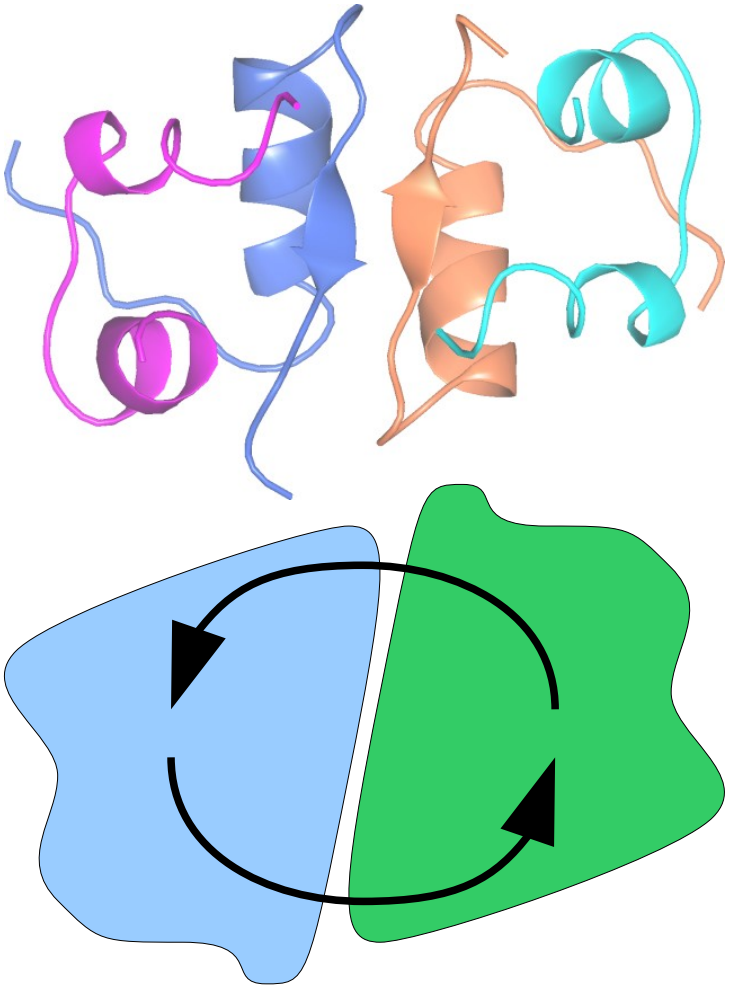




K Cowtan, kevin.cowtan@york.ac.uk

Non-crystallographic symmetry

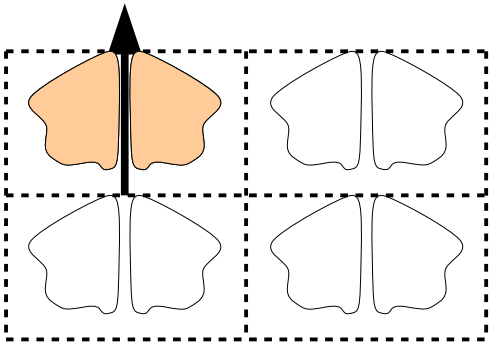
- If the molecule has internal symmetry, we can average together related regions.
- In the averaged map, the signal-noise level is improved.
- If a full density modification calculation is performed, powerful phase relationships are formed.
- With 4-fold NCS, can phase from random!



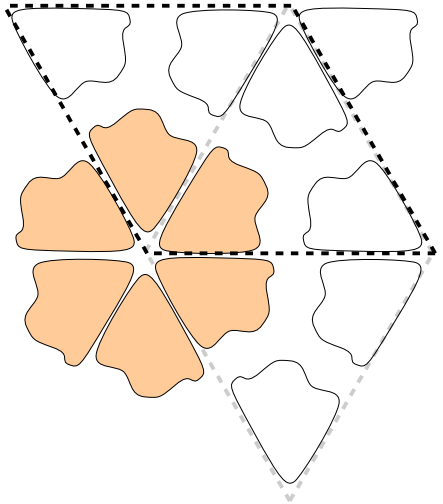
K Cowtan, kevin.cowtan@york.ac.uk

Non-crystallographic symmetry

Crystallographic

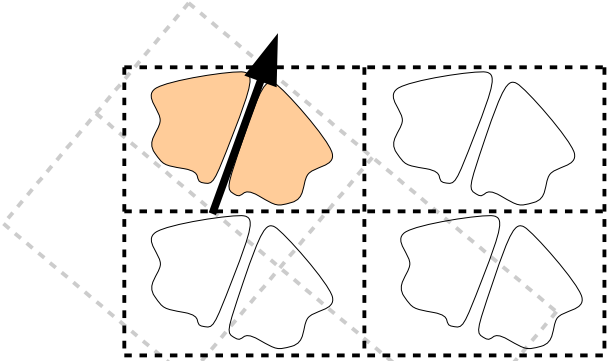


Aligned
2-fold

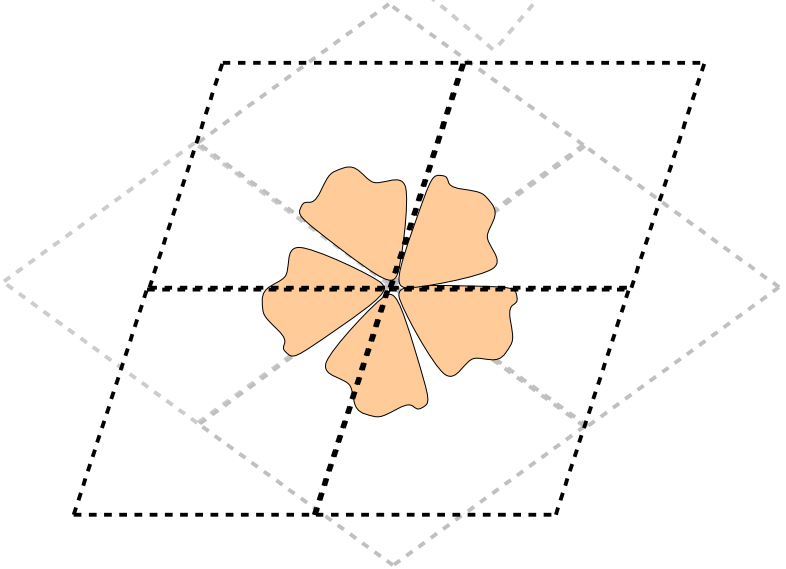


Aligned
6-fold

Non-crystallograpic



Unaligned
2-fold



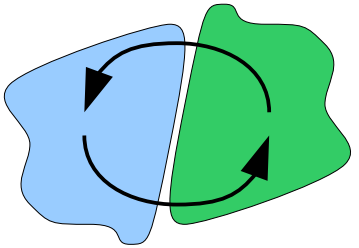
Aligned
5-fold

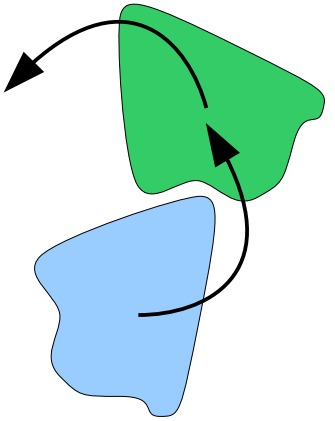
K Cowtan, kevin.cowtan@york.ac.uk

Non-crystallographic symmetry

Useful terms:

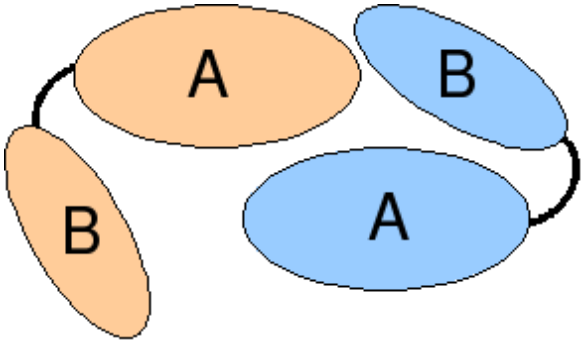
- Proper and improper NCS: (closed and open)

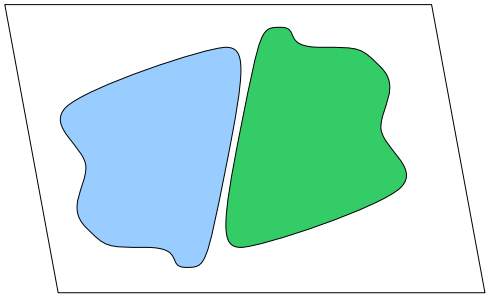


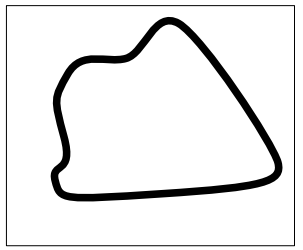


- Multi-domain averaging:

- Multi-crystal averaging:



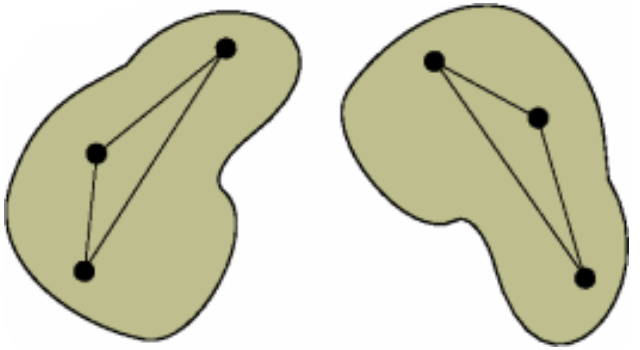




K Cowtan, kevin.cowtan@york.ac.uk

Non-crystallographic symmetry

- How do you know if you have NCS?
  - Cell content analysis – how many monomers in ASU?
  - Self-rotation function.
  - Difference Pattersons (pseudo-translation only).
- How do you determine the NCS?
  - **From heavy atoms.**
  - From initial model building.
  - From molecular replacement.
  - *From density MR (hard).*
- Mask determined automatically.

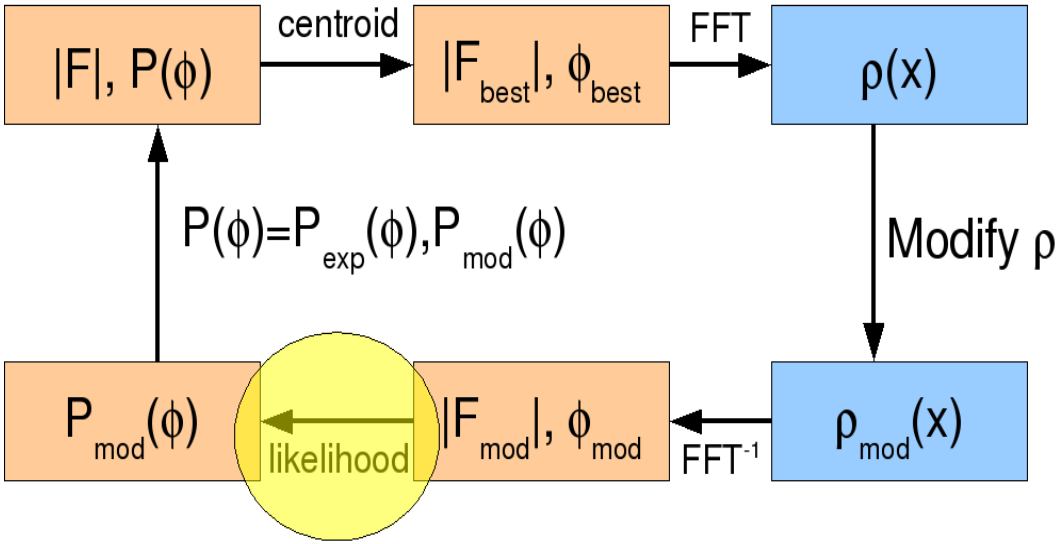


K Cowtan, kevin.cowtan@york.ac.uk

Estimating phase probabilities

Problem: How do we go from a single phase estimate to a full phase probability distribution?

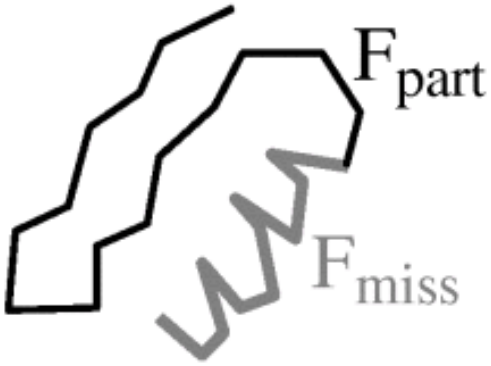
- We need to make an estimate of the error in the estimated phase.
- The errors in the phases are a parameter of the model itself, and may be estimated by likelihood methods.



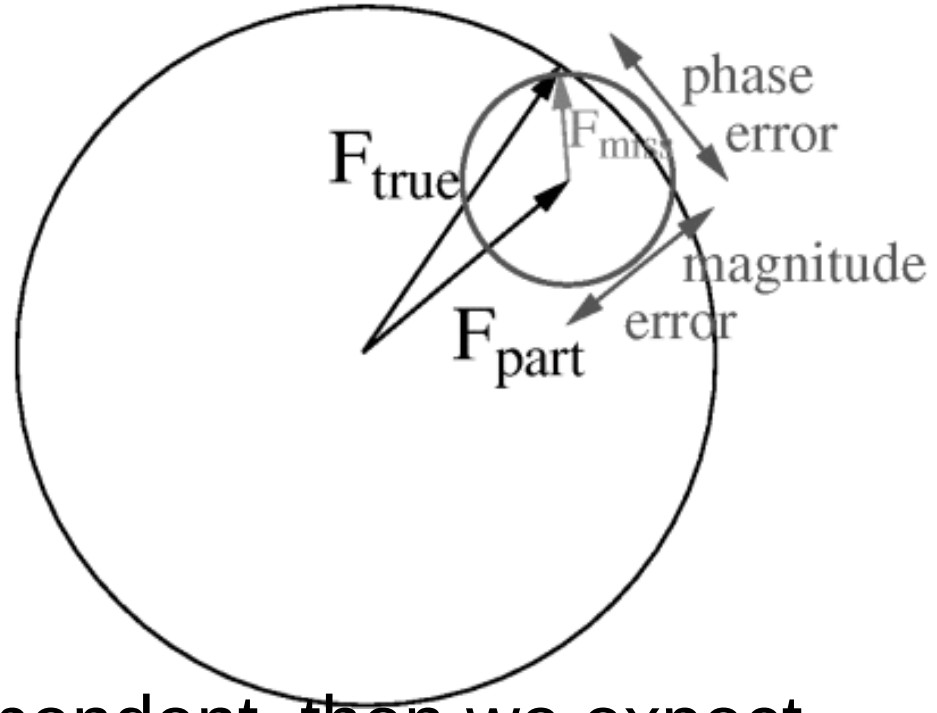
K Cowtan, kevin.cowtan@york.ac.uk

Estimating phase probabilities

Sim/σ_A weighting:



$F_{true} = F_{part} + F_{miss}$

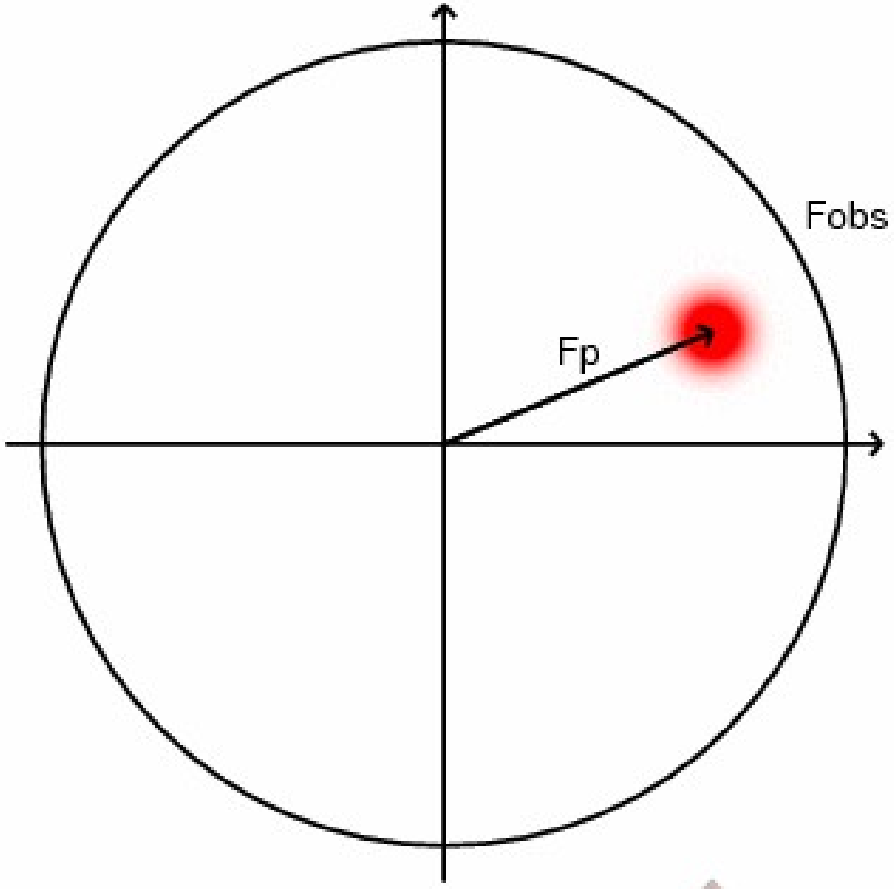


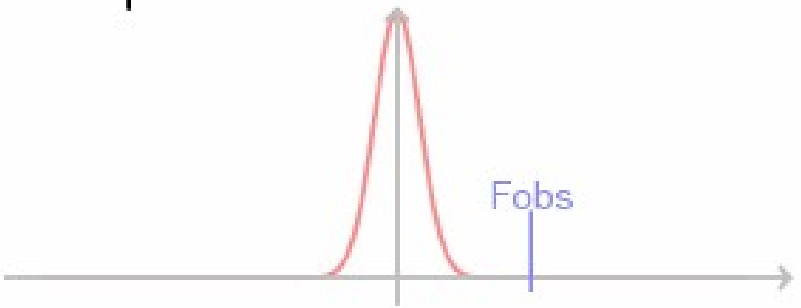
We know $|F_{true}|$, $|F_{part}|$, ϕ_{part}

Assuming ϕ_{part}, ϕ_{miss} are independent, then we expect the difference in magnitudes between $|F_{true}|$ and $|F_{part}|$, averaged over reflections, to give an indication of the phase error.

K Cowtan, kevin.cowtan@york.ac.uk

Estimating phase probabilities





K Cowtan, kevin.cowtan@york.ac.uk

Combining phase probabilities

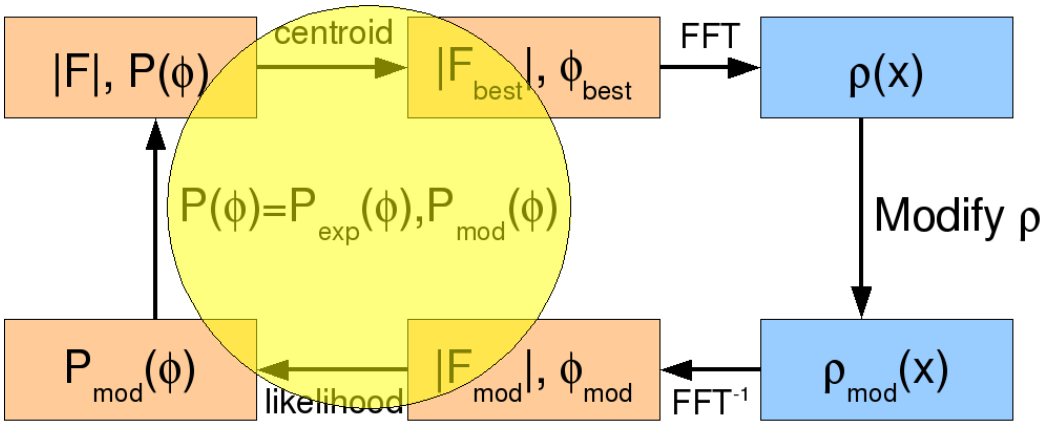
Once we have an estimate for the error in ϕ_{mod}, we can construct a probability distribution $P_{mod}(\phi)$.

The the next cycle can be started with

$$P_{new}(\phi) = P_{exp}(\phi)P_{mod}(\phi)$$

Problem: $P_{exp}(\phi)$ and $P_{mod}(\phi)$ are not independent.

The result is bias, increasing with cycle.



K Cowtan, kevin.cowtan@york.ac.uk

Bias reduction

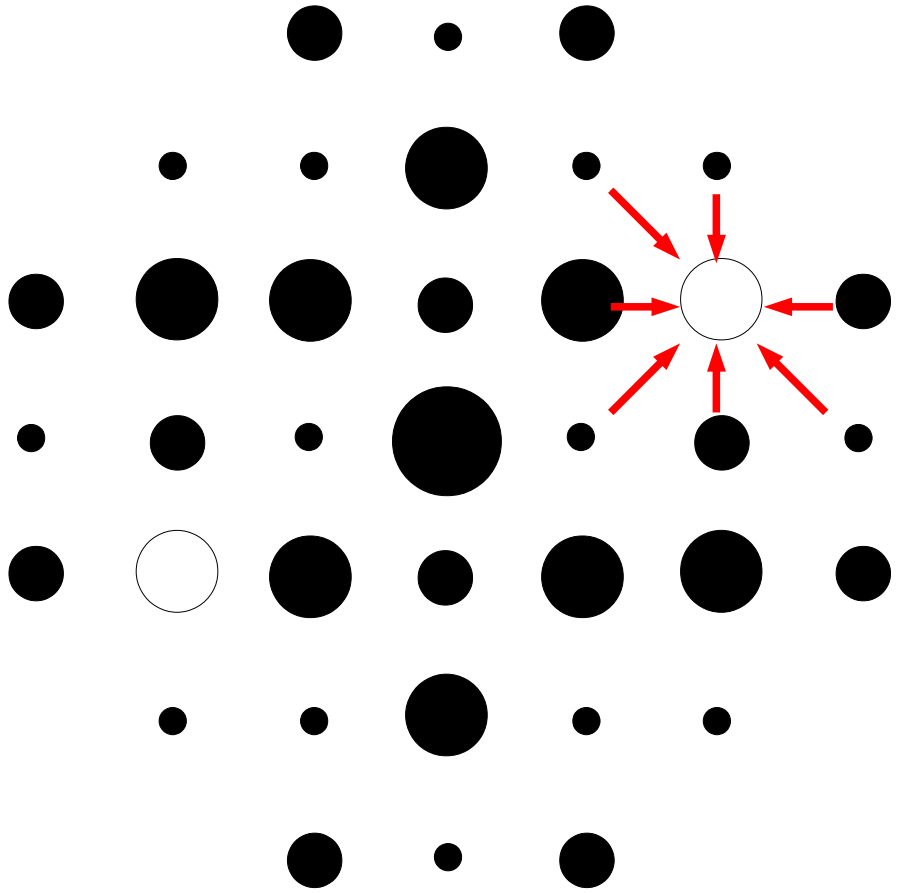
Solution:

Make each reflection only dependent on the other reflections in the diffraction pattern, and not on its own initial value.

Omit one reflection at a time, and use only the modified value of the omitted reflection. (Very slow.)

But can be implemented efficiently:

- Solvent flipping
- The γ-correction



K Cowtan, kevin.cowtan@york.ac.uk

Density modification in Parrot

Builds on existing ideas:

- DM:
  - Solvent flattening
  - Histogram matching
  - NCS averaging
  - Perturbation gamma
- Solomon:
  - Gamma correction
  - Local variance solvent mask
  - Weighted averaging mask

K Cowtan, kevin.cowtan@york.ac.uk

Density modification in Parrot

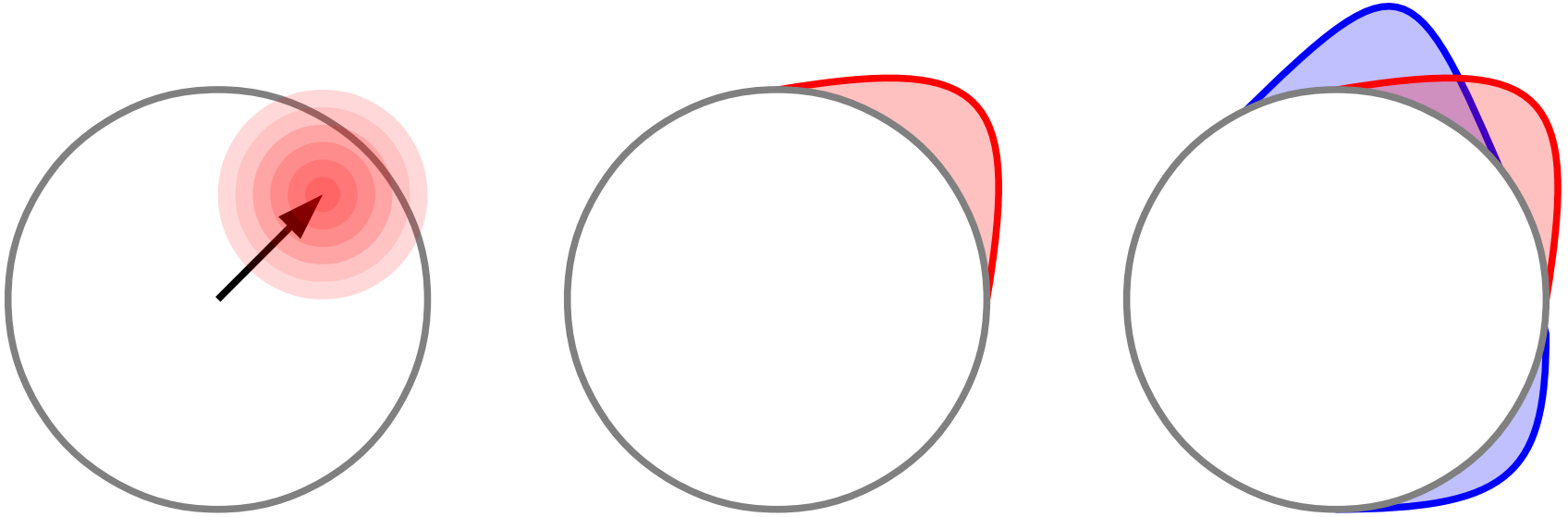
New developments:

- MLHL phase combination
  - (as used in refinement: *refmac*, *phenix.refine*)
- Anisotropy correction
- Problem-specific density histograms
  - (rather than a standard library)
- Pairwise-weighted NCS averaging...

K Cowtan, kevin.cowtan@york.ac.uk

Estimating phase probabilities

Traditional approach: Rice likelihood function



Estimate the accuracy of the modified F/phase

Turn this into a phase probability distribution

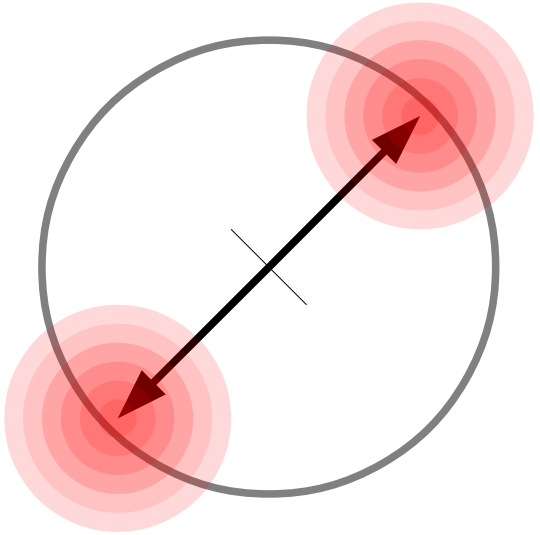
Combine with the experimental phase probability

The estimate for the accuracy of the modified F/phase come from the agreement between the modified F and the observed F. **Source of bias.**

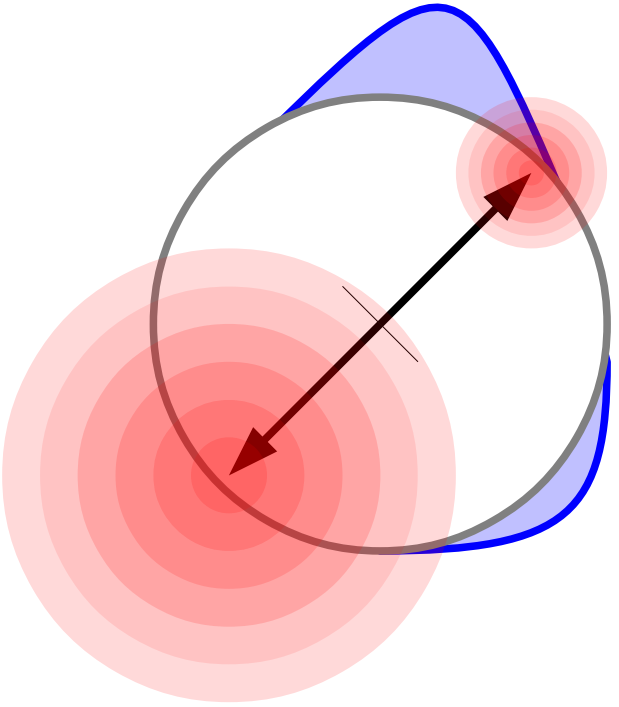
K Cowtan, kevin.cowtan@york.ac.uk

Estimating phase probabilities

Problem:



Error estimation does not take into account experimental phase information



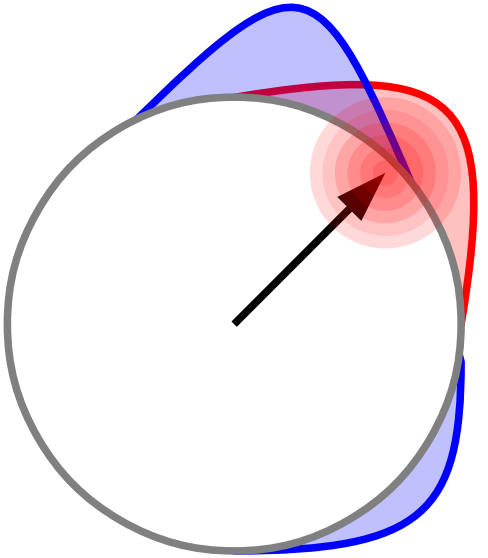
The experimental data tells us that the probable error is different in the two cases

Using the additional information from the phases improves the error model and reduces bias.

K Cowtan, kevin.cowtan@york.ac.uk

Estimating phase probabilities

Solution:
MLHL-type likelihood target function.



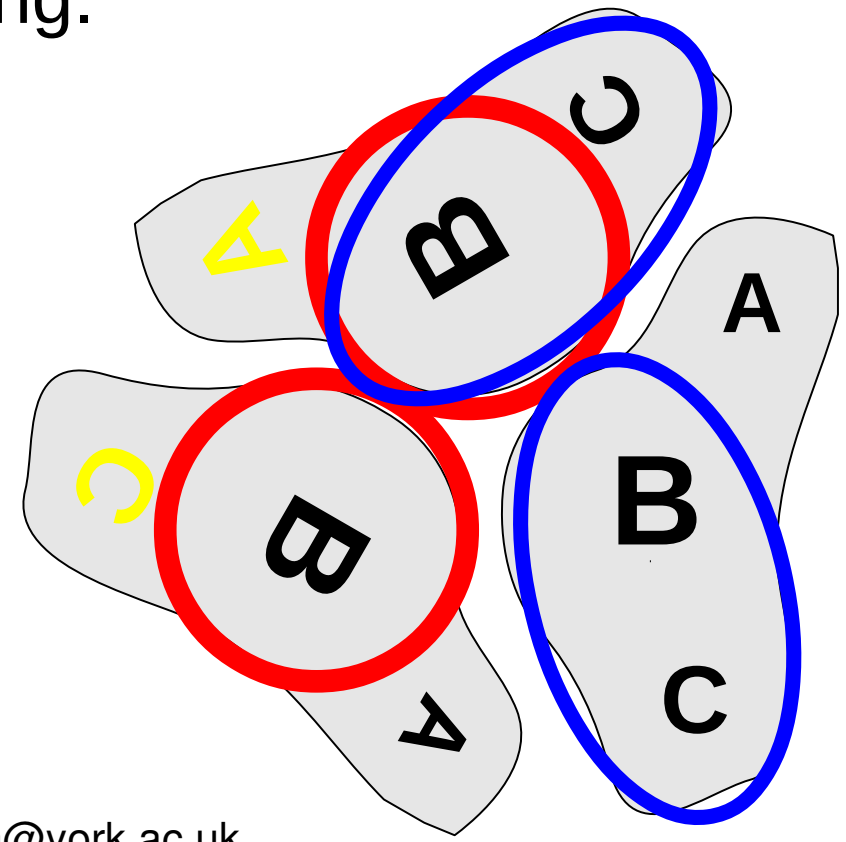
Perform the error estimation and phase combination in a single step, using a likelihood function which incorporates the experimental phase information as a prior.
This is the same MLHL-type like likelihood refinement target used in modern refinement software such as *refmac* or *phenix.refine*.

K Cowtan, kevin.cowtan@york.ac.uk

Recent Developments:

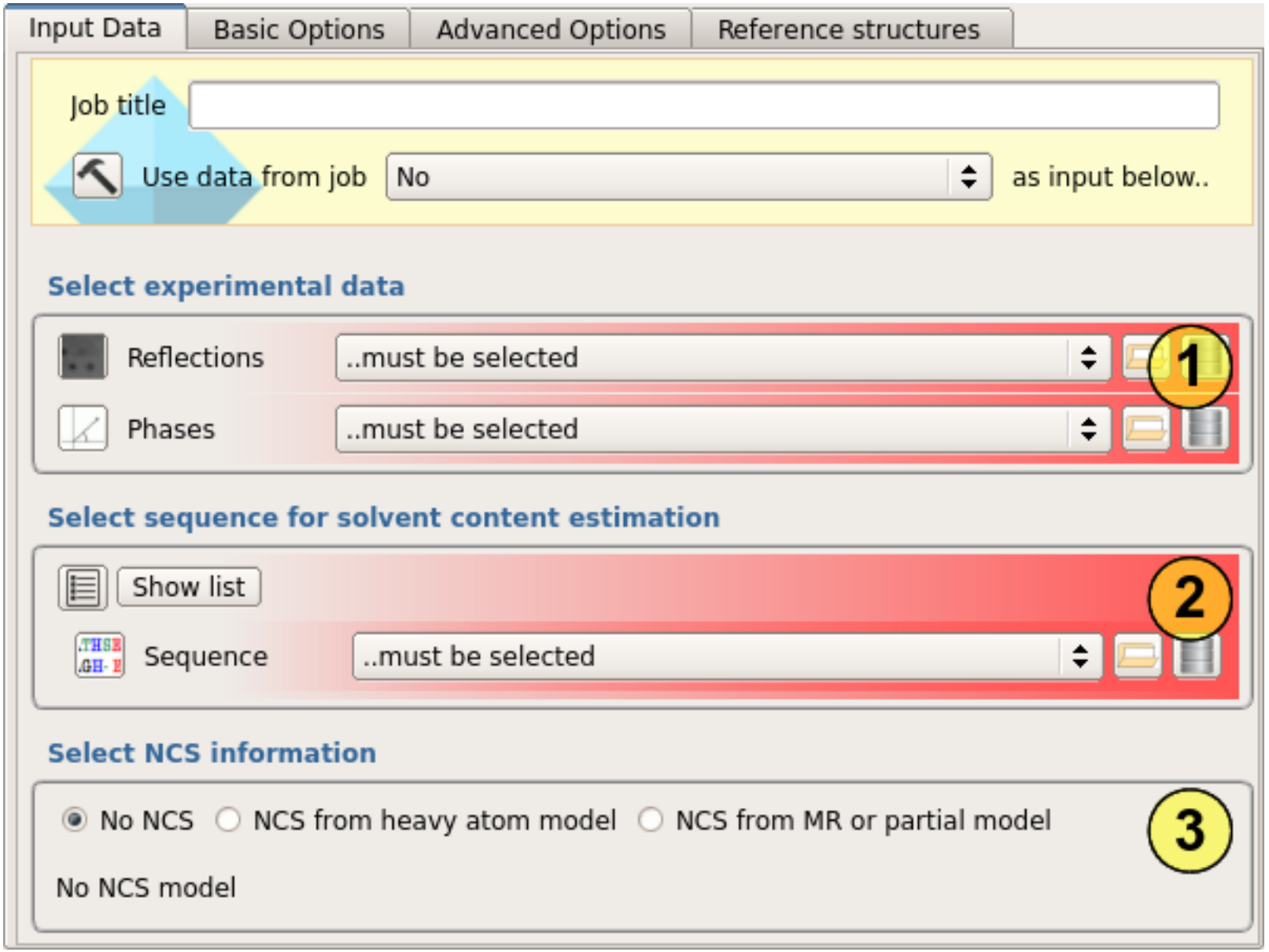
Pairwise-weighted NCS averaging:

- Average each pair of NCS related molecules separately with its own mask.
- Generalisation and automation of multi-domain averaging.



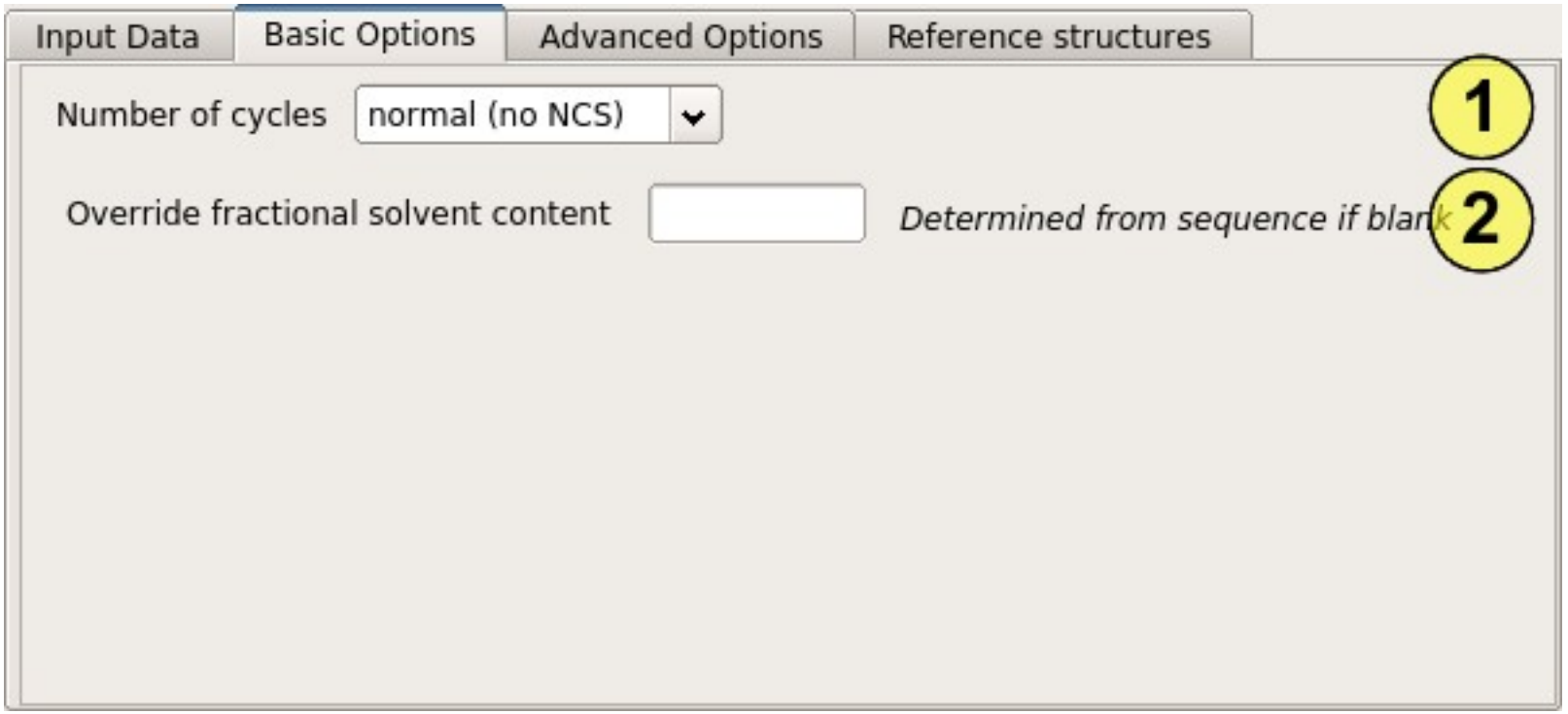
K Cowtan, kevin.cowtan@york.ac.uk

Parrot



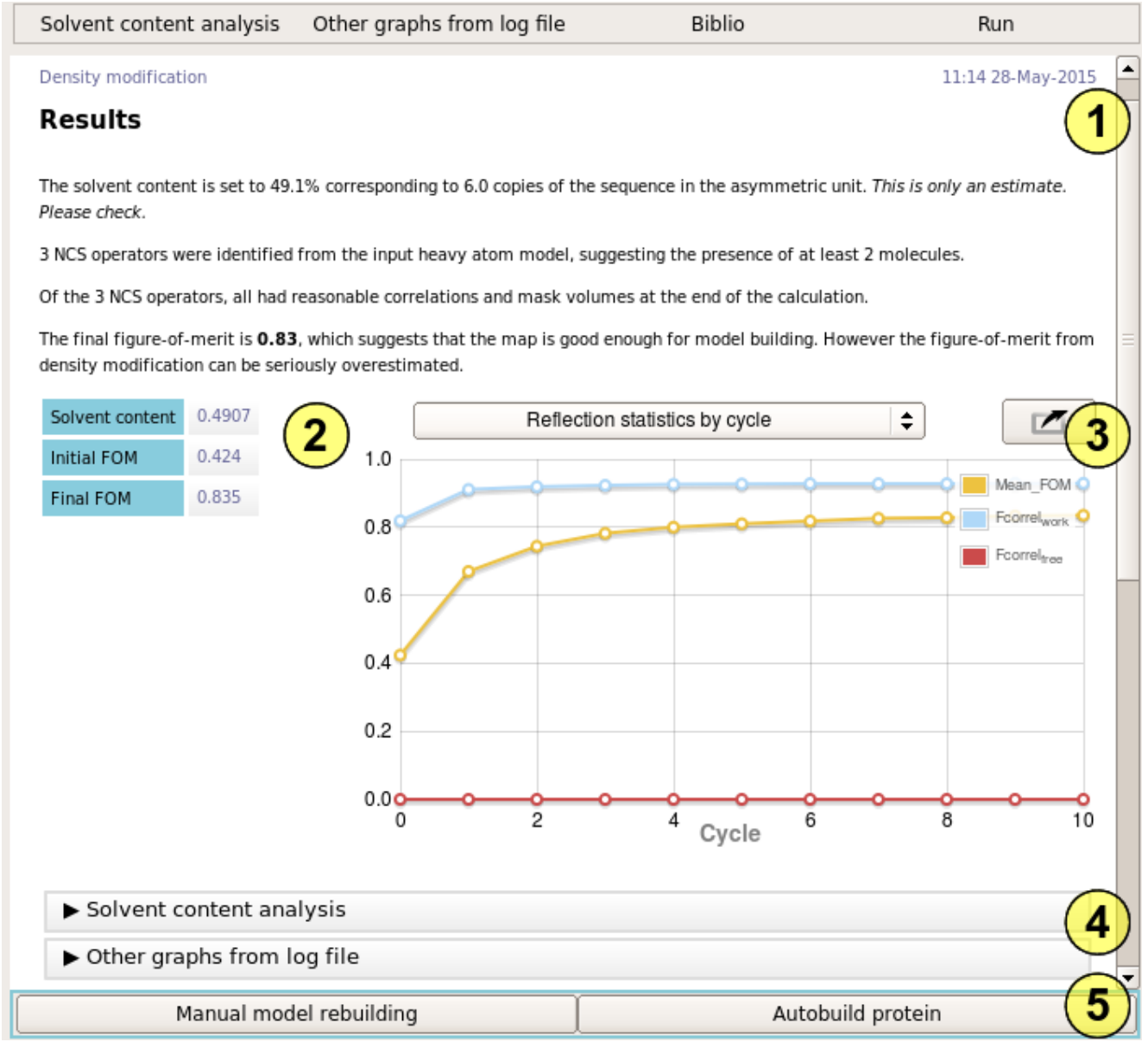
K Cowtan, kevin.cowtan@york.ac.uk

Parrot



K Cowtan, kevin.cowtan@york.ac.uk

Parrot



K Cowtan, kevin.cowtan@york.ac.uk

Parrot

Summary:

A new classical density modification program, employing the latest techniques.

- Fully automated
- Fast

K Cowtan, kevin.cowtan@york.ac.uk

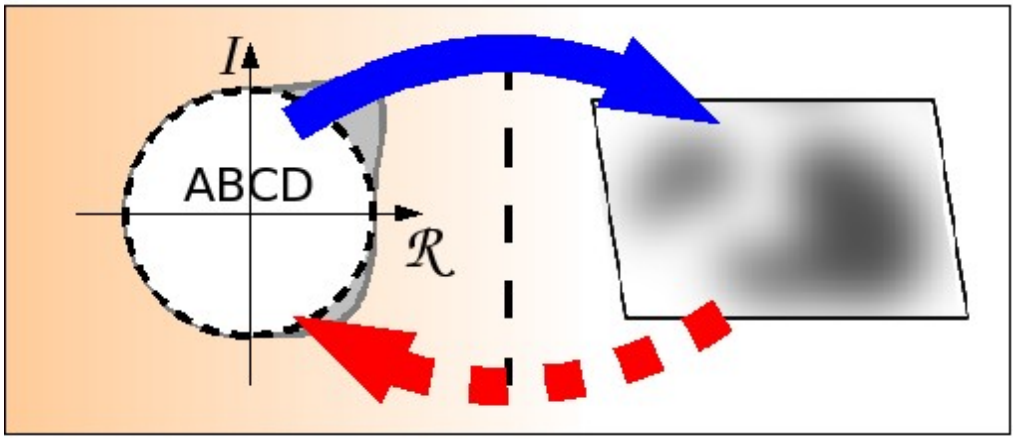
Density Modification

Kevin Cowtan, York.

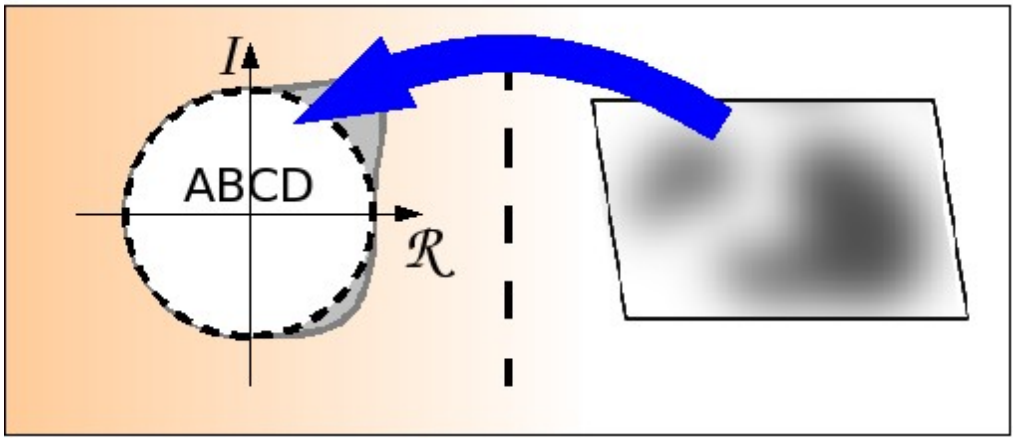
Statistical density modification:
e.g. Resolve

Density modification

- Traditional density modification:
  *Take the **phases** to the **mask**.*
  Use them to calculate a map.
  But how do we get back to:
  - reciprocal space?
  - probabilities?



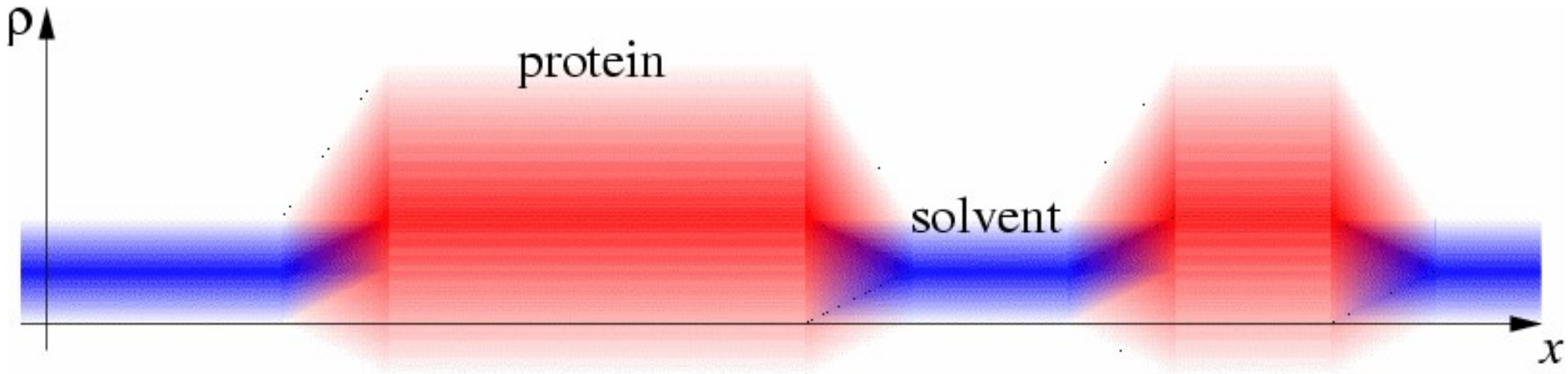
- Statistical density modification:
  *Take the **mask** to the **phases**.*
  - First convert mask to probability.
  - Then transform that probability.



K Cowtan, kevin.cowtan@york.ac.uk

Statistical density modification

- Form a statistical description of expected map features.

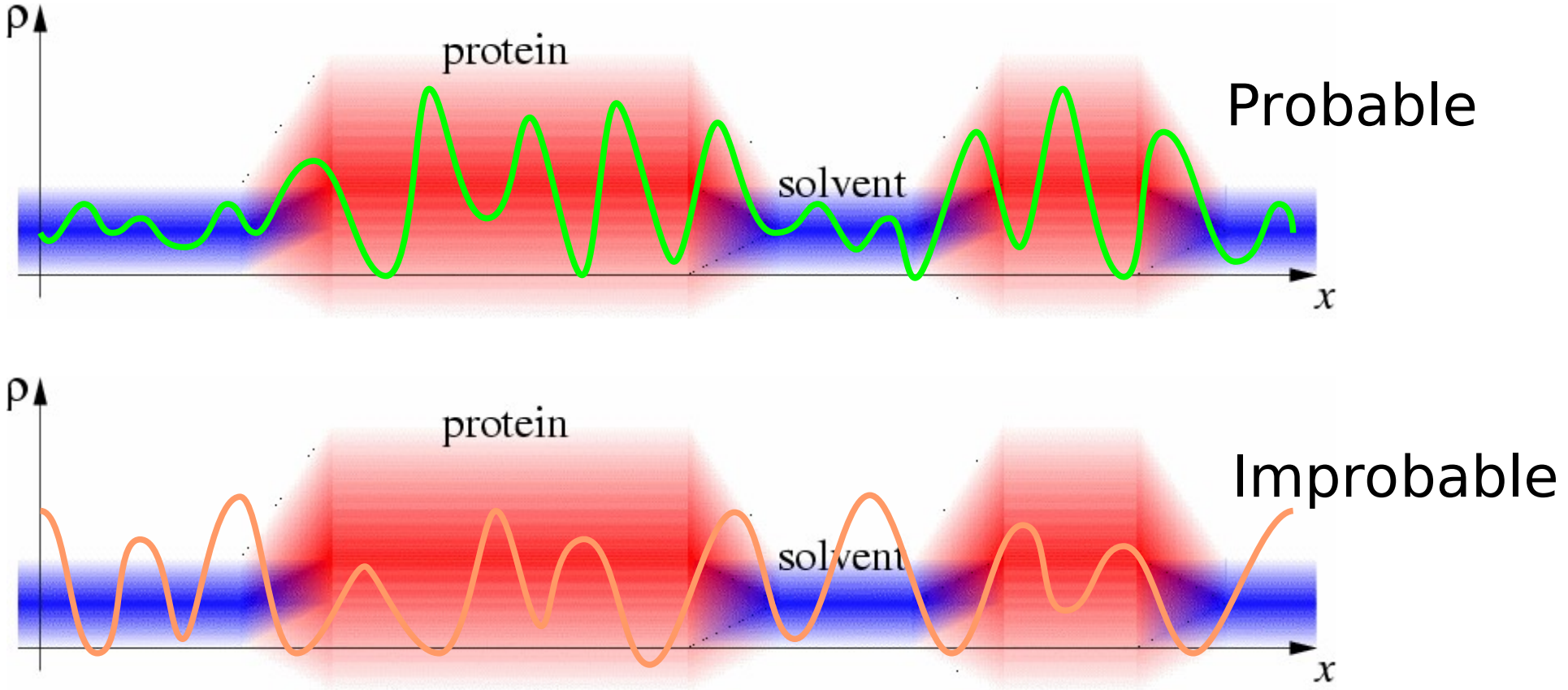


- e.g.
  - Protein has higher mean, and is more peaky (higher variance)
  - Solvent has lower mean, and is flatter (lower variance)

K Cowtan, kevin.cowtan@york.ac.uk

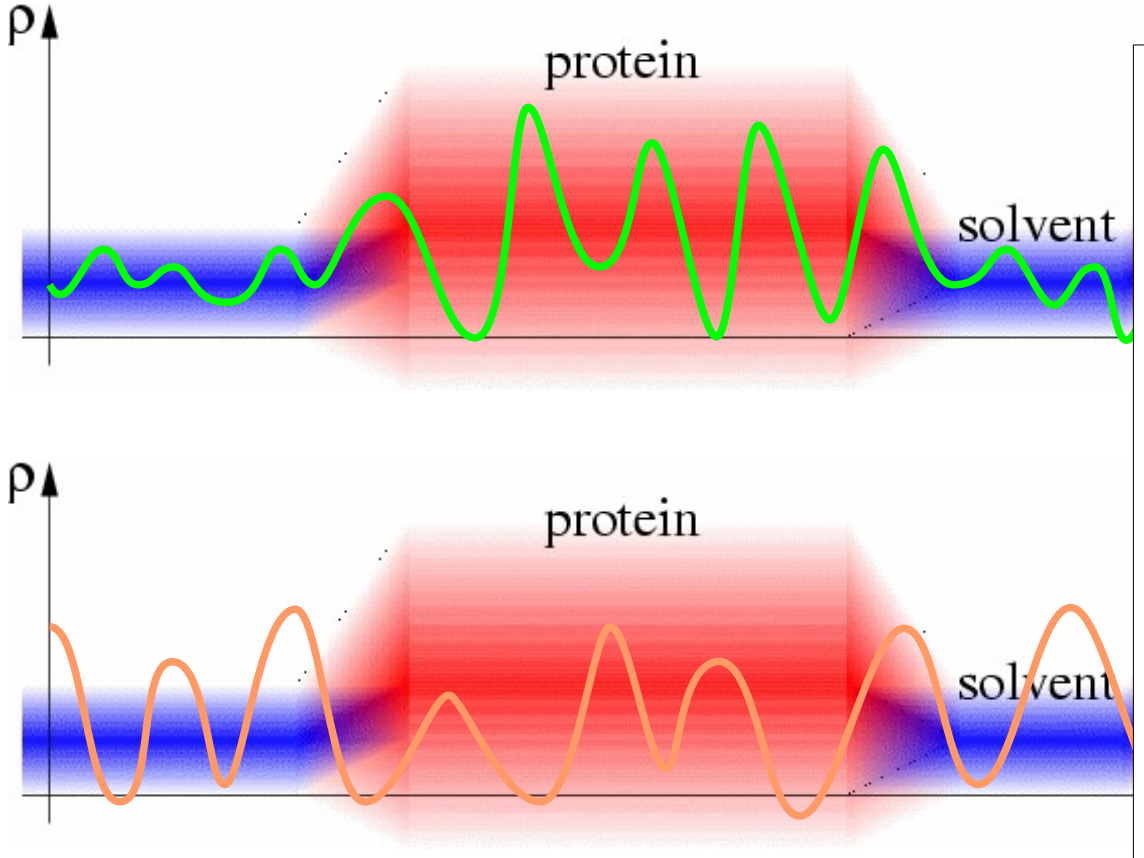
Statistical density modification

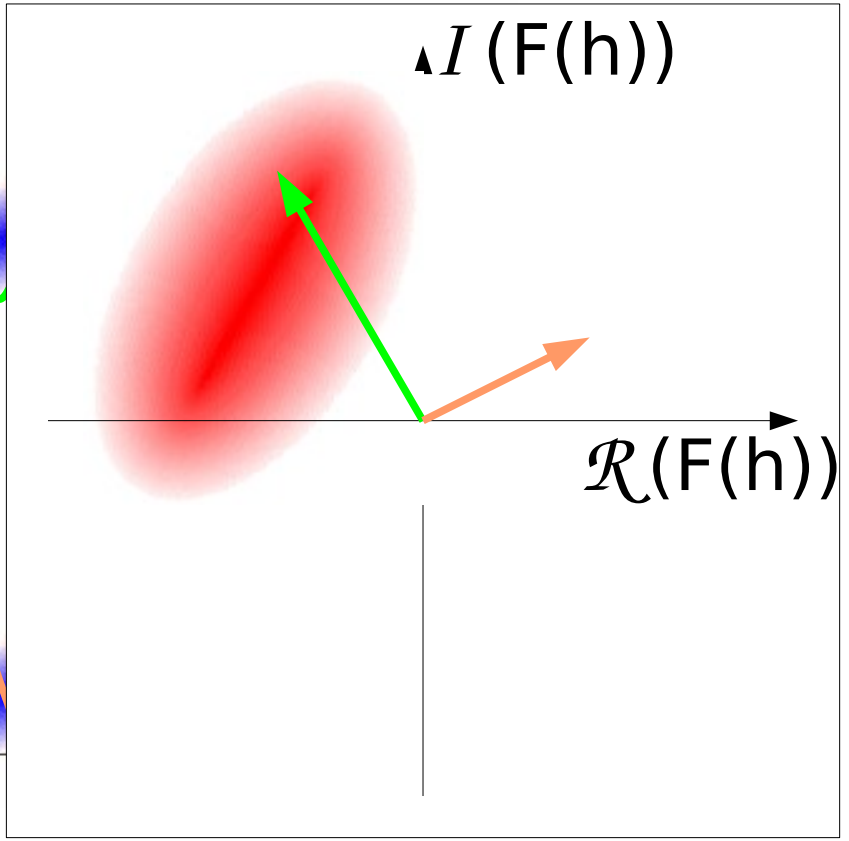
- Probability of a map is determined by how well it fits these distributions:



Statistical density modification

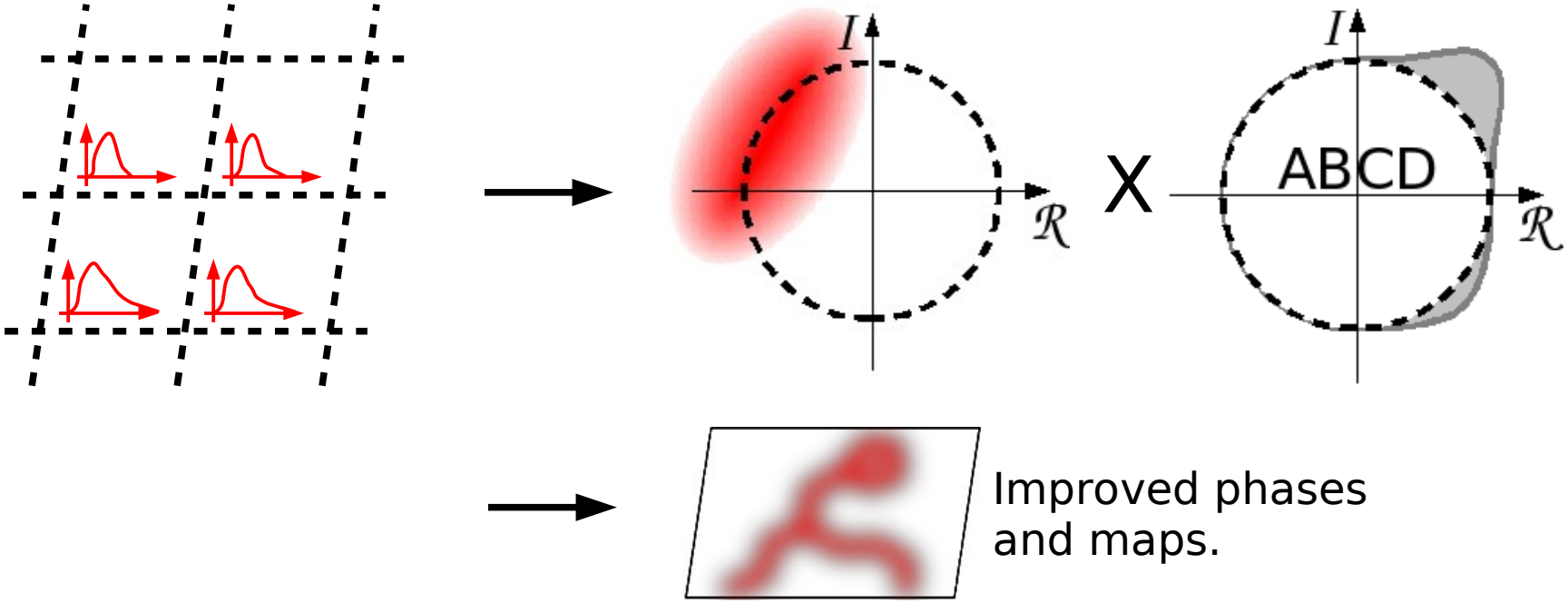
- Probability of each structure factor is given by the probability of the corresponding map.





Statistical density modification

- Obtain per-grid density probability distributions.
- Transform to reciprocal space.
- Combine with experimental phases.
  - Map probability becomes phase probability distribution.



Bricogne (1992) Proc. CCP4 Study Weekend
Bricogne (1997) Methods in Enzymology

K Cowtan, kevin.cowtan@york.ac.uk

Statistical density modification

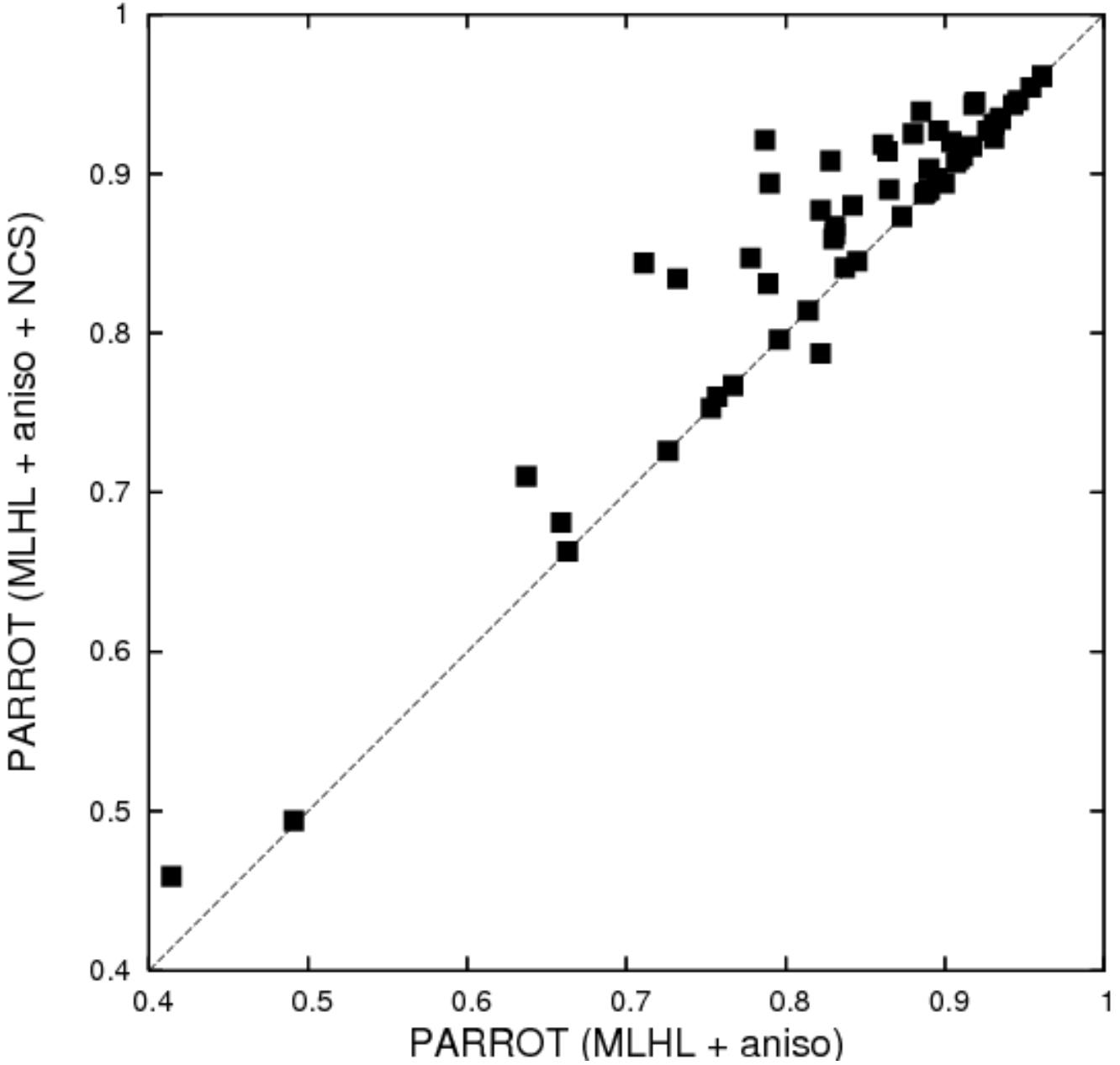
Advantages:

- Reduced bias.
- Better phases.

Disadvantages:

- Slow.
- Latest classical methods comparable.

Parrot: simple vs NCS averaged



Map correlations

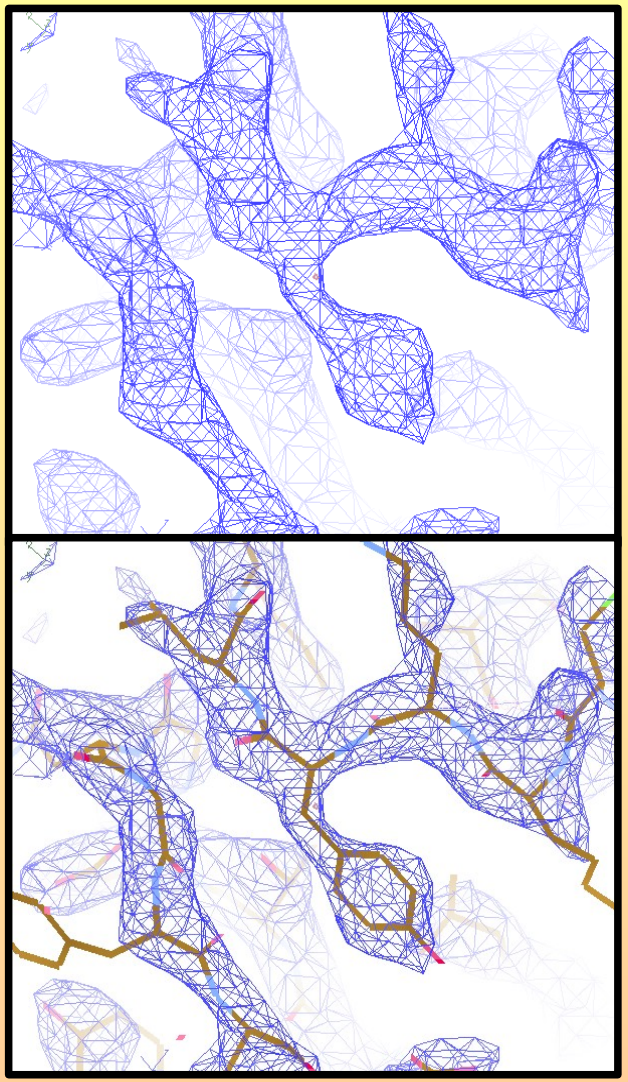
Comparing with and without NCS averaging.

K Cowtan, kevin.cowtan@york.ac.uk

Model Building

Model building software:

- Proteins:
  - Buccaneer
  - ARP/wARP
  - Phenix autobuild
- Nucleic acids:
  - Nautilus/Coot
  - ARP/wARP
  - Phenix autobuild



K Cowtan, kevin.cowtan@york.ac.uk

Acknowledgements

Help:

- JCSG data archive: www.jcsg.org
- Garib Murshudov, Eleanor Dodson, Paul Emsley, Randy Read, Clemens Vonrhein and many others

Funding:

- The Royal Society, BBSRC, CCP4

K Cowtan, kevin.cowtan@york.ac.uk