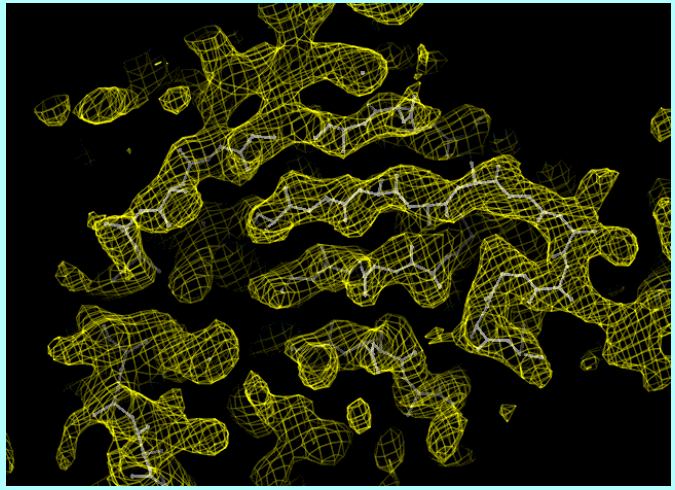
Phenix

Statistical density modification in macromolecular crystallography

Tom Terwilliger
Los Alamos National Laboratory

Statistical density modification (RESOLVE)

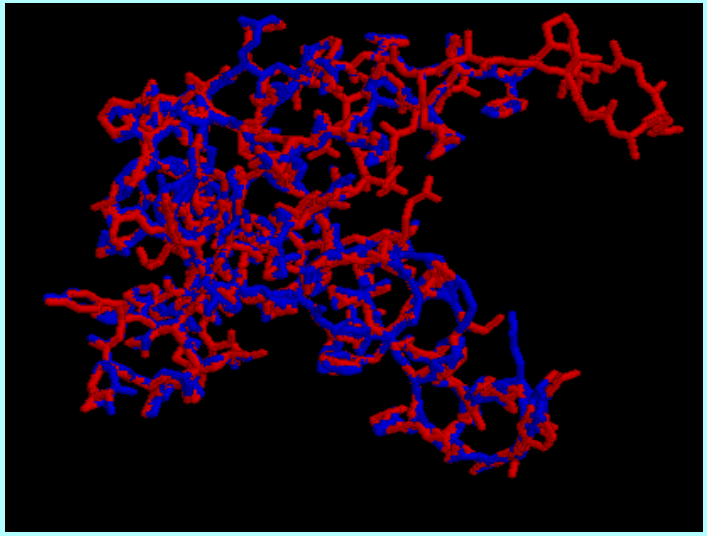


•Principle: phase probability information from probability of the map and from experiment:

•$P(\phi) = P_{\text{map probability}}(\phi)\, P_{\text{experiment}}(\phi)$

•"Phases that lead to a believable map are more probable than those that do not"

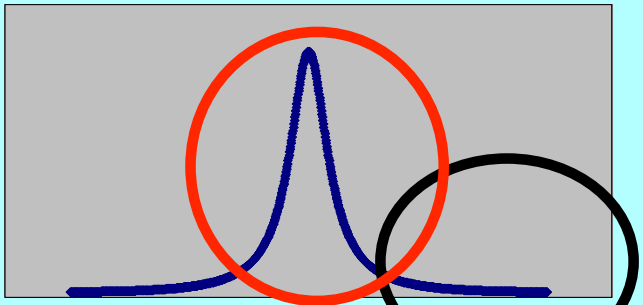
•A believable map is a map that has…
•a relatively flat solvent region
•NCS (if appropriate)
•A distribution of densities like those of model proteins



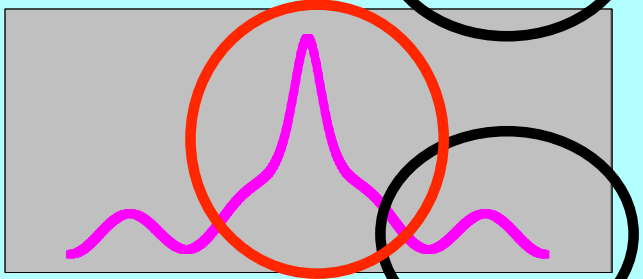
•Method:
•calculate how map probability varies with electron density ρ
•deduce how map probability varies with phase ϕ
•combine with experimental phase information

Map probability phasing: Getting a new probability distribution for each phase given estimates of all others

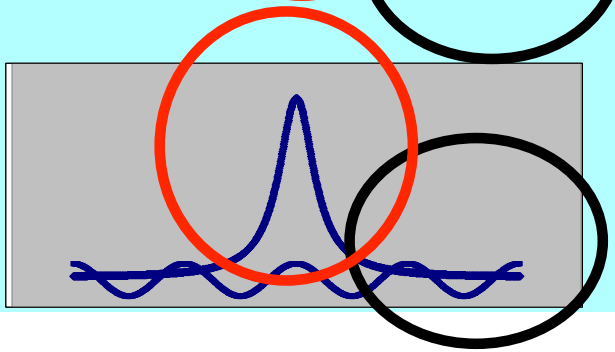
1. Identify expected features of map (flat far from center)
2. Calculate map with current estimates of all structure factors except one (k)
3. Test all possible phases ϕ for structure factor k (for each phase, calculate new map including k)
4. Probability of phase ϕ estimated from agreement of map with expectations
5. Phase probability of reflection k from map is independent of starting phase probability because reflection k is omitted from the map



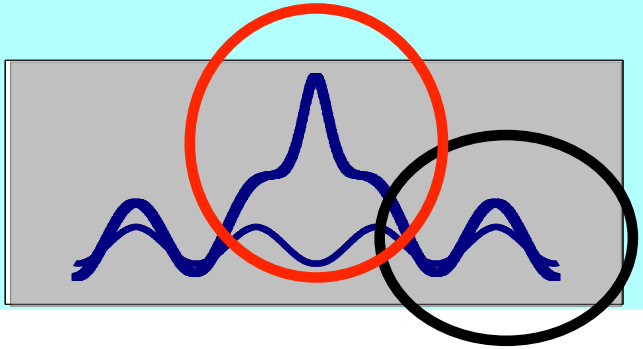
A function that is (relatively) flat far from the origin



Function calculated from estimates of all structure factors but one (k)



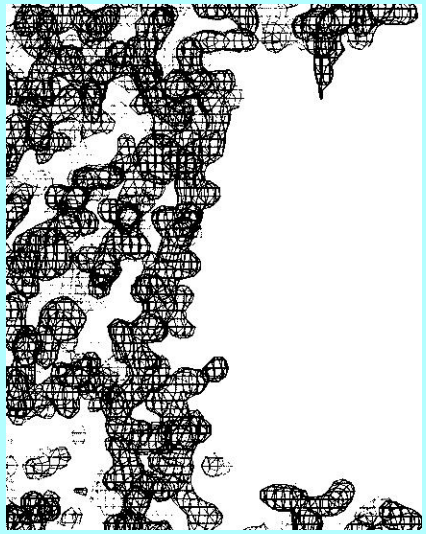
Test each possible phase of structure factor k. P(ϕ) is high for phase that leads to flat region



A map-probability function – allowing different weighting of information from different parts of the map

Log-probability of the map is sum over all points in map of local log-probability

$$LL^{MAP}(\{\mathbf{F_h}\}) \approx \frac{\mathrm{N_{REF}}}{\mathrm{V}} \int_{\mathbf{V}} LL(\rho(\mathbf{x}, \{\mathbf{F_h}\}))\mathrm{d}^3\mathbf{x}$$



A map with a flat (blank) solvent region is a likely map

Local log-probability is believability of the value of electron density (ρ(x)) found at this point

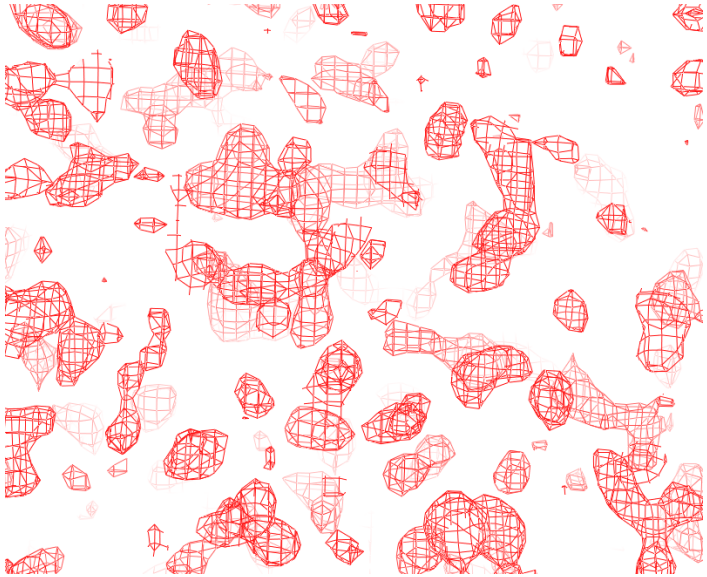
$$LL(\rho(\mathbf{x}, \{\mathbf{F_h}\})) = \ln[p(\rho(\mathbf{x})|PROT)p_{PROT}(\mathbf{x}) + p(\rho(\mathbf{x})|SOLV)p_{SOLV}(\mathbf{x})]$$

If the point is in the PROTEIN region, most values of electron density (ρ(x)) are believable

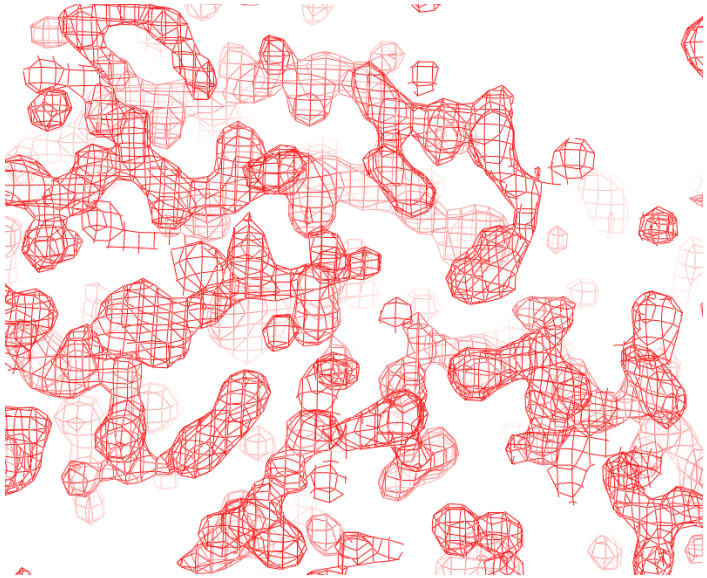
If the point is in the SOLVENT region, only values of electron density near zero are believable

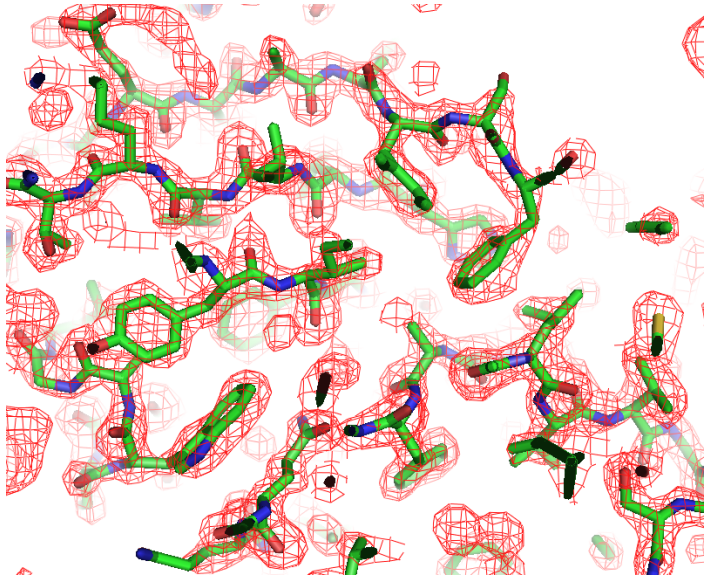
Statistical density modification
(nsf-N SAD map , 2Å, no NCS, 50% solvent)

Phaser SAD map
(CC=0.43)



Phaser +RESOLVE
(CC=0.79)





Statistical density modification with cross-crystal averaging

Cell receptor at 3.5/3.7 Å. Data courtesy of J. Zhu

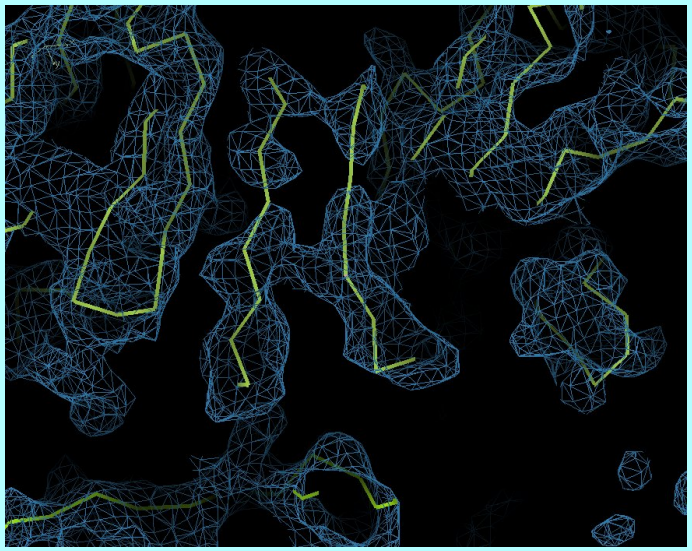
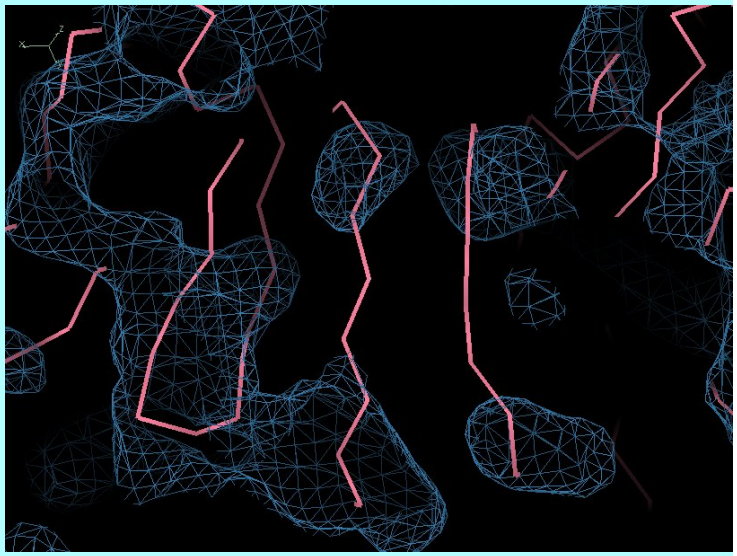
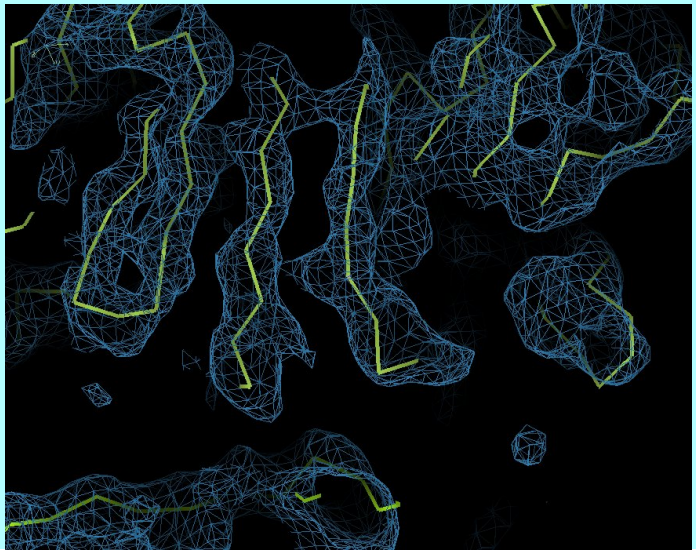
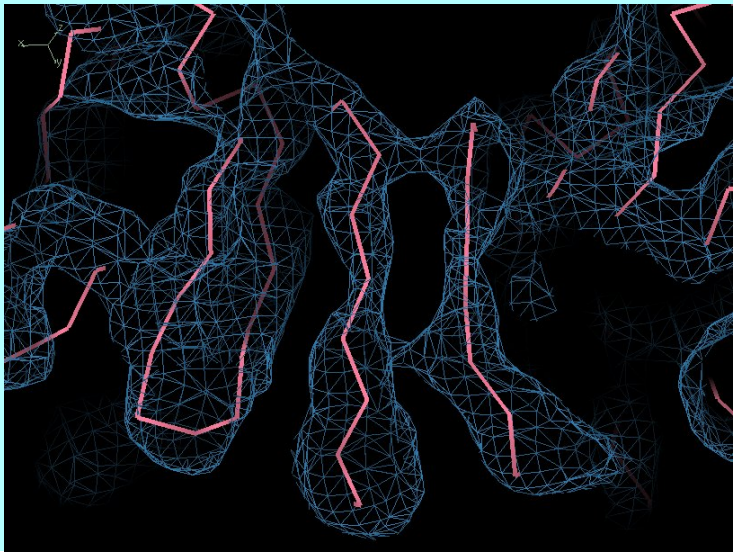
	Crystal 1 (4 copies)	Crystal 2 (2 copies)
RESOLVE density modification		
PHENIX Multi-crystal averaging		

Image enhancement using local feature recognition

Electron density maps of proteins have many features in common

- Connected density
- Preferred distances for spacing between regions of high density
- Preferred shapes of density

Starting image in red

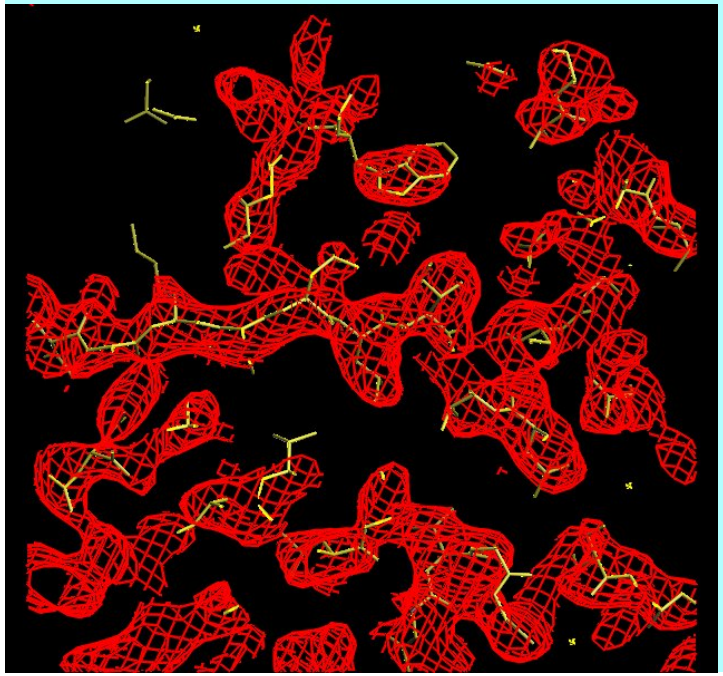


Image improved using expectations about local features

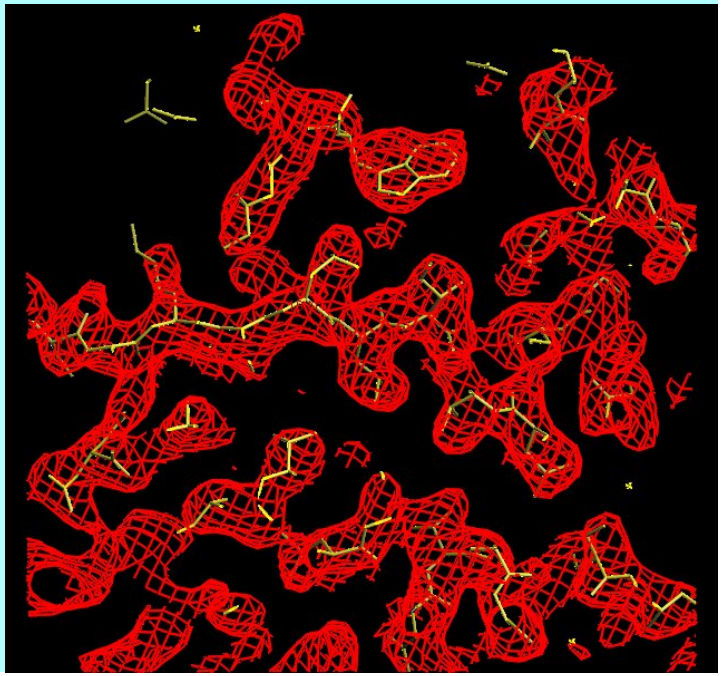


Image enhancement using local feature recognition

Approach:

- Use the pattern of density near a point x to estimate the value of density at x
- Combine new estimate of density with previous one to improve the overall image

Starting image in red

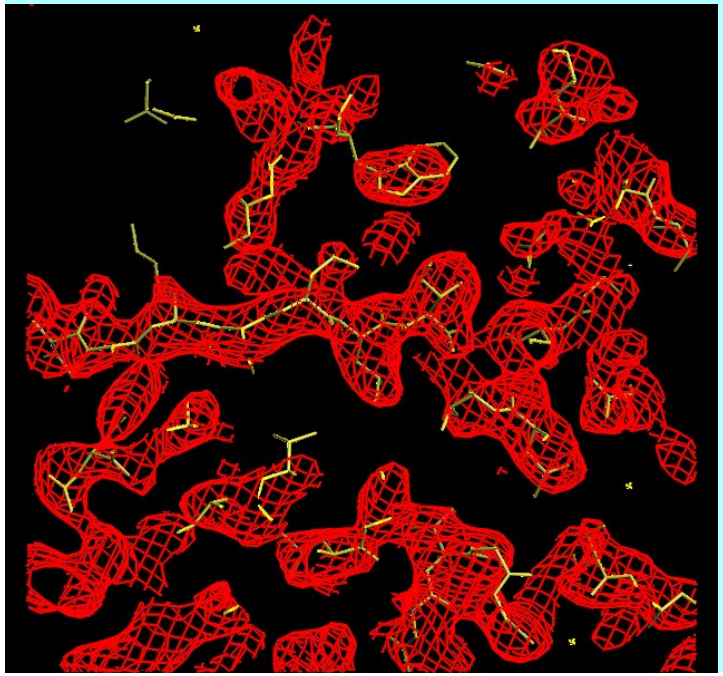


Image improved using expectations about local features

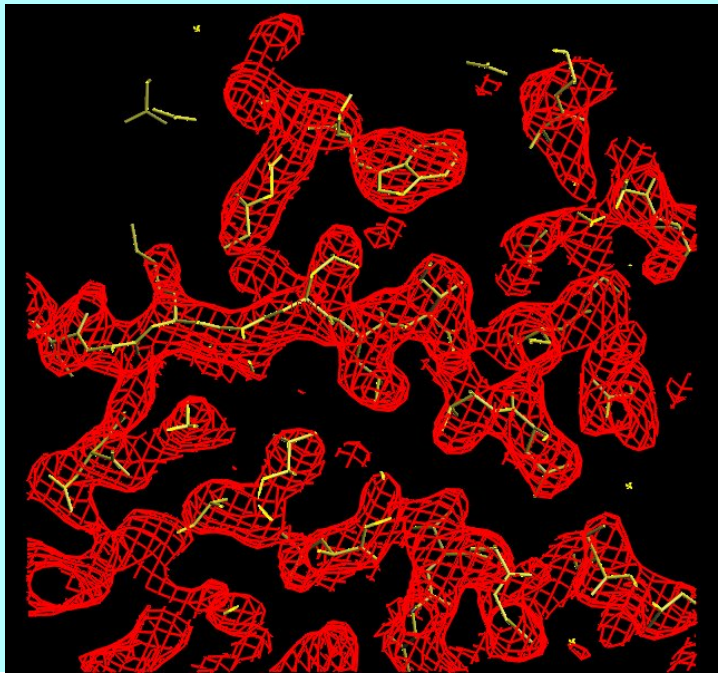


Image enhancement using local feature recognition

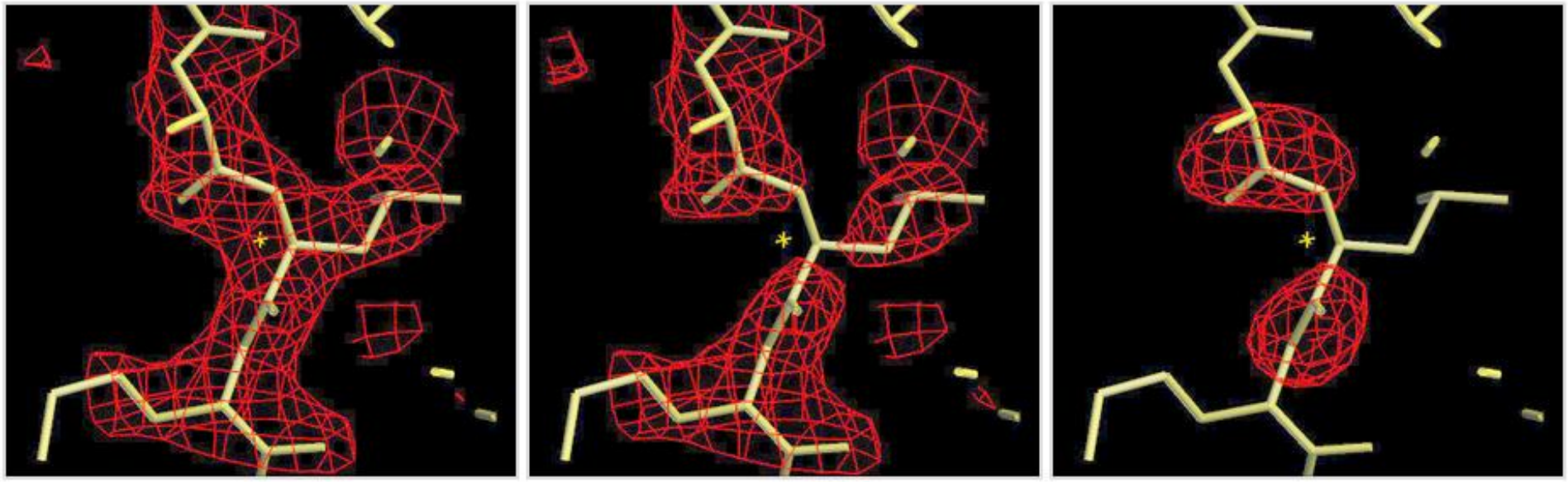
Approach:

- Create N templates of local density using model data
  - Examine density near each point x in image (within 2 Å)
  - Exclude region very close to x (about 1 Å)
  - Cluster and average local patterns of density (after rotation to maximize CC)
- Identify relationship between finding pattern k of density near x, and density at x
  - Find all locations in the image where template k best matches the local density near x
  - Calculate average value of density at x for these cases = $\rho_{mean}(k)$
- Identify pattern near each point in actual map and use it to estimate density at that point
  - For each point x in the image, identify which template k best matches the local density near x
  - Use $\rho_{mean}(k)$ as estimate of density at x

Constructing a feature library

Identify averaged local features of model images

Go through map point-by-point. For each point x (marked by * at center of pictures)
Examine density in map near x : $\rho_{current}(x+\Delta x)$
Remove all information about density at x from $\rho(x+\Delta x)$
-> $g(x+\Delta x)$, local pattern at x, without information from x
Group local density patterns by correlation coefficient into clusters
Sort clusters according to average density at x for that cluster



$\rho_{current}(x+\Delta x)$

Local density near x in model image

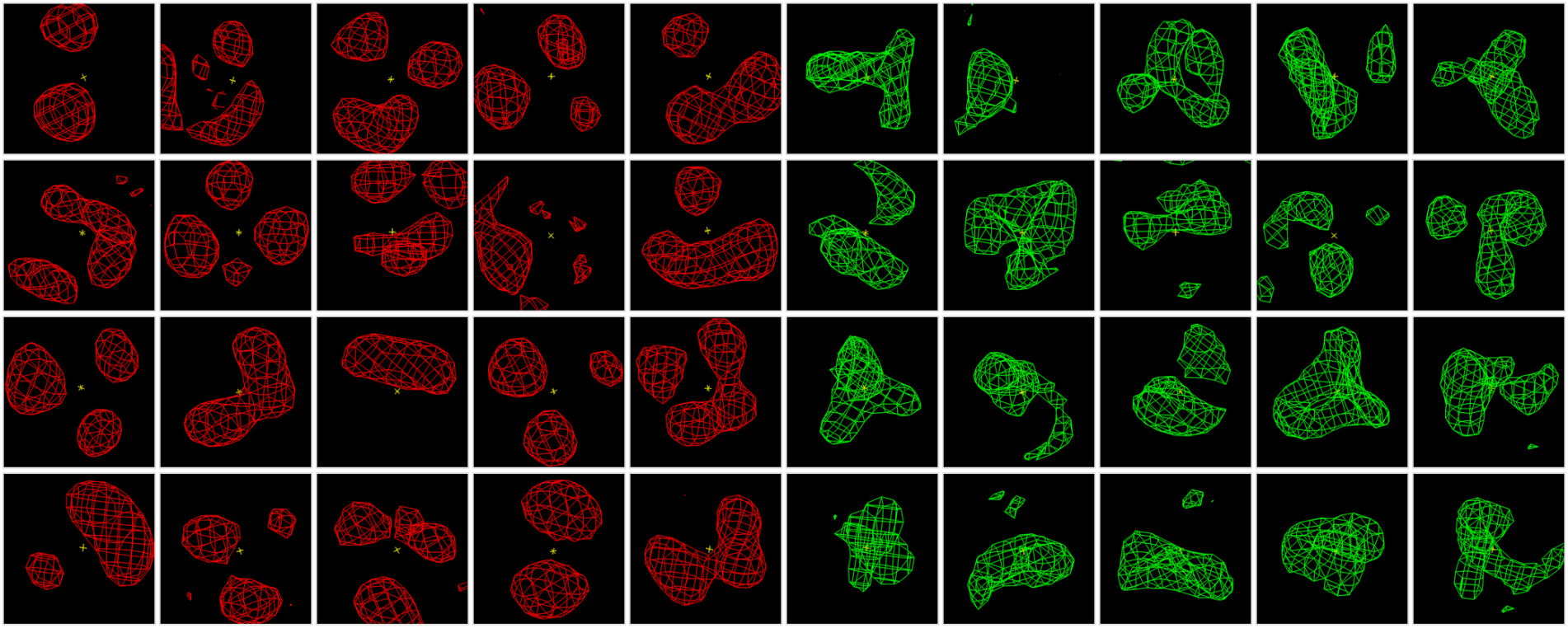
$g(x+\Delta x)$

local density…setting points near x to near zero

Average of local density for one cluster associated with positive density at x

Constructing a feature library

20 Templates of average local density



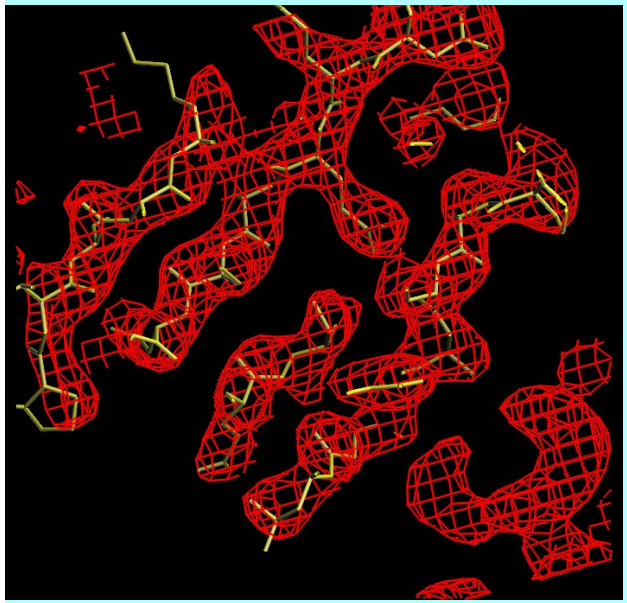
Contours at +1.5 σ

Contours at -1.5 σ

Image enhancement using local feature recognition

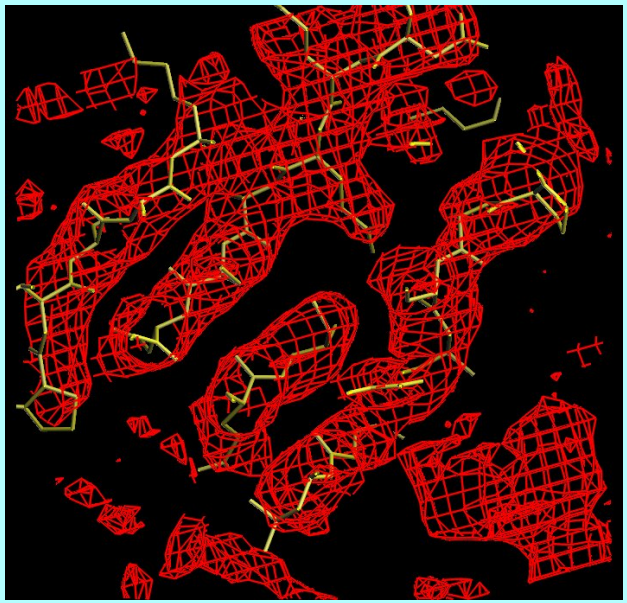
Image recovery from a good map…

- gives an image that has (mostly) correct features
- errors are (almost) uncorrelated with original errors



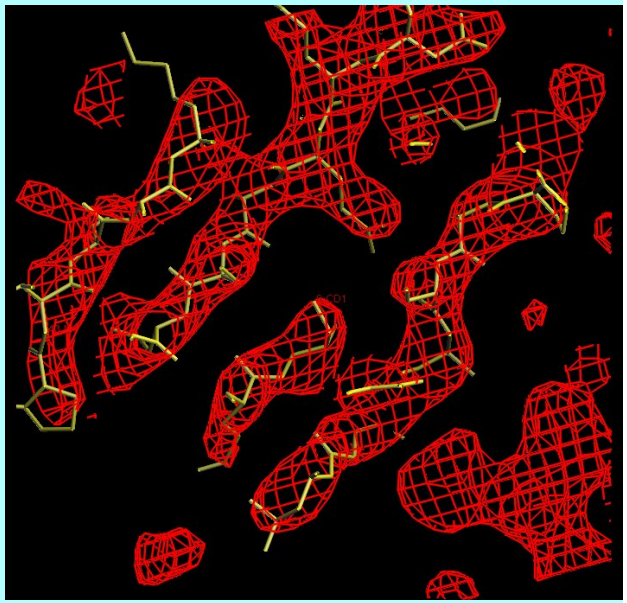
RESOLVE map gene 5 protein at 2.6 A

CC to perfect map = 0.8



Recovered image derived from RESOLVE map

CC to perfect map = 0.36



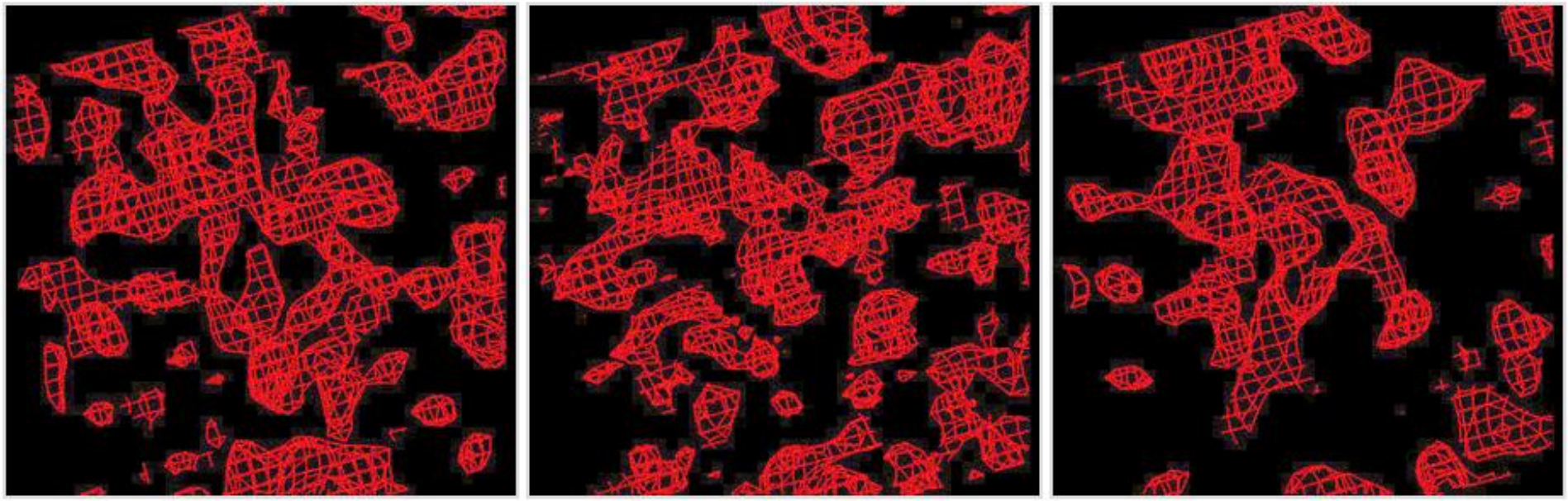
Map phased using only using information from recovered image

CC to perfect map = 0.64

CC of errors with errors in RESOLVE map = 0.11

Image enhancement using local feature recognition

Image recovery from a random map…gives an uncorrelated image



Random map at 2.6 A

Recovered image
derived from random map

CC to original
random map=-0.01

Map phased using only
recovered image

CC to original
random map=-0.04

Iterative procedure for image enhancement using local feature recognition

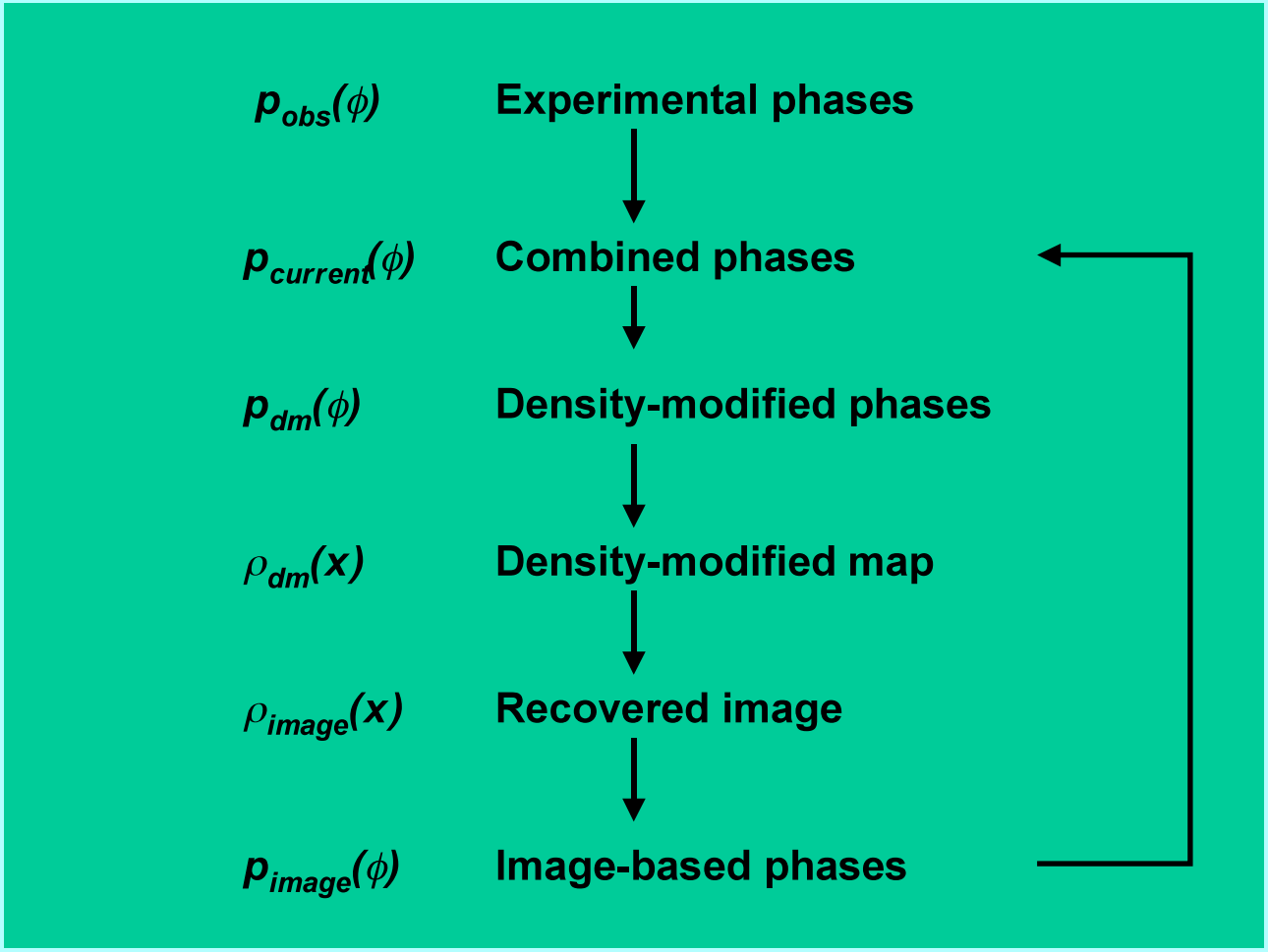
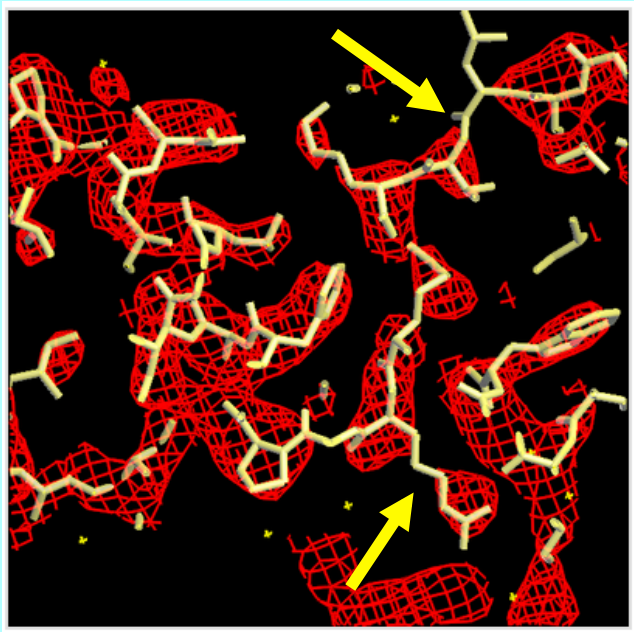


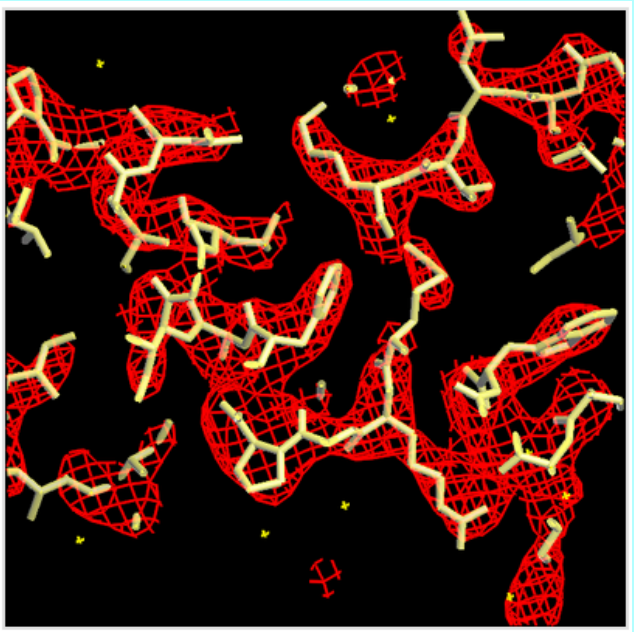
Image enhancement using local feature recognition (nusA protein structure)

Starting map
CC=0.65



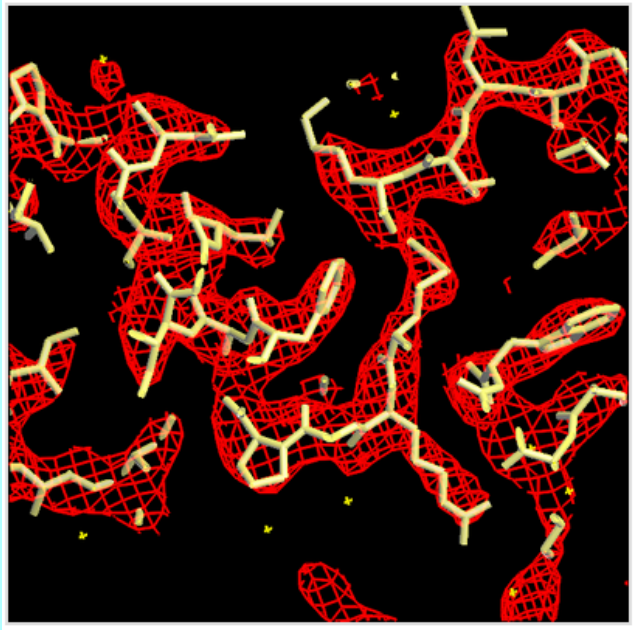
Cycle 1

CC=0.75



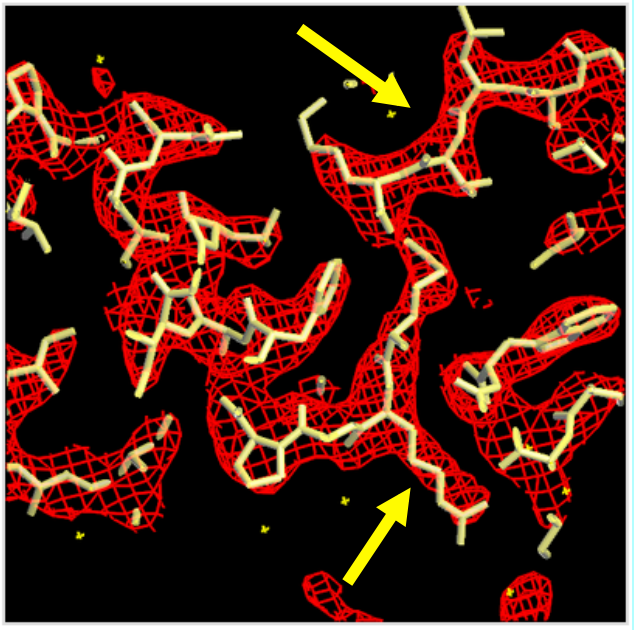
Cycle 3

CC=0.84



Cycle 5

CC=0.85

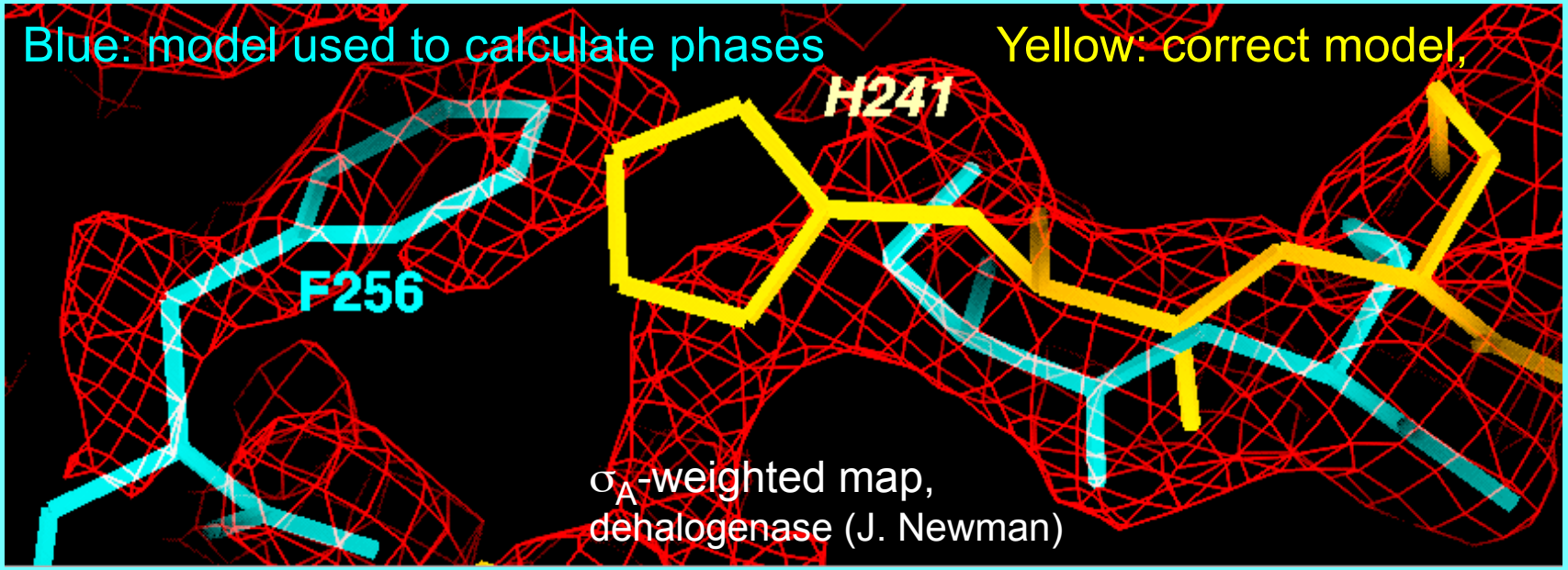


Removing model bias with prime-and-switch phasing

The problem:

Atomic model used to calculate phases -> map looks like the model

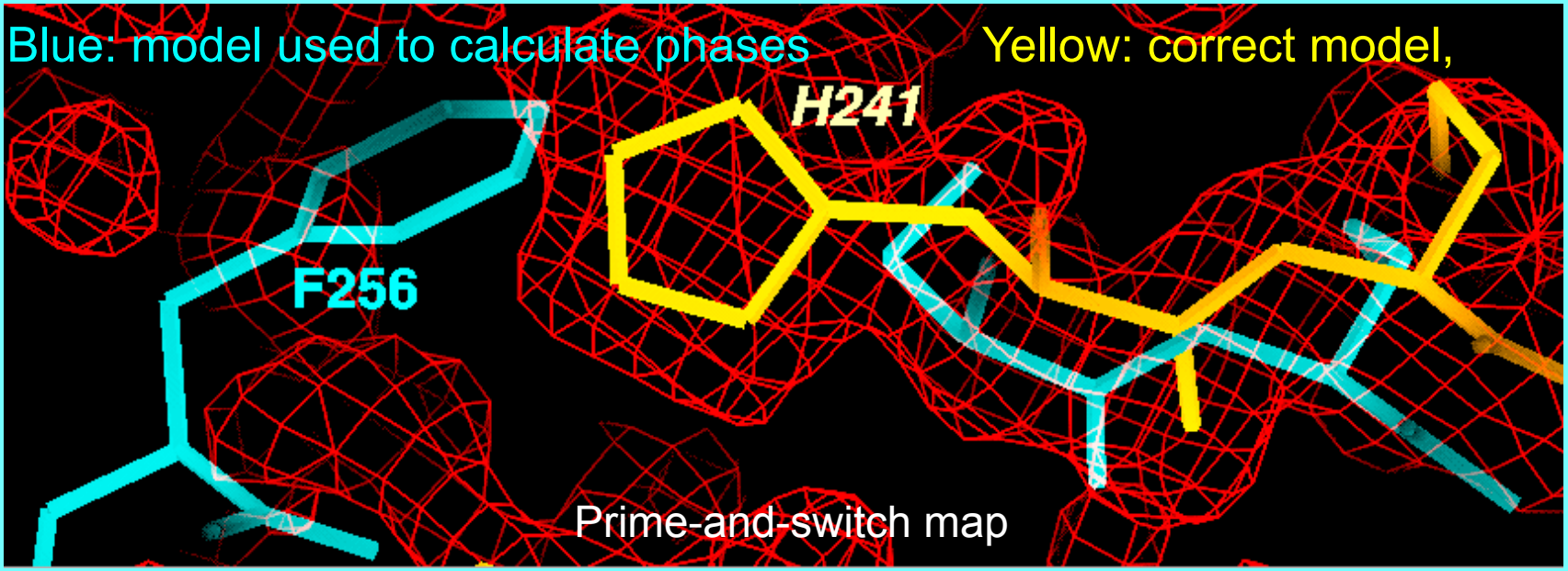
Best current solution: σ_A-weighted phases



Prime-and-switch phasing

A solution:

Start with σ_A-weighted map
Identify solvent region (or other features of map)
Adjust the phases to maximize the probability of the map – without biasing towards the model phases

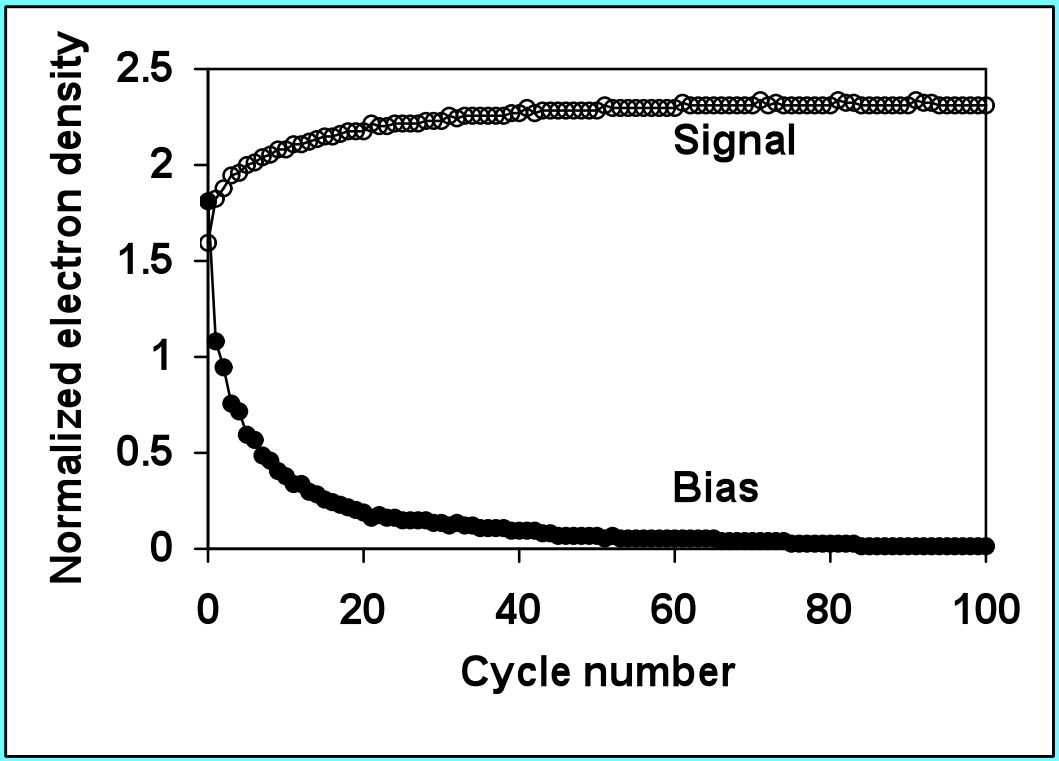


Prime-and-switch phasing

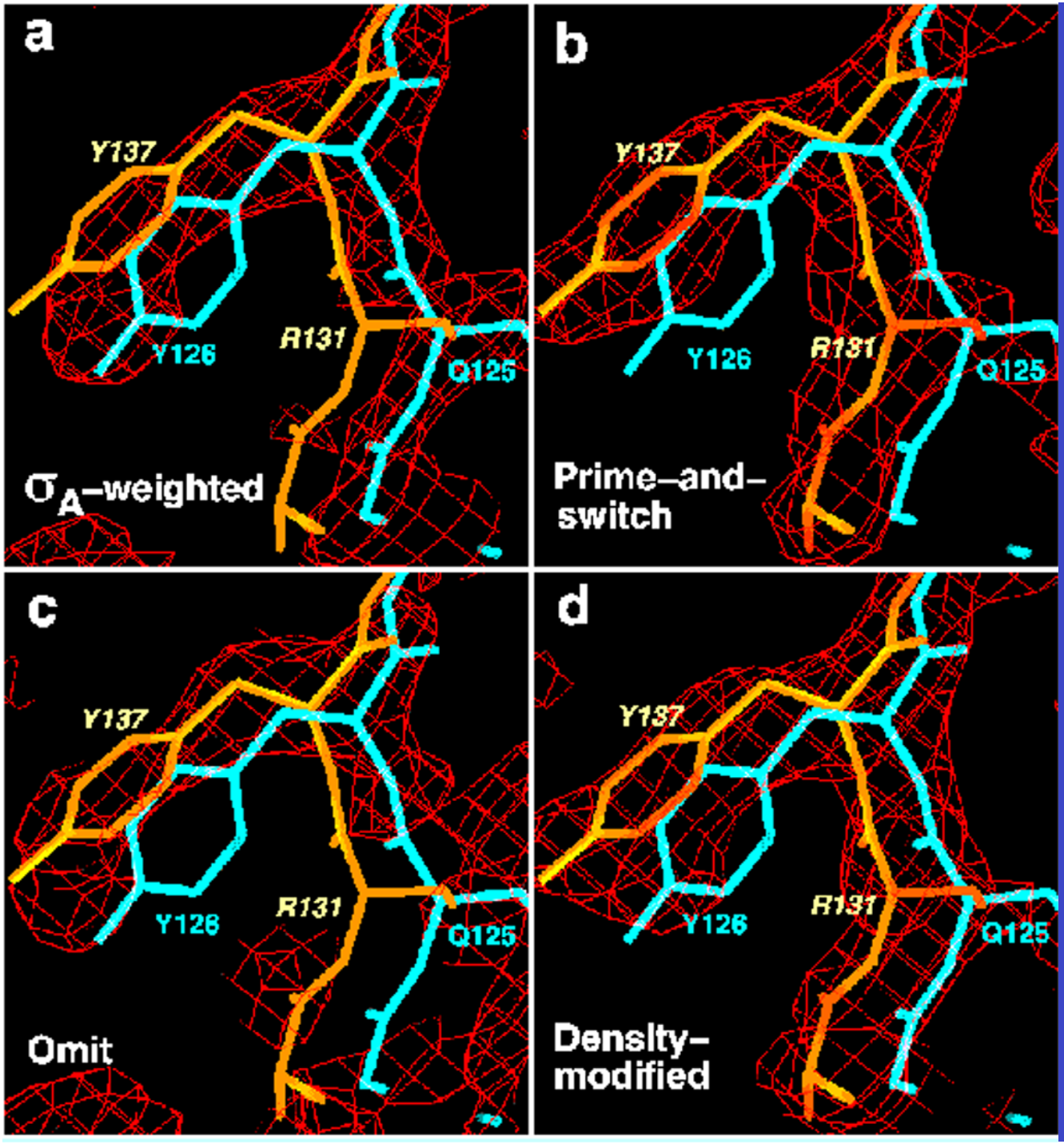
Why it should work…

Priming: Starting phases are close to correct ones…but have bias towards misplaced atoms

Switching: Map-probability phase information comes from a different source…which reinforces just the correct phase information



Signal: peak height at correct atomic positions
Bias: peak height at incorrect atoms in starting model



Prime-and-switch example

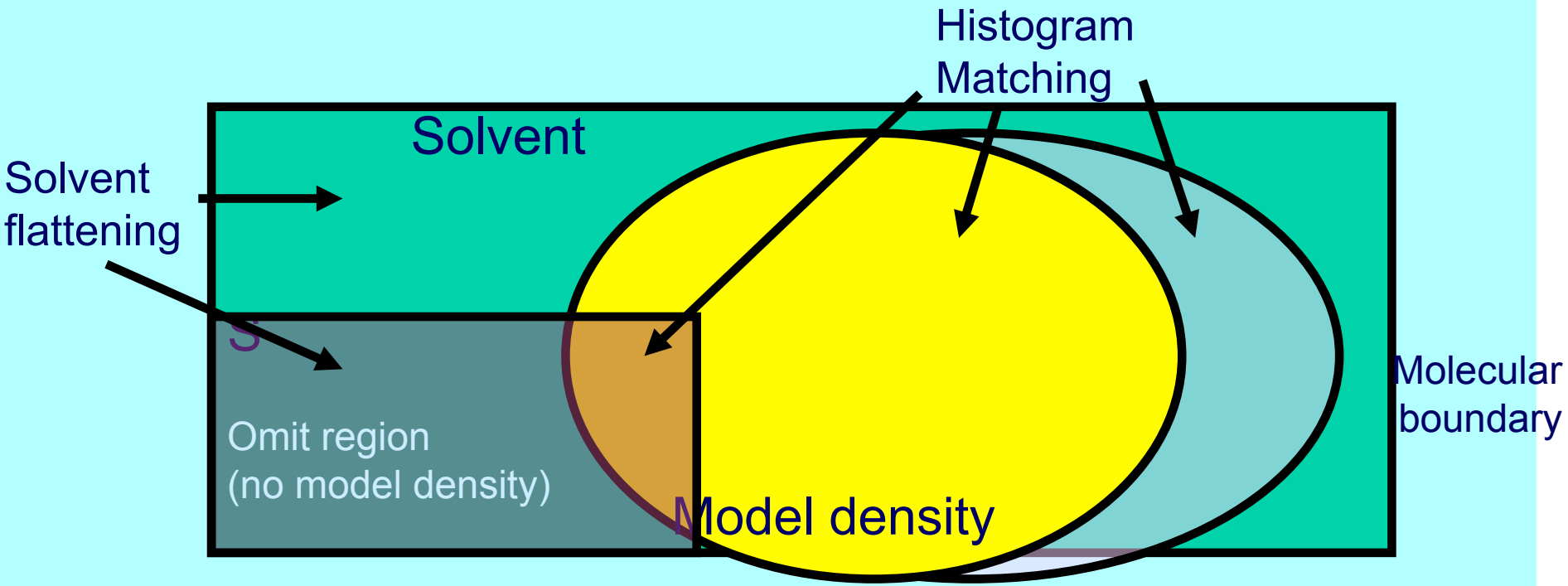
(IF5A, T. Peat)

Orange: correct model

Blue: model used to calculate phases

Composite omit map with statistical density modification

Statistical density modification allows a separate probability distribution for electron density at each point in the map: can specify that "missing" density is within molecular boundary

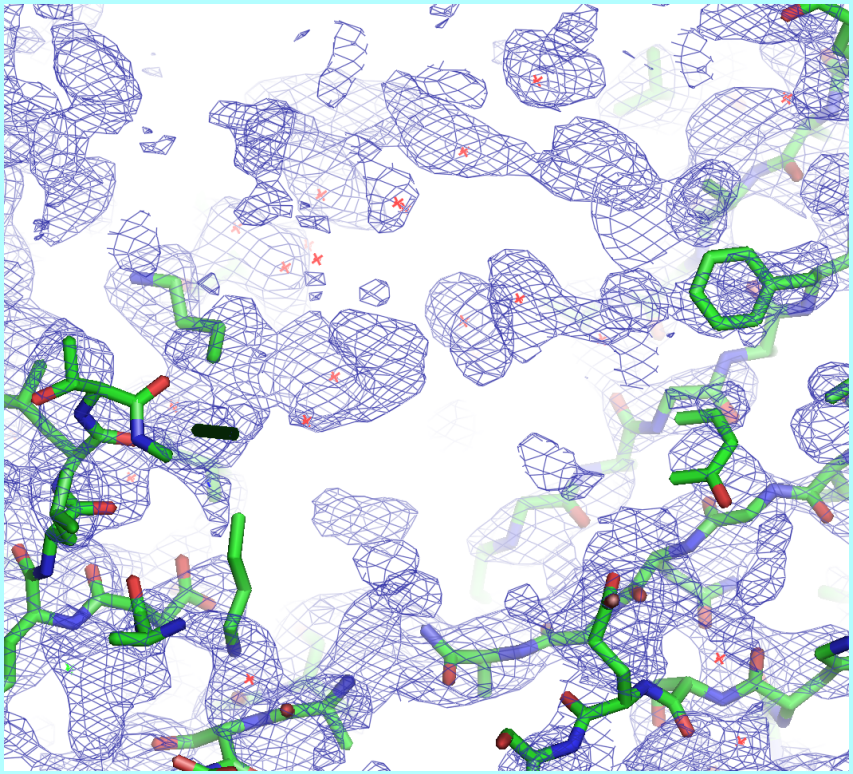


Can be used with or without experimental phases…with or without omit

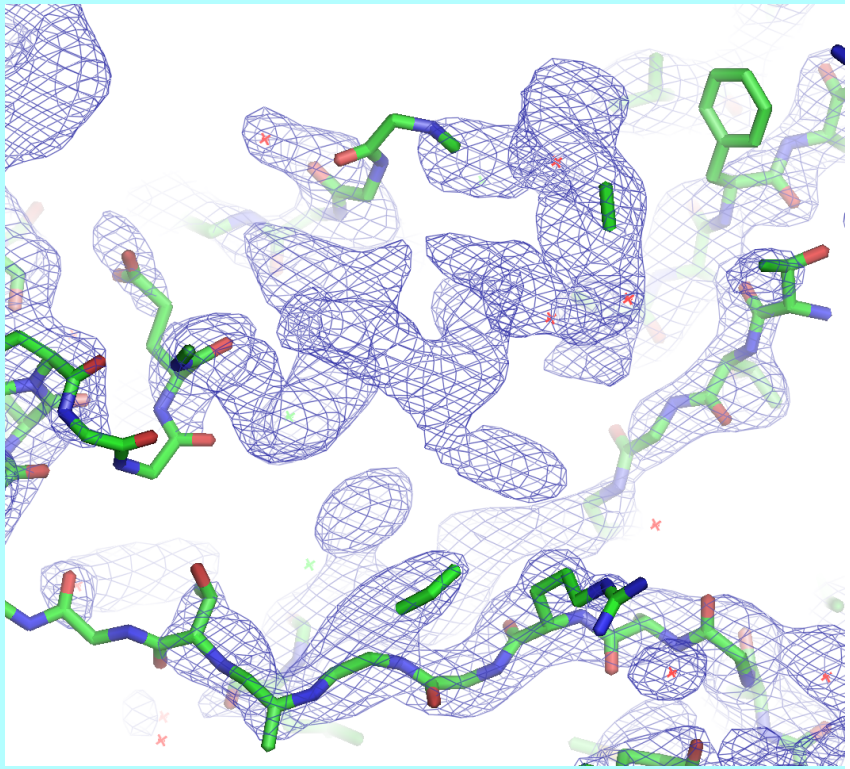
Iterative-Build OMIT procedure

"Is the density in my map biased by the model?"

2mFo-DFc omit map



After building outside OMIT region 10 cycles



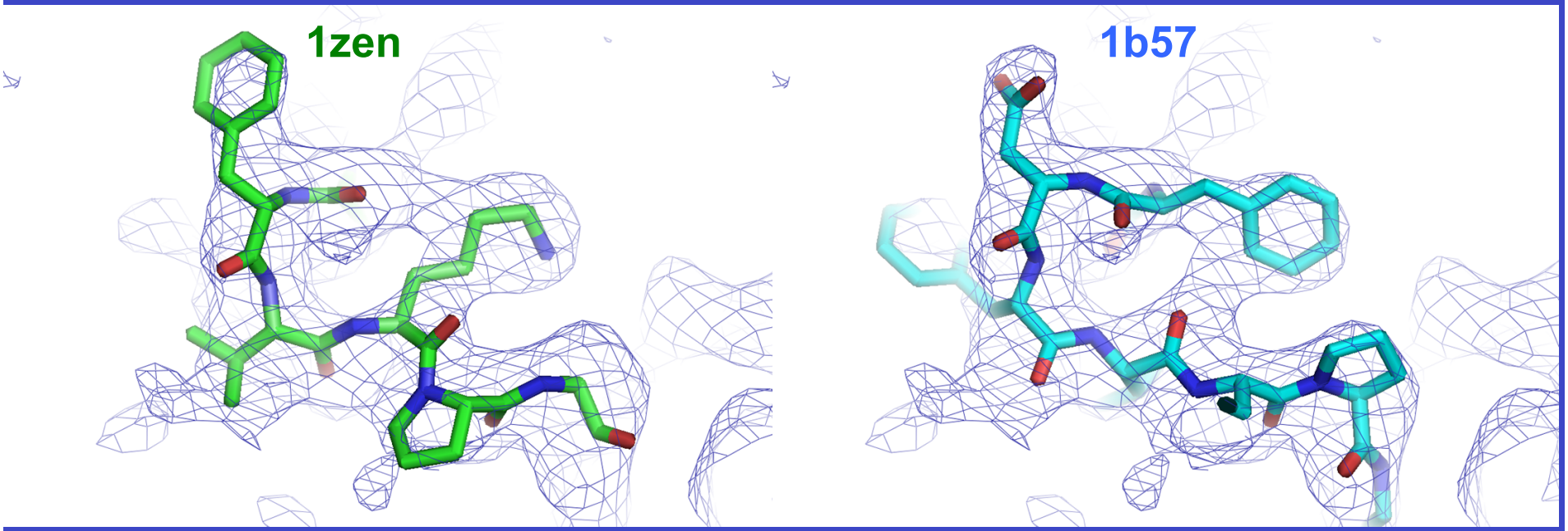
1HP7 molecular replacement with 1AS4

R/Rfree after initial refinement: 0.41/0.48

Iterative-Build OMIT procedure

"Removing model bias"

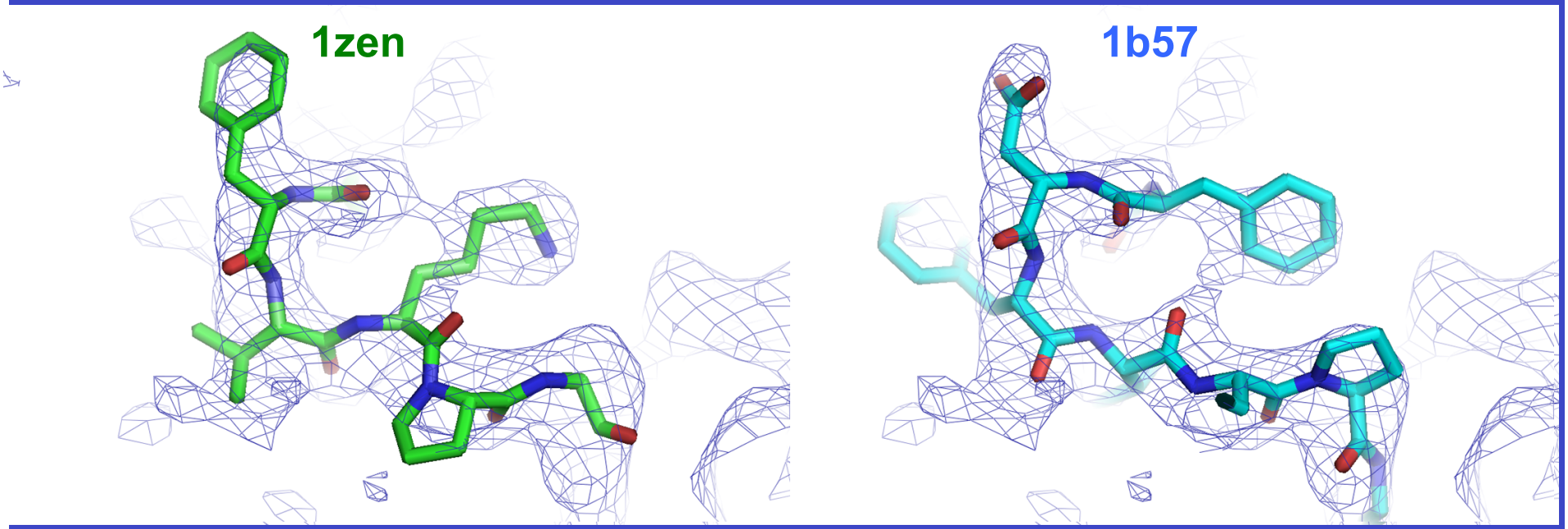
2mFo-DFc map
Phased with 1zen model



Iterative-Build OMIT procedure

"Removing model bias"

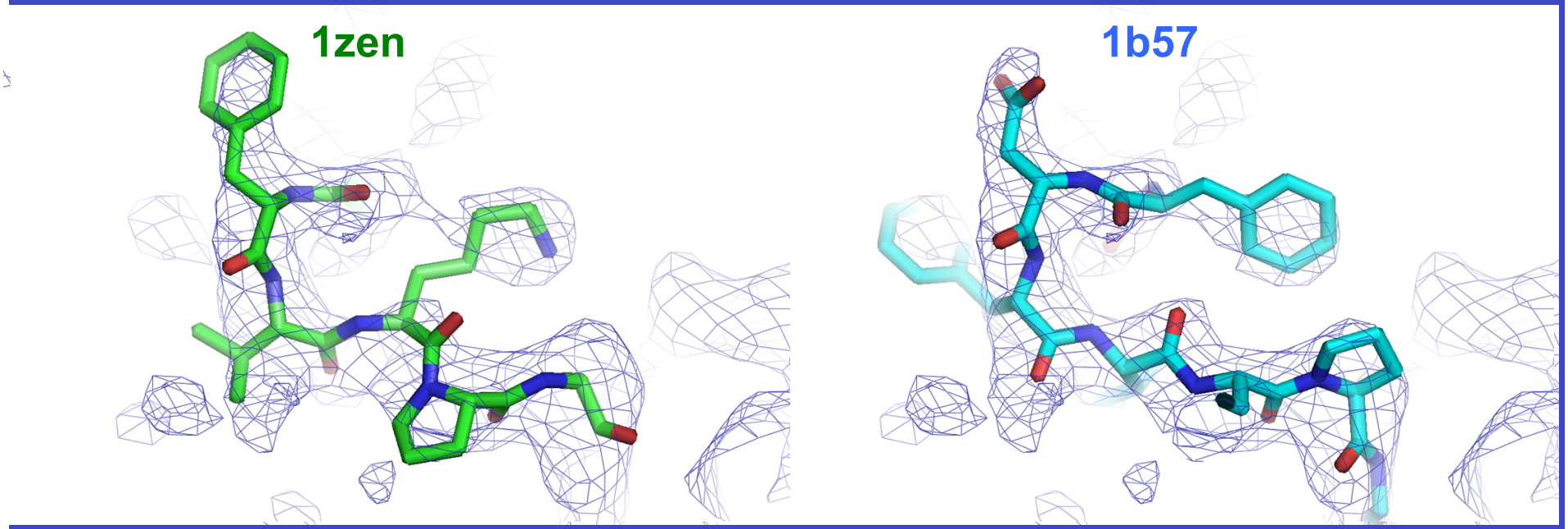
2mFo-DFc omit map
Phased with 1zen model



Iterative-Build OMIT procedure

"Removing model bias"

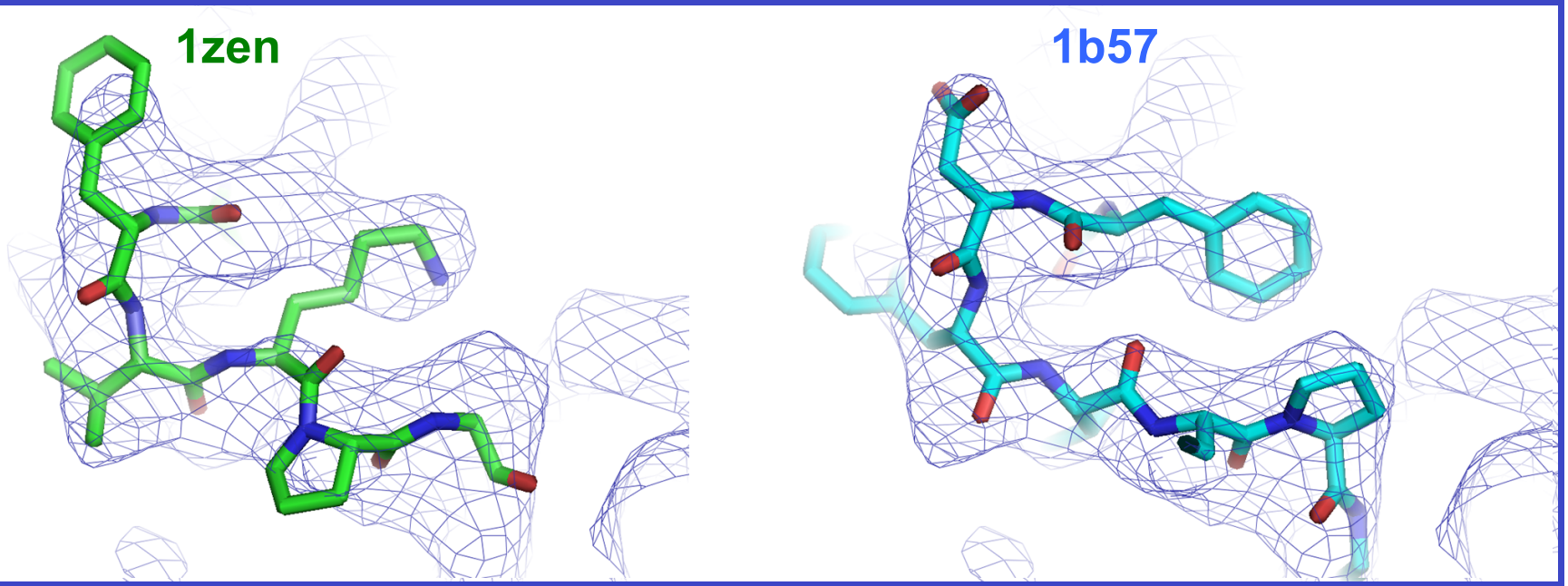
2mFo-DFc SA-omit map
Phased starting with 1zen model



Iterative-Build OMIT procedure

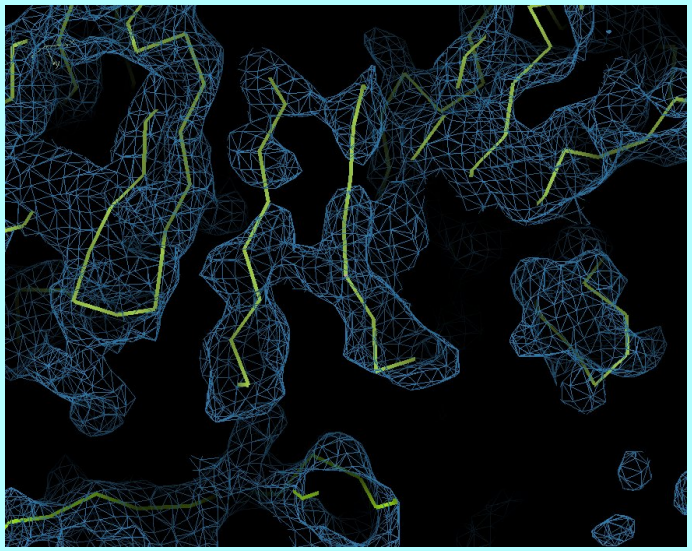
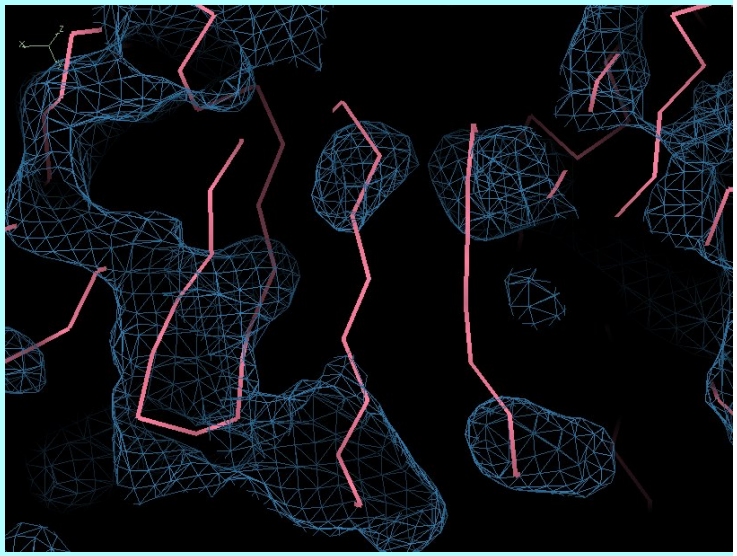
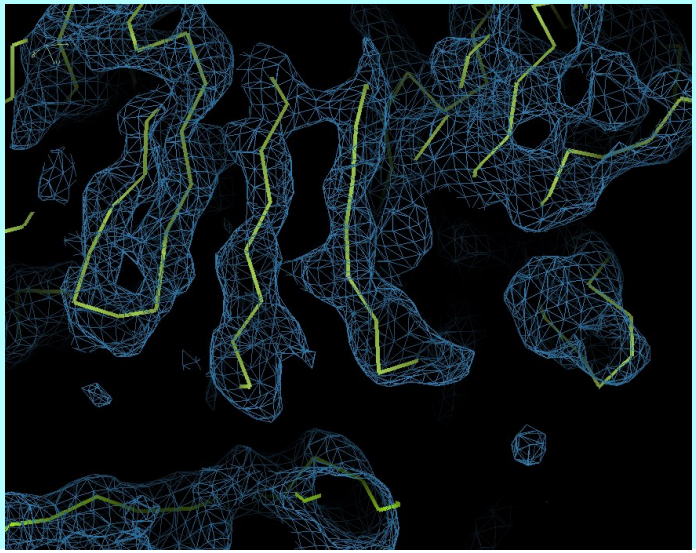
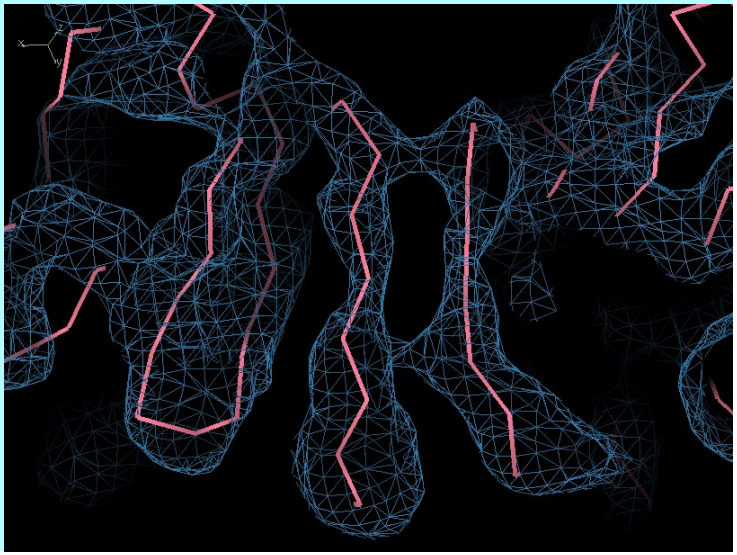
"Removing model bias"

2mFo-DFc iterative-build omit map
Phased starting with 1zen model



Statistical density modification with cross-crystal averaging

Cell receptor at 3.5/3.7 Å. Data courtesy of J. Zhu

	Crystal 1 (4 copies)	Crystal 2 (2 copies)
RESOLVE density modification		
PHENIX Multi-crystal averaging		

The PHENIX Project

Phenix

Lawrence Berkeley Laboratory

Paul Adams, Ralf Grosse-Kunstleve, Pavel Afonine, Nat Echols, Nigel Moriarty, Jeff Headd, Nicholas Sauter, Peter Zwart

Los Alamos National Laboratory

Tom Terwilliger, Li-Wei Hung

Cambridge University

Randy Read, Airlie McCoy, Gabor Bunkoczi, Rob Oeffner

Duke University

Jane & David Richardson, Vincent Chen, Chris Williams, Bryan Arendall, Laura Murray

An NIH/NIGMS funded Program Project