



Missoula, June 27, 2024

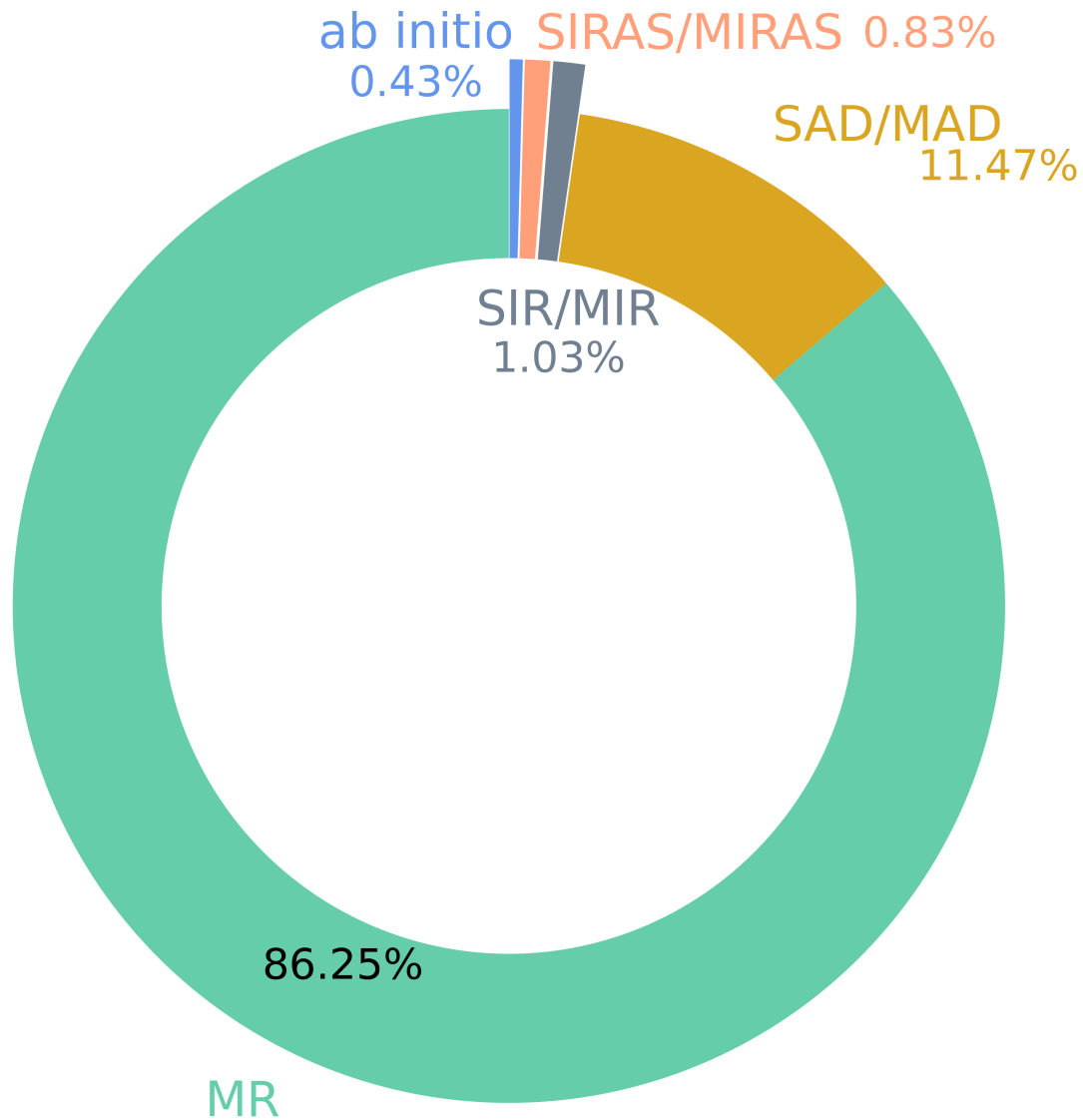
Planning and carrying out
automated structure determination
using **SAD phasing**

Single wavelength **Anomalous Diffraction**

Dorothee Liebschner, Lawrence Berkeley National Laboratory
Tom Terwilliger, Los Alamos National Laboratory



Phasing methods in the PDB



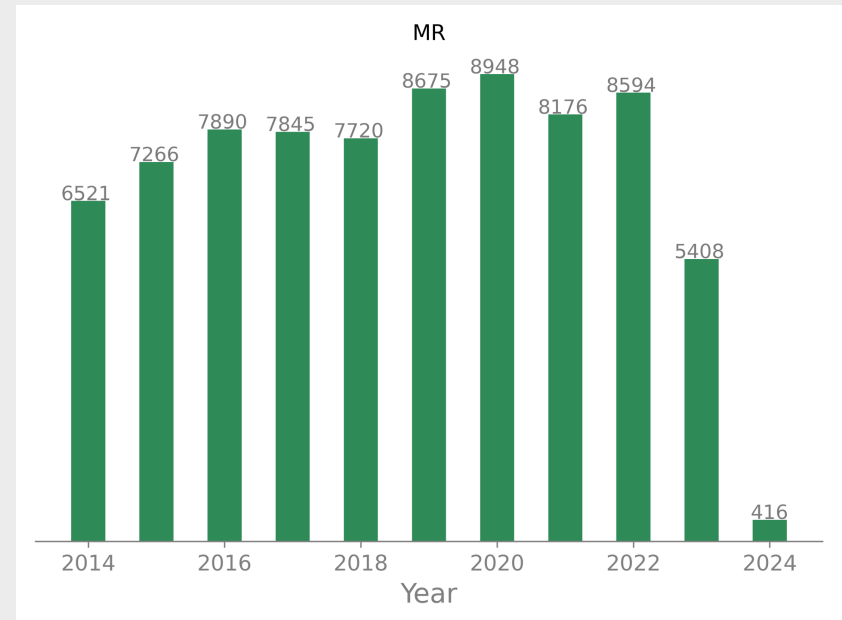
Note: Not all models in the PDB have (correct) info

Phasing methods in the PDB

SAD



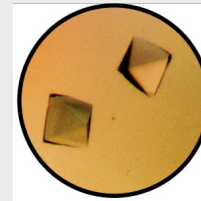
Molecular replacement



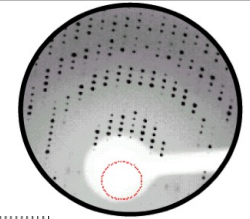
- Less experimental phasing
- More and more MR structures
- Trend will continue with predicted models (AlphaFold etc.)

Solving a structure with SAD phasing (Se)

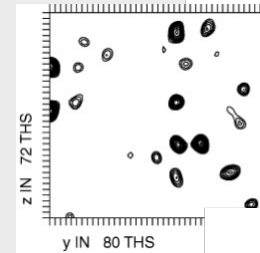
Crystals with SeMet



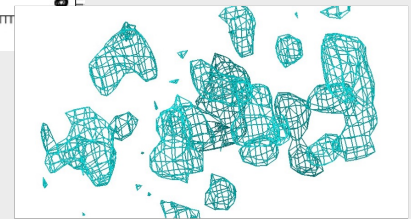
Collect anomalous SAD data



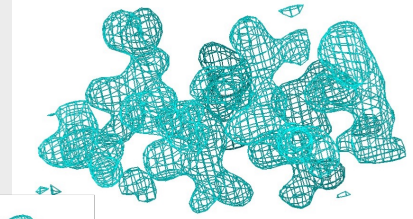
Locate Se atoms



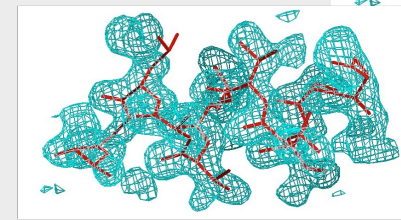
Phasing (calculate density map)



Density modification (improve map)



Model building



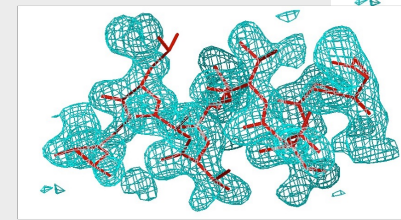
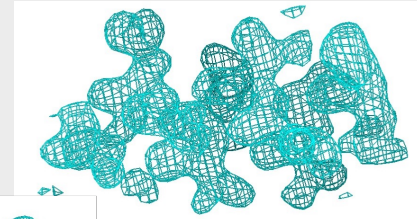
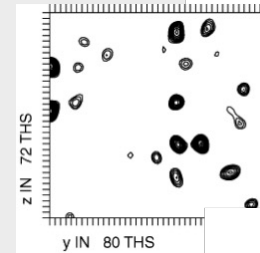
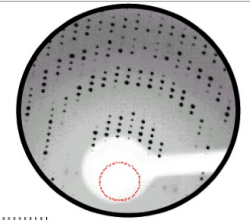
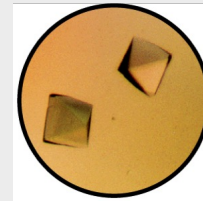
Solving a structure with SAD phasing (Se)

Planning the experiment

Automating the analysis

Improving the map

Building a model



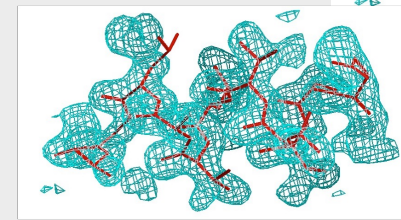
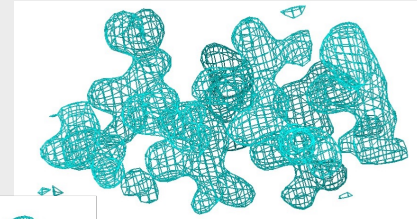
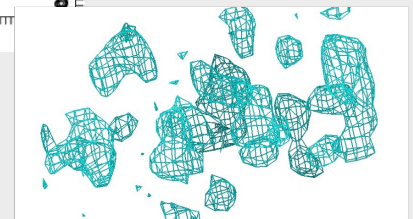
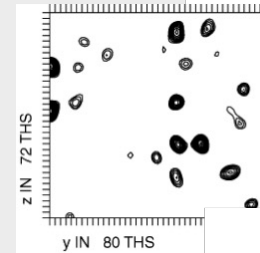
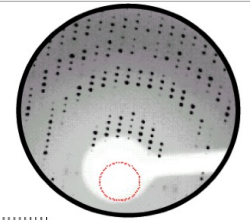
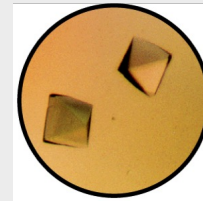
Will I solve my SAD structure?

Planning the experiment

Automating the analysis

Improving the map

Building a model



Will I find the anomalous substructure?

What is important for the SAD experiment?

How many sites?

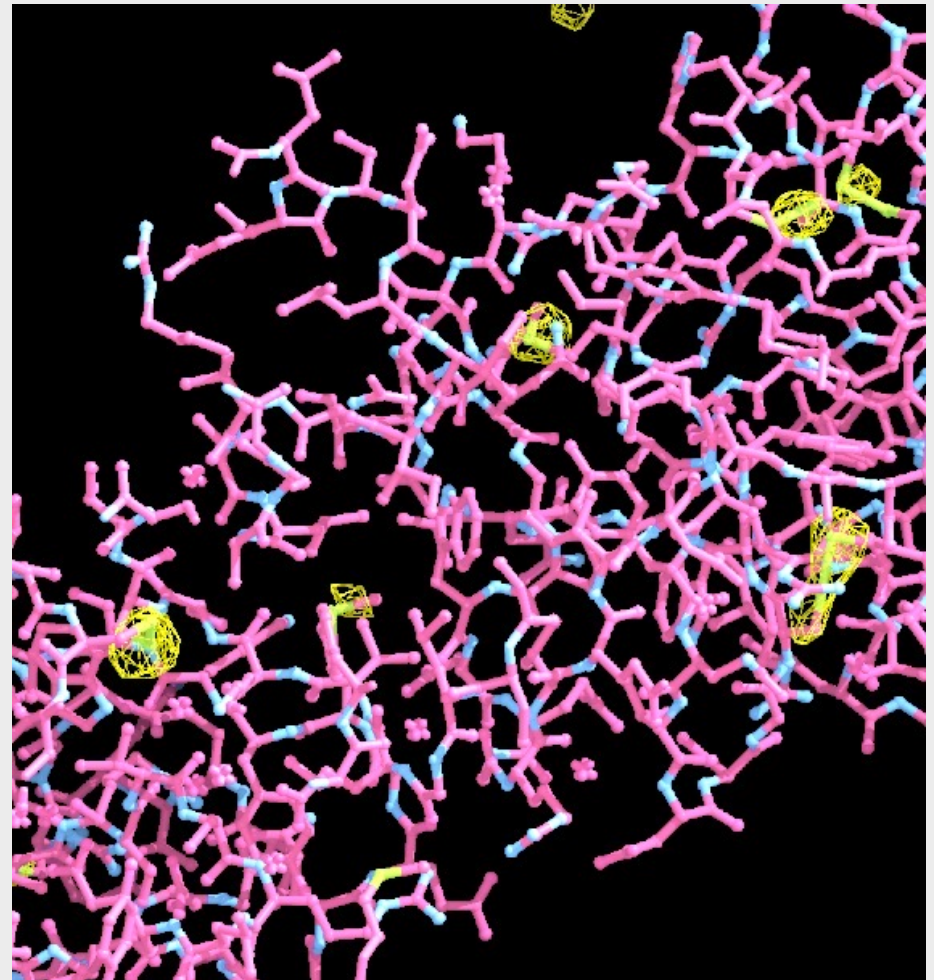
Are sites ordered?

Anomalous atom?

Wavelength?

Accurate data?

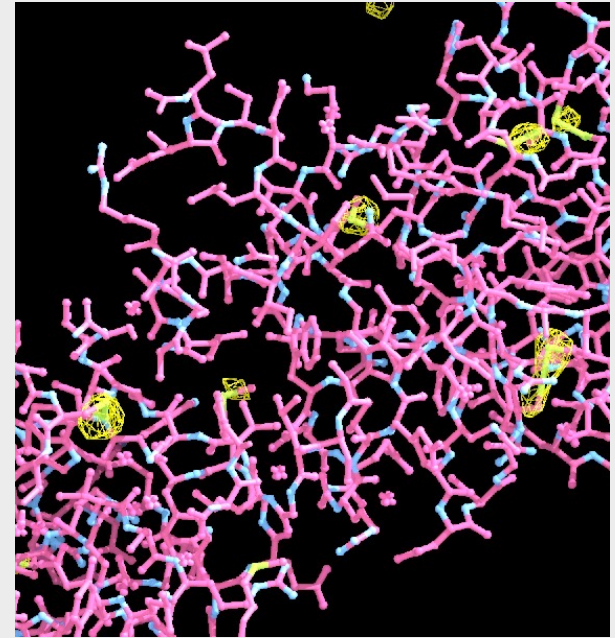
How many reflections?



Key steps for SAD phasing

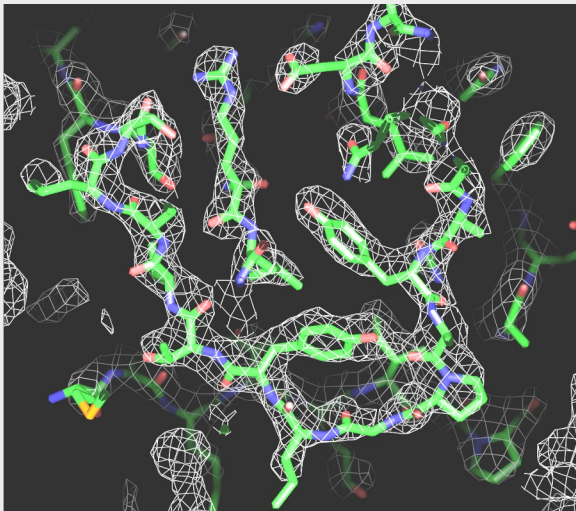
1. Find the substructure

Anomalous
signal S



2. Calculate an interpretable map

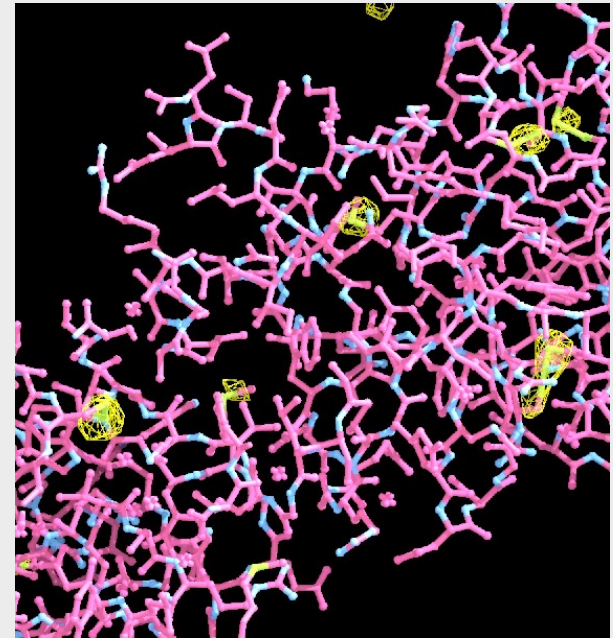
Anomalous
correlation CC_{ano}



Anomalous signal

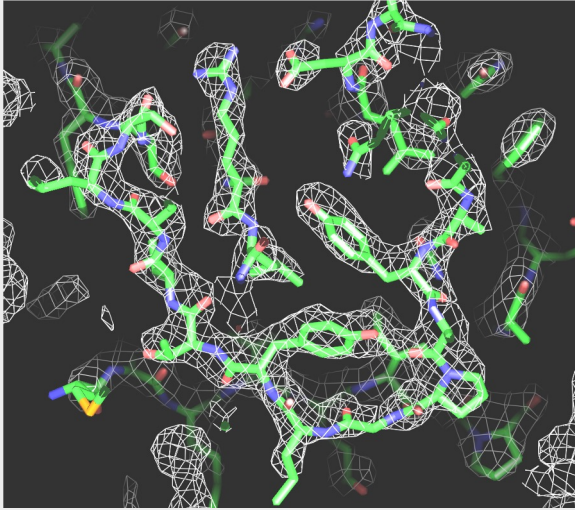
1. Find the substructure

Anomalous
signal S



- Peak height in anomalous difference Fourier
- “Information per site”
- Substructure likely to be found if $S > 10$

Anomalous correlation



2. Calculate an interpretable map

Anomalous
correlation CC_{ano}

- Correlation of anomalous differences with ideal
- Accuracy of anomalous data
- Accuracy of phasing

Anomalous signal: key to finding substructure

Anomalous
signal S

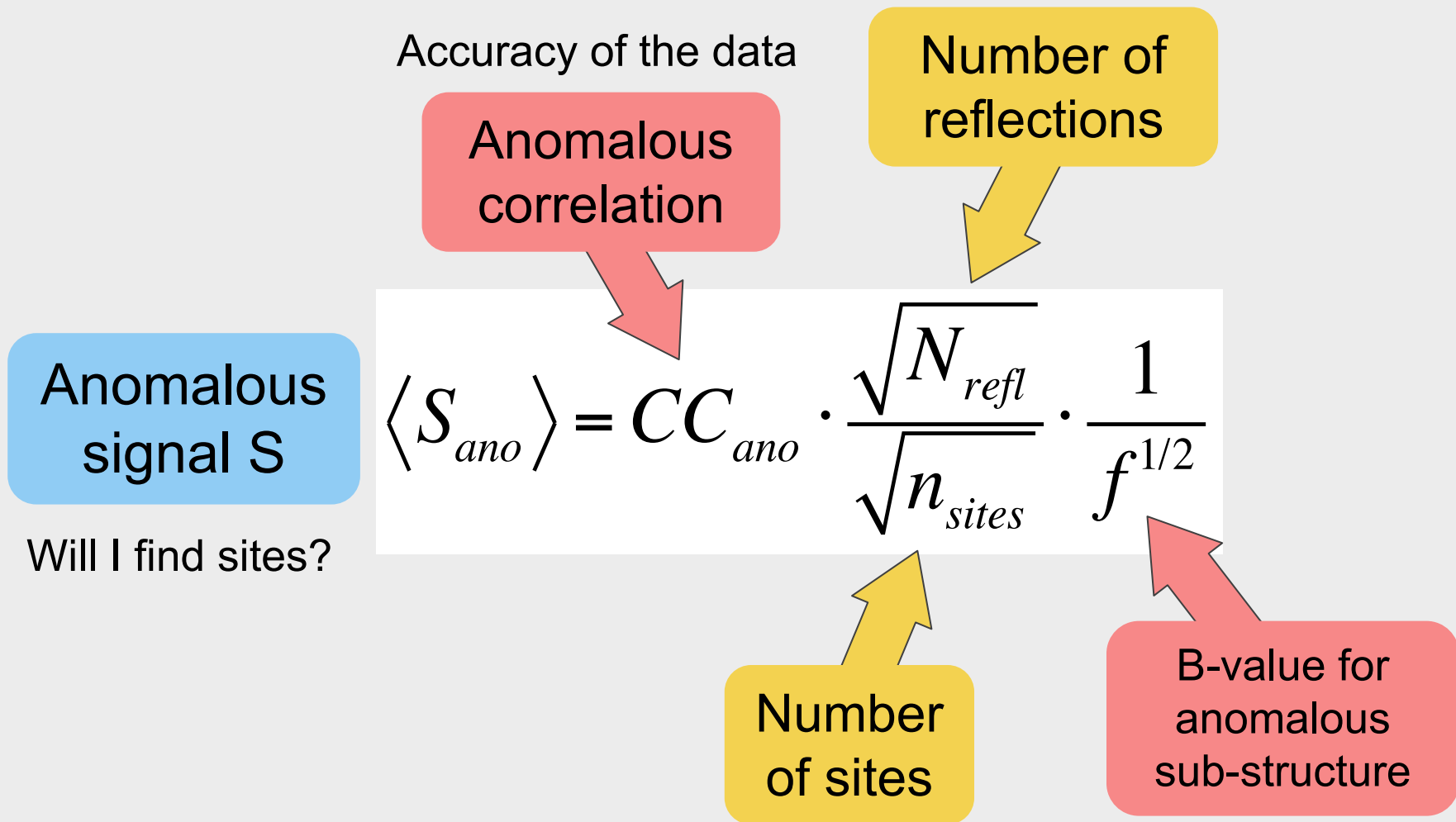
Will I find sites?

← Peak height in anomalous
difference Fourier

We can only calculate this once
we solved the structure.

→ Estimate it

Anomalous signal: key to finding substructure



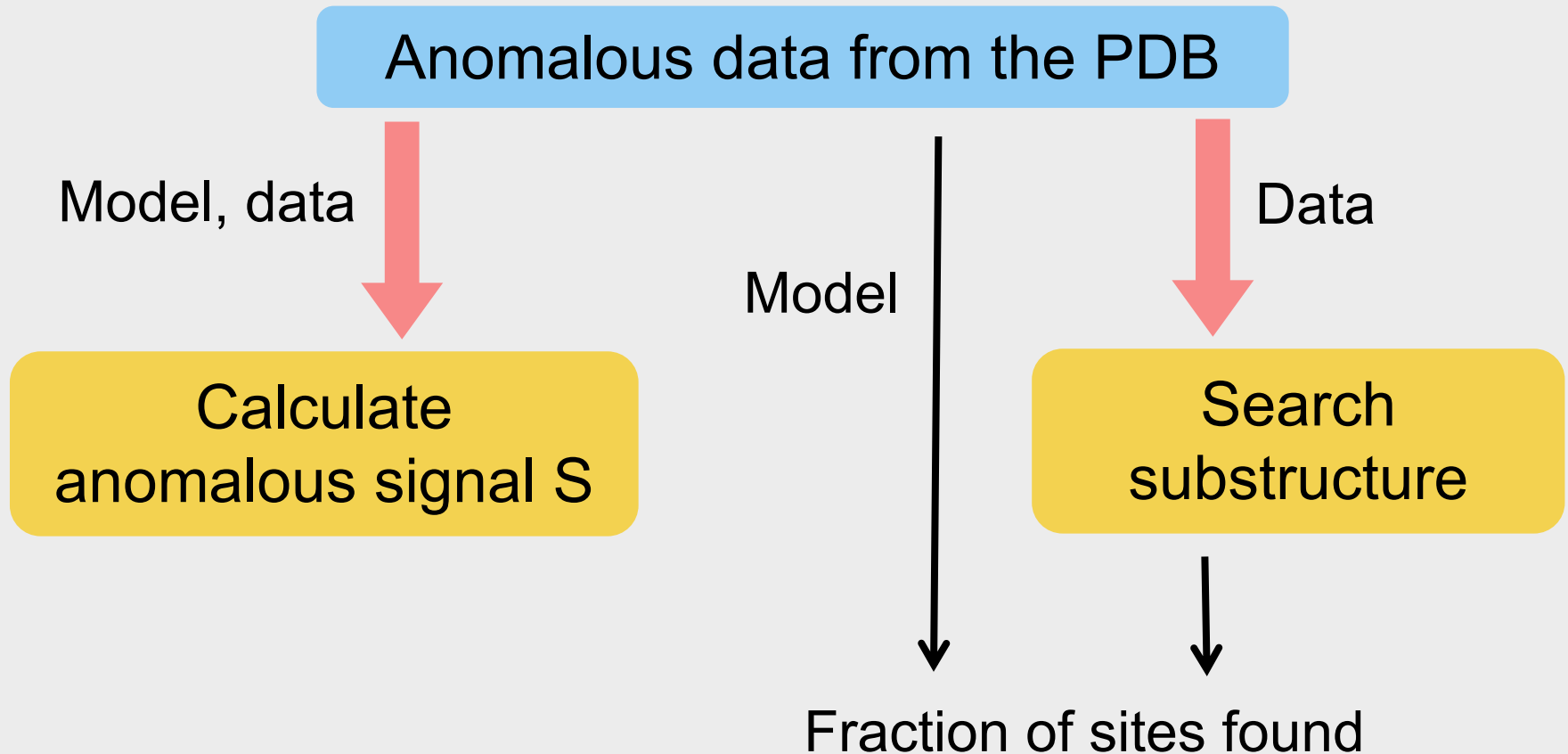
Anomalous signal: key to finding substructure

Relationship between the anomalous signal and the solution of the anomalous substructure

“What value of S (anomalous signal) do I need to solve the substructure?”

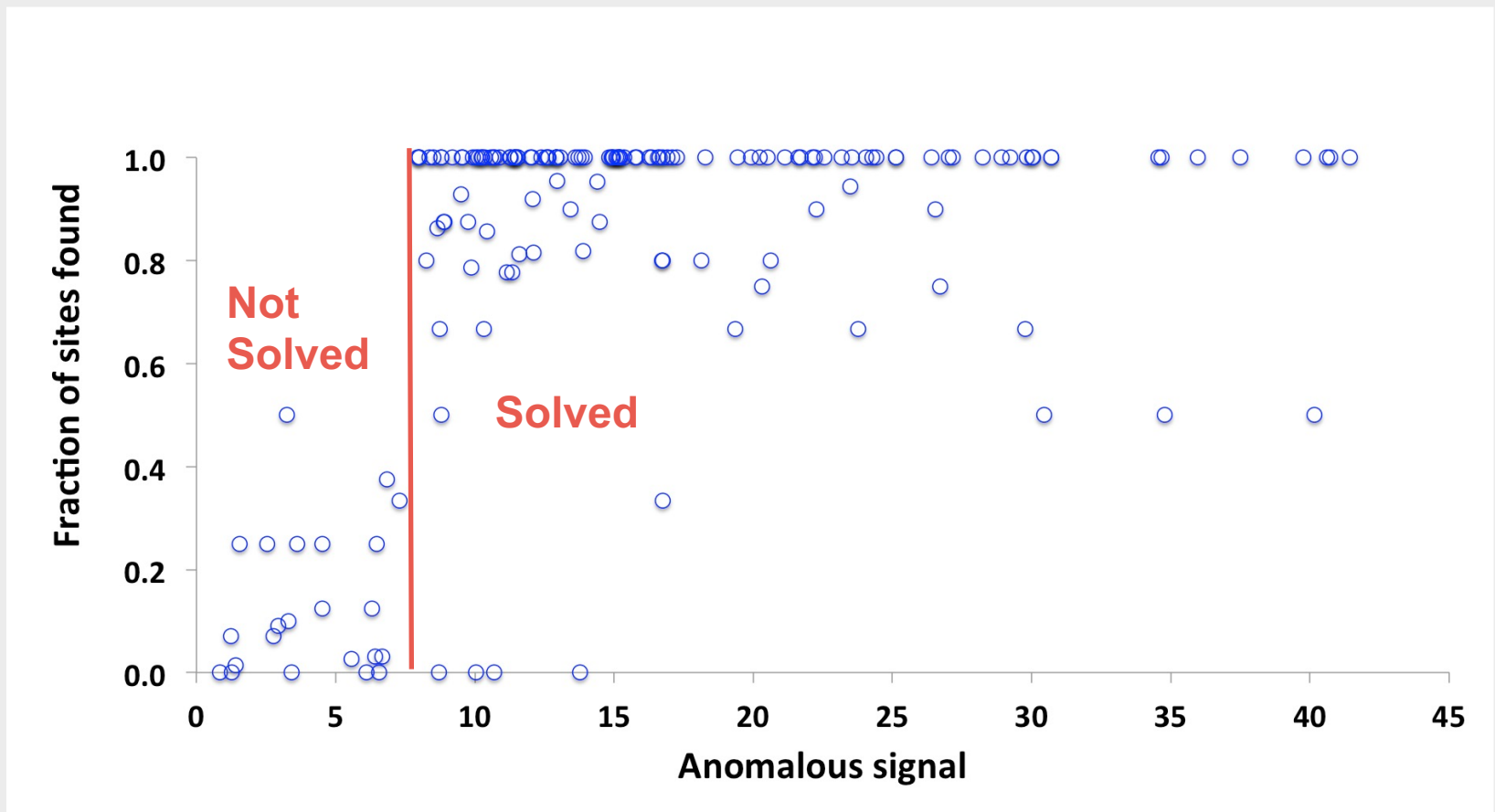
Anomalous signal: key to finding substructure

Relationship between the anomalous signal and the solution of the anomalous substructure



Anomalous signal: key to finding substructure

Relationship between the anomalous signal and the solution of the anomalous substructure



Simulating the anomalous signal

I/σ (accuracy of data)

Anomalous atom
(wavelength, f'')

Number of sites

Resolution

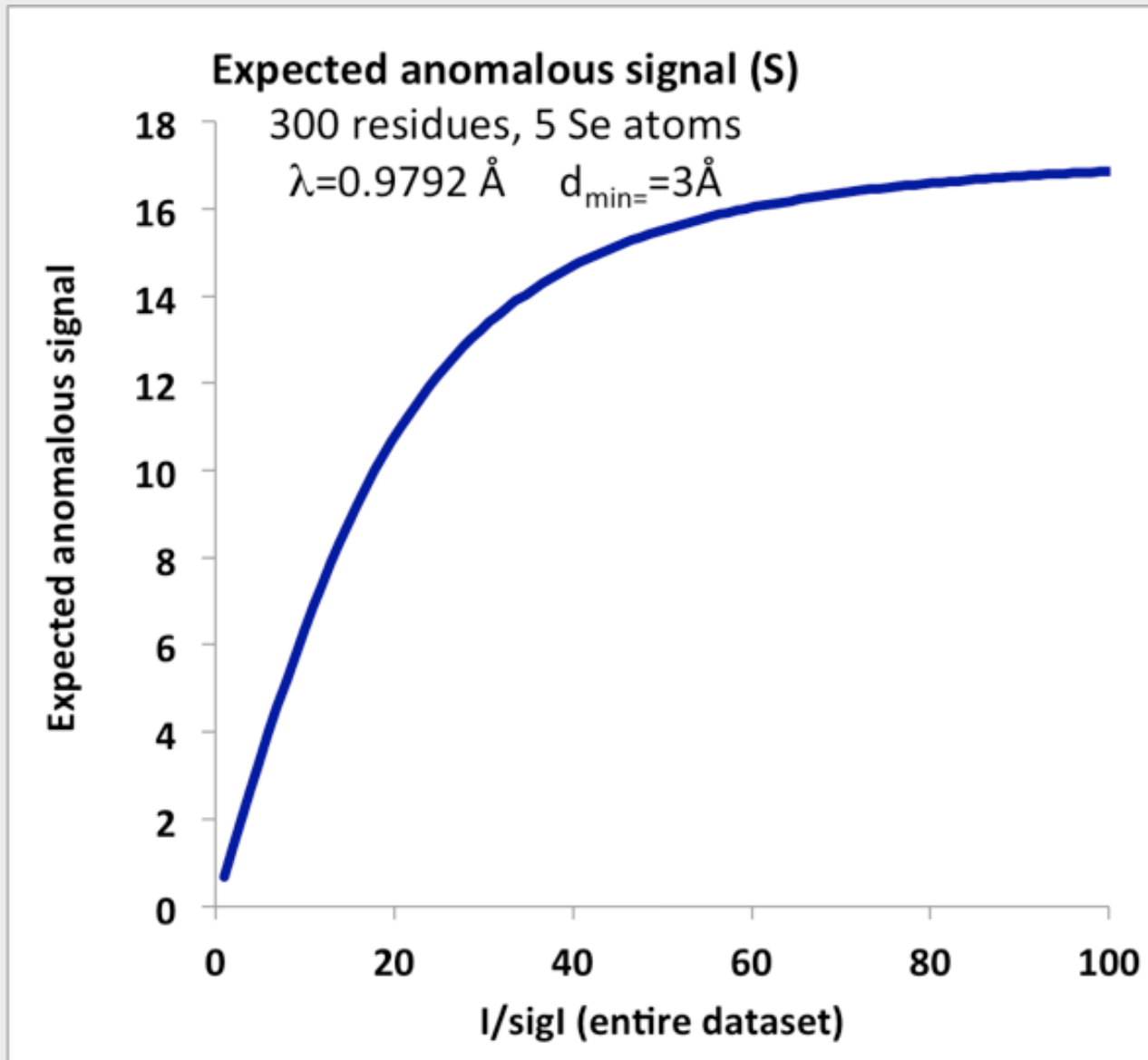
Sequence



Anomalous signal S

Can I solve my structure by SAD phasing? Planning an experiment, scaling data and evaluating the useful anomalous correlation and anomalous signal. Terwilliger TC, Bunkóczi G, Hung L-W, Zwart PH, Smith JL, Akey D, Adams PD Acta Cryst. D72, 359-374 (2016).

Estimating the anomalous signal



Se-SAD

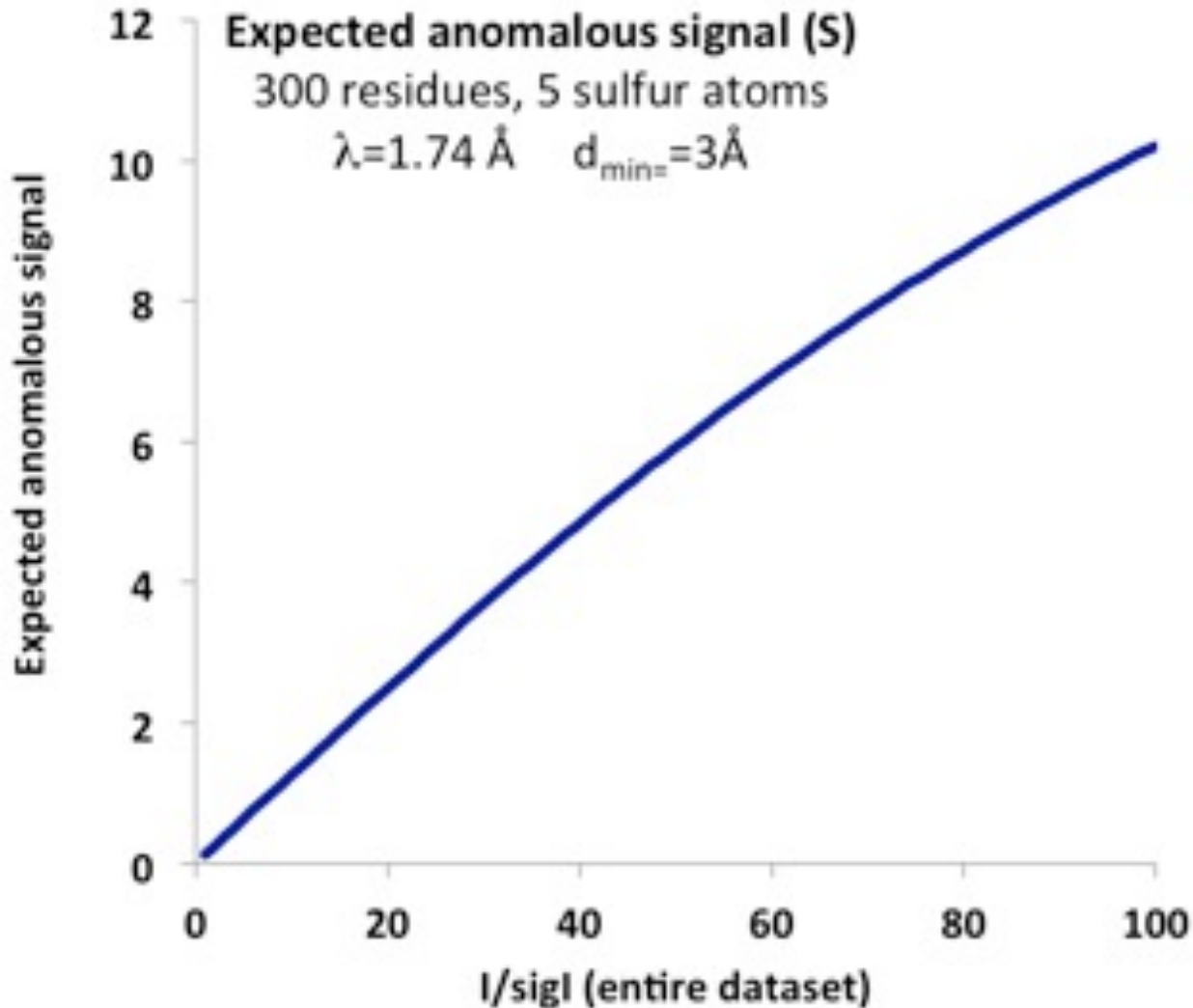
5 sites

3 Å data

300 residues

0.9792 Å

Estimating the anomalous signal



S-SAD

5 sites

3 Å data

300 residues

1.74 Å

Estimating the anomalous signal

What can be changed to increase the chances of success?

- Experimental settings
- Sample?

S-SAD

5 sites

3 Å data

300 residues

1.74 Å

Summary

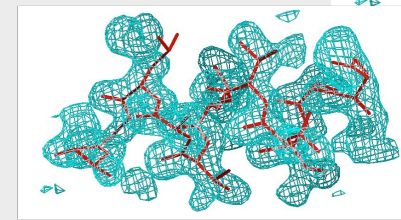
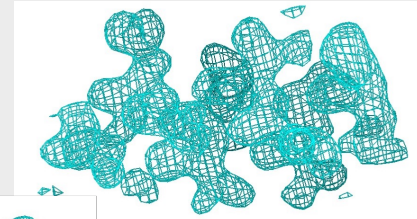
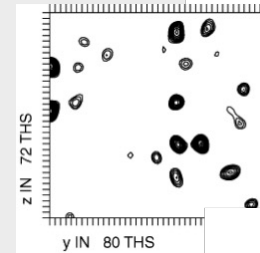
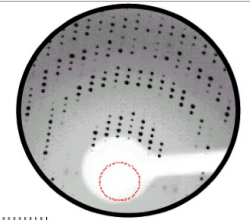
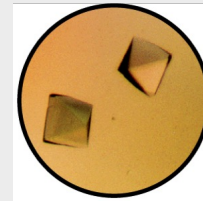
- We can estimate the anomalous signal S from the data
- If $S > 10 \rightarrow$ substructure is likely to be found
- We can simulate the anomalous signal (before doing the experiment)

Will I solve my SAD structure?

Planning the experiment

Automating the analysis

Improving the map



Why automate structure determination?

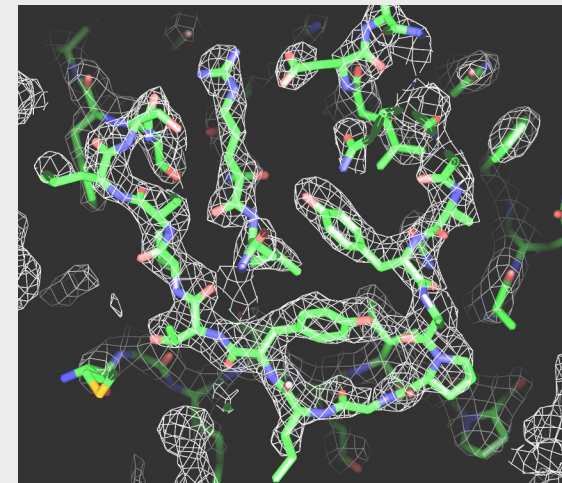
Makes straightforward cases easier

... and difficult cases feasible for experts

Speeds up the process

Reduces errors

Allows you to try more possibilities



Decision-making in automation

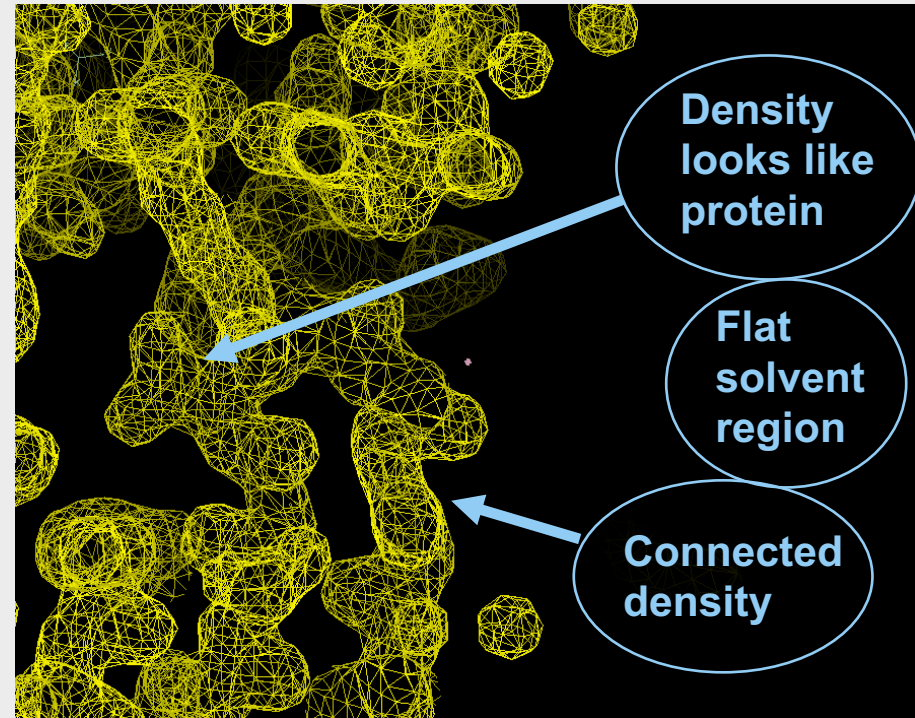
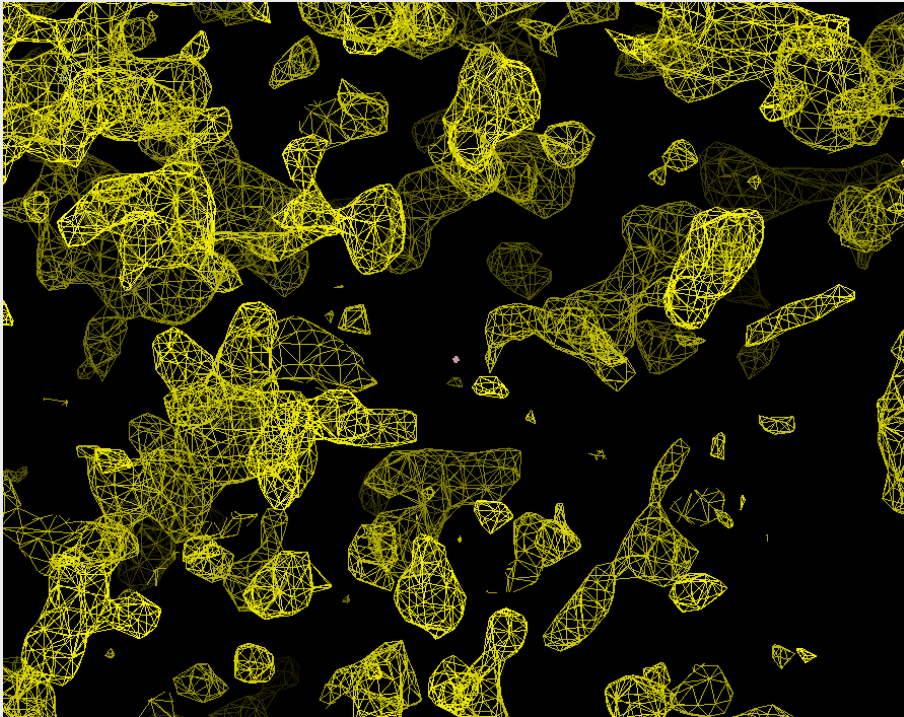
What does a good electron density map look like?



Using expected features of maps to make decisions

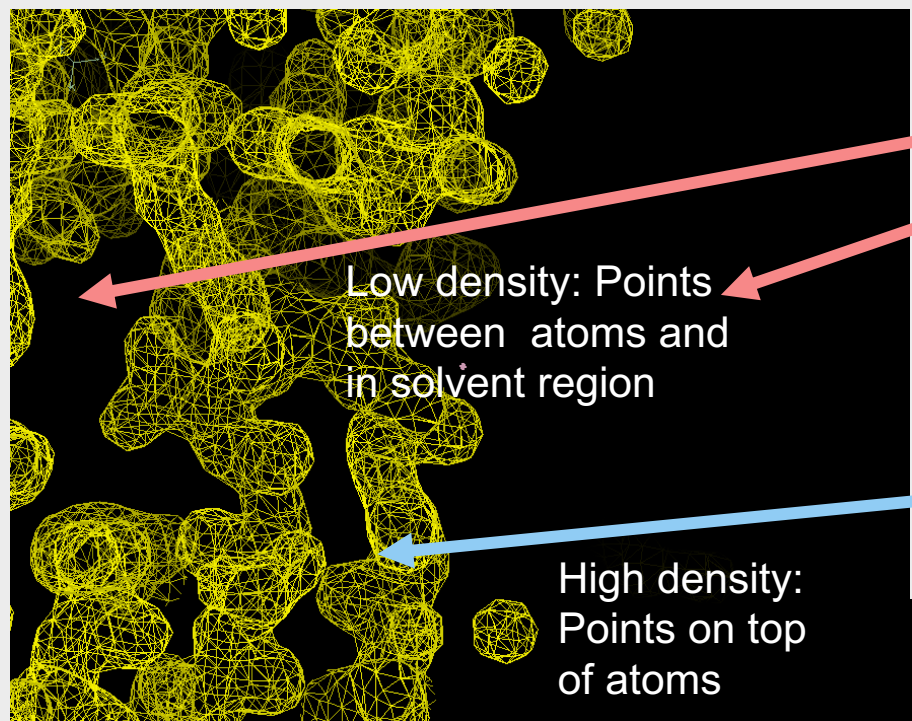
Decision-making in automation

Which map is better?

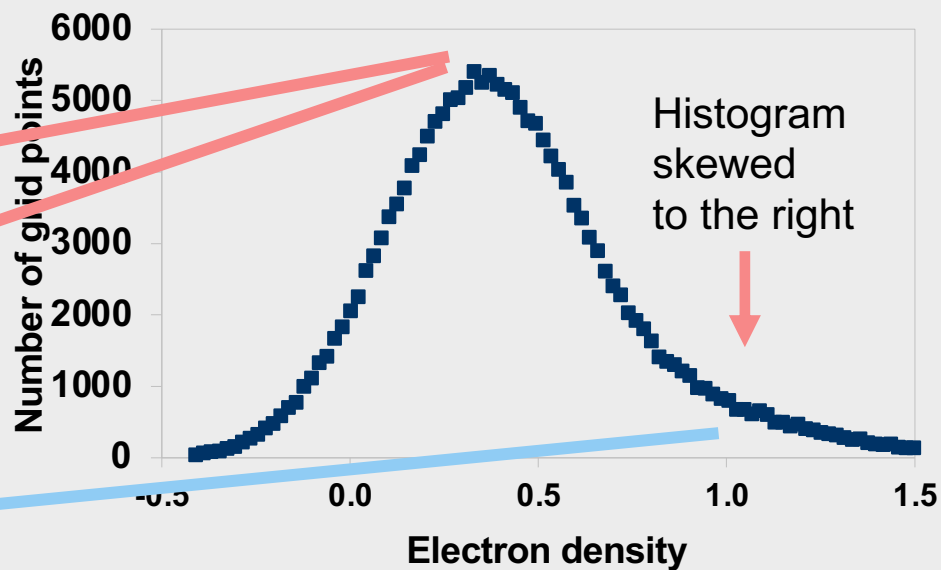


Histograms of density have positive skew

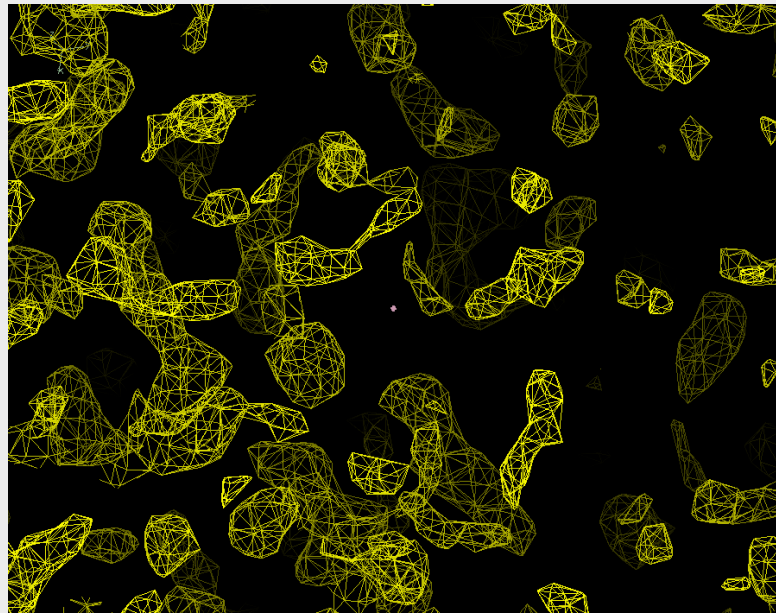
Skew = measure of the asymmetry



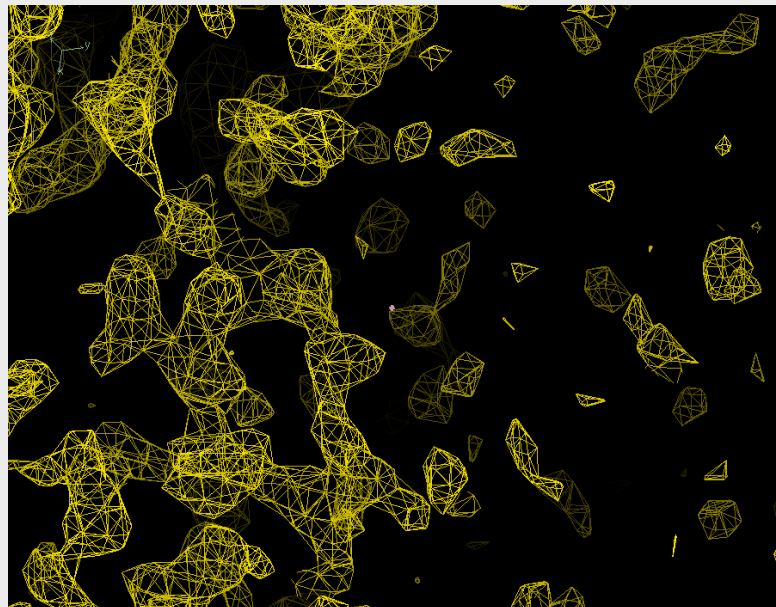
Typical histogram of electron density



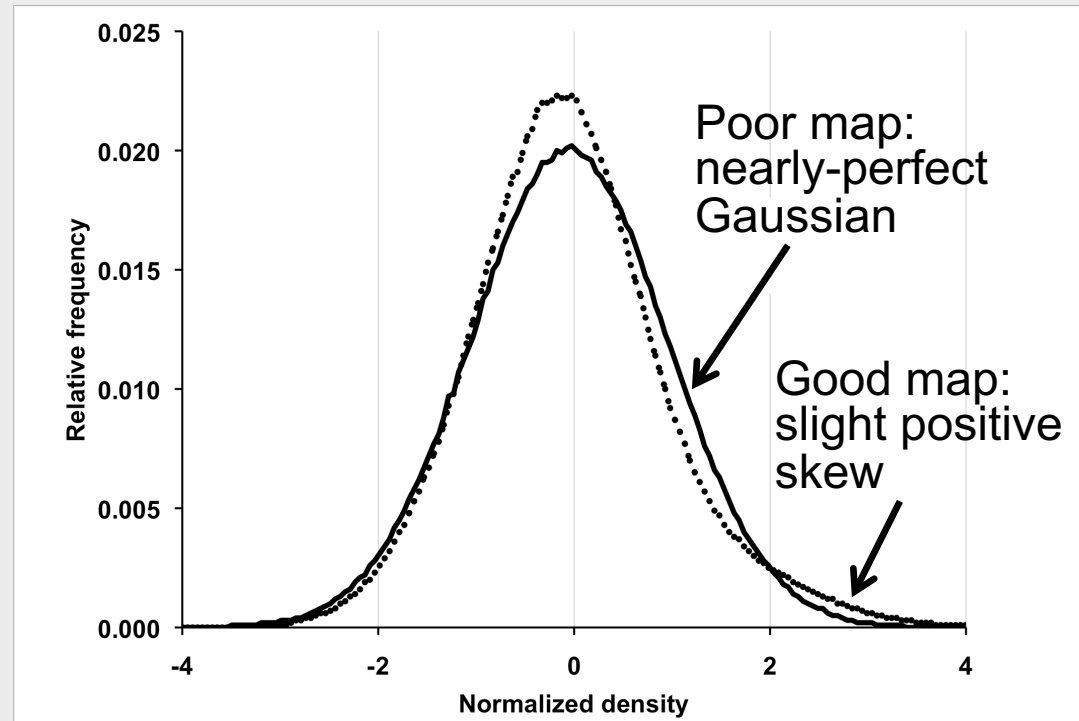
Histograms of density have positive skew



Poor map
(inverse hand)



Good map

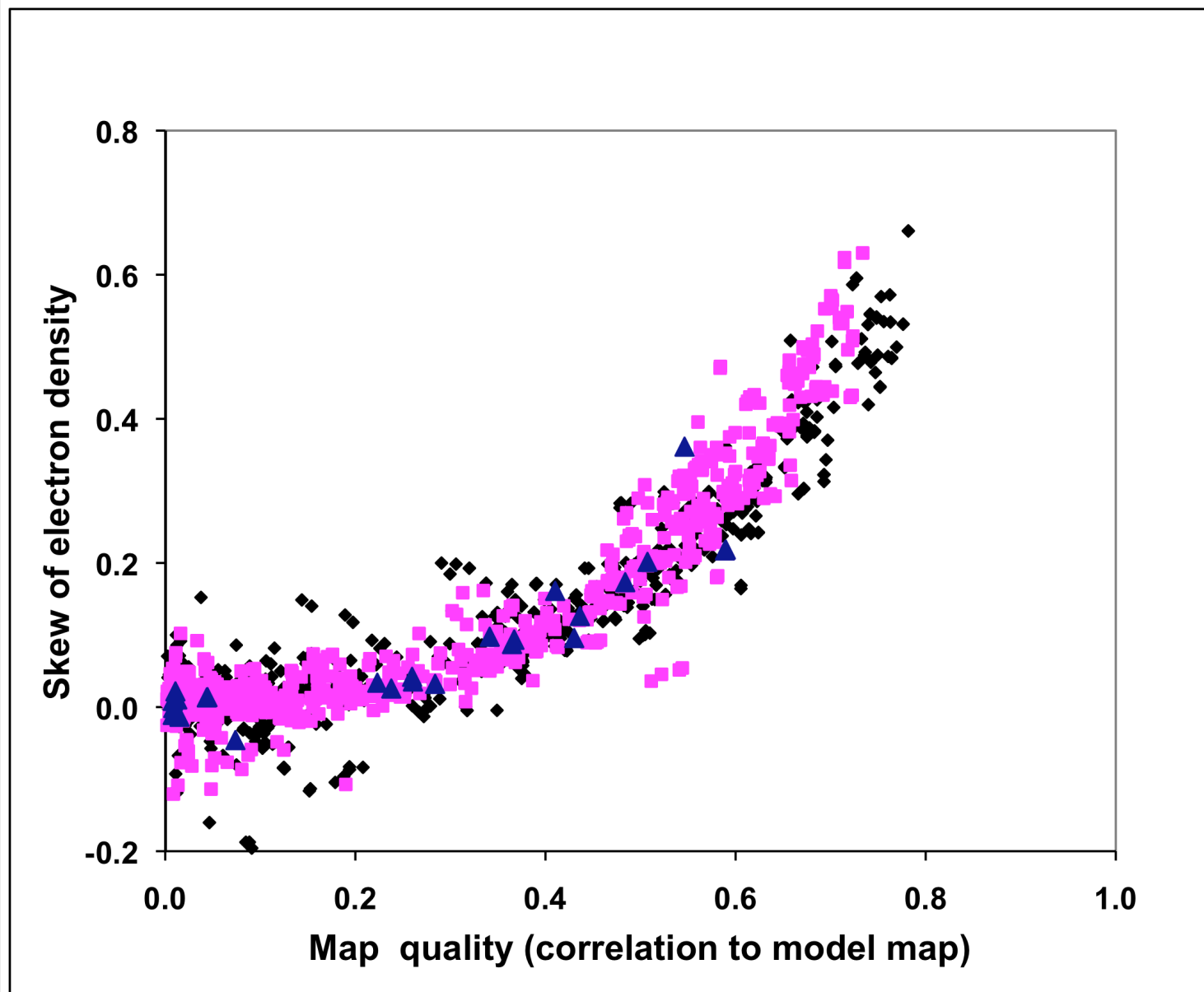


Histograms of density have positive skew

How well does the skew reflect map quality?

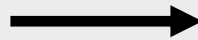
- 247 MAD, SAD, MIR datasets with final model available
- Run *phenix.autosol* on each dataset → maps
- Score the maps based on skew
- Compare the scores with the actual quality of the maps (correlation to model map)

Positive skew in good maps

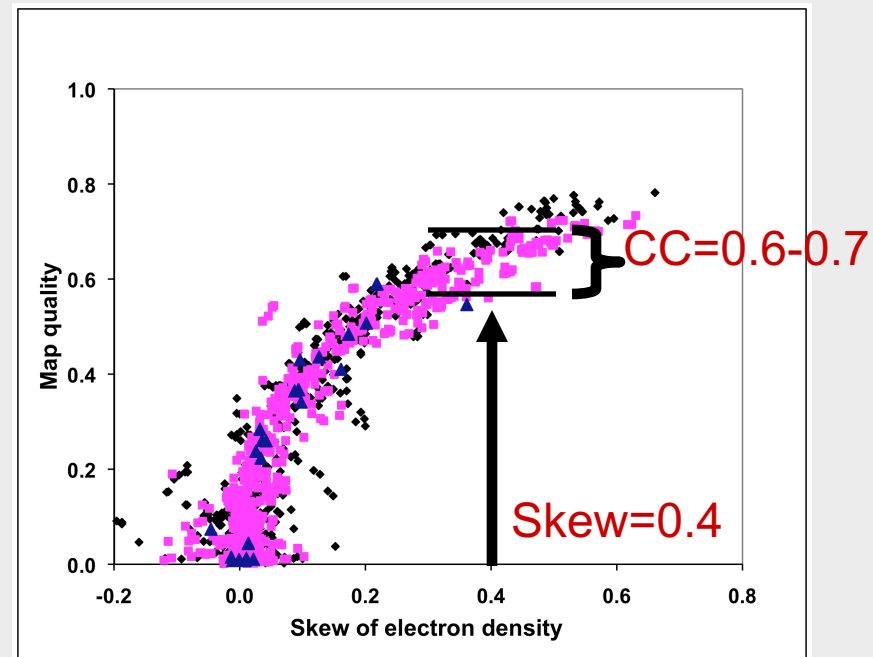
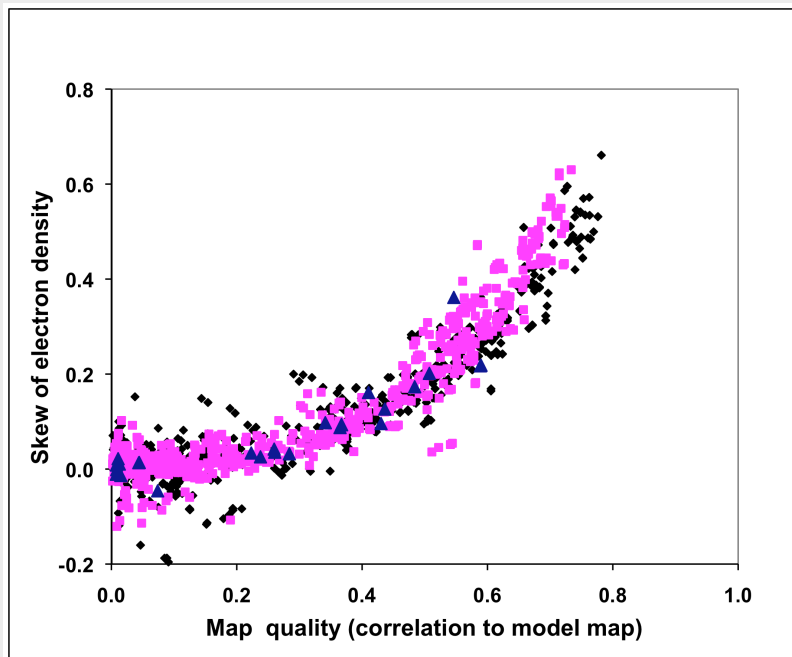


Estimate map quality from skew

Skew depends on
map quality



Estimate map quality
from skew



Summary

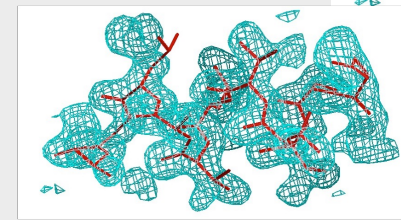
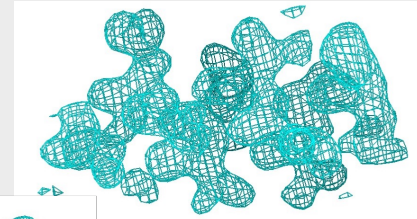
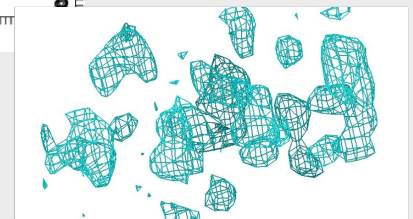
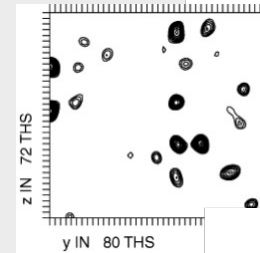
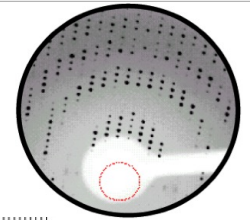
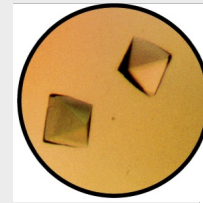
- The skew reflects map quality
- We can estimate map quality from the skew
- Use the skew for decision making (automation)

Will I solve my SAD structure?

Planning the experiment

Automating the analysis

Improving the map



Map improvement by density modification

What does a good electron density map look like?



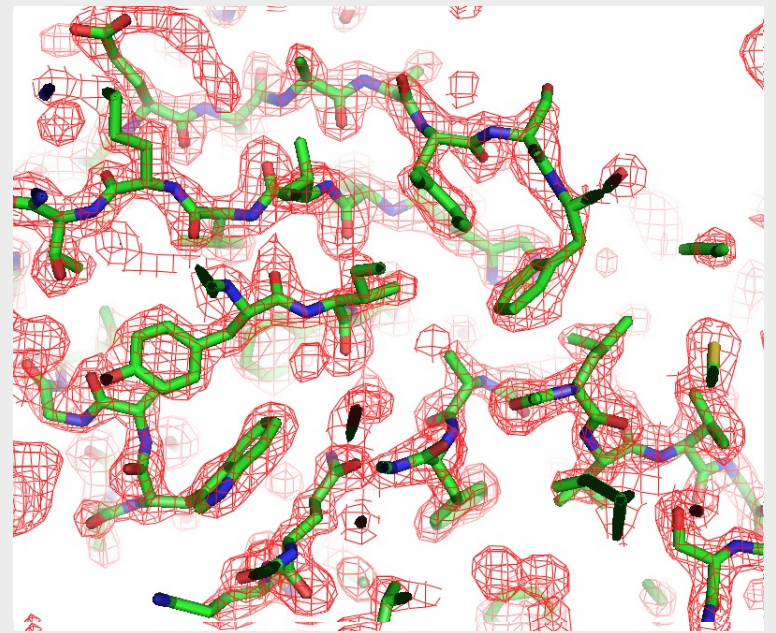
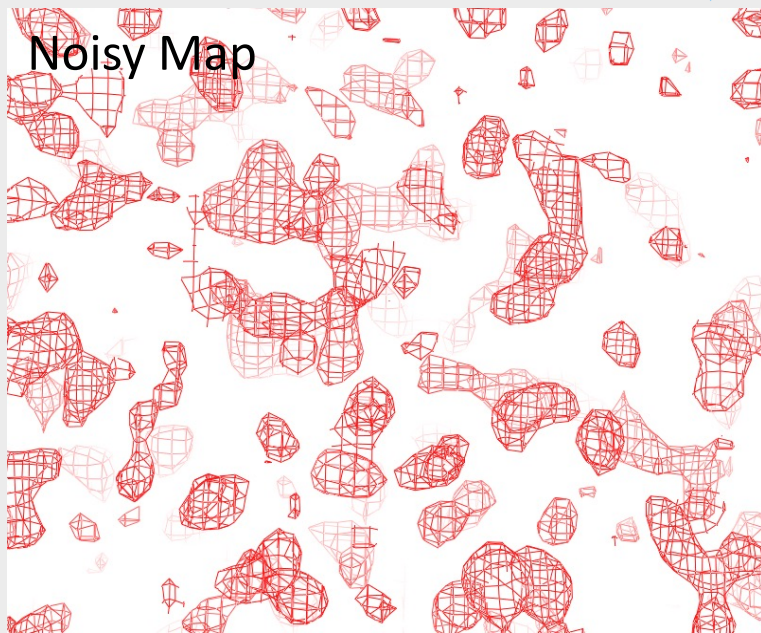
Using expected features of maps to improve maps

Density modification = “phase improvement”

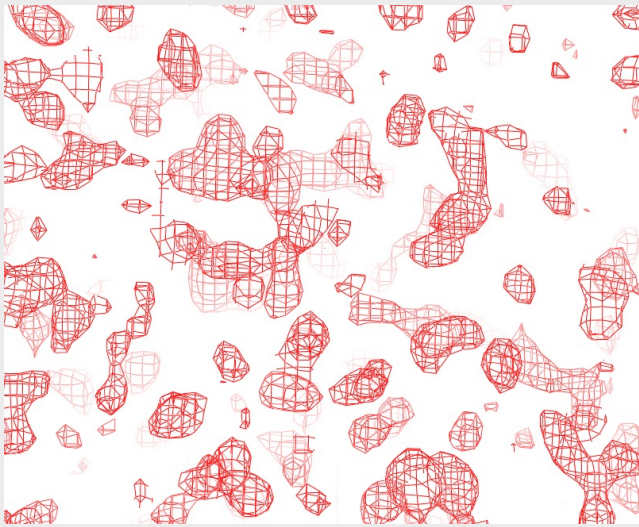
Experimental Data

Initial phases

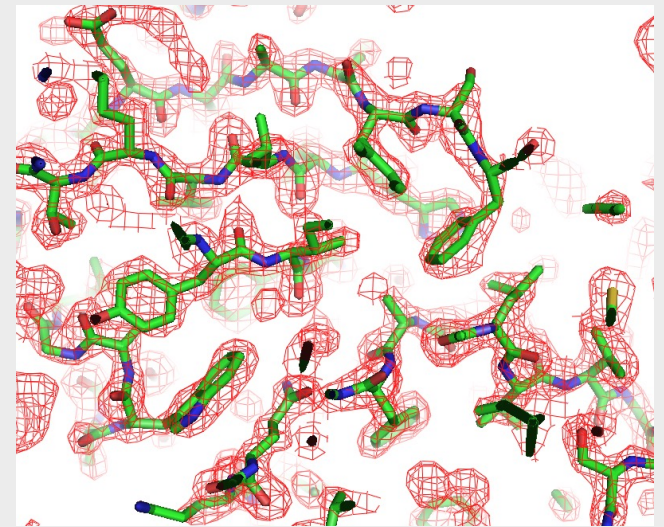
Improved phases



Basis of density modification



Noisy map



Clear map

Improve the noisy map to create the clear map using two key facts:

1. We know a good map when we see it

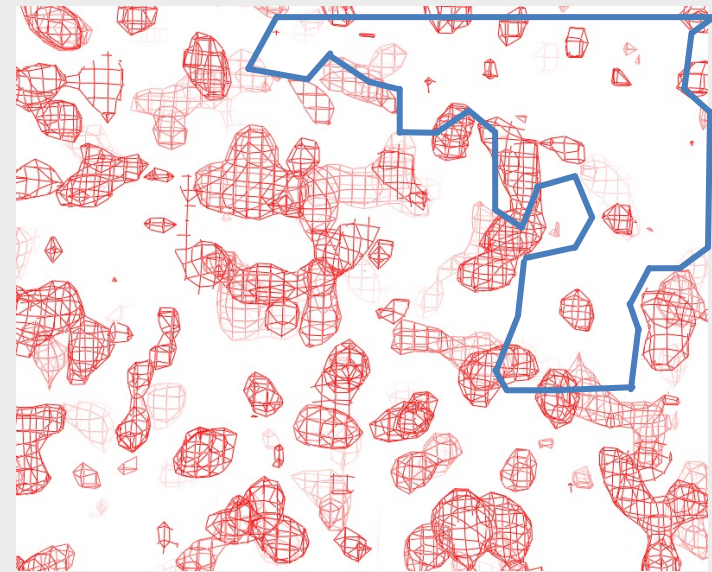
2. Improvement anywhere means improvement everywhere

Density modification: strategy

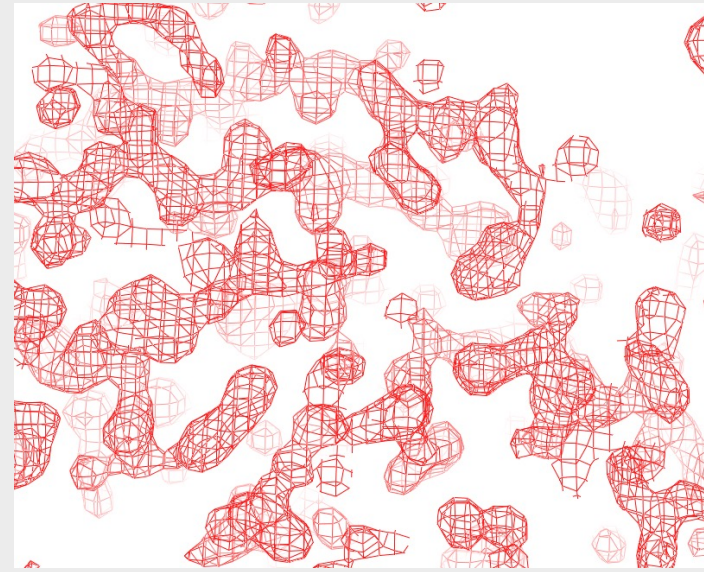
Identify local
expected density

Find phases
consistent with
experiment that
lead to **expected**
density

Density
everywhere is
improved



Noisy map

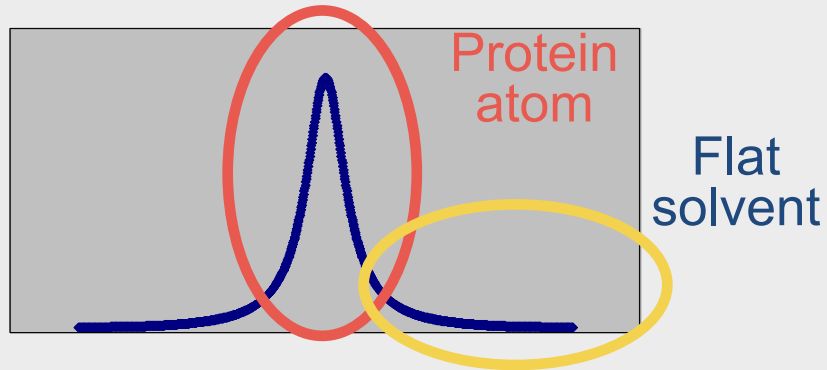


Clear map

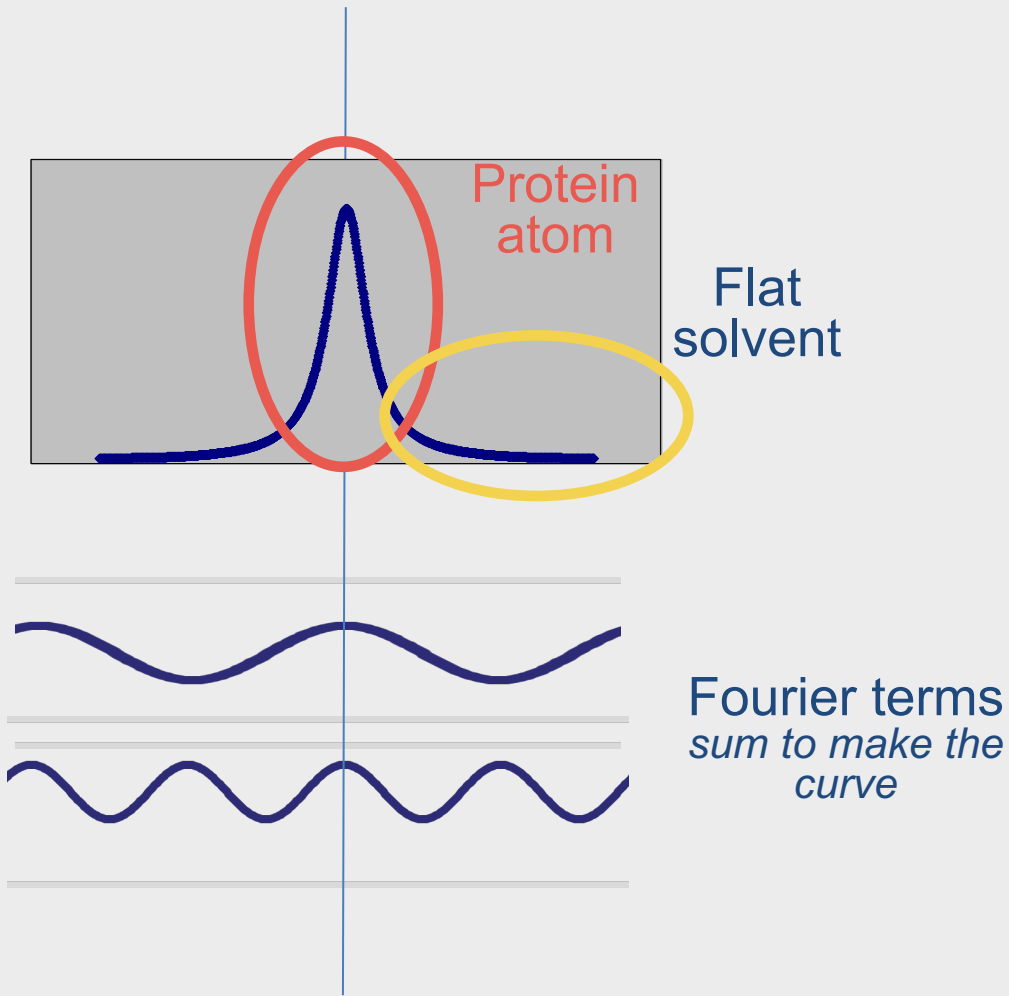
One atom and a flat solvent region

1-dimensional example to illustrate the details of statistical density modification

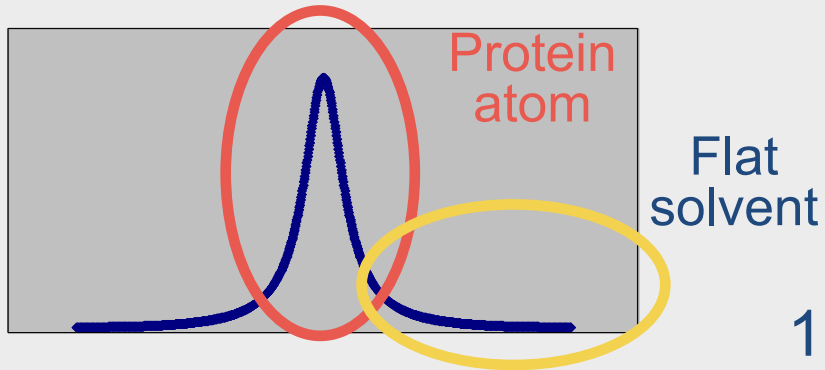
crystal with one protein atom



A Fourier sum of sines and cosines



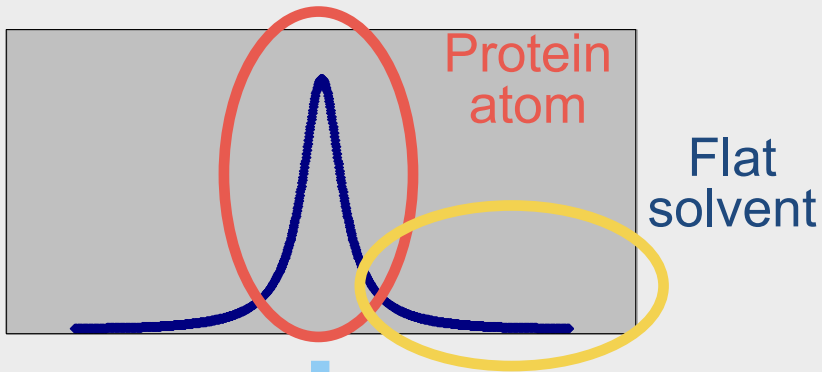
A Fourier sum of sines and cosines



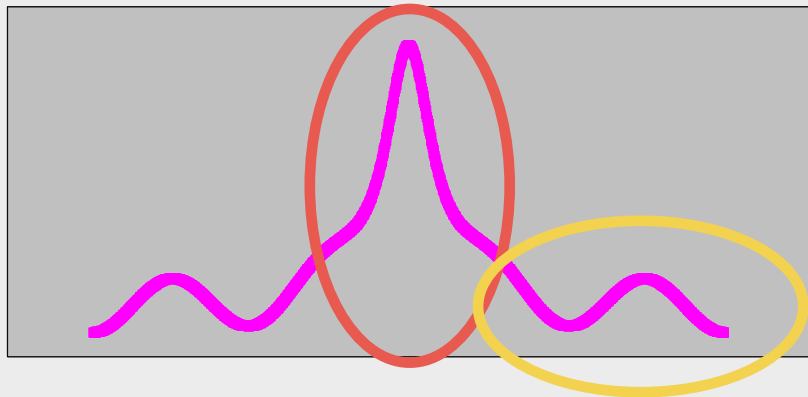
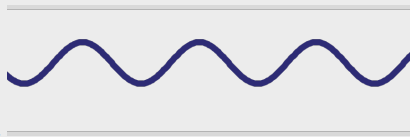
Find out the phase of one Fourier term using:

- 1) Phases of all other Fourier terms
- 2) Flat solvent

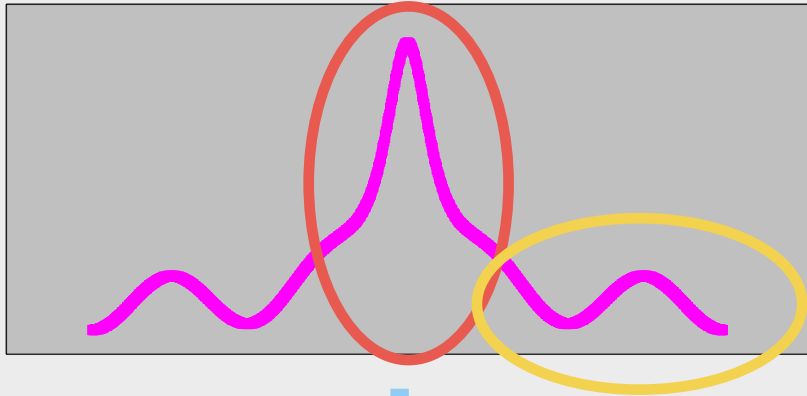
A Fourier sum of sines and cosines



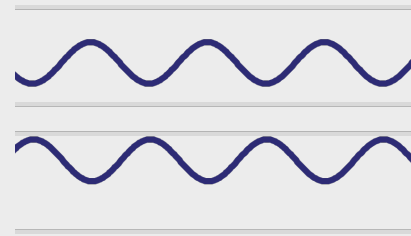
Take out one
Fourier term:



Using flat solvent to identify phase of one term



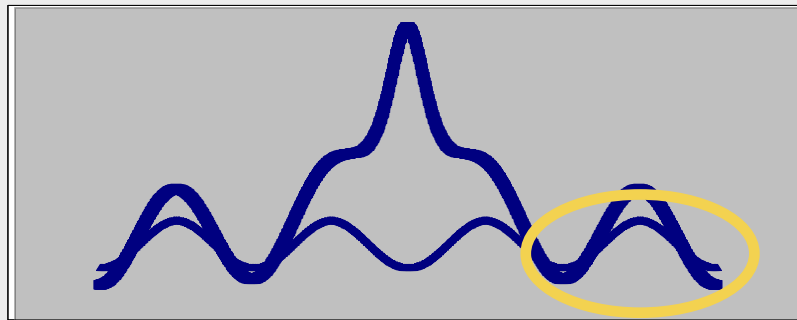
One Fourier term:



Correct phase

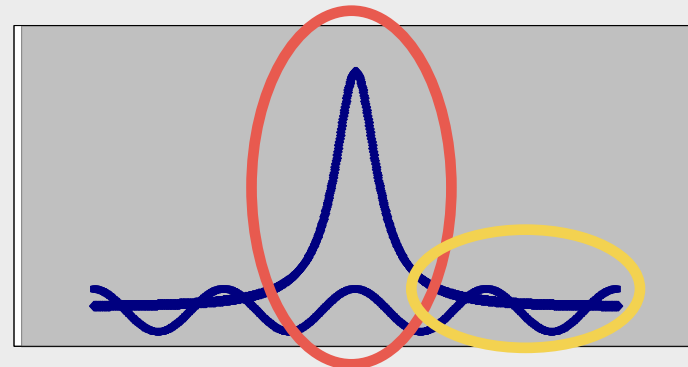
Phase 180° off

Adding the incorrect phase

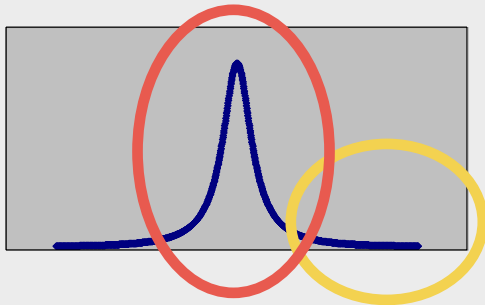


Solvent gets worse

Adding the correct phase

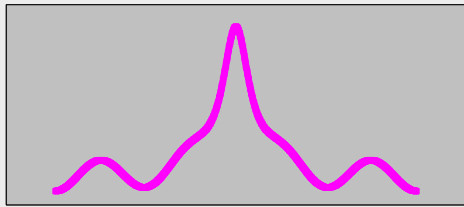


Density modification of real maps

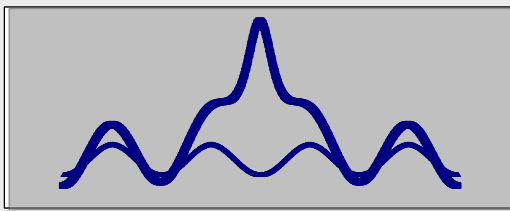


Expectation about the flatness of the solvent.

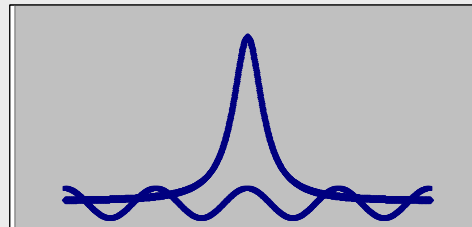
- Identify the phase of one Fourier term
- Improve the map in the protein region



= Transfer information from one part of the map to another.

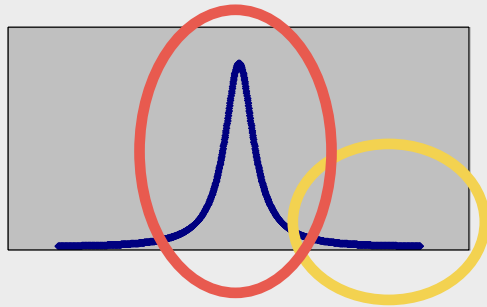


Incorrect
phase



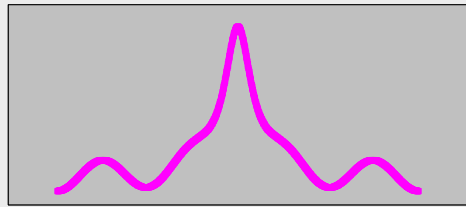
Correct
phase

Density modification of real maps



Real world:

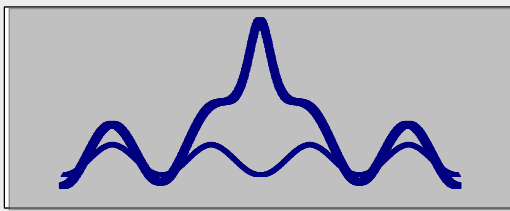
Correct phase $\rightarrow \rho_{map}(\varphi)$



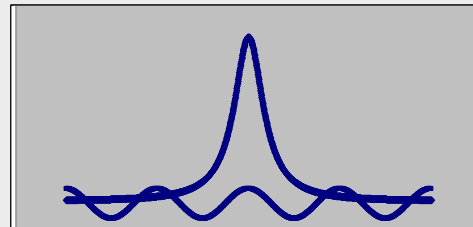
Experimental phase information = $\rho_{exp}(\varphi)$

Density modification phase probability:

$$\rho(\varphi) = \rho_{exp}(\varphi) \rho_{map}(\varphi)$$

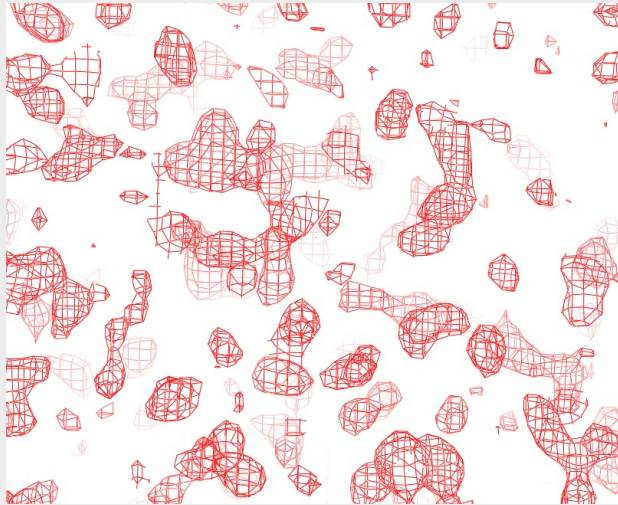


Incorrect
phase



Correct
phase

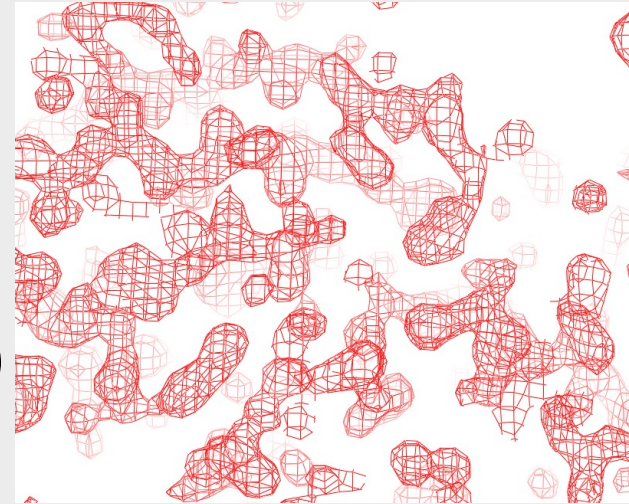
Summary



Improved
phases



$$\rho(\varphi) = \rho_{exp}(\varphi) \rho_{map}(\varphi)$$



We know a good
map when we see it

Improvement
anywhere means
improvement
everywhere

Density modification
transfers information
from one part of the
map to another

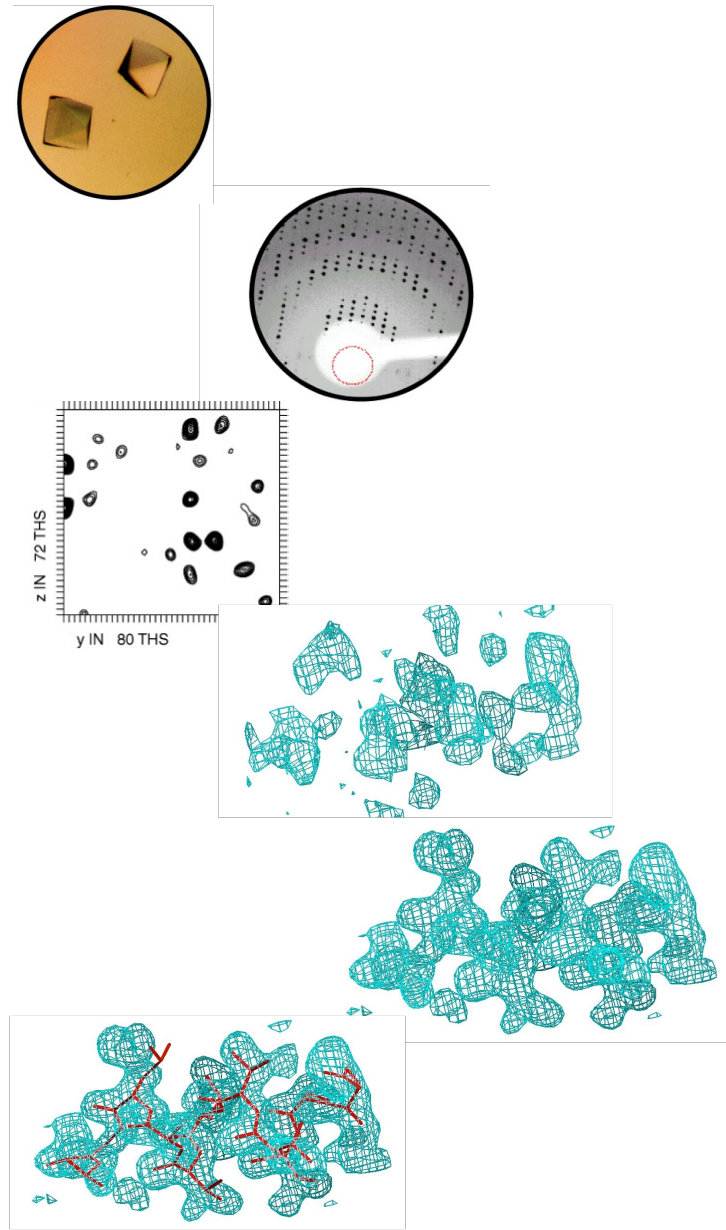
Automated model-building

Planning the experiment

Automating the analysis

Improving the map

Building a model

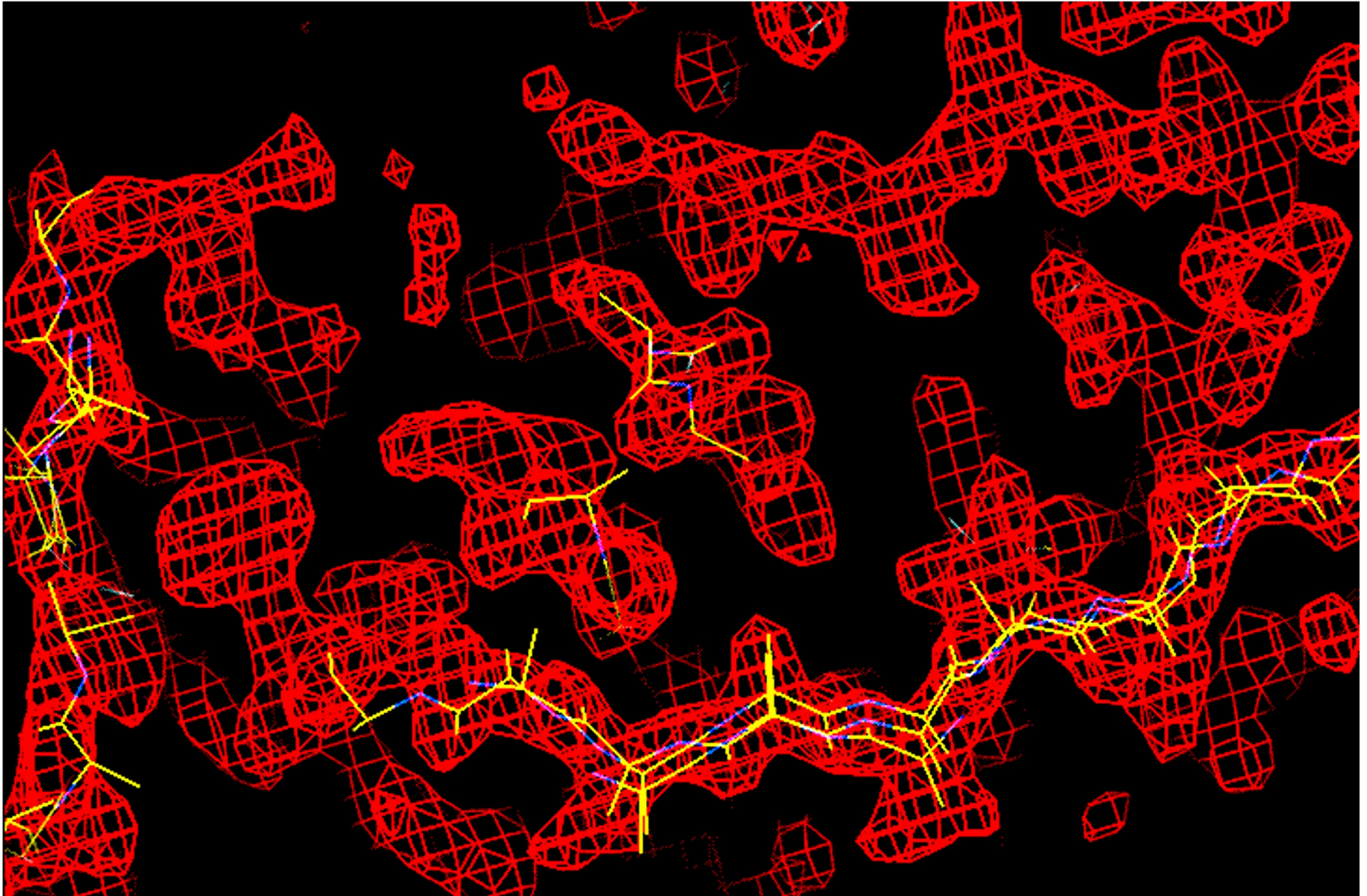


Automated model-building

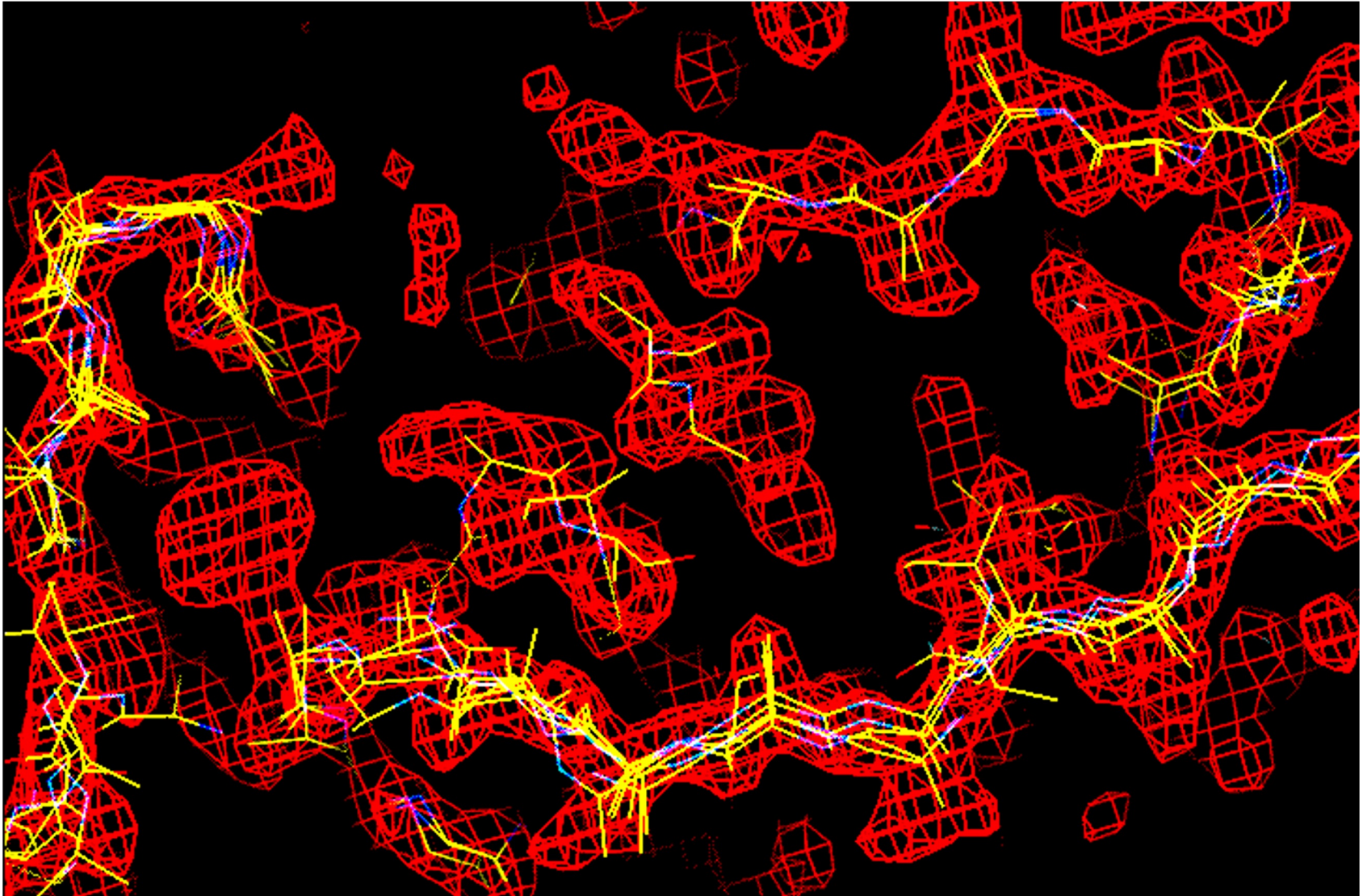
Multi-step procedure:

- Locate helices and strands
- Extend helices and strands iteratively with tripeptides from libraries
- Assemble fragments into a poly-ala chain
- Build side chains and align them to the protein sequence

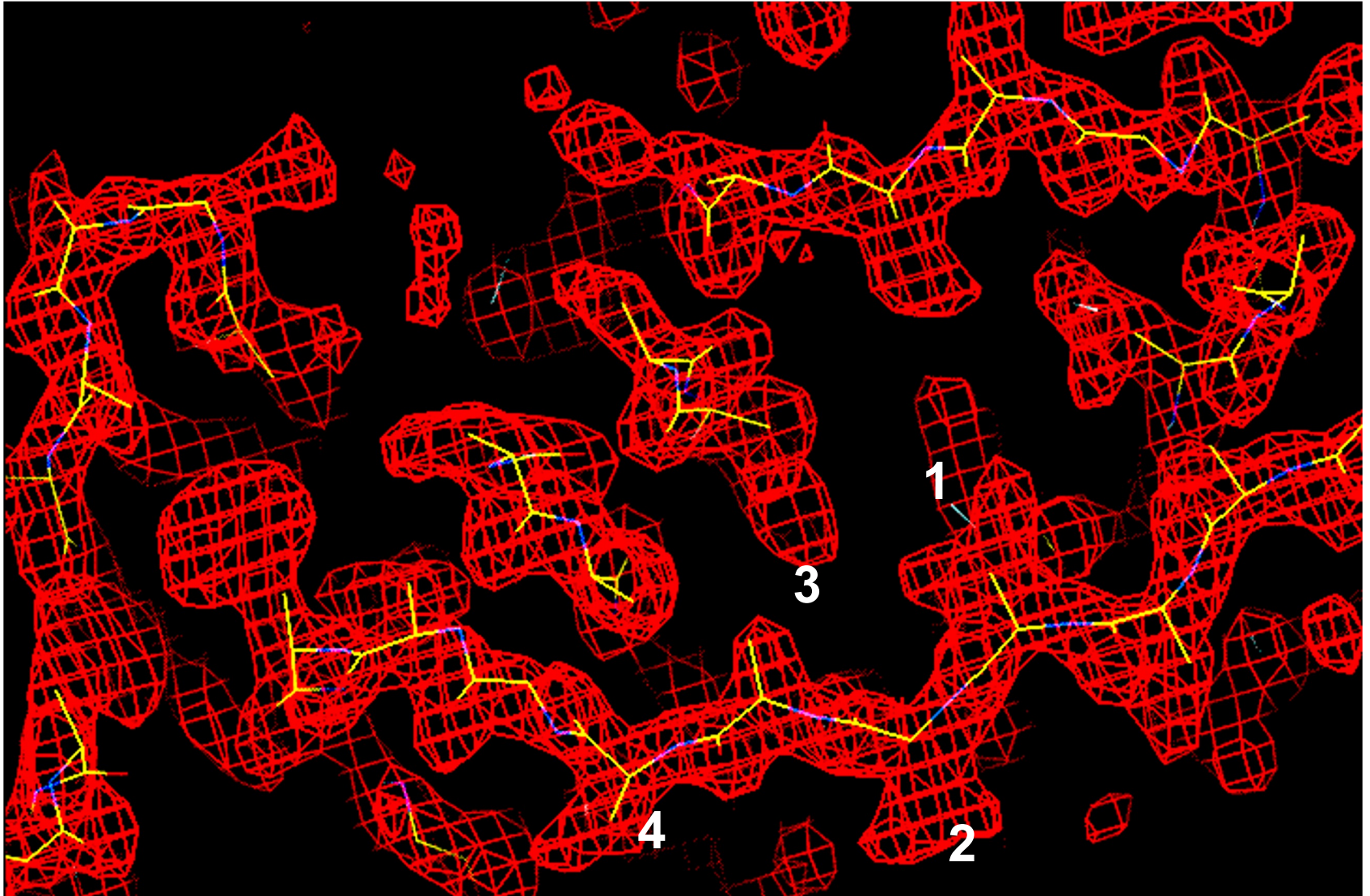
Finding regular protein structure



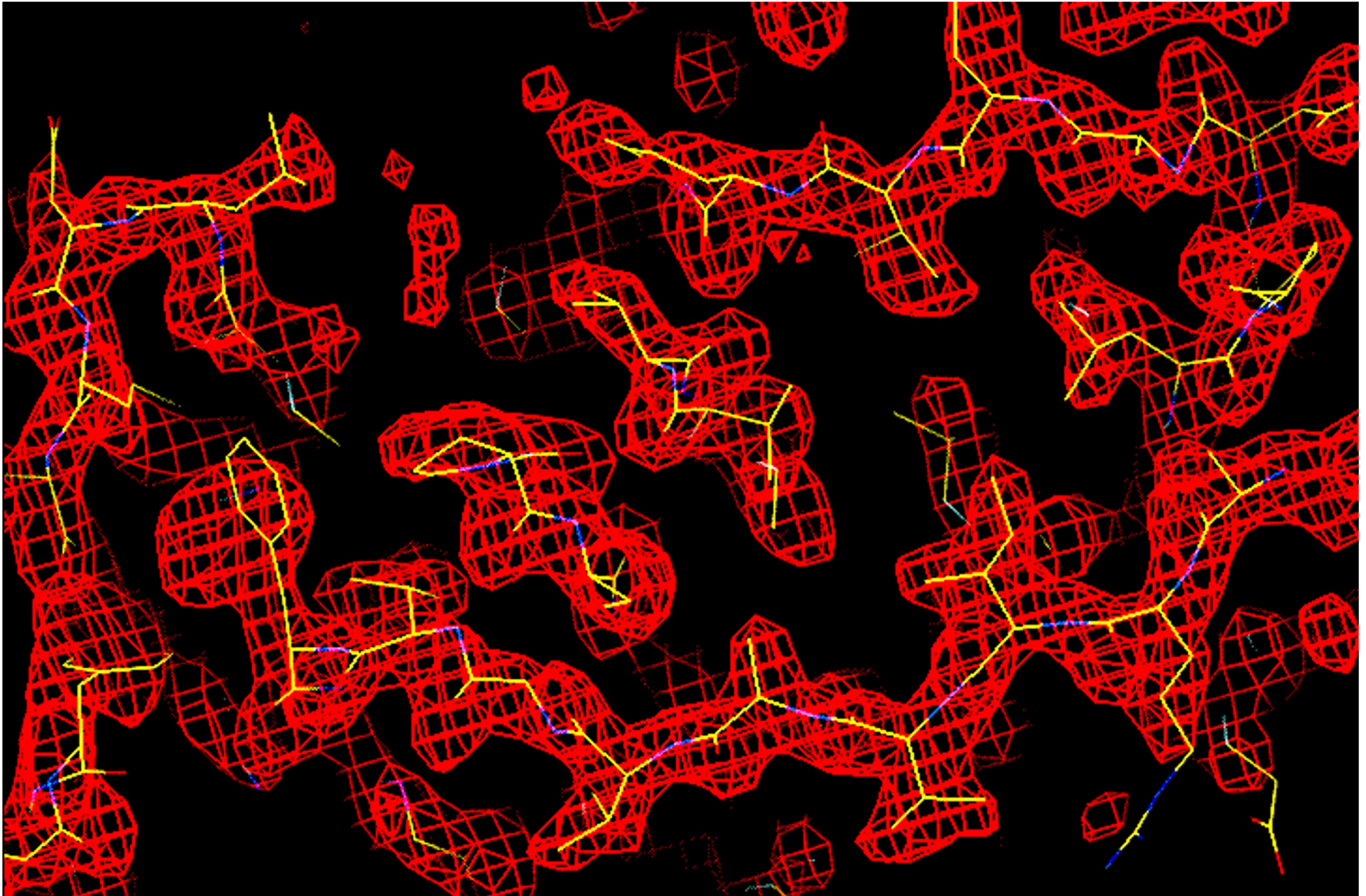
Extending with short fragments from PDB



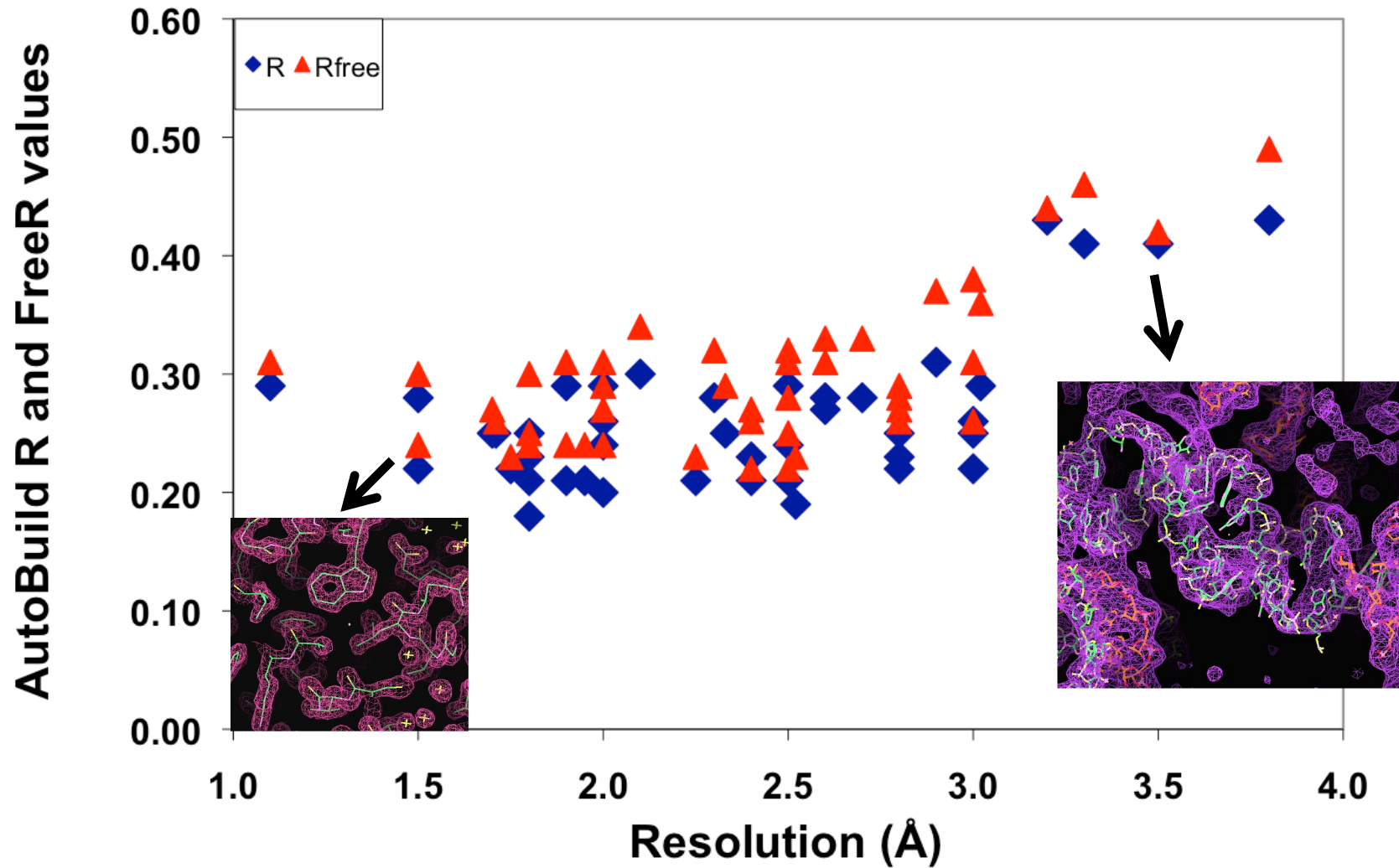
Assembling best model



Inserting side chains based on sequence



Automated structure determination



Summary

- You can simulate your SAD experiment → you can plan your experiment.
- Use prior expectations about density maps to automate the analysis.
- When you improve the map *anywhere*, the map will get better *everywhere*.

References

phenix-online.org

Phenix documentation

Tutorials with sample data



Video tutorials

www.youtube.com/c/phenixtutorials

Terwilliger, T. C. (2000). *Acta Cryst. D.* **56**, 965–972.

Terwilliger, T. C. (2002). *Acta Cryst. D.* **58**, 1937–1940

Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). *Acta Cryst. D.* **64**, 61–69.

Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H. & Hung, L.-W. (2009). *Acta Cryst. D* **65**, 582–601.

Terwilliger, T. C., Bunkóczi, G., Hung, L.-W., Zwart, P. H., Smith, J. L., Akey, D. L. & Adams, P. D. (2016). *Acta Cryst. D* **72**, 346–358.

Terwilliger, T. C., Bunkóczi, G., Hung, L.-W., Zwart, P. H., Smith, J. L., Akey, D. L. & Adams, P. D. (2016). *Acta Cryst. D* **72**, 359–374.

The Phenix Project

Lawrence Berkeley Laboratory

Paul Adams, Pavel Afonine,
Dorothee Liebschner, Nigel
Moriarty, Billy Poon,
Christopher Schlicksup,
Oleg Sobolev



University of Cambridge

Randy Read, Airlie McCoy,
Alisia Fadini



Los Alamos National Laboratory New Mexico Consortium

Tom Terwilliger, Li-Wei Hung



UTHealth

Matt Baker



Duke University

Jane Richardson, Christopher
Williams, Vincent Chen,
Michael Prisant



An NIH/NIGMS funded
Program Project

Liebschner D, *et al.*, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. Acta Cryst. 2019 D75:861–877