

Density Modification

Kevin Cowtan, York.

Traditional density modification:

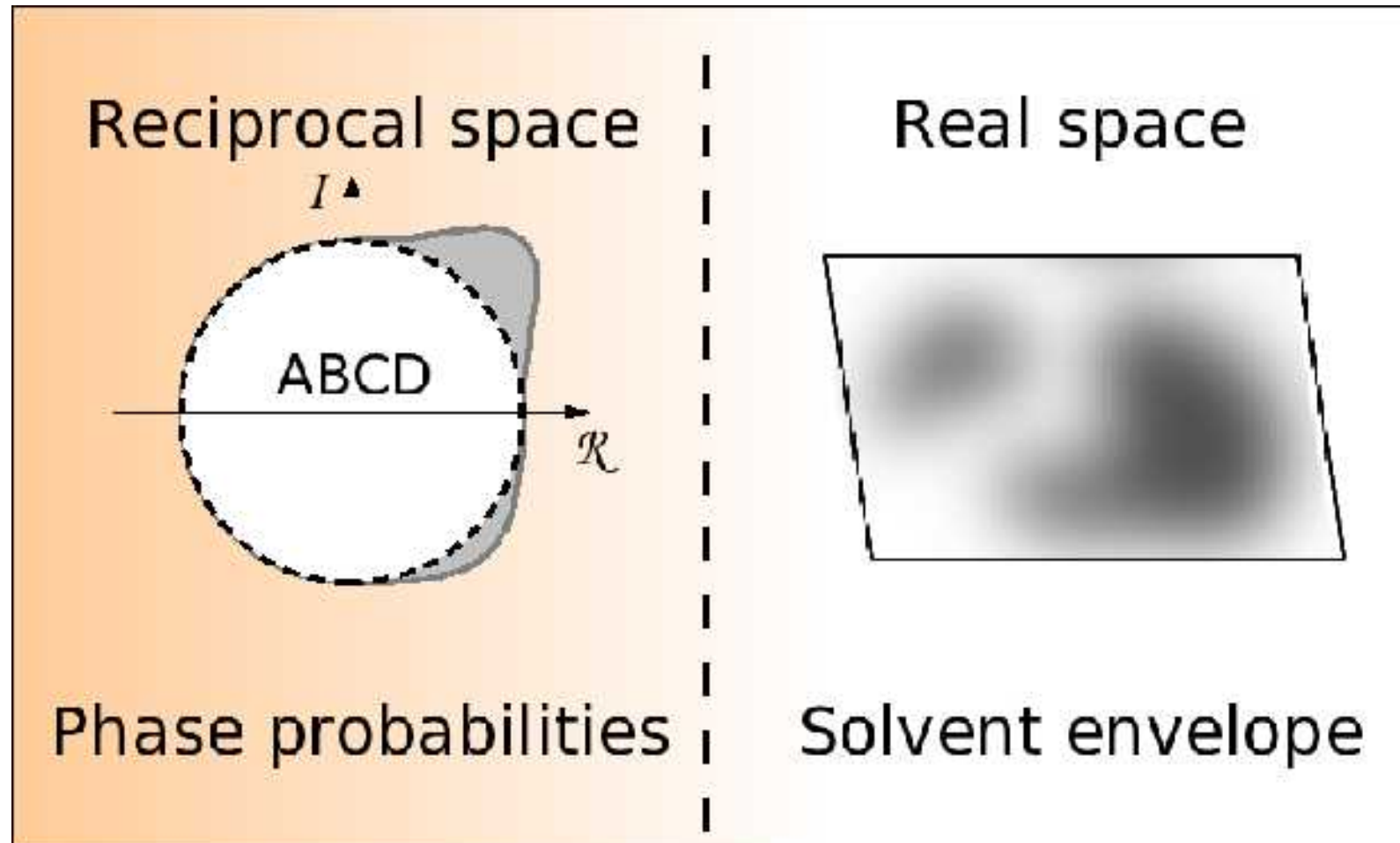
e.g. 'dm', 'solomon', CNS

Statistical density modification:

e.g. Resolve, Pirate

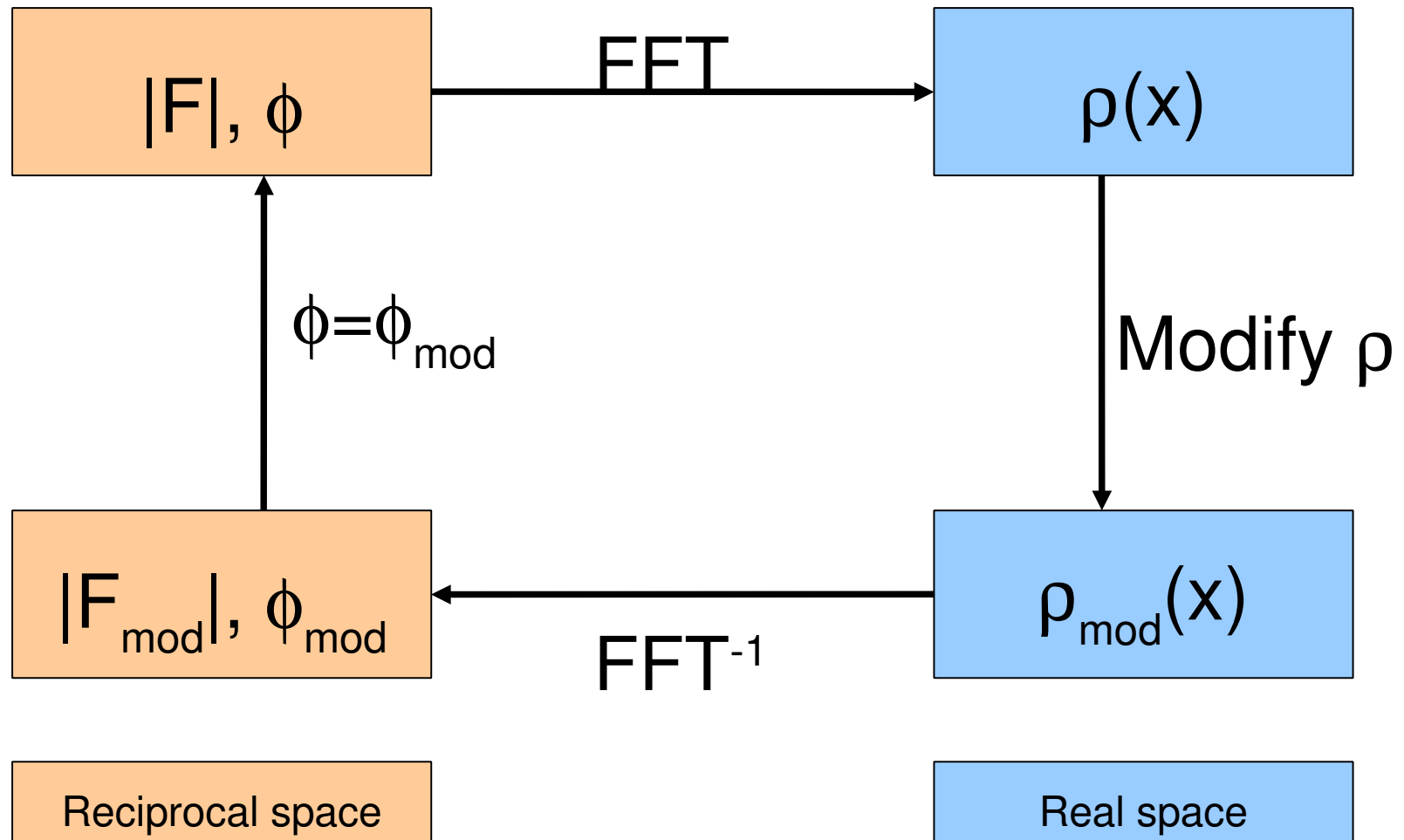
Density modification

- Density modification is a problem in combining information:



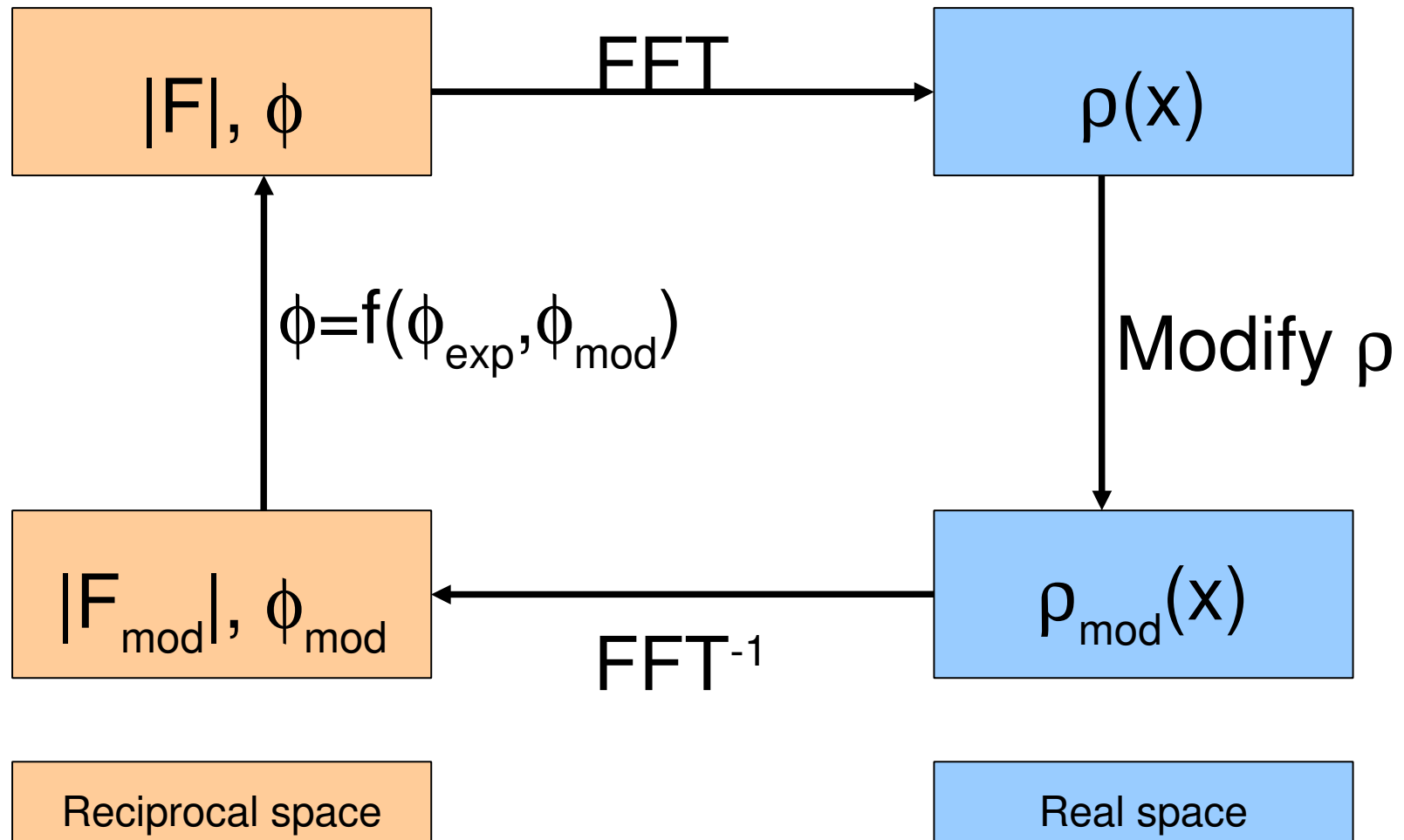
Density modification

1. Rudimentary calculation:



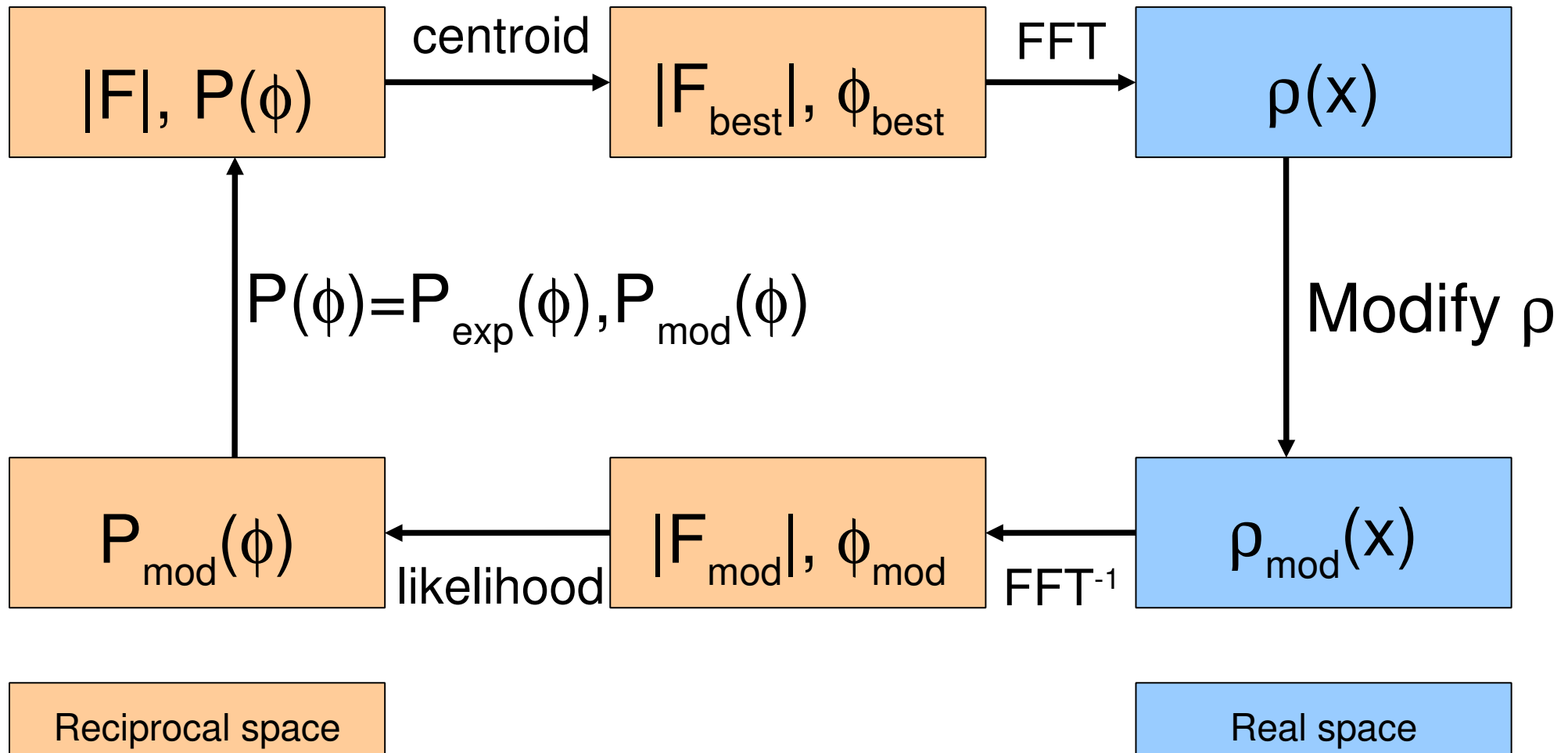
Density modification

2. Phase weighting:



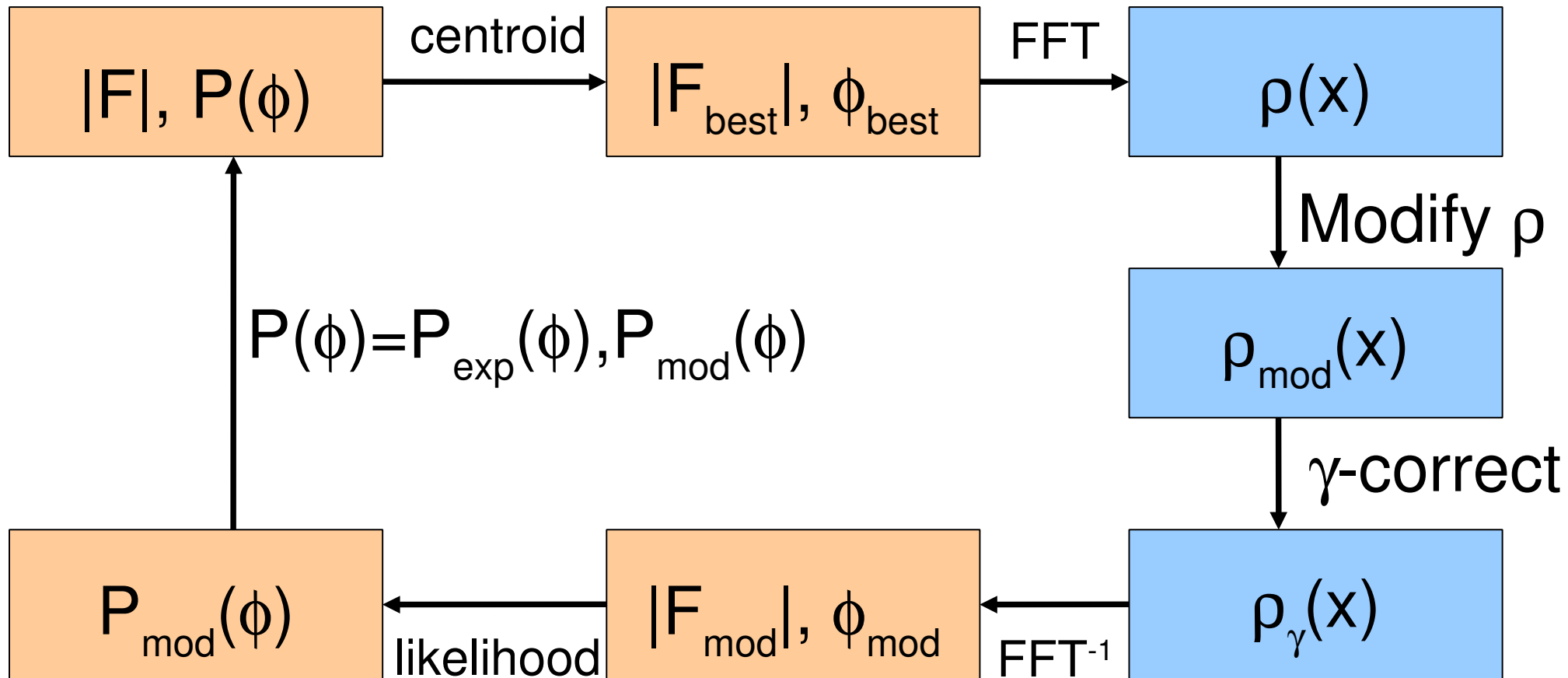
Density modification

3. Phase probability distributions:



Density modification

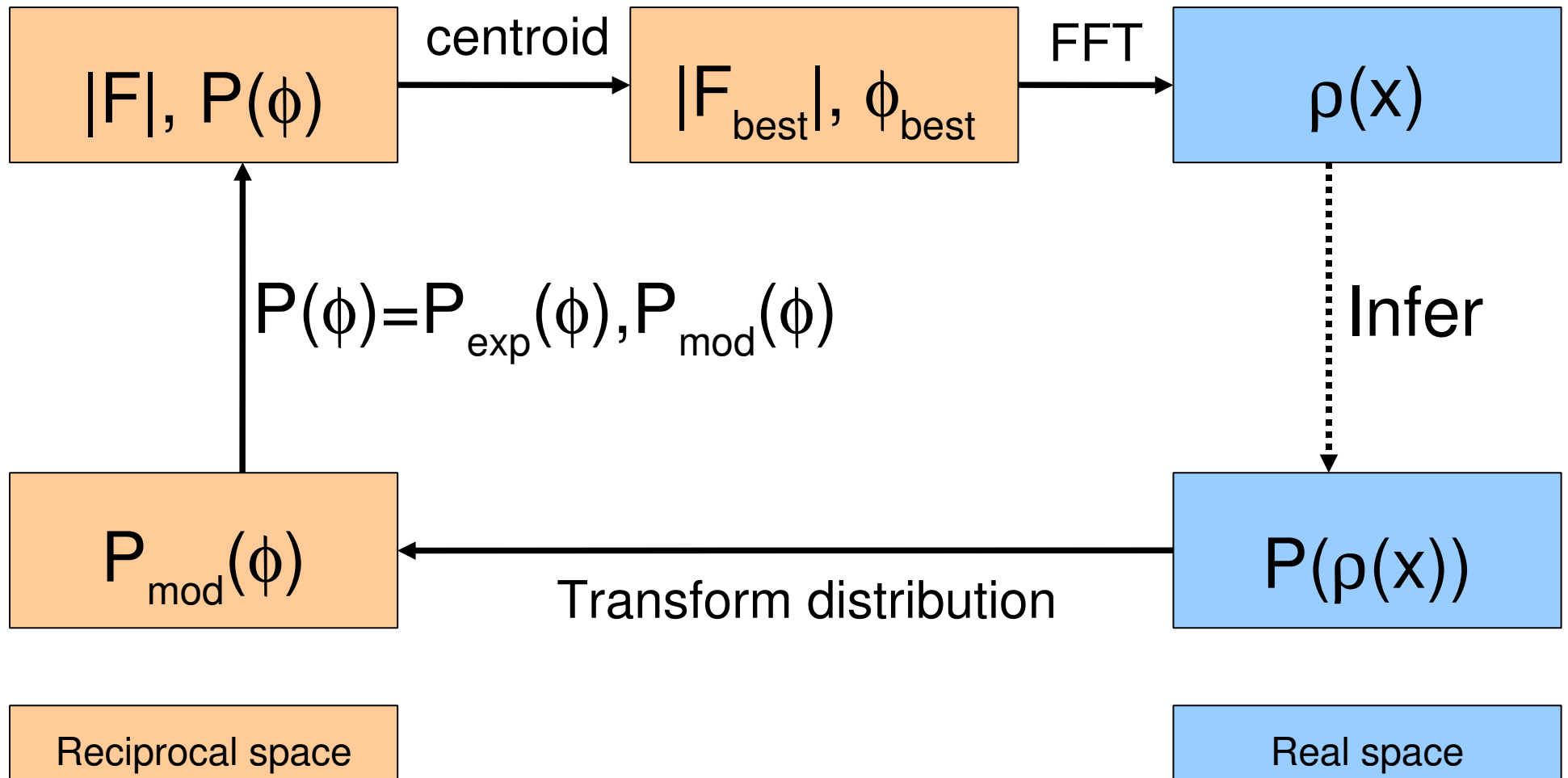
4. Bias reduction (gamma-correction):



J.P.Abrahams

Density modification

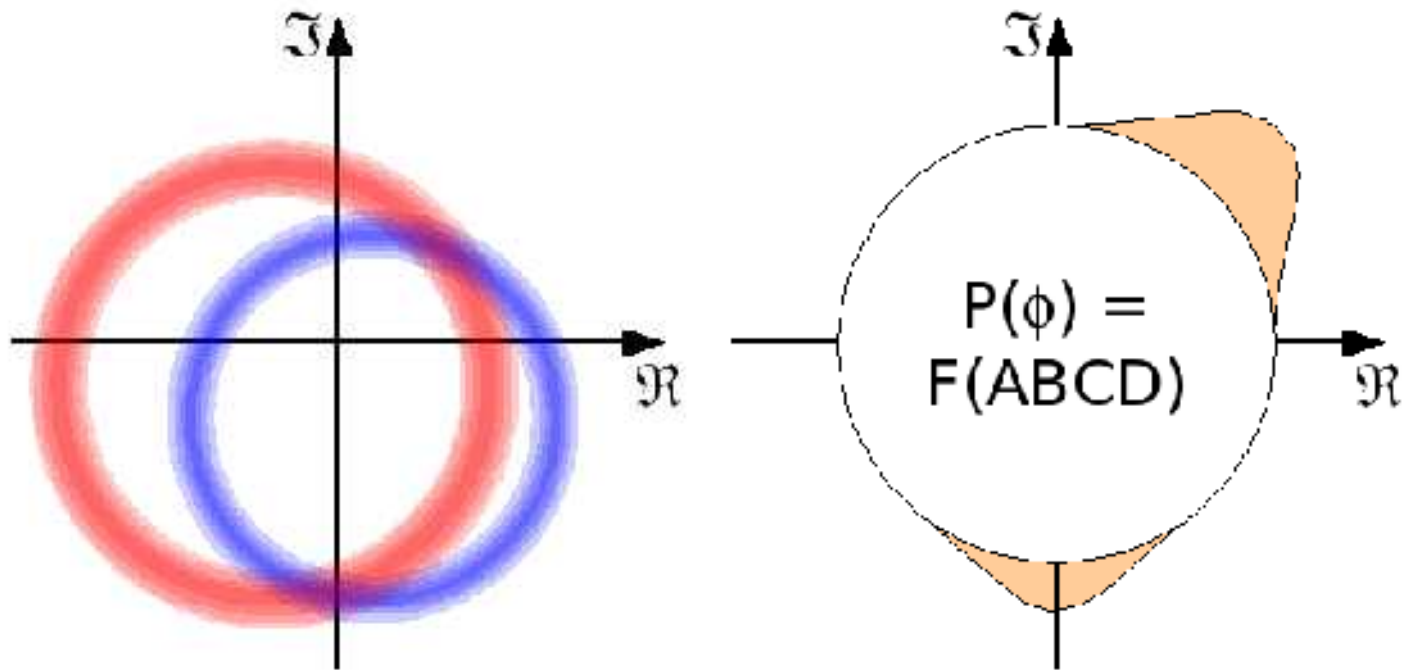
Statistical density modification:



Density modification

How do we represent phase probabilities?

Henrickson-Lattman coeffs: 4 numbers - A,B,C,D representing a bimodal distribution in phase angle:

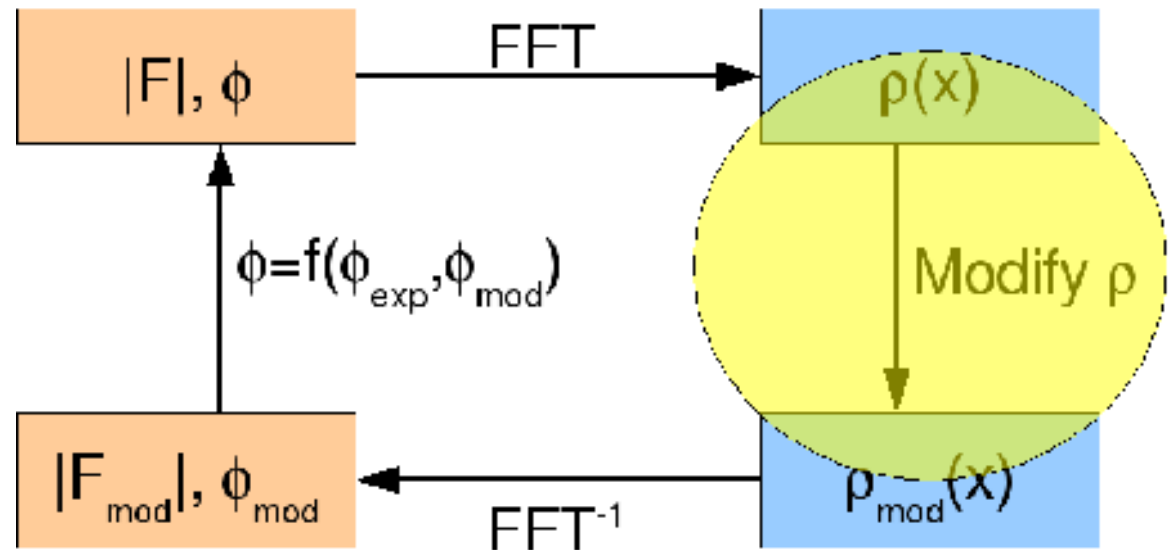


A,B represent a unimodal distribution (equivalent to ϕ , FOM)
C,D represent the superimposed bimodality.

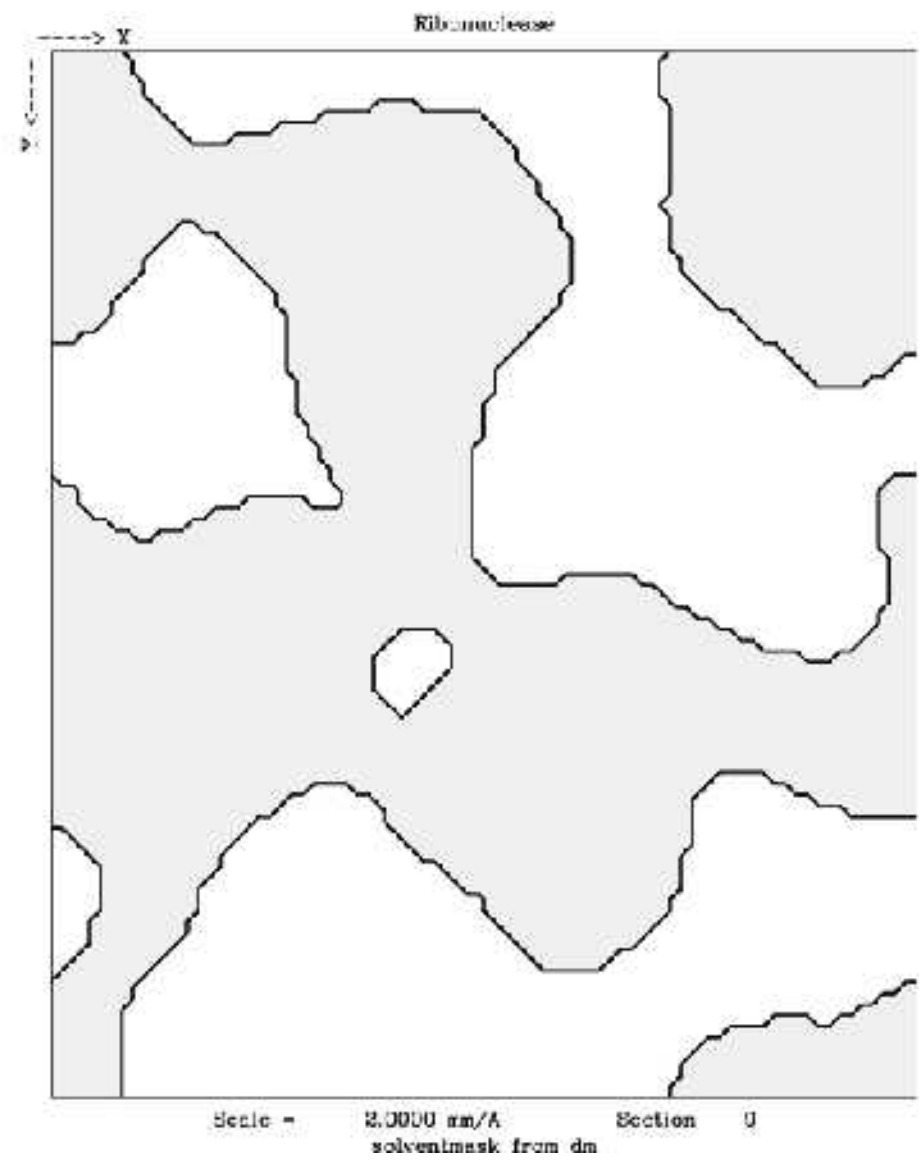
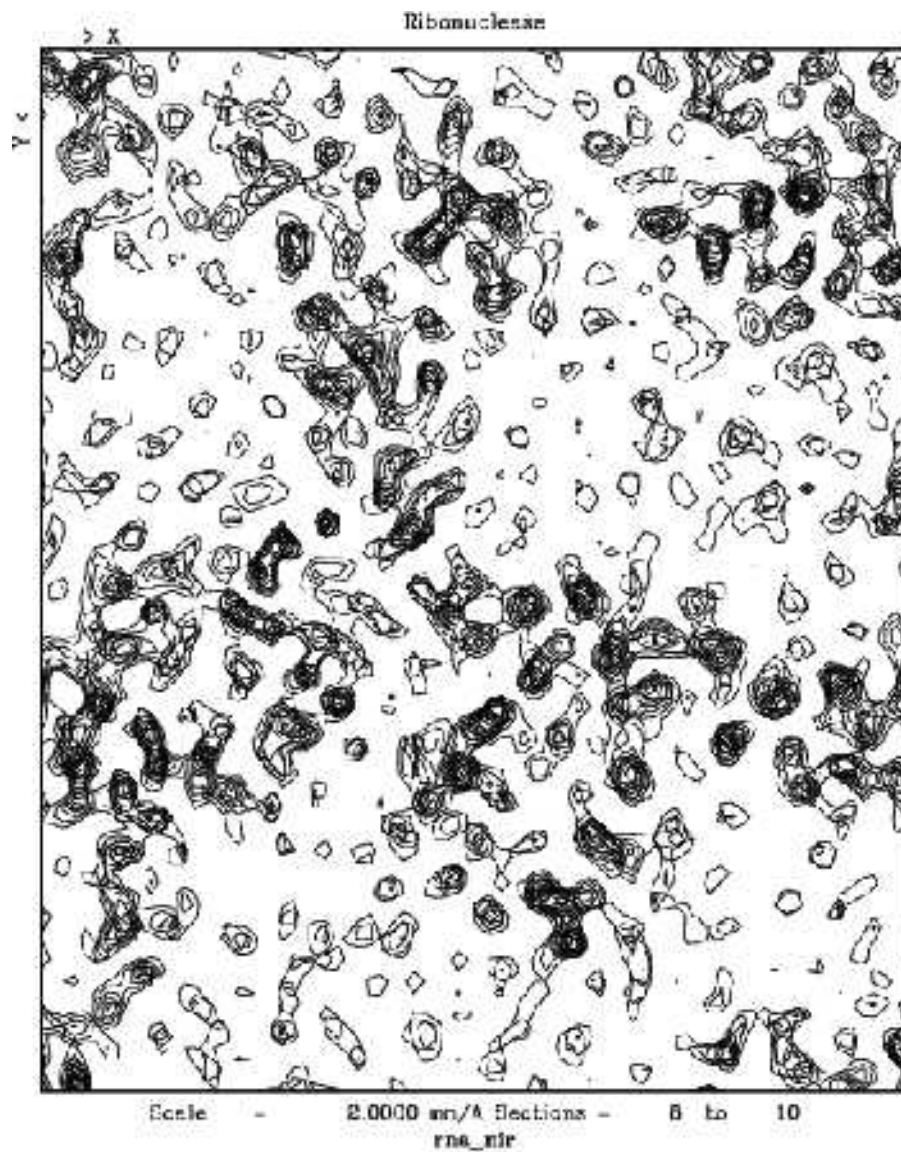
Density modification

Traditional density modification techniques:

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



Solvent flattening



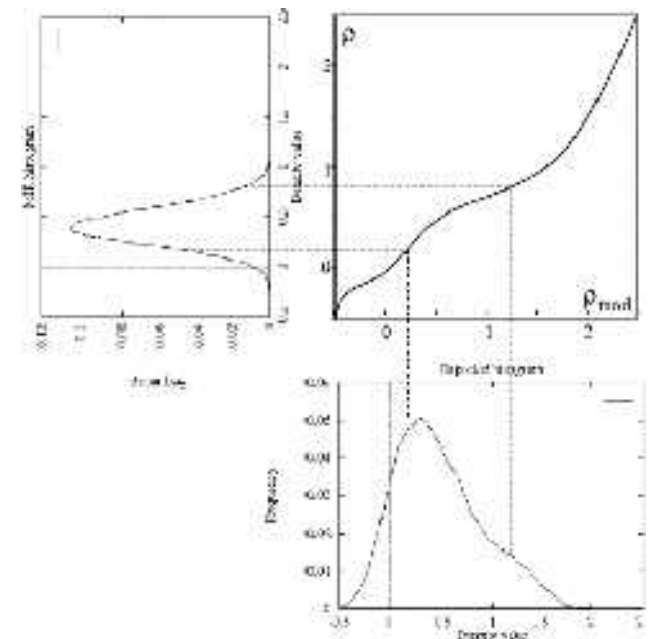
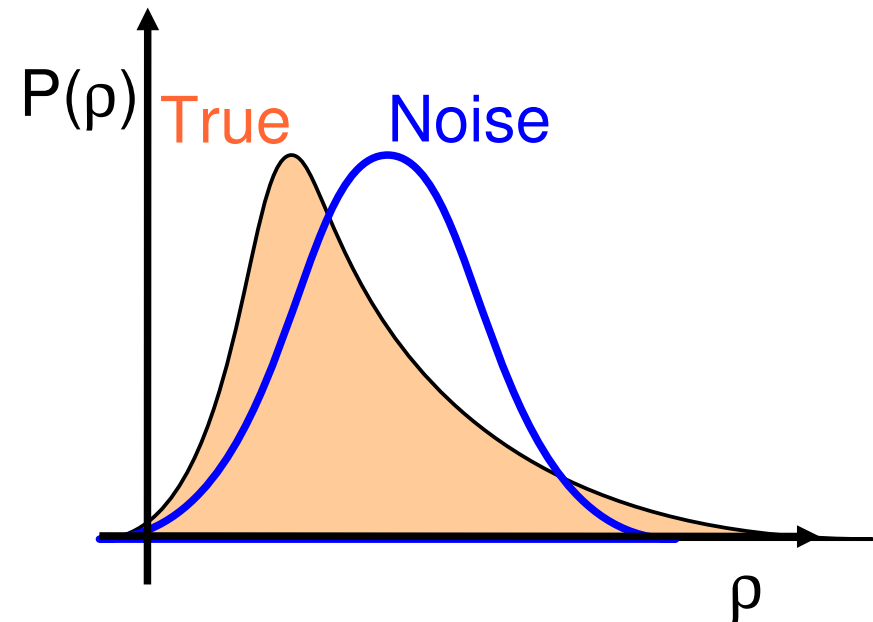
Histogram matching

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.

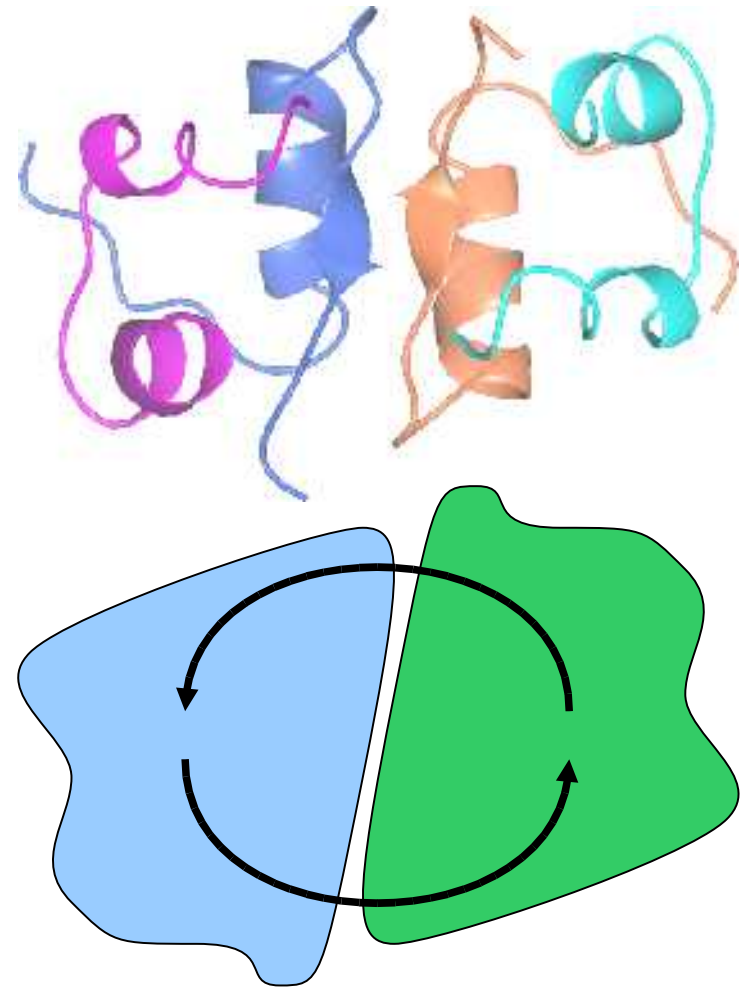
Sharpen the protein density by a transform which matches the histogram of a well phased map.

Useful at better than 4Å.



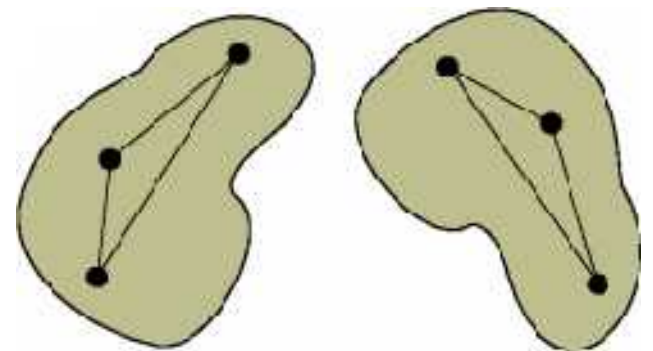
Non-crystallographic symmetry

- 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!



Non-crystallographic symmetry

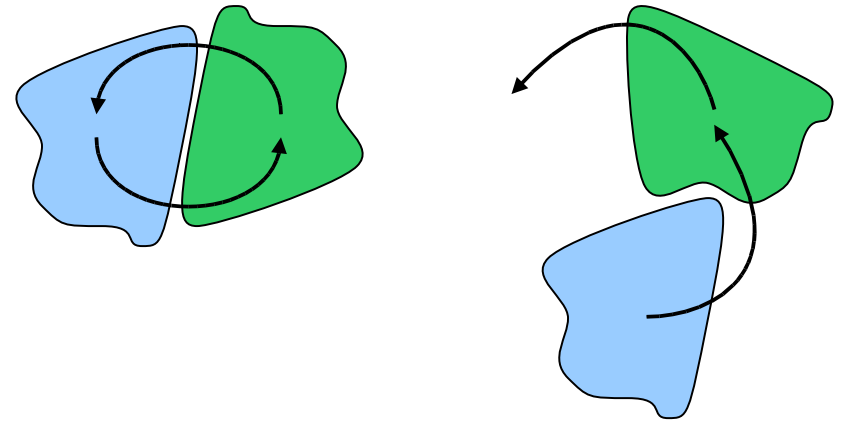
- How do you know if you have NCS?
 - Cell content analysis – how many monomers in ASU?
 - Self-rotation function.
 - Difference Pattersons (pseudo-translation only).
- How do you determine the NCS?
 - **From heavy atoms.**
 - From initial model building.
 - From molecular replacement.
 - *From density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



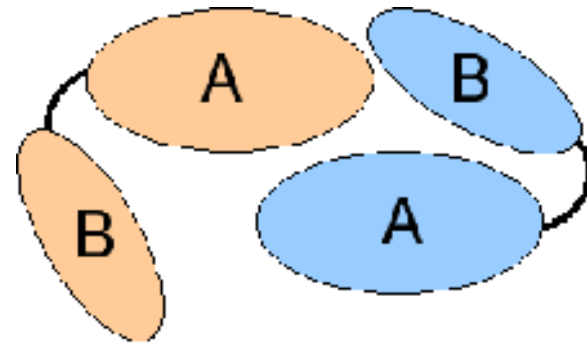
Non-crystallographic symmetry

Useful terms:

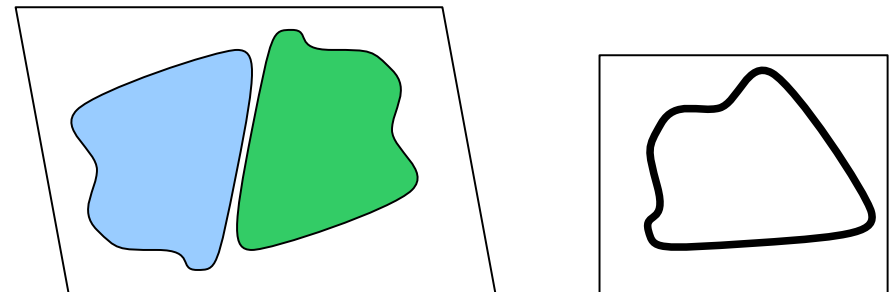
- Proper and improper NCS:
(closed and open)



- Multi-domain averaging:



- Multi-crystal averaging:



Combining phase probabilities

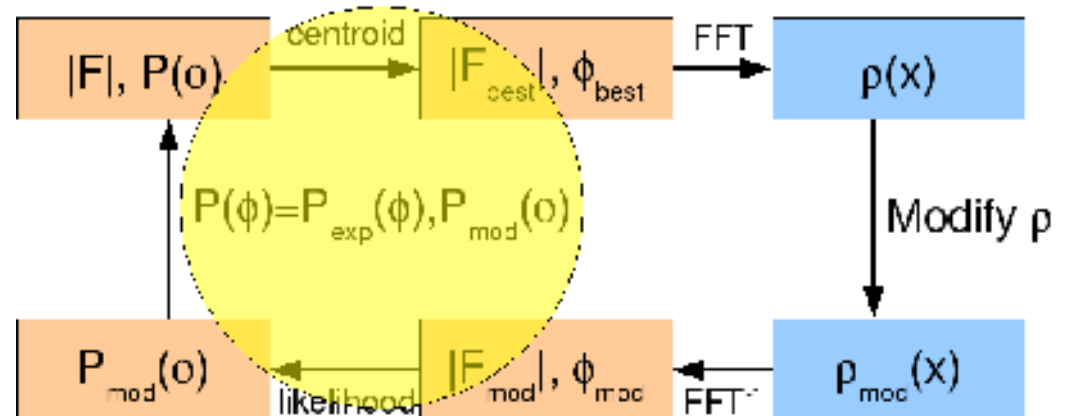
Once we have an estimate for the error in ϕ_{mod} , we can construct a probability distribution $P_{\text{mod}}(\phi)$.

The the next cycle can be started with

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

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

The result is bias, increasing with cycle.



Bias reduction

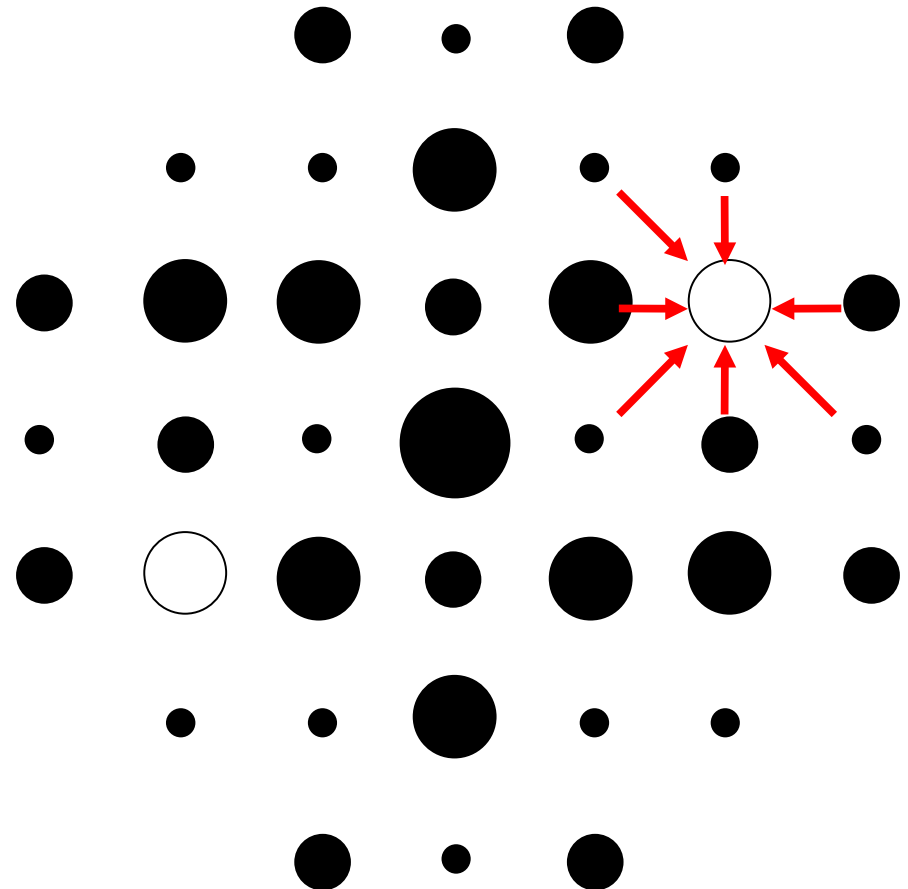
Solution:

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

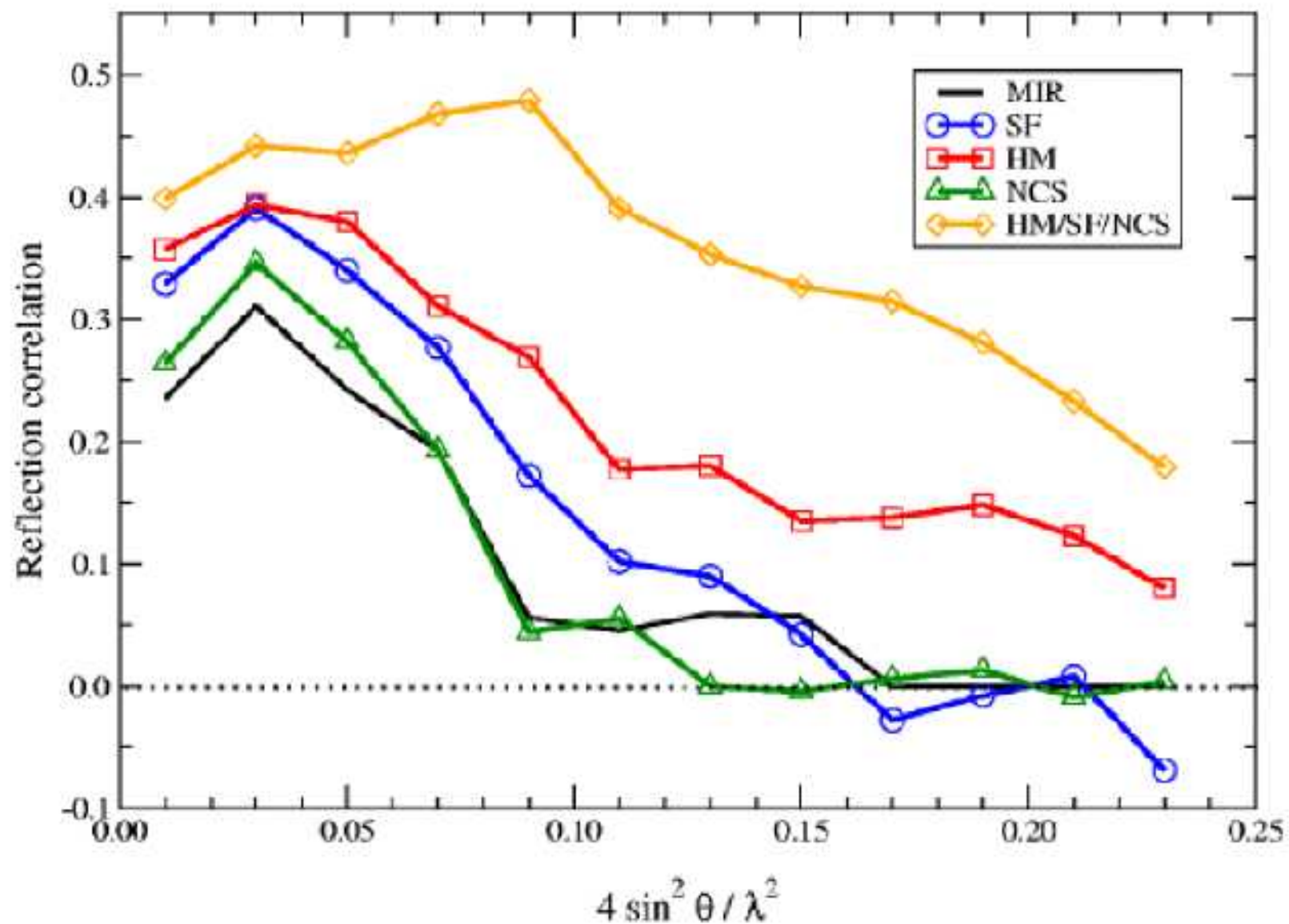
Omit one reflection at a time, and use only the modified value of the omitted reflection. (Very slow.)

But can be implemented efficiently:

- Solvent flipping
- The γ -correction



Density Modification Results:



5-carboxymethyl-2-hydroxymuconate isomerase

Wigley D.B., Roper D.I., Cooper R.A.

(1989) *J. Mol. Biol.*, **210**, 881



The Royal Society
York Structural Biology Laboratory

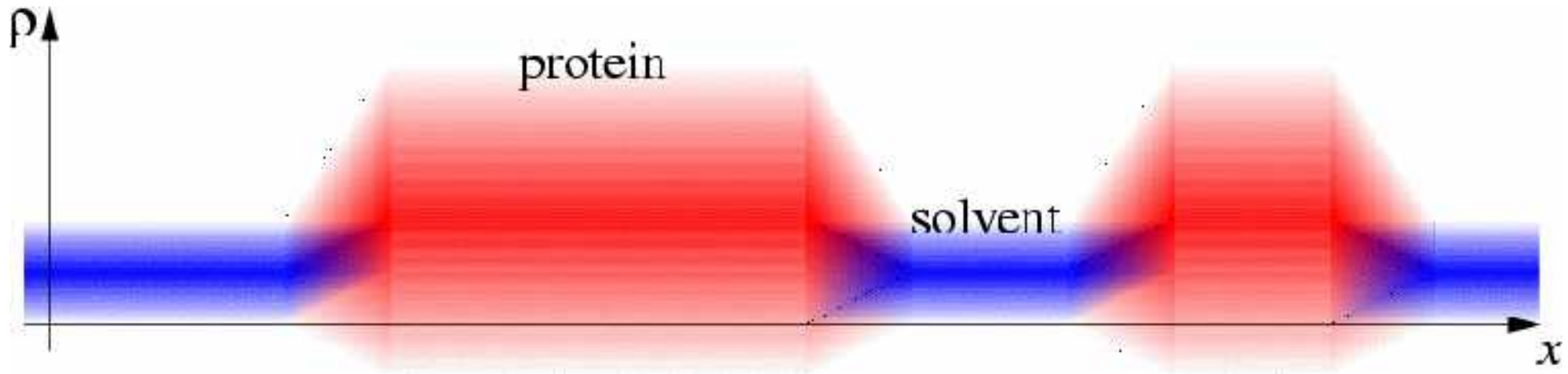
Density Modification

Kevin Cowtan, York.

Statistical density modification:
e.g. Resolve, Pirate

Statistical density modification

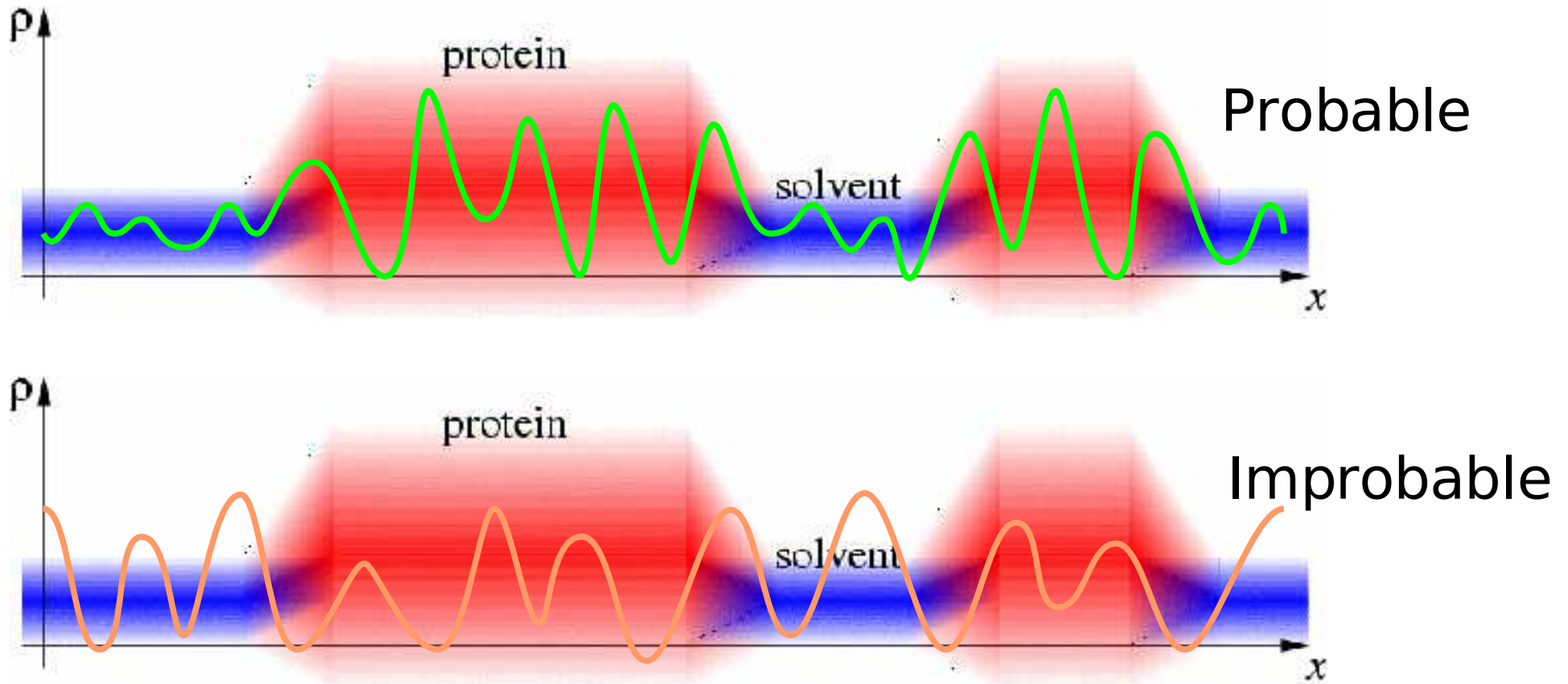
- Form a statistical description of expected map features.



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

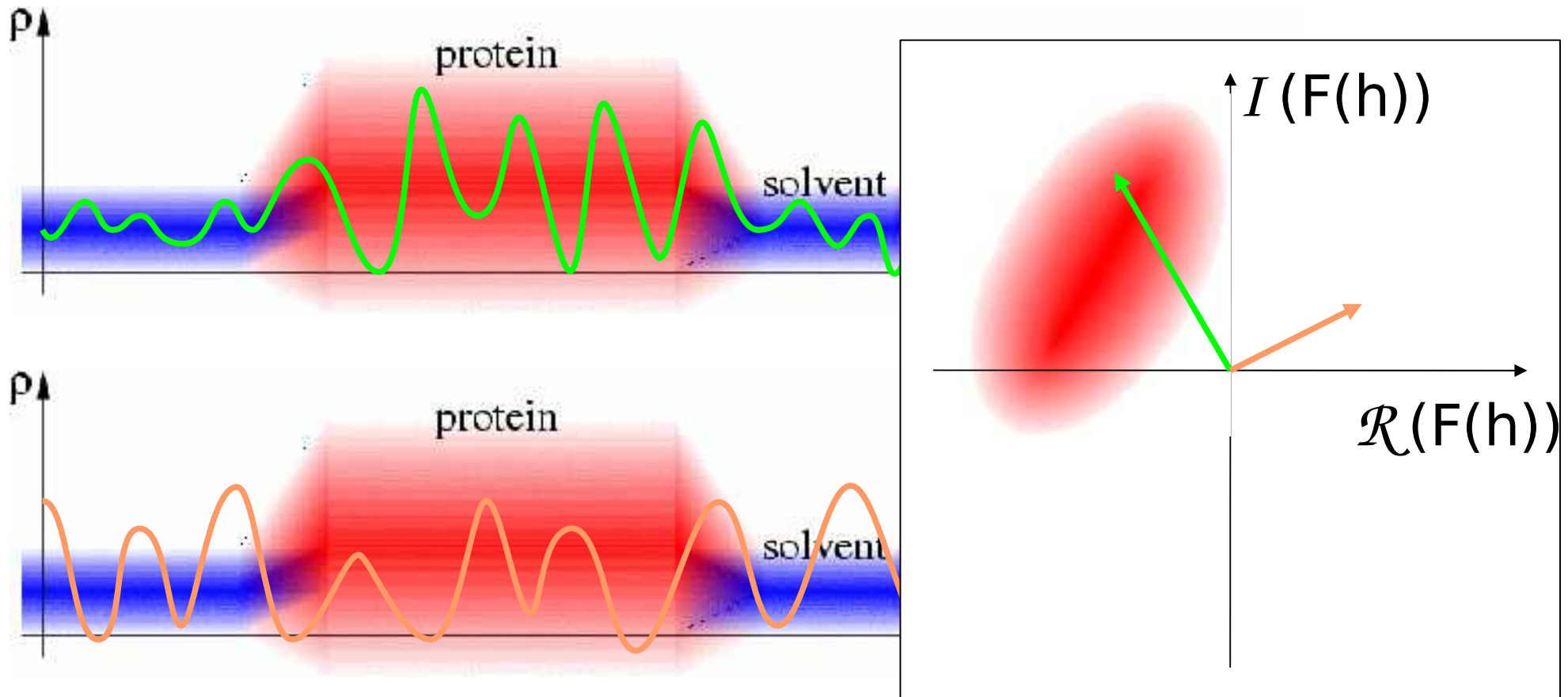
Statistical density modification

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



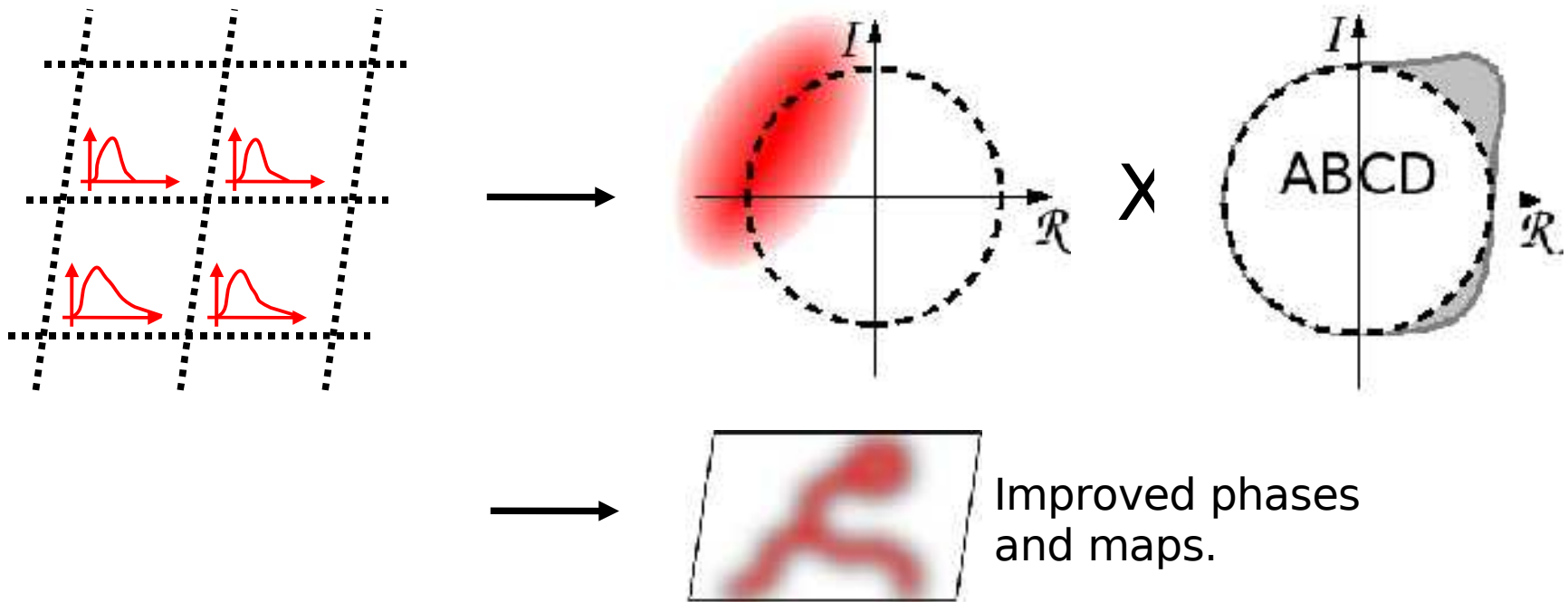
Statistical density modification

- 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

Statistical density modification

Advantages:

- Reduced bias.
- Better phases?

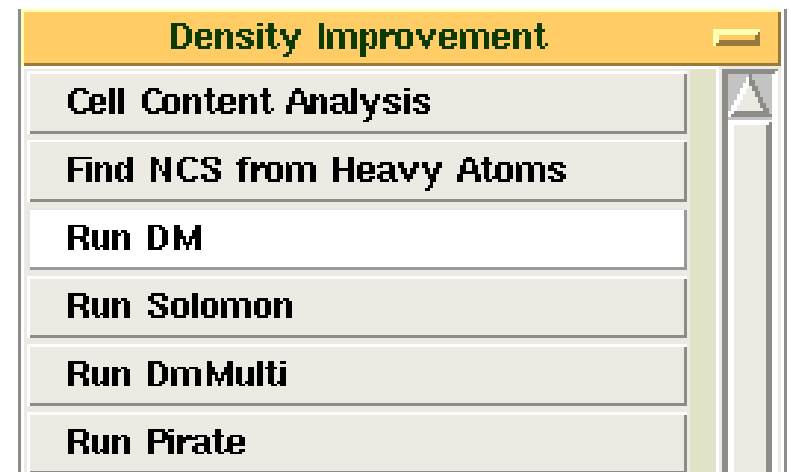
Disadvantages:

- Transforming a probability distribution from one space to another is complex and slow.
- More dependent on good error estimates from the phasing calculation.

Density modification

Useful tasks in CCP4:

- Density Improvement module:
 - 'Cell content analysis'.
 - 'DM' (trad. density modification).
 - 'DMMULTI' (multi-xtal averaging).
 - 'Pirate' (statistical density modification).
 - 'Find NCS from heavy atoms'.
- Coordinate utilities:
 - 'Superpose molecules'.



Buccaneer

The buccaneer software for automated model building of protein structures across a broad range of resolutions.

Kevin Cowtan

YSBL, University of York

cowtan@ysbl.york.ac.uk

Buccaneer

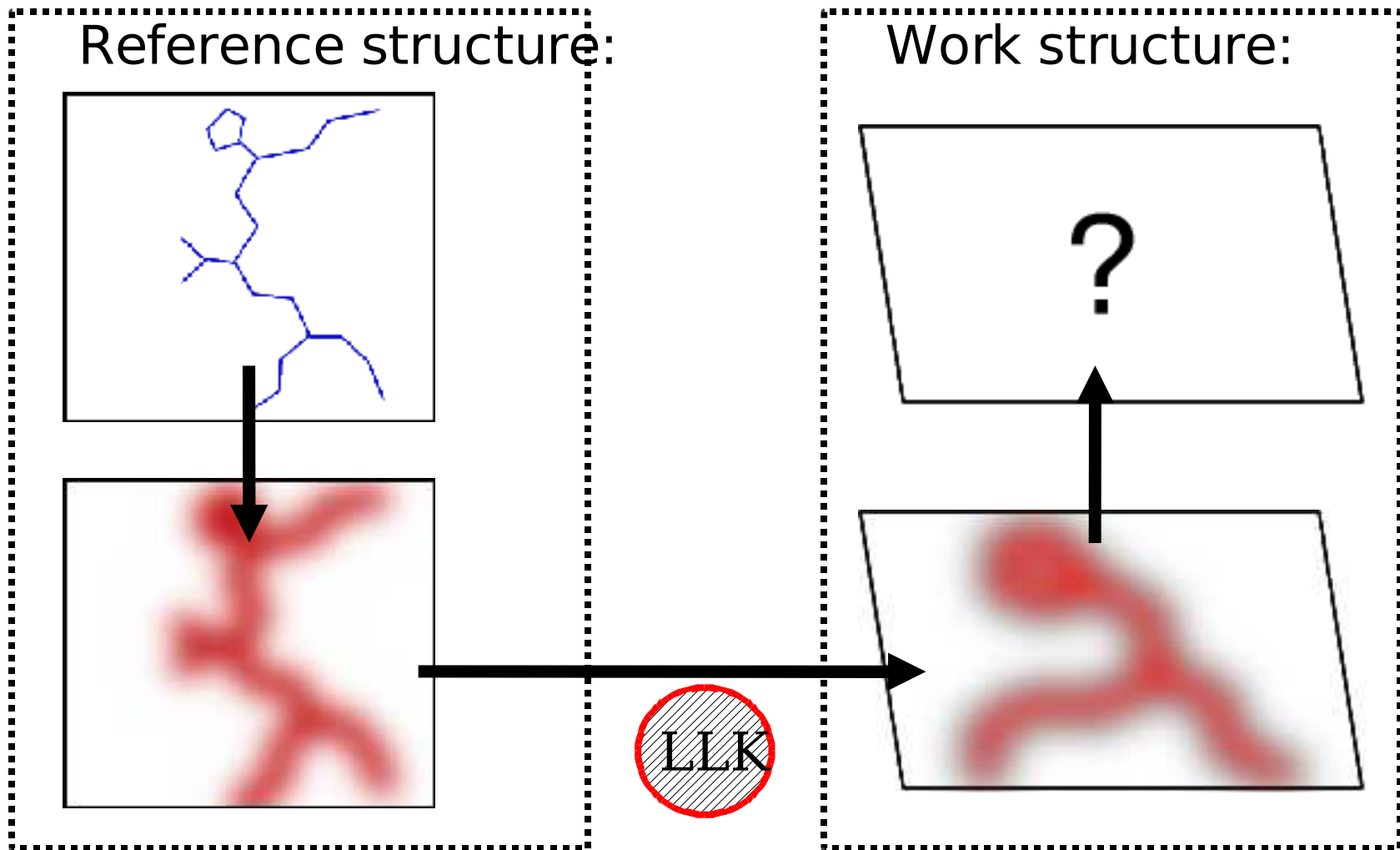
Statistical model building software based on the use of a reference structure to construct likelihood targets for protein features.

New since 2007: Version 1.0 for CCP4 v6.1

- Buccaneer-Refmac pipeline
- NCS auto-completion
- Improved sequencing

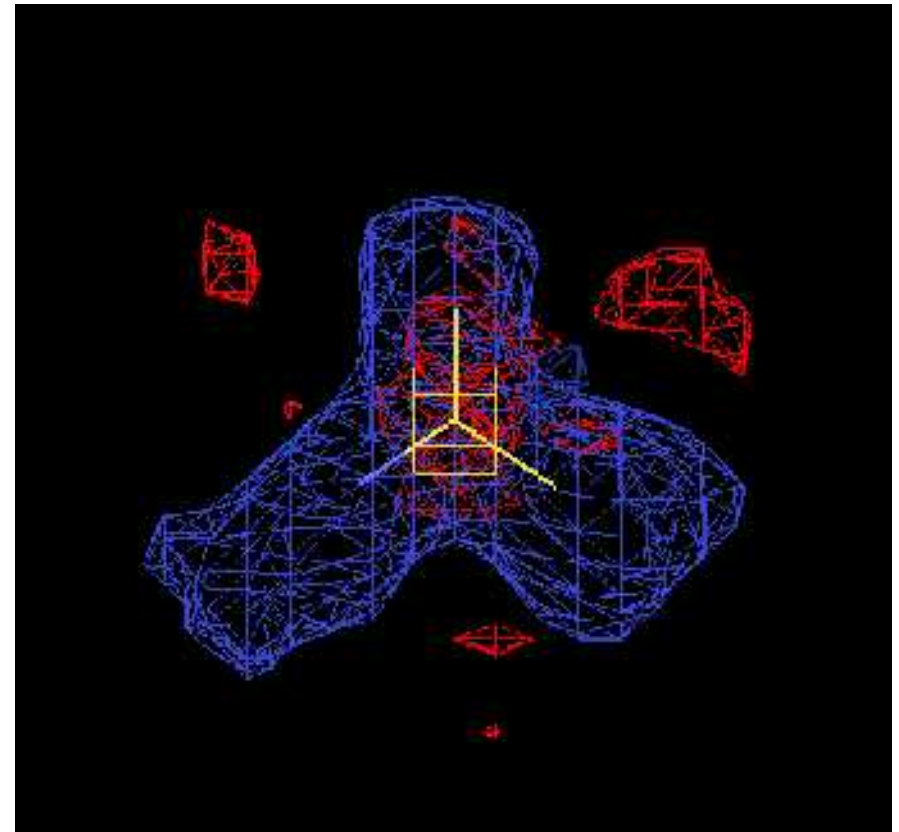
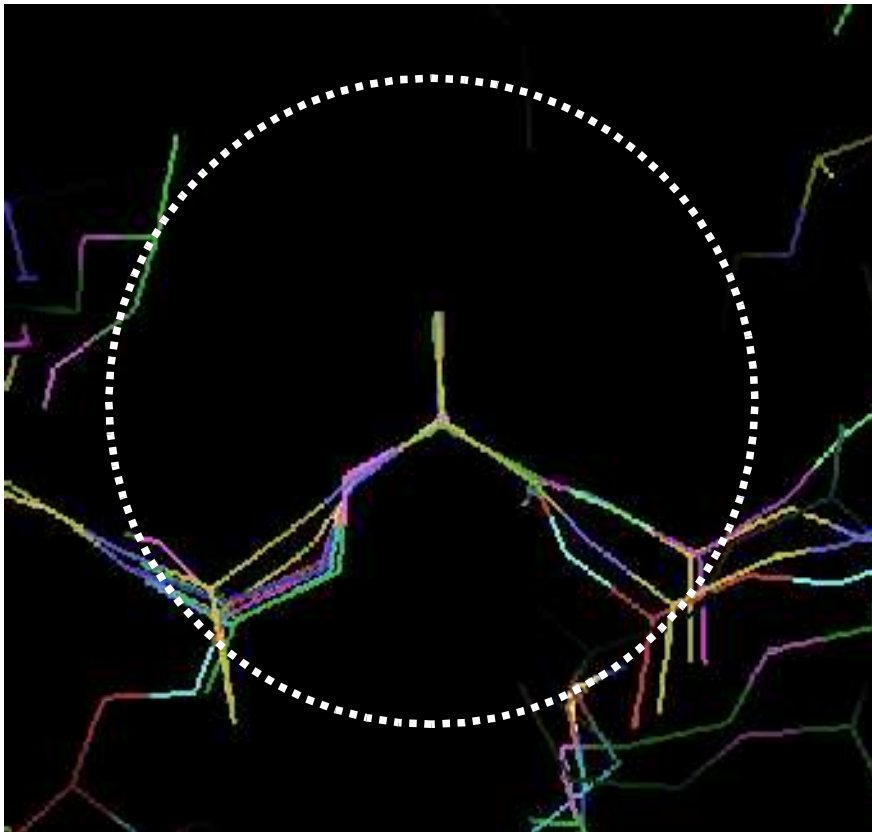
Buccaneer: Method

- Compare simulated map and known model to obtain likelihood target, then search for this target in the unknown map.



Buccaneer: Method

- Compile statistics for reference map in 4Å sphere about $C\alpha$ => LLK target.



- Use mean/variance.

4Å sphere about Ca also used by 'CAPRA' loeger et al. (but different target function).

Buccaneer

10 stages:

- **Find** candidate C-alpha positions
- **Grow** them into chain fragments
- **Join** and merge the fragments, resolving branches
- **Link** nearby N and C termini (if possible)
- **Sequence** the chains (i.e. dock sequence)
- **Correct** insertions/deletions
- **Filter** based on poor density
- **NCS Rebuild** to complete NCS copies of chains
- **Prune** any remaining clashing chains
- **Rebuild** side chains

Buccaneer

Use a likelihood function based on conserved density features.

The same likelihood function is used several times. This makes the program very simple (<3000 lines), and the whole calculation works over a range of resolutions.

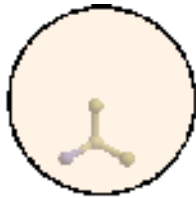
Finding, growing: Look for C-alpha environment



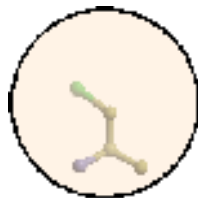
(4.0Å sphere about Cα)

Sequencing:

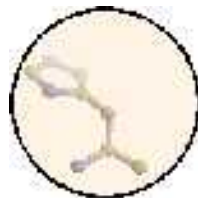
Look for C-beta environment



ALA



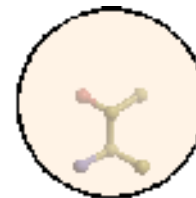
CYS



HIS



MET



THR

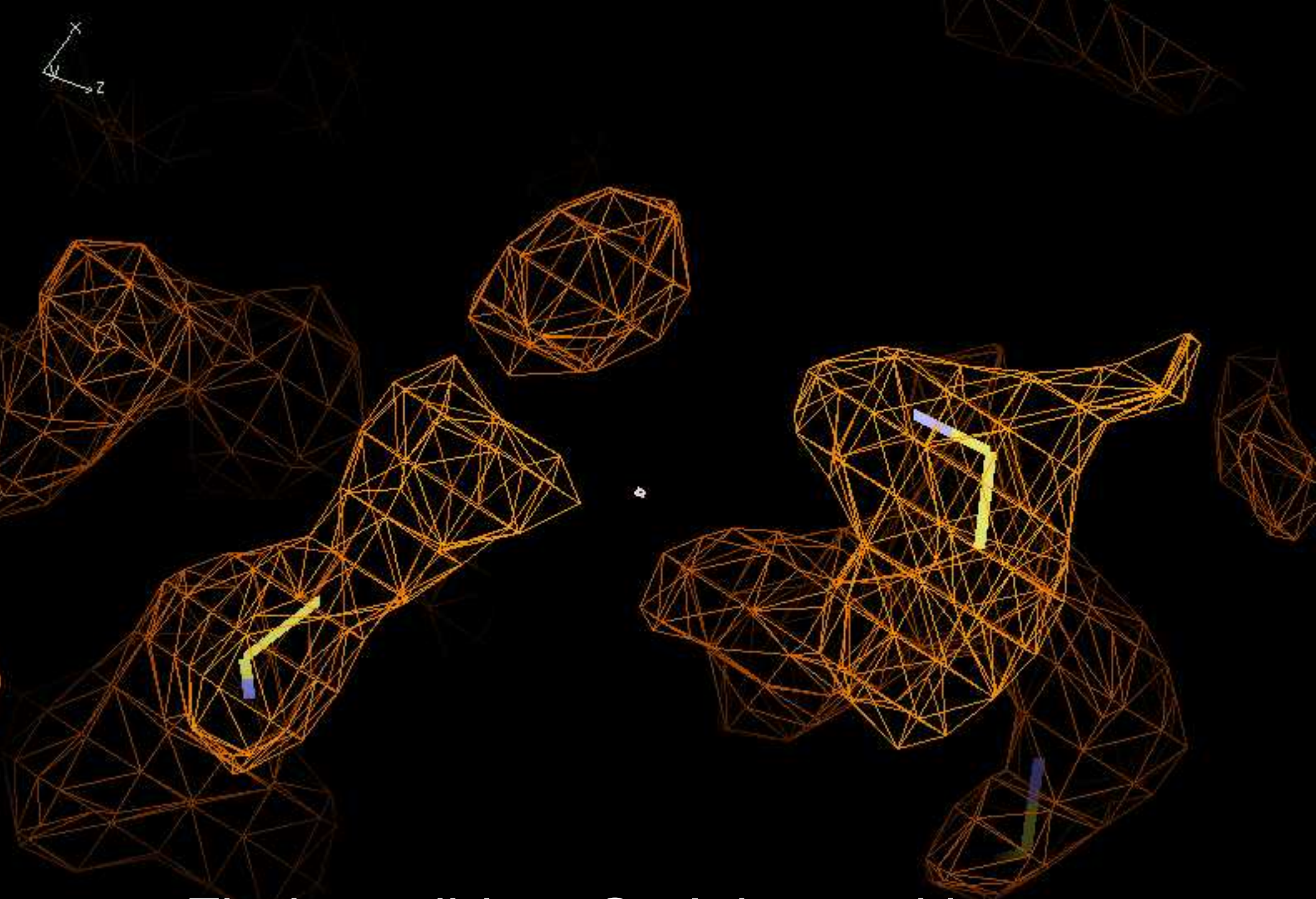
(5.5Å sphere about Cβ)

... x20

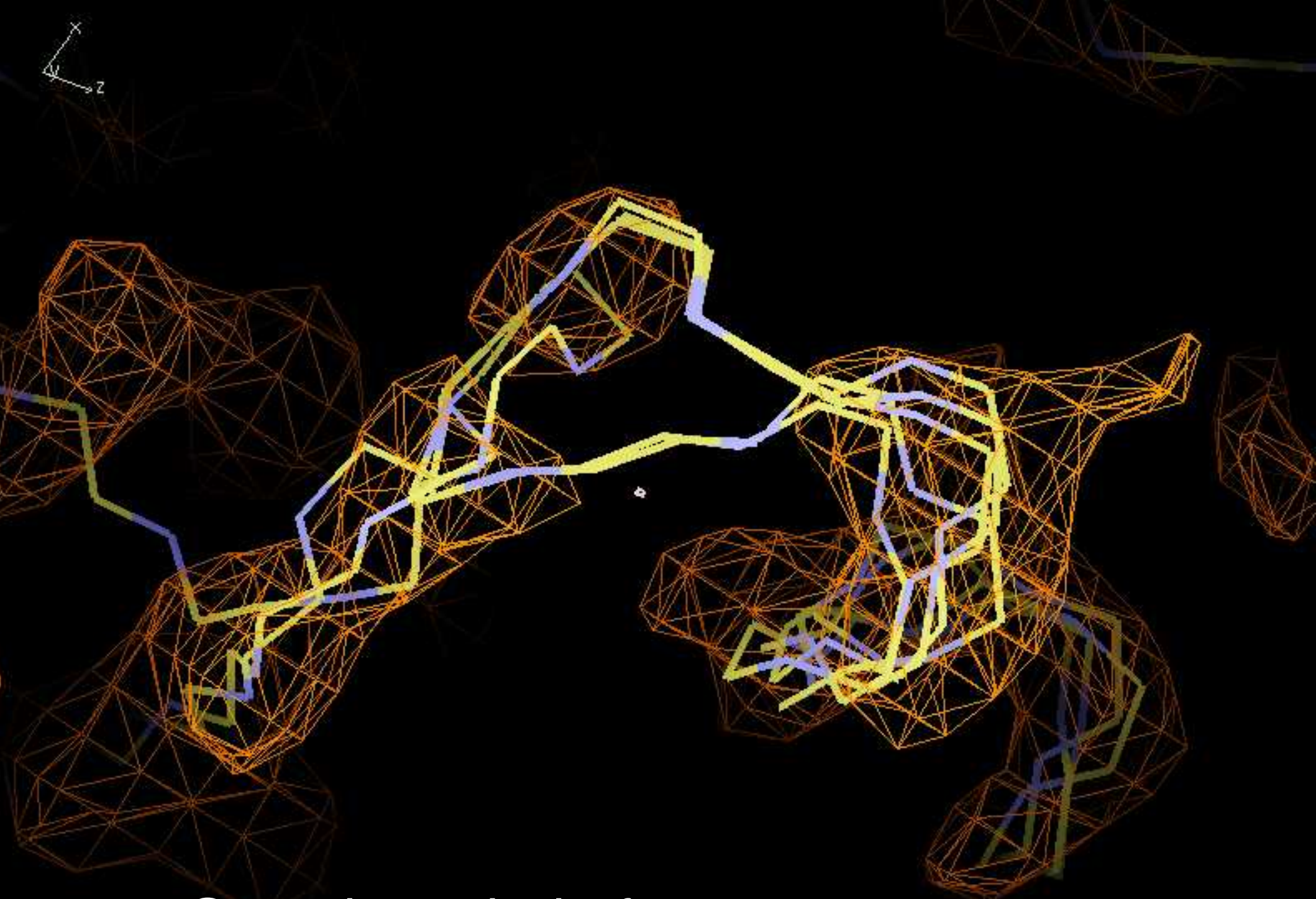
Buccaneer

Case Study:

A difficult loop in a 2.9Å map, calculated using real data from the JCSG.



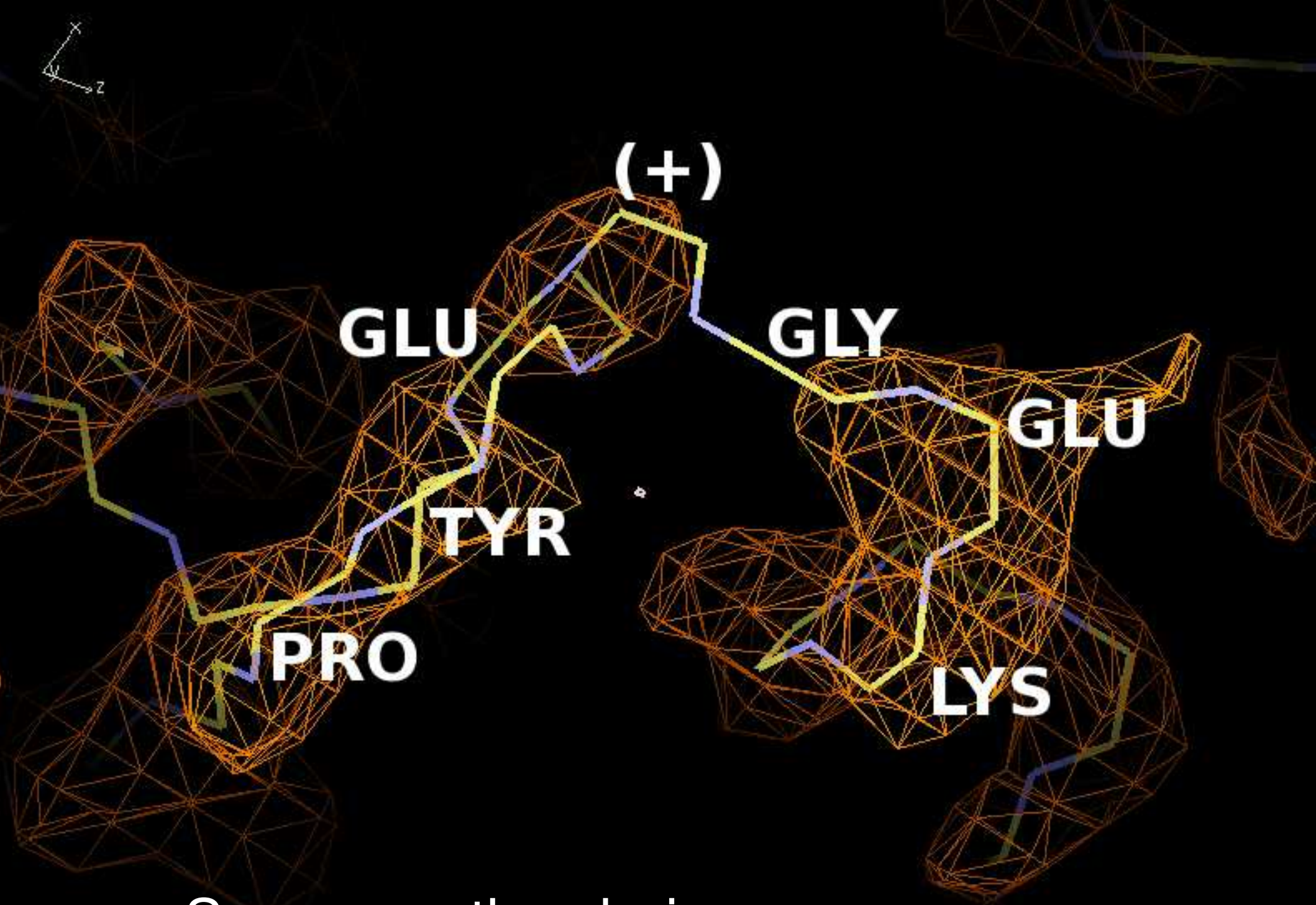
Find candidate C-alpha positions



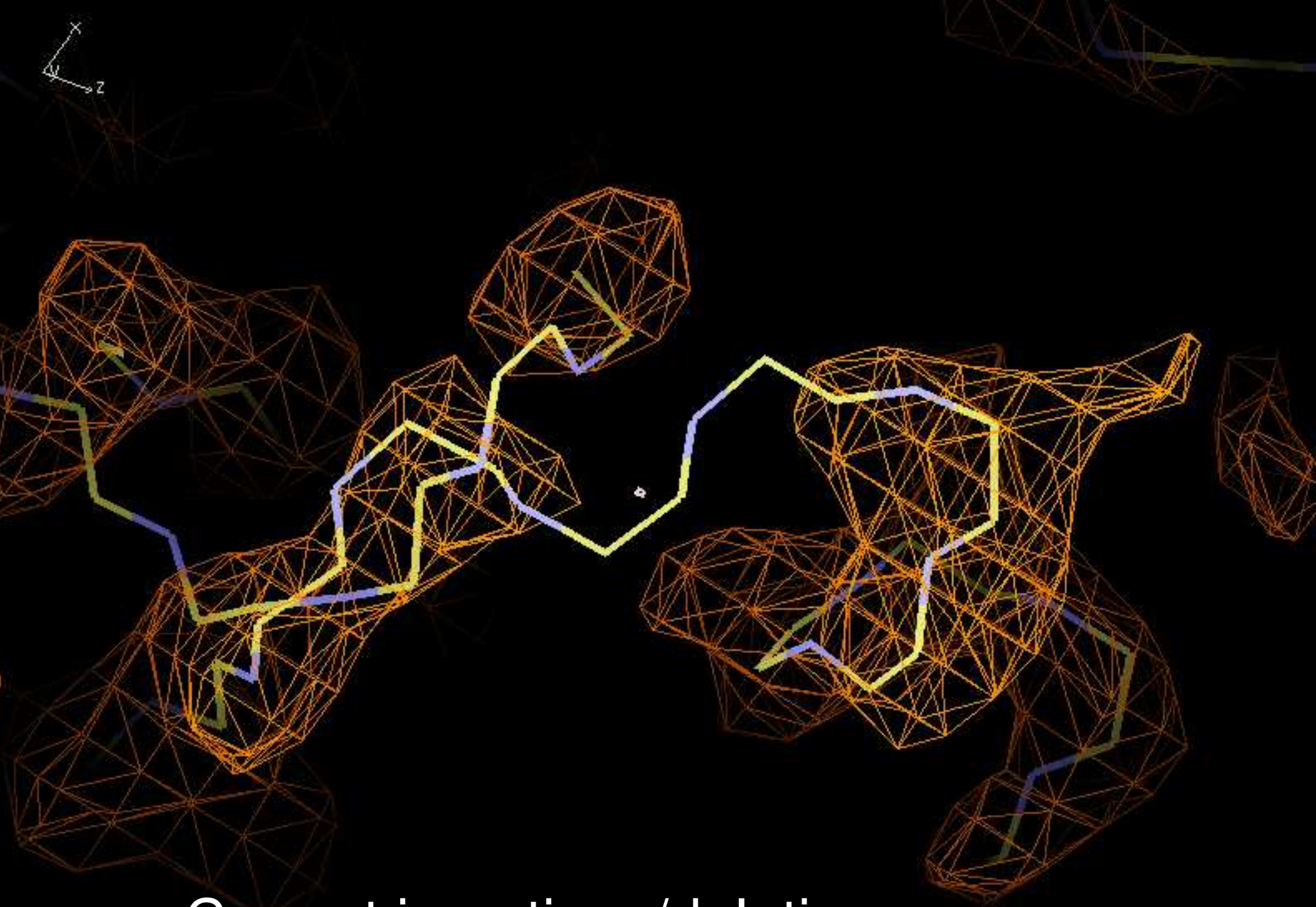
Grow into chain fragments



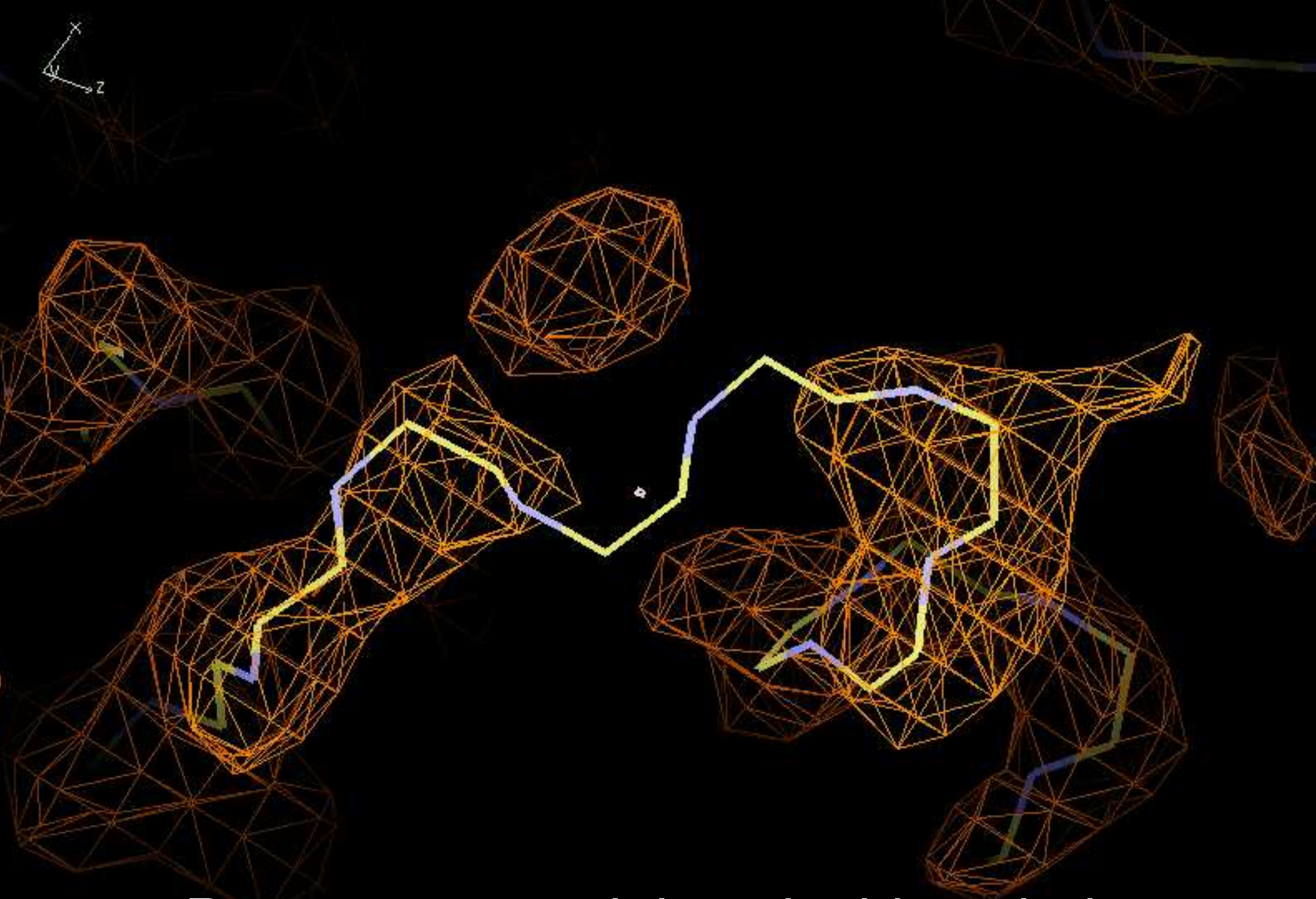
Join and merge chain fragments



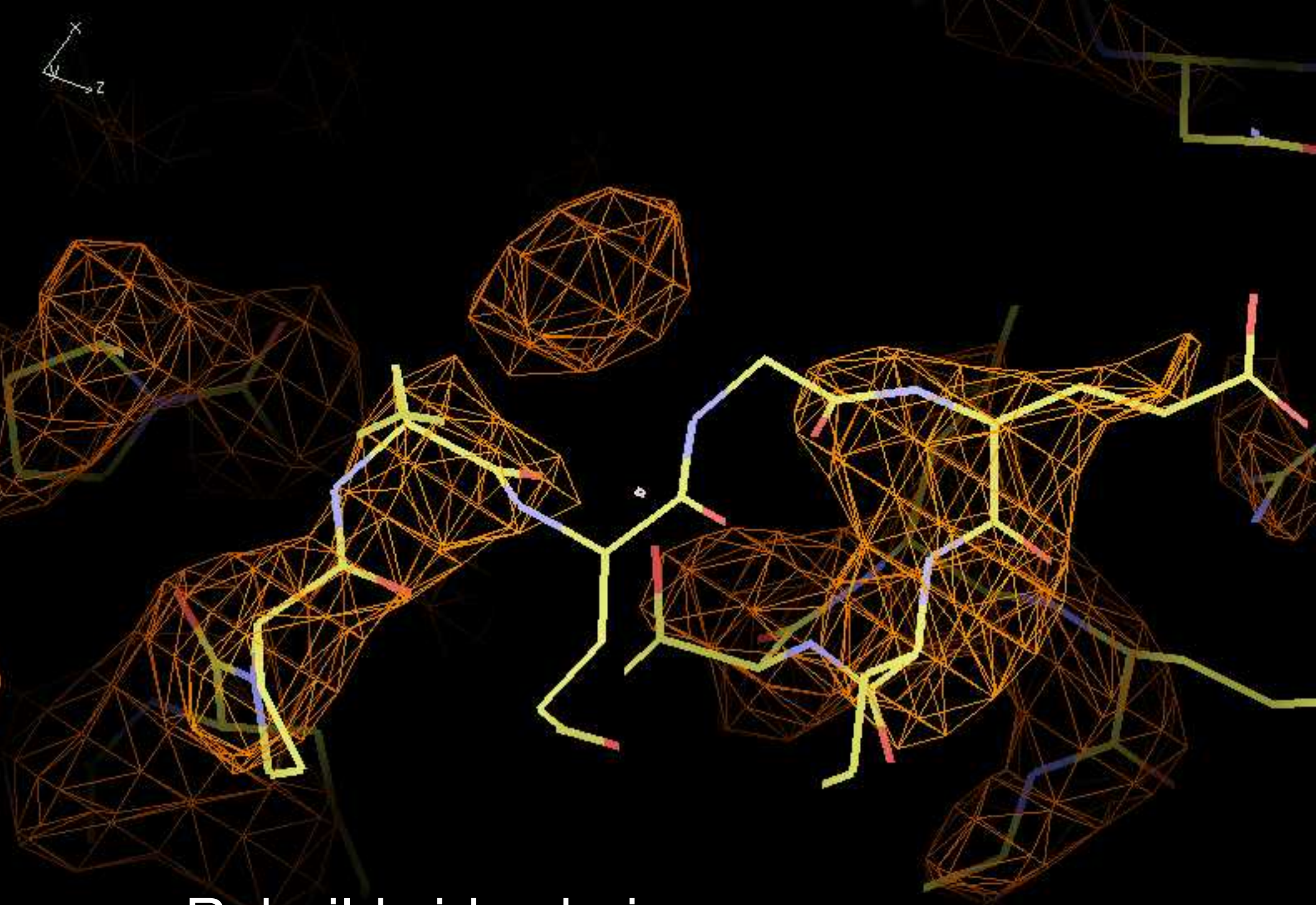
Sequence the chains



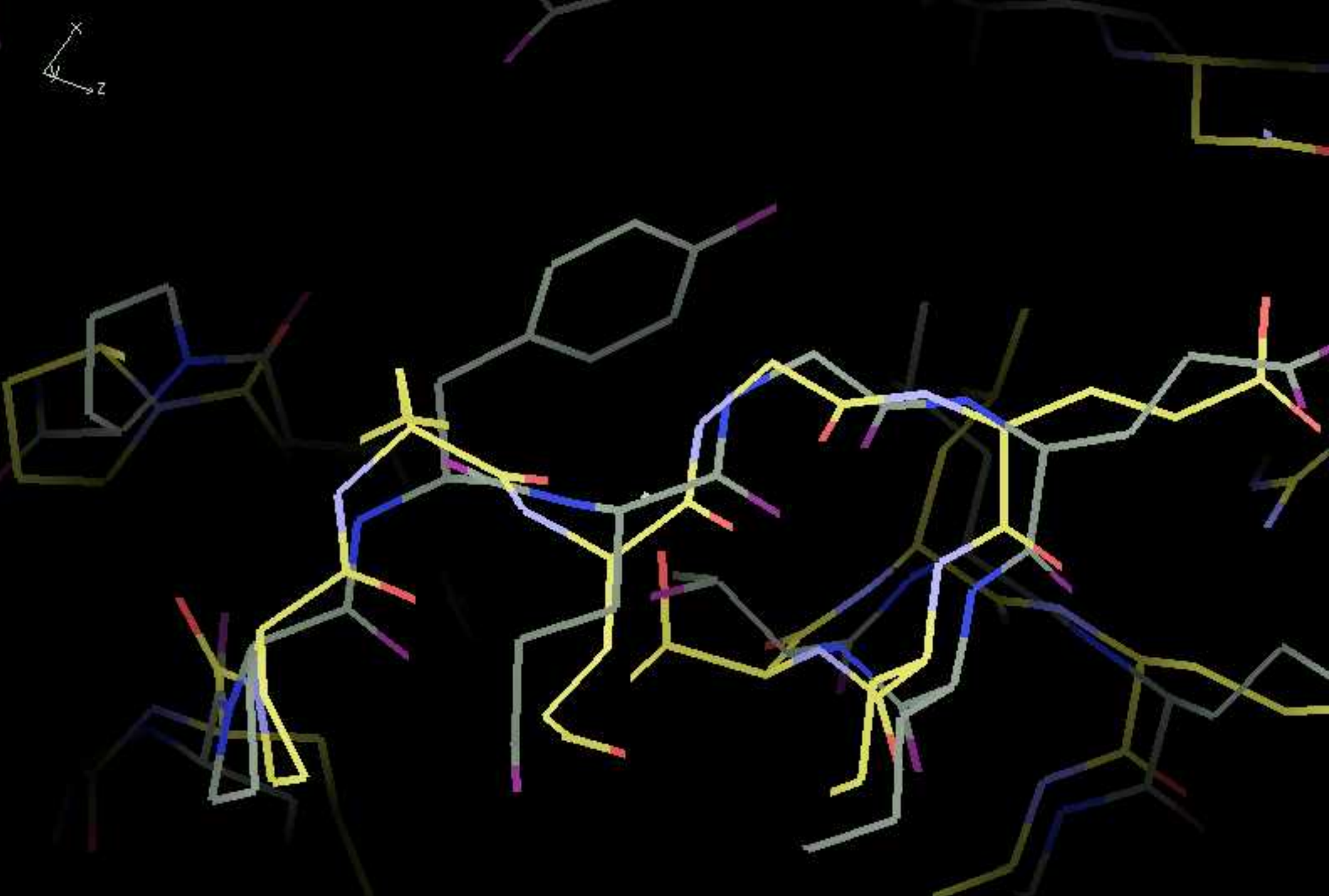
Correct insertions/deletions



Prune any remaining clashing chains



Rebuild side chains

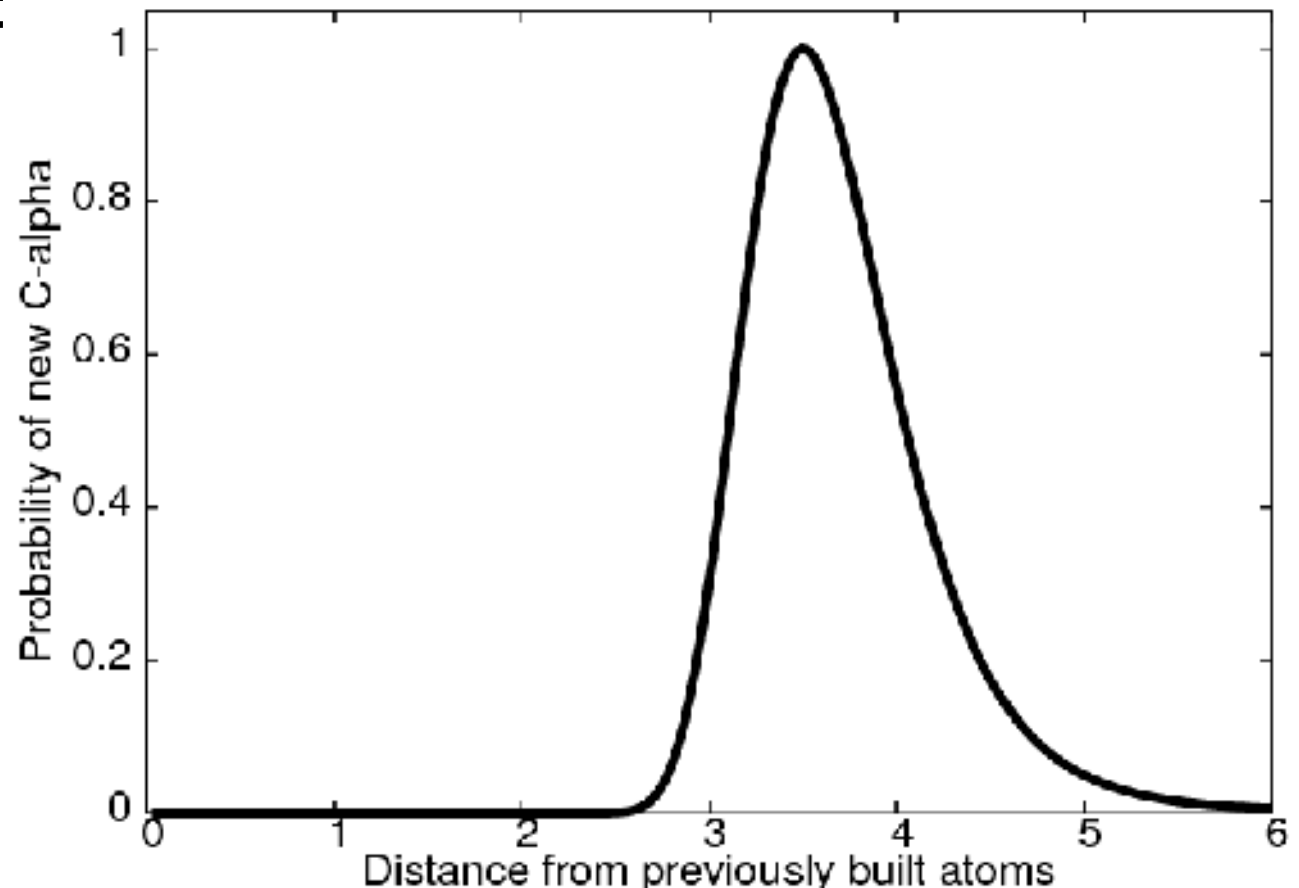


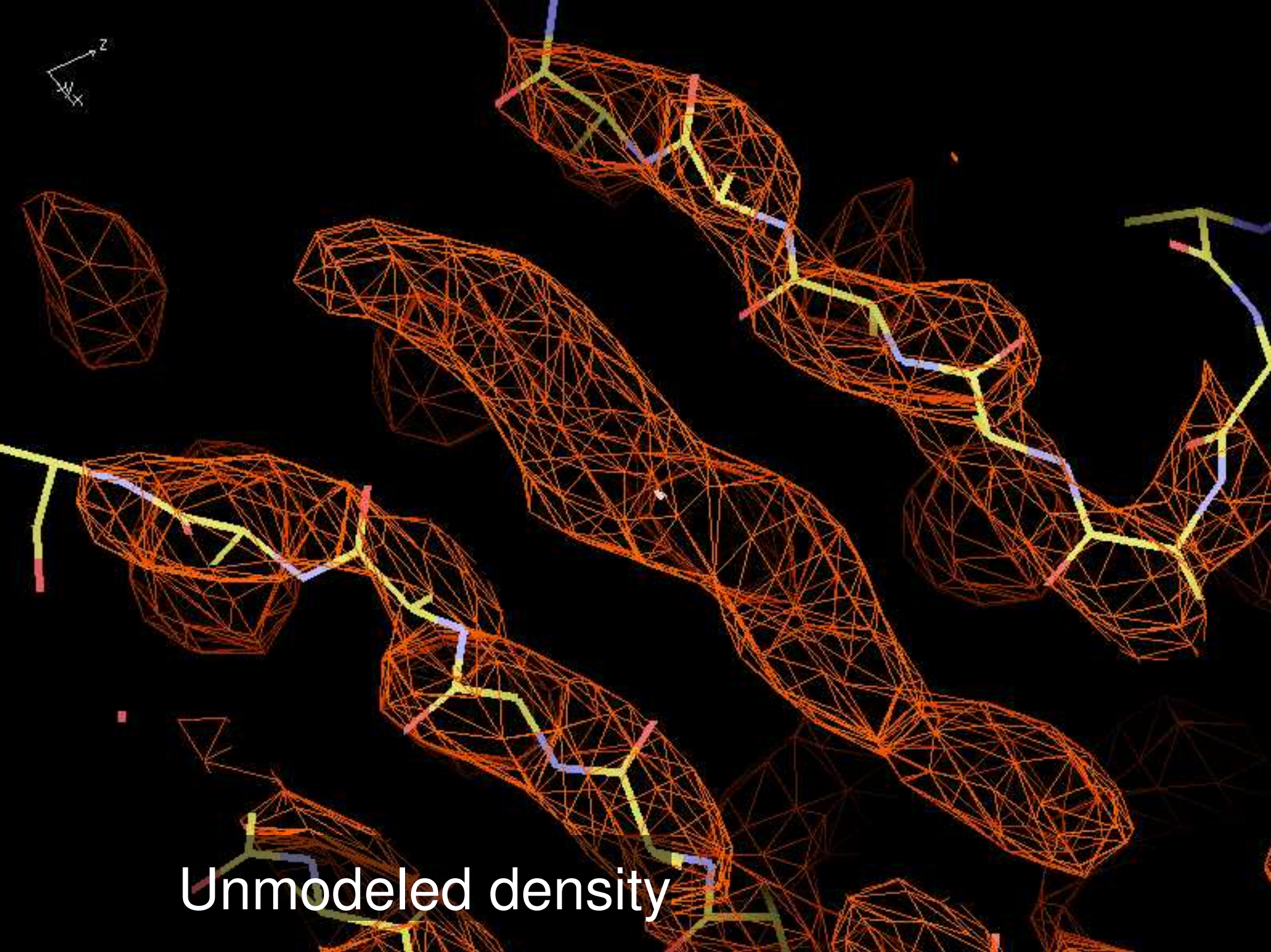
Comparison to the final model

Buccaneer

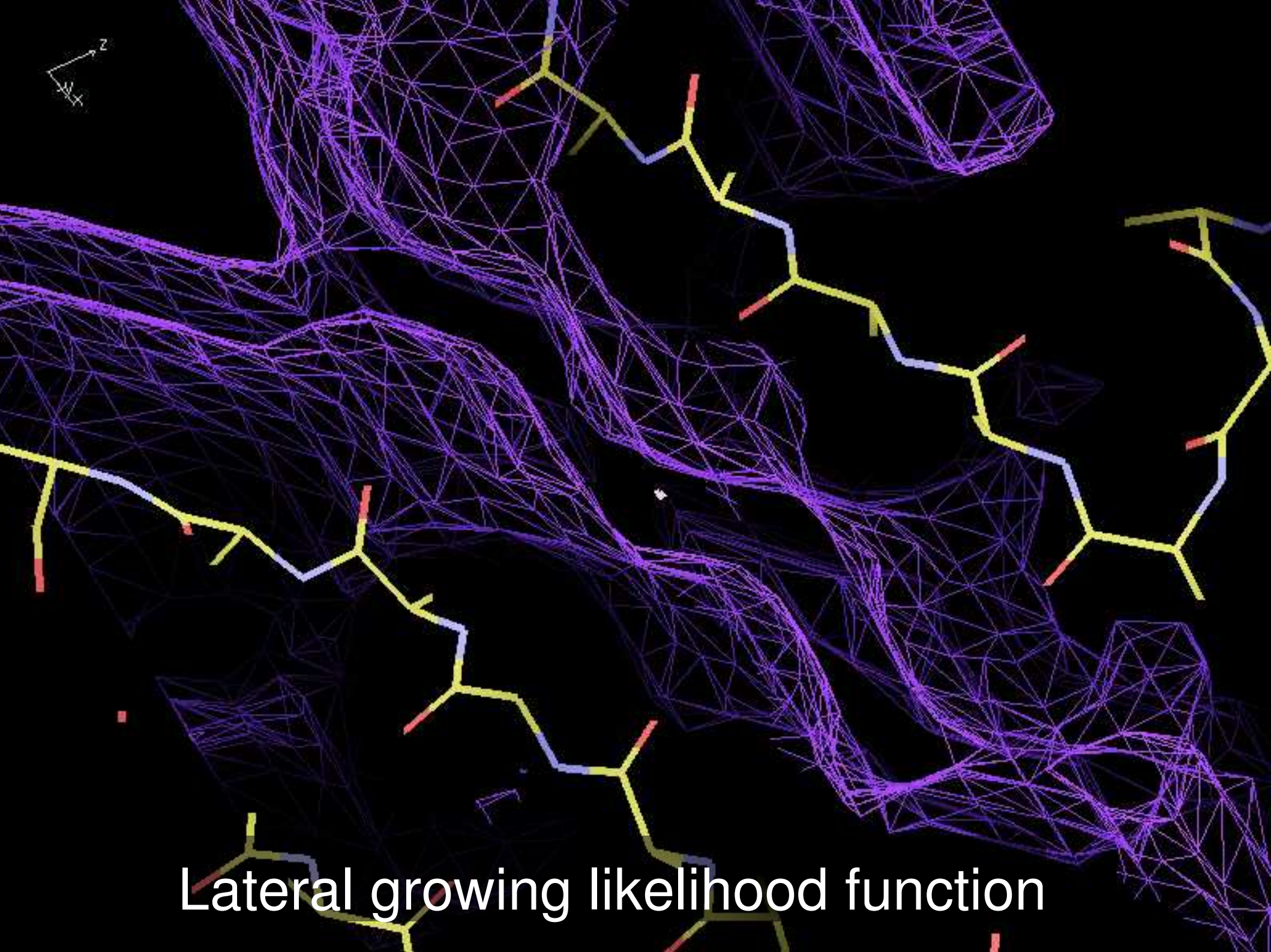
Model completion uses “**Lateral growing**”:

Grow sideways from existing chain fragments by looking for new C-alphas at an appropriate distance “sideways” from the existing chain:

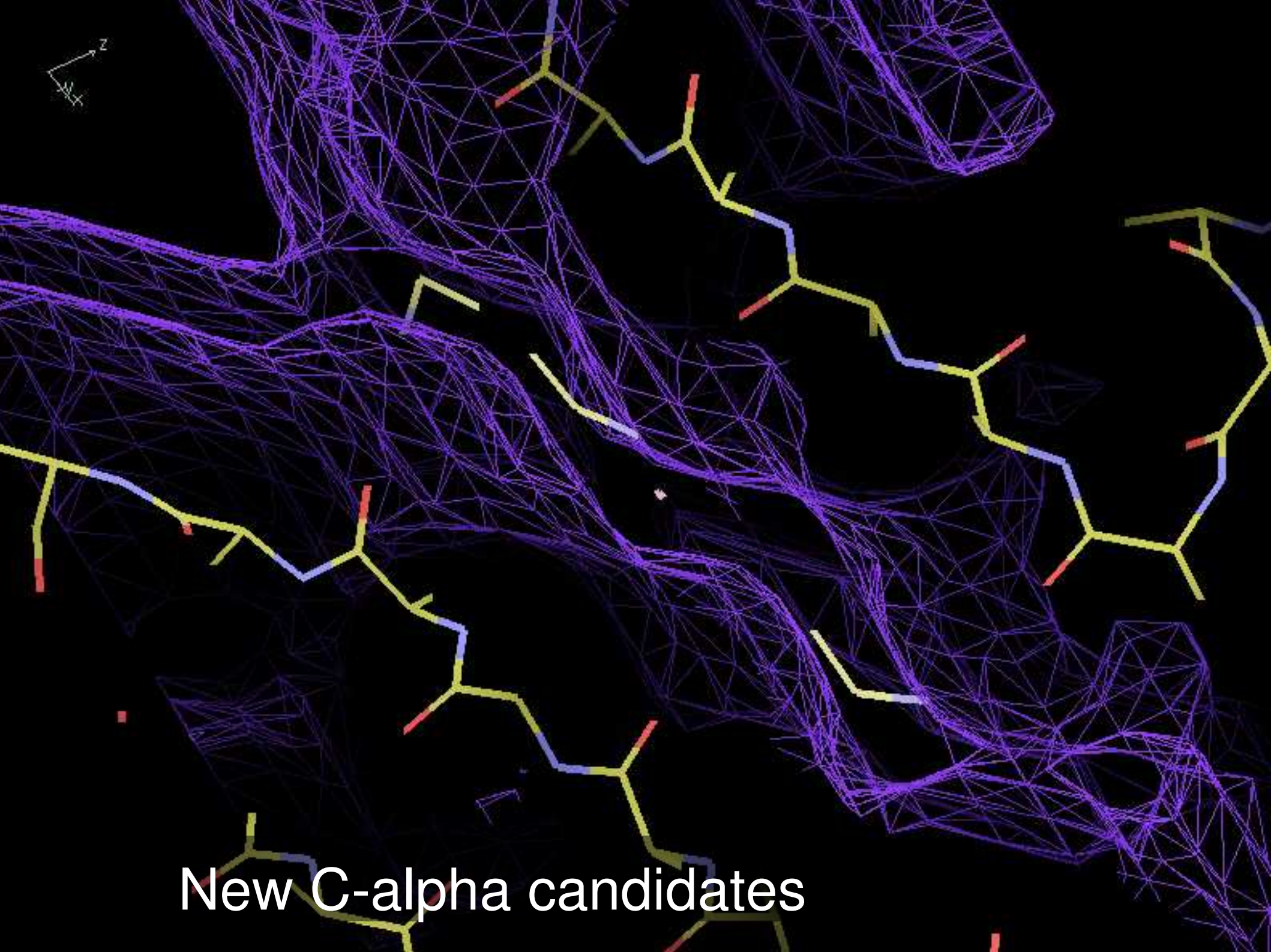




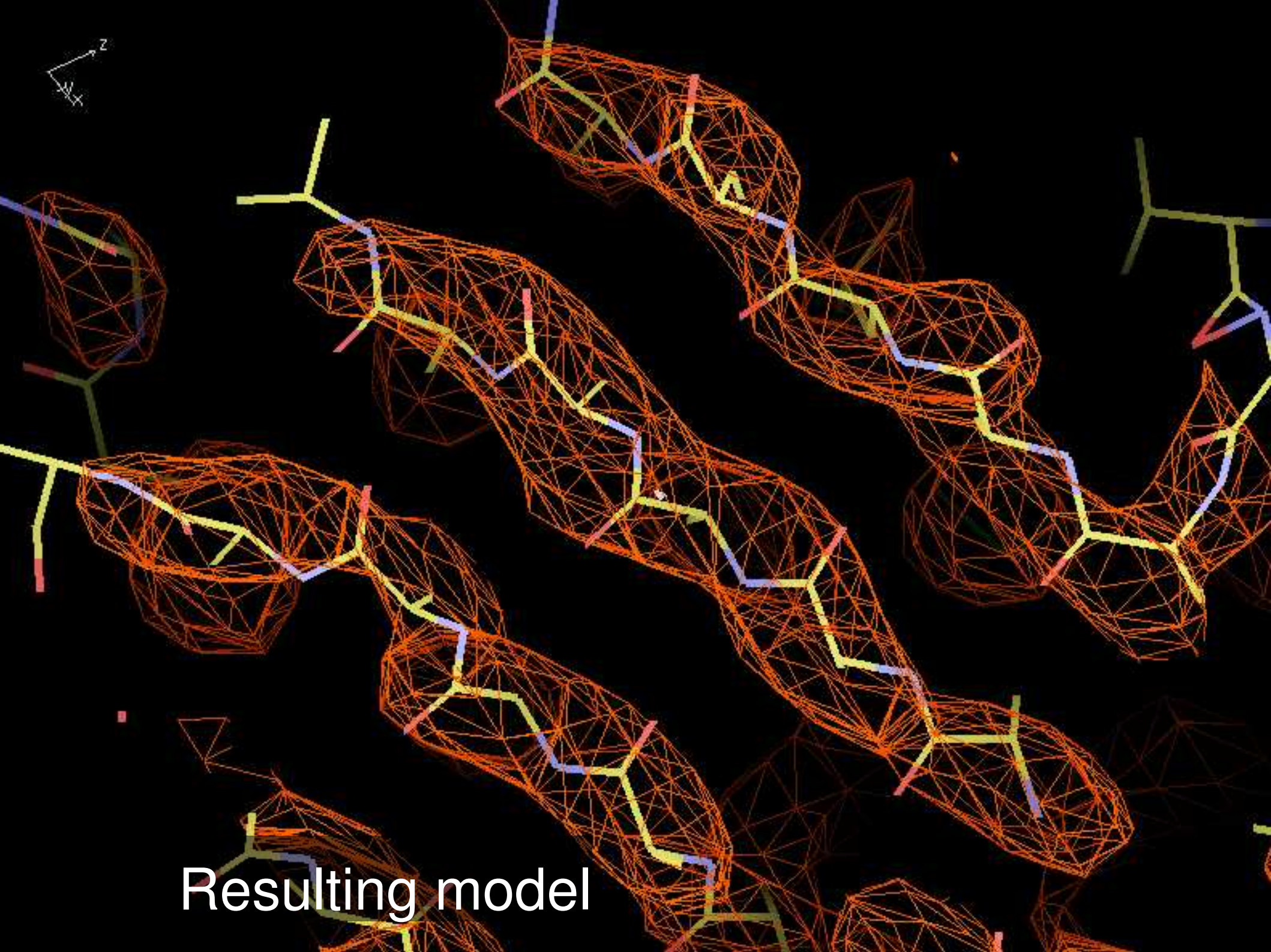
Unmodeled density



Lateral growing likelihood function



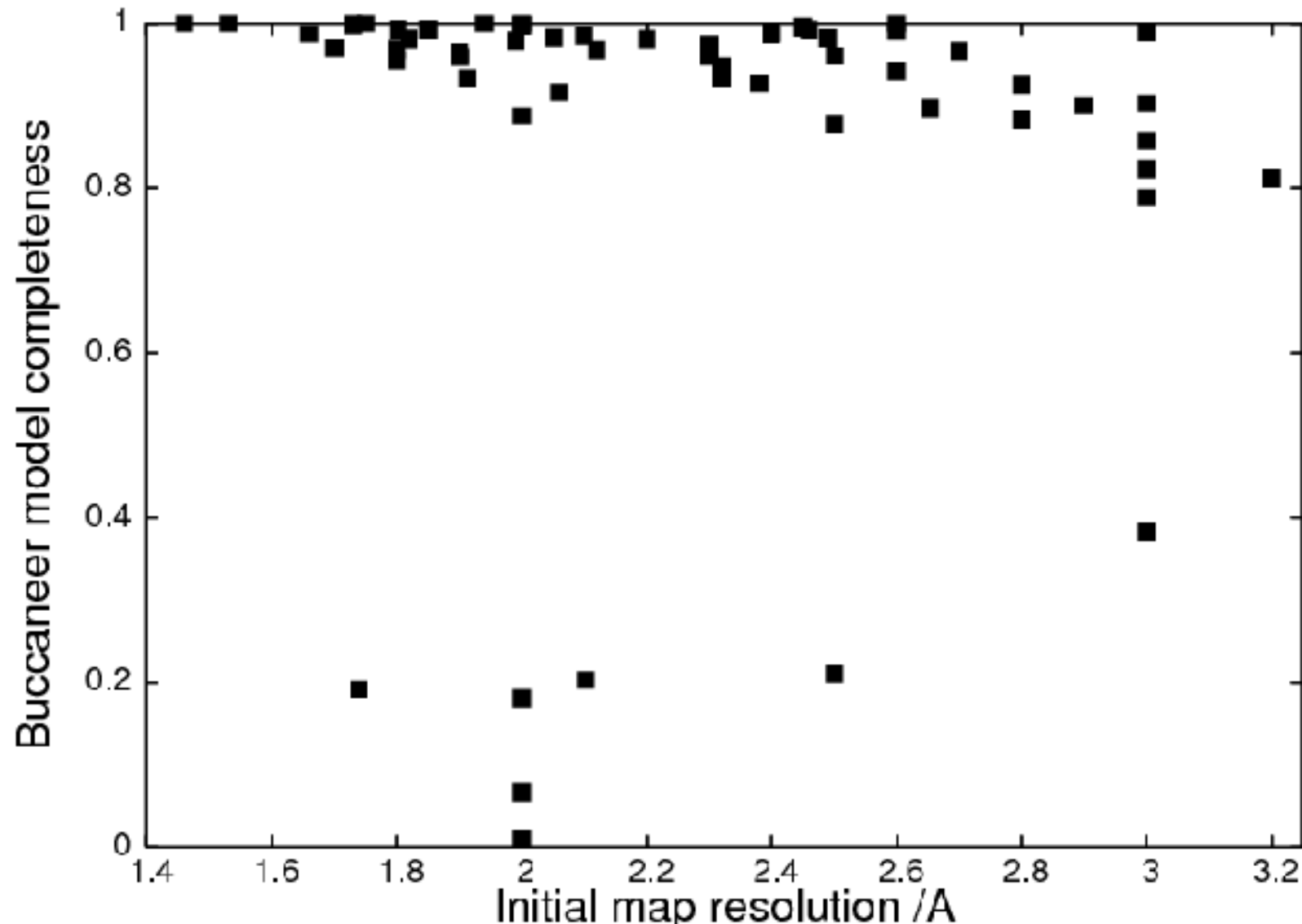
New C-alpha candidates



Resulting model

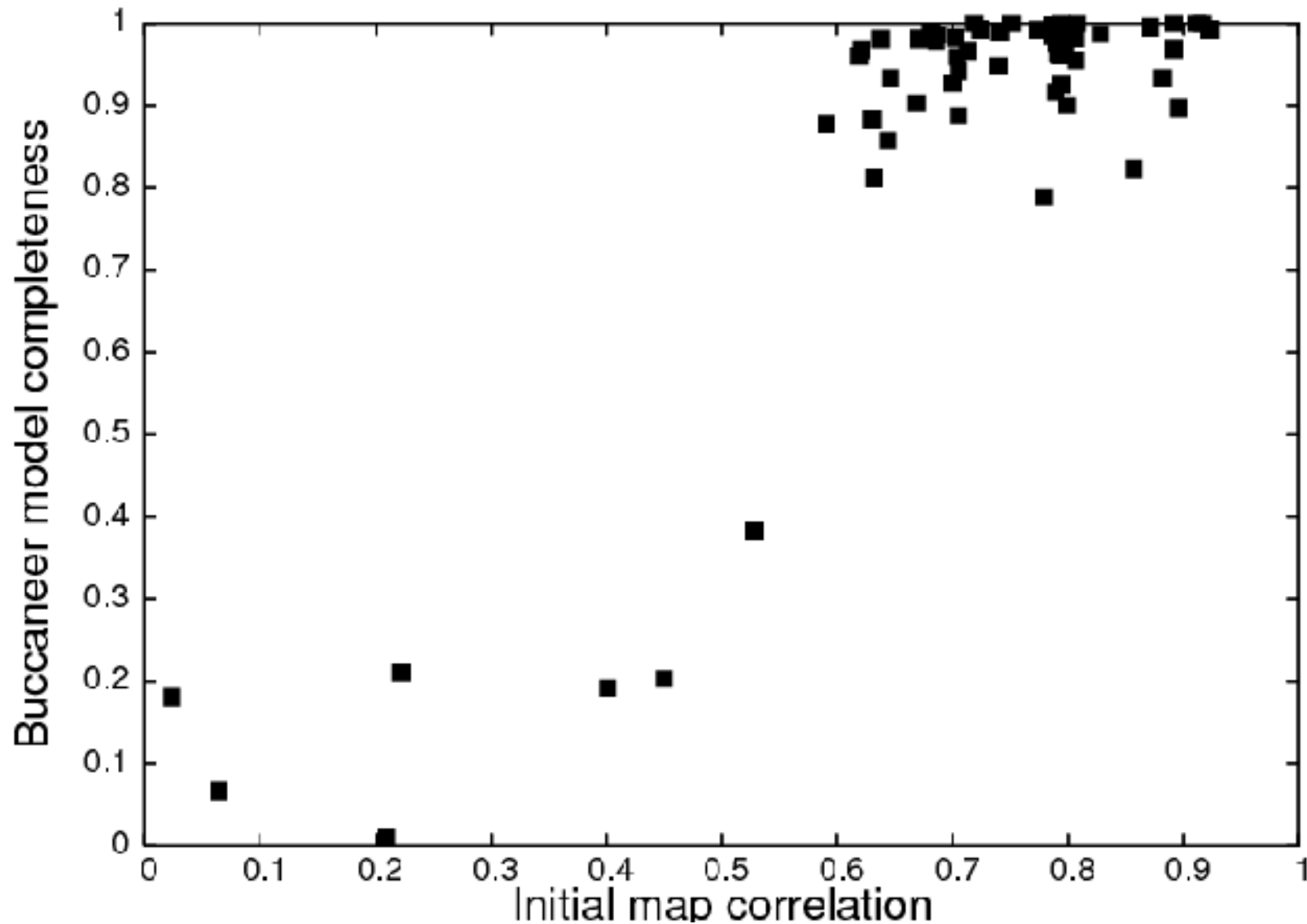
Buccaneer: Results

Model completeness not very dependent on resolution:



Buccaneer: Results

Model completeness dependent on initial phases:



Buccaneer

Chain tracing/refinement using Buccaneer/Refmac

Help

Job title

Data for (unsolved) work structure: (Note: perform phase improvement/density modification first)

☐ Specify an initial model to be extended.

Work SEQ in PROJECT Browse View

Work MTZ in PROJECT Browse View

FP SIGFP

HLA HLB

HLC HLD

Free R flag

Use Free-R flag: ☒ Use map coefficients: ☐ Use PHI/FOM instead of HL coefficients: ☐

Work PDB out PROJECT buccaneer.pdb Browse View

Options ☒

Number of cycles of building/refinement to run: 3

Buccaneer parameters ☐

Refmac parameters ☐

Run Save or Restore Close

Buccaneer: Comparisons

With ARP/wARP:

- From my tests and user reports buccaneer will build most structures ARP/wARP will build.
 - But ARP/wARP gets a lower R-factor.
 - The benefit of using individual atoms.
- Buccaneer will also build some structures where ARP/wARP fails.
 - Particularly with good data but poor resolution.
 - The benefit of using larger rigid groups.
- Best approach: Buccaneer then ARP/wARP.

Buccaneer: Comparisons

User reports:

- 2 novel structures build at 3.6A
- 1 commercial MR structure rebuilt at York at 3.1A (Buccaneer/Coot/Quanta)
- Numerous other reports (but few citations).

Buccaneer: Future

Buccaneer 1.1

- Improved numbering of output sequences (ins/del)
- Favour more probable sidechain rotamers
- Prune clashing side chains
- Optionally fix the model in the ASU
- Use Se and S atoms to help sequencing
- Performance improvements (33%)
- Multi-threaded

Buccaneer: Future

Buccaneer 1.2+

- Rebuilding cis-peptides, pep-flips, pep-rots.
- Sequence verification.
- Loop building.
- Low resolution model refinement.
- B-factor refinement.
- Integrated density modification.

Buccaneer: Summary

A simple, fast, easy to use (i.e. MTZ and sequence) method of model building which is robust against resolution.

Results can be further improved by iterating with refinement in refmac (and in future, density modification).

Proven on real world problems.

The Tutorial

Experimental data to model with marginal 3A data:

- Phase with 'crank'.
- Phase improvement with 'dm'.
- Build with 'buccaneer'.
- Find NCS.
- Redo phase improvement.
- Redo model building.
- Finish in 'Coot'. Who can get the lowest R_{free} ?

Use CCP4 documentation wiki: <http://www.ccp4wiki.org>

Achnowledgements

Help:

- JCSG data archive: www.jcsg.org
- Paul Emsley
- Eleanor Dodson

Funding:

- The Royal Society