Phenix

Phasing (MR and experimental), density modification and map interpretation

Macromolecular Crystallography & Cryo-EM School

Instituto de Química-Física "Rocasolano", CSIC 10 May 2023

Tom Terwilliger

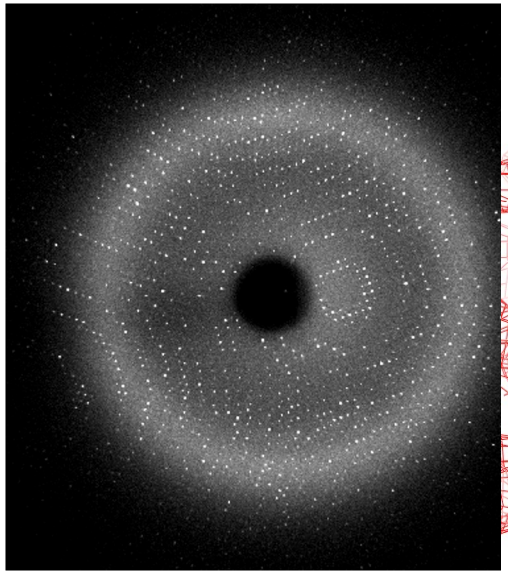
Los Alamos National Laboratory/New Mexico Consortium

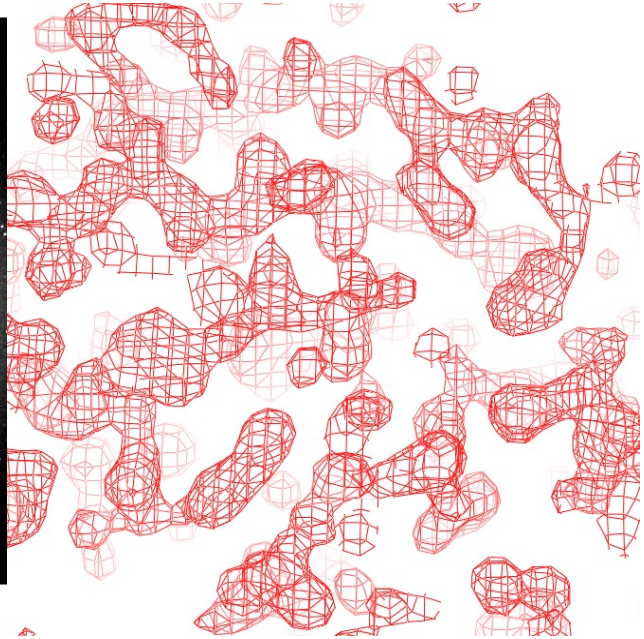
Los Alamos NATIONAL LABORATORY

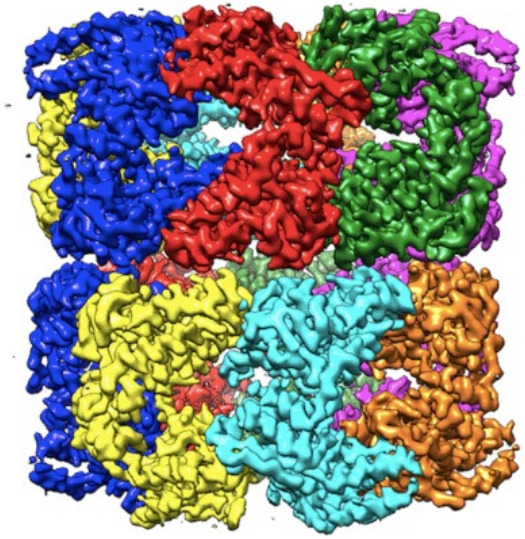
BERKELEY LAB

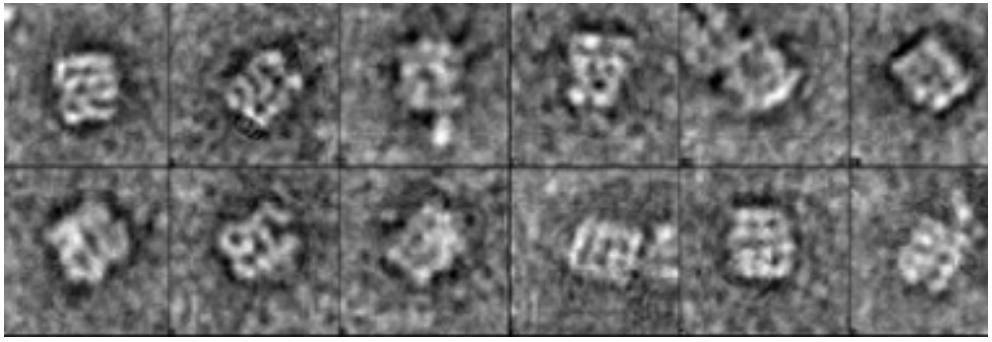
UNIVERSITY OF CAMBRIDGE

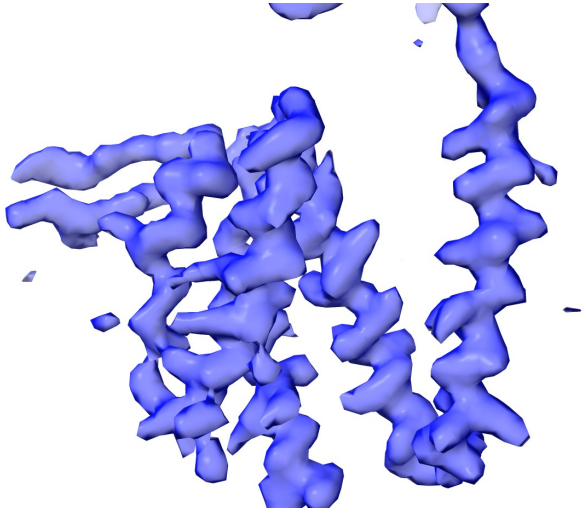
X-ray and cryo-EM data and maps





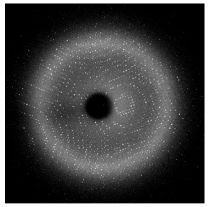






X-ray and cryo-EM data → density map

Data



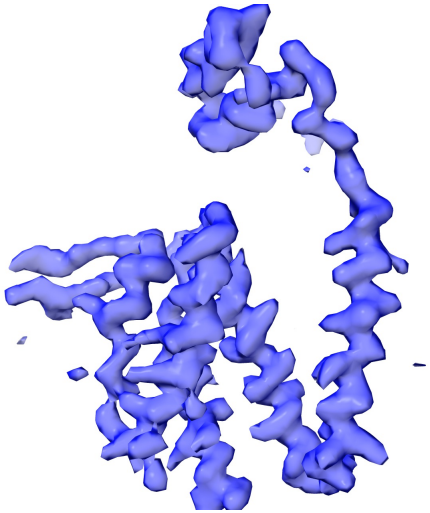
X-ray

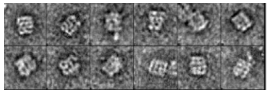
Experimental phasing

Molecular replacement

Phases

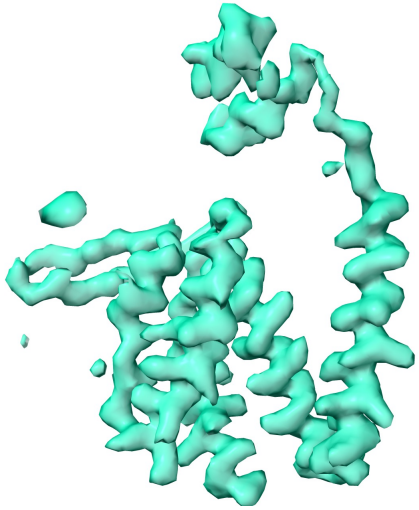
Density map



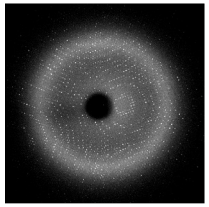


CryoEM

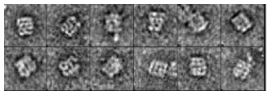
Combine information from images



X-ray and cryo-EM density map interpretation

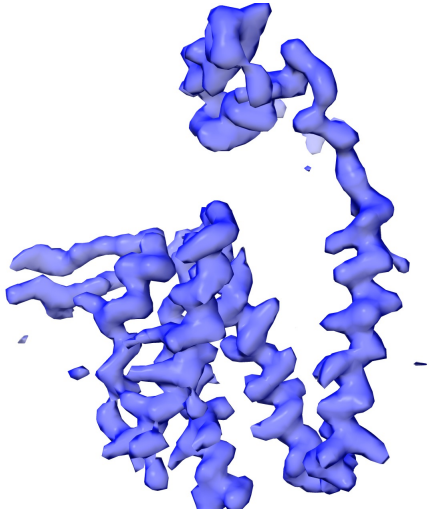


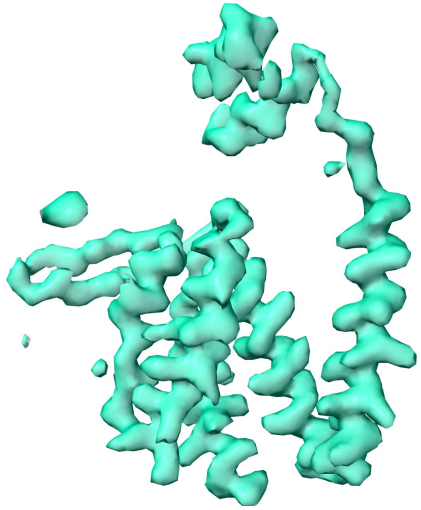
X-ray



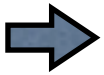
CryoEM

Density map

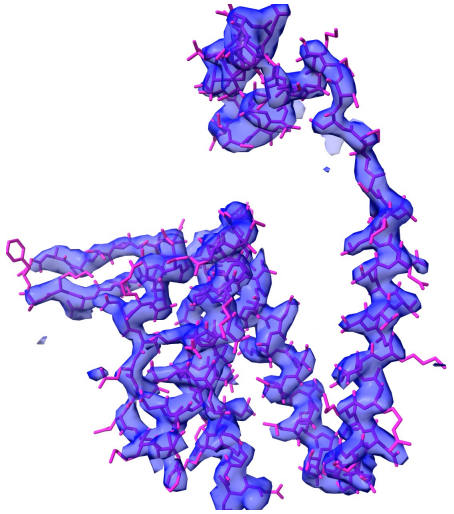


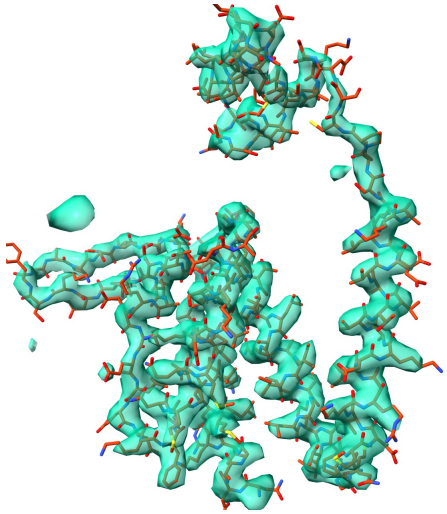


Dock or build model



Model and map

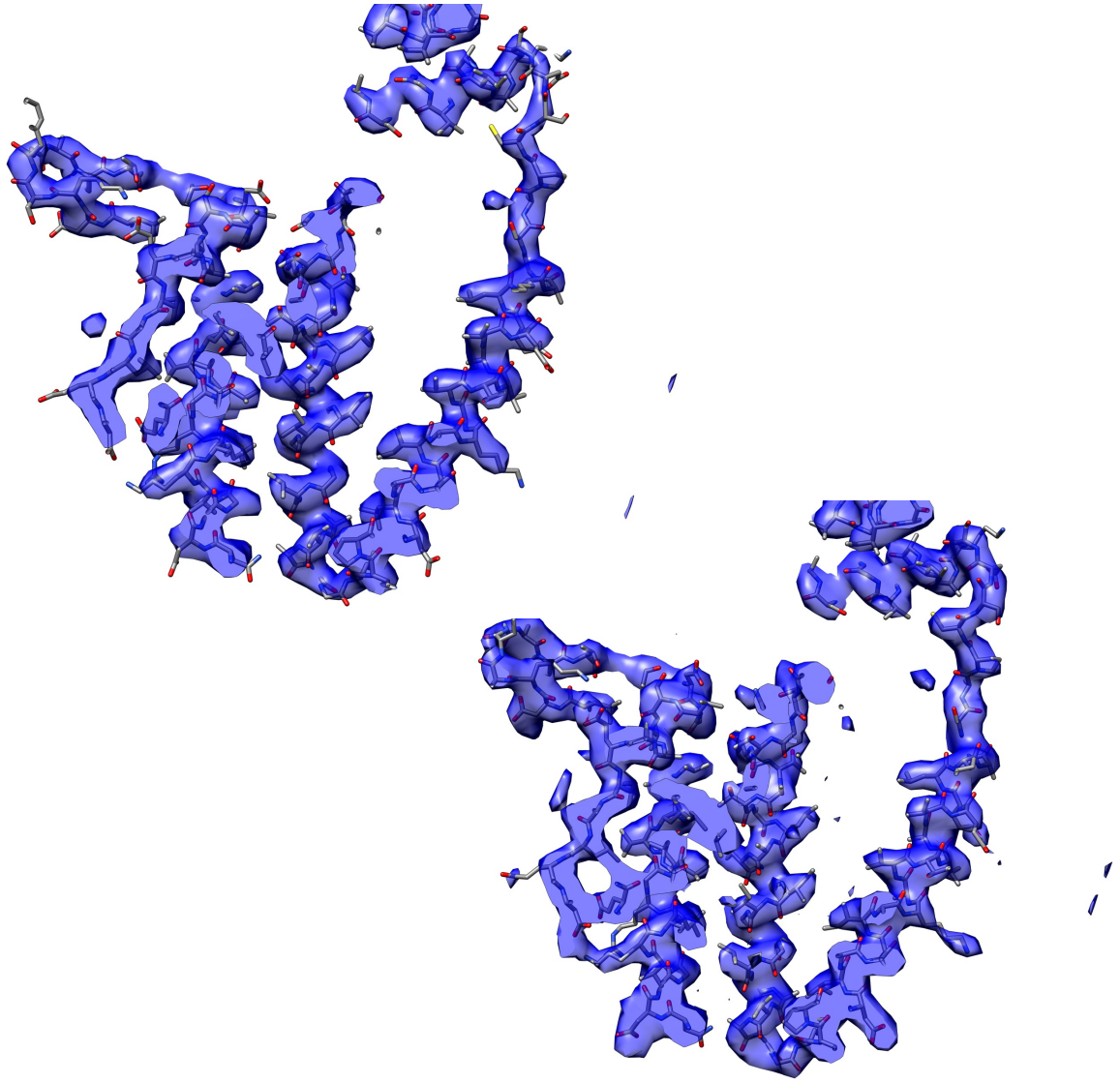




Crystallographic and cryo-EM maps are similar in many ways

X-ray map (3.8 Å resolution)

Cryo-EM map (3.5 Å resolution)

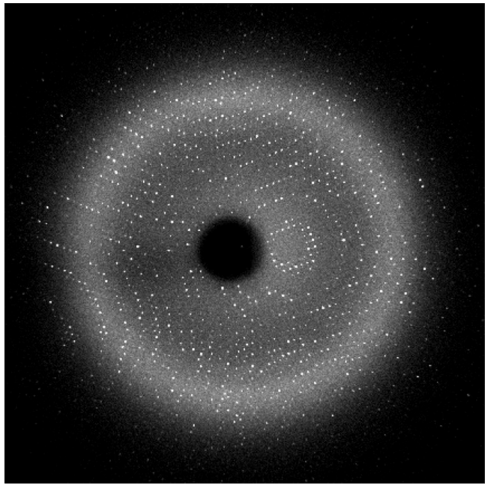


GroEL residues 10-170

X-ray and cryo-EM maps as Fourier transforms

(X-ray data are missing phases)

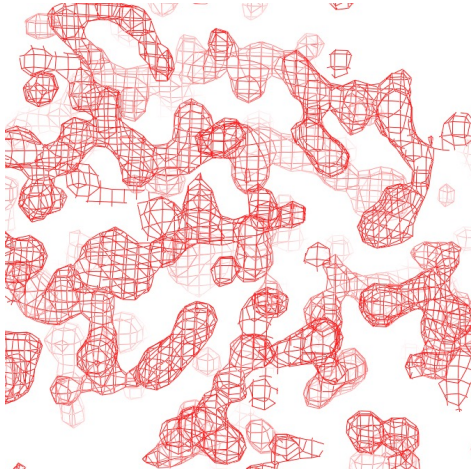
X-ray



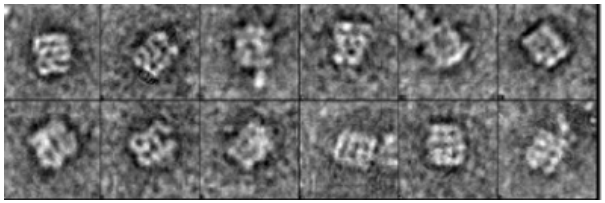
Diffraction pattern is (more or less) section through FT of structure (amplitudes only)

Combine FT sections

Inverse FT with phases



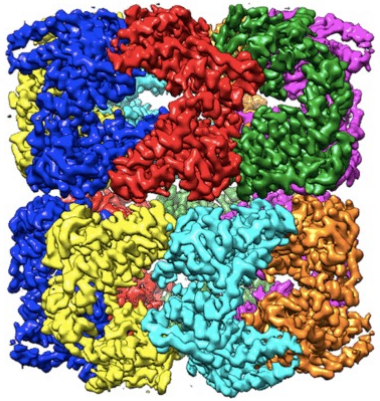
CryoEM



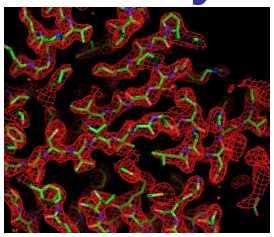
FT of one image is a section through FT of structure (amplitudes and phases)

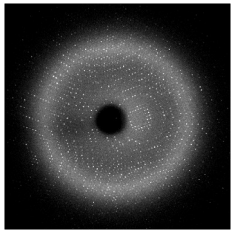
Combine FT sections

Inverse FT



Many ways to find the phases in crystallography

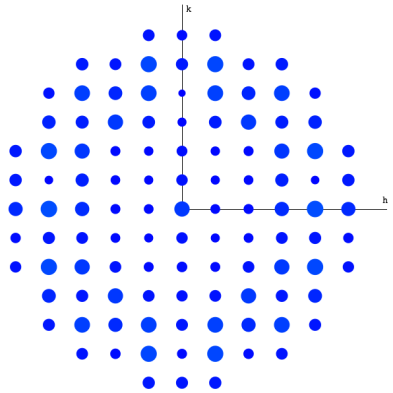


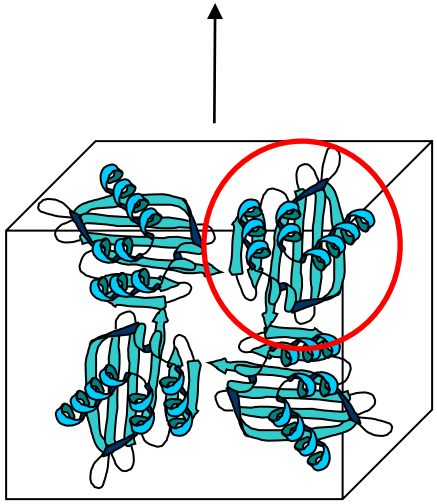


Method	Source of phasing information
SIR – single isomorphous replacement	A few heavy atoms (e.g., Hg, Au) in "derivative" contribute to differences from "native"
SAD – single-wavelength anomalous diffraction	A few atoms (e.g., Se, I, Hg atoms) contribute to "anomalous" differences in diffraction between spot *h* and spot *-h*
MAD – multiple-wavelength anomalous diffraction	A few atoms contribute to anomalous and wavelength-dependent "dispersive" differences
SIRAS, MIR	Combinations of SIR and SAD
Molecular replacement	Molecular location and phases are found using a related molecule as a template
Direct methods	Guess where atoms are, good guesses match the measured structure factors

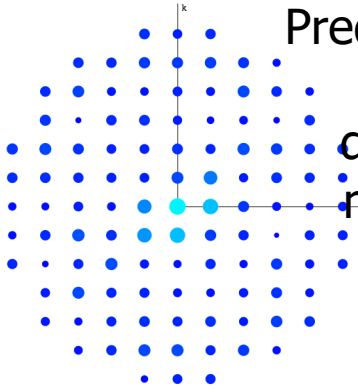
X-ray phases from molecular replacement

X-ray data

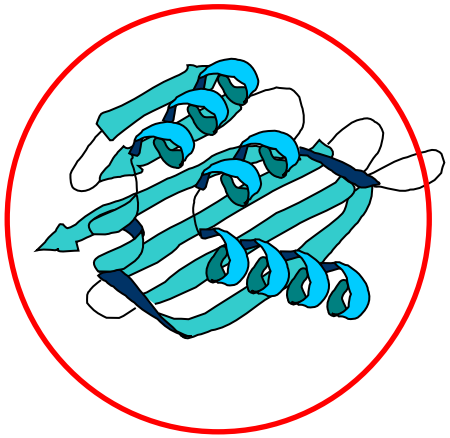




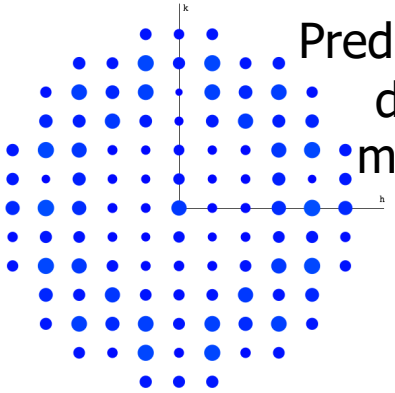
True structure



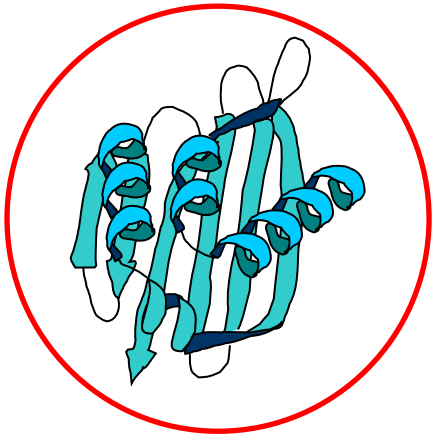
Predicted data do not match



Incorrect orientation and location



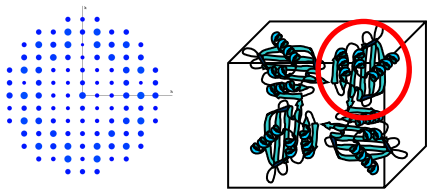
Predicted data match



Correct orientation and location

Molecular replacement

Two-stage search for orientation and position



Rotation search

Likelihood: *probability of measuring these data if this rotation were correct, averaging over all possible translations*

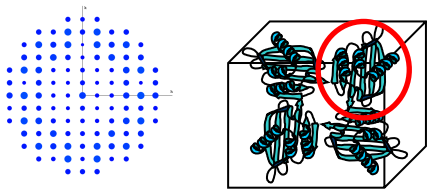
Translation search

Likelihood: *probability of measuring these data if this rotation/translation were correct*

Likelihood scoring

How likely is it that I would measure these data if this solution were correct?

Did MR work?



LLG: log-likelihood gain

How much better this solution explains the data than a random one.

(LLG=50 or greater: usually correct)

TFZ: Translation function Z-score

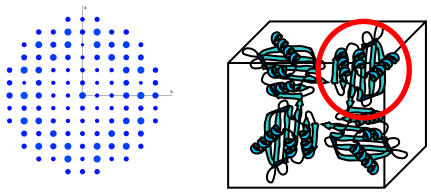
How much better this translation is than average for this orientation

(TFZ=7 or greater: usually correct)

Likelihood scoring

How likely is it that I would measure these data if this solution were correct?

Will MR work?



eLLG: estimated log-likelihood gain

Based on accuracy of model, number of reflections

(eLLG=50 or greater: usually can be solved)

Likelihood scoring

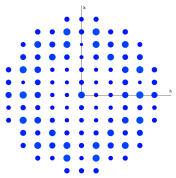
How likely is it that I would measure these data if this solution were correct?

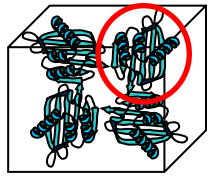
Getting the most from MR

Collect the best data possible

Higher resolution helps

A better search model helps





The likelihood calculation is affected by crystal pathologies:

translational non-crystallographic symmetry

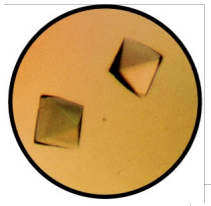
twinning

anisotropy of the data

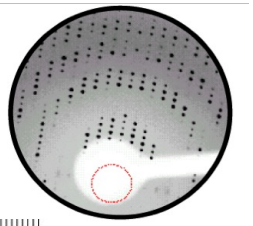
MR can be combined with experimental phasing (anomalous data)

Obtaining experimental X-ray phases with Se-SAD

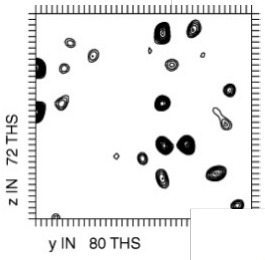
Crystals with SeMet



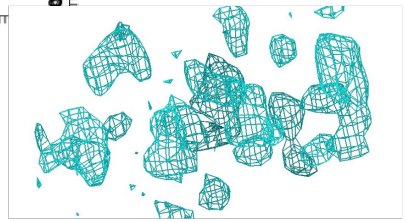
Collect anomalous SAD data



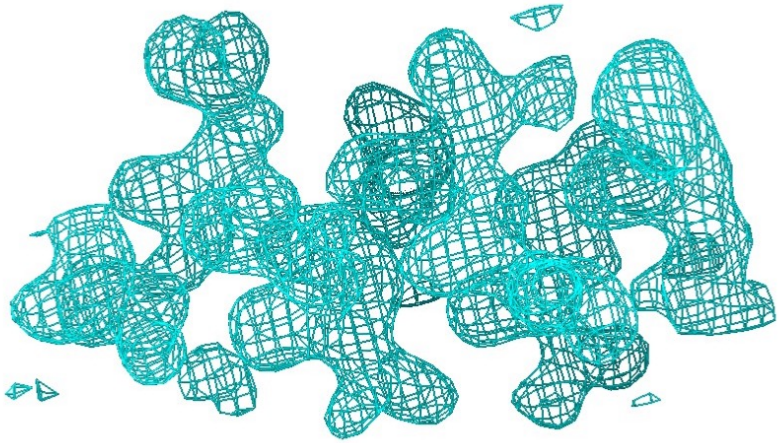
Locate Se atoms

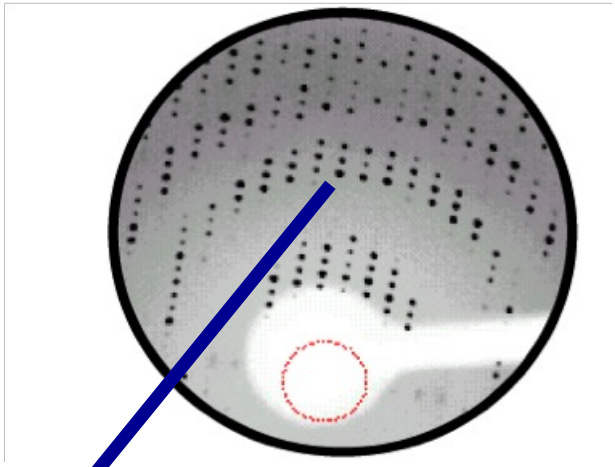


Phasing (calculate density map)



If we knew the phases (ϕ_h) we could calculate a map





$\rho(x)$

(Where the atoms are)

F_h is square root of measured intensity I_h of spot h

$$\rho(x) = \sum_h F_h e^{i\phi_h} e^{-2\pi i h x}$$

We do not know the phase (ϕ_h)

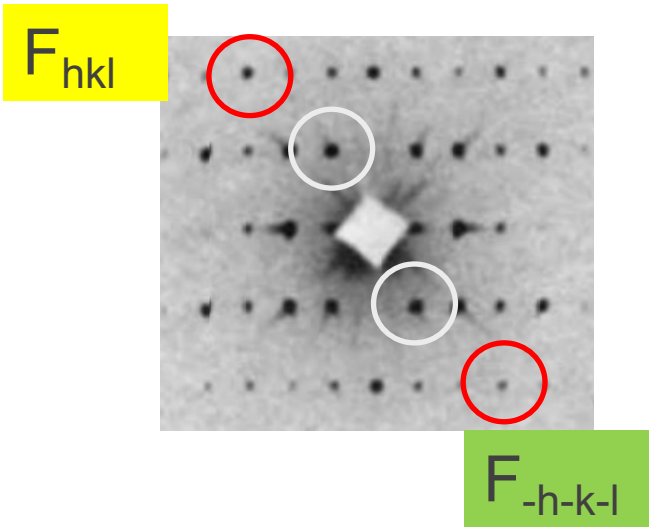
SAD phasing (single-wavelength anomalous diffraction)

If no anomalous scattering:

$$F_{hkl} = F_{-h,-k,-l}$$

Anomalous differences:

$$F_{hkl} \neq F_{-h,-k,-l}$$



Anomalous diffraction from Fe and S in HiPIP.

White pair: small difference.
Red pair: large difference.

Holden et al., J.Biol. Chem 261, 14746 (1986)

Where do anomalous differences come from?

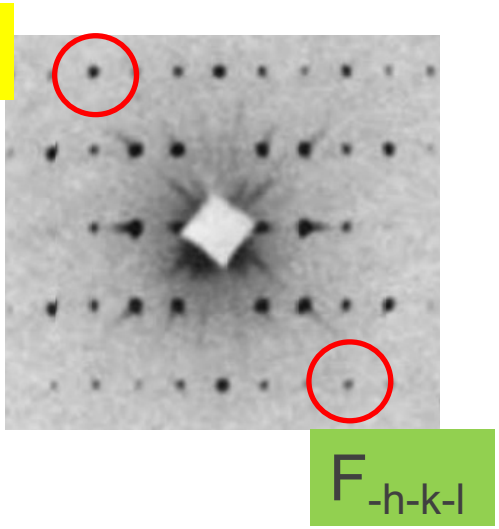
One reflection in Se-SAD: F_{hkl}

Protein: $F_{protein}$

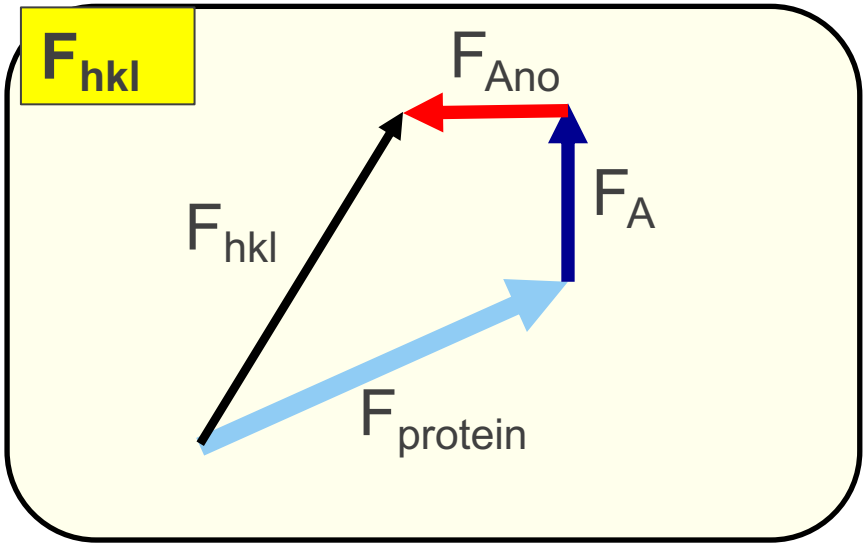
Se atoms: F_A

Anomalous scattering from Se: F_{ano}

Total structure factor $F_{hkl} = F_{protein} + F_A + F_{ano}$



Key fact for anomalous scattering from one type of atom: phase of F_{ano} is always +90° from F_A

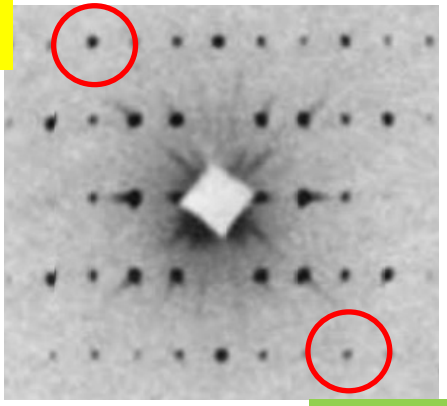


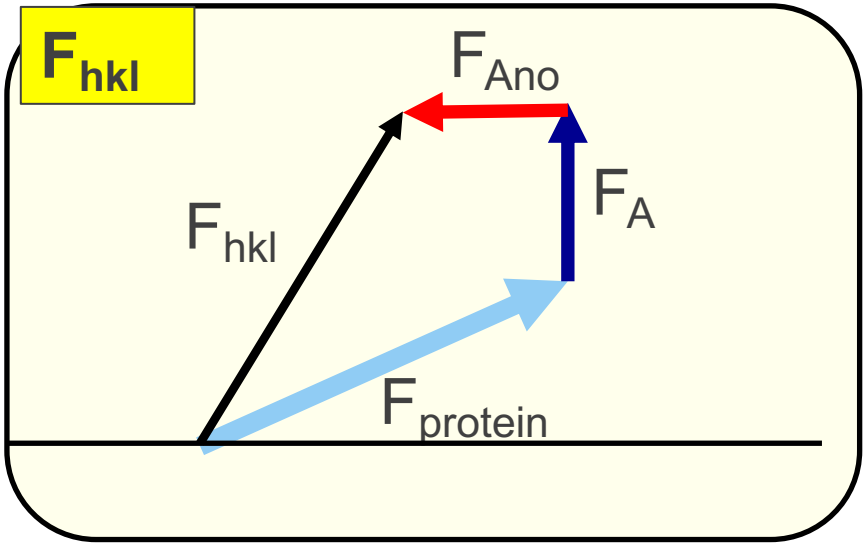
Where do anomalous differences come from?

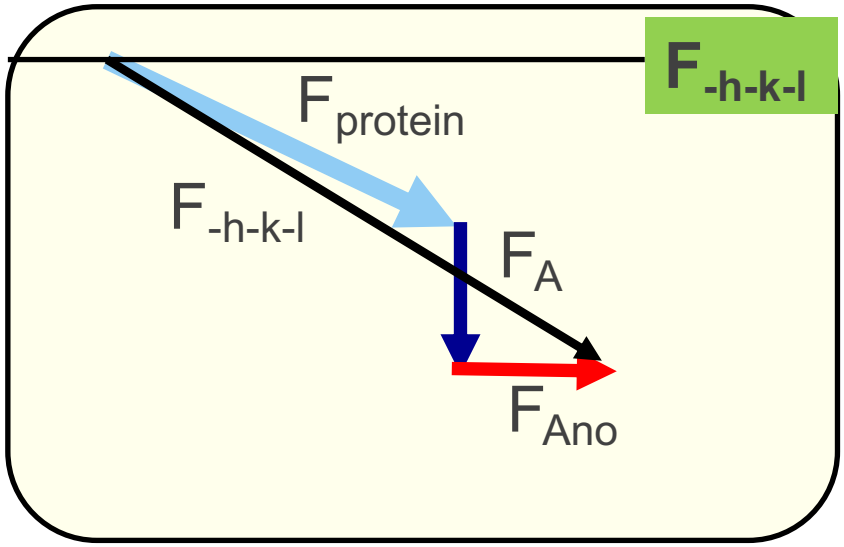
Compare F_{hkl} and F_{-h-k-l}

Normal scattering: phase of F_{-h-k-l} is negative of phase of F_{hkl} for $F_{protein}$ and F_A

Anomalous scattering: phase of F_{ano} is always +90° from F_A





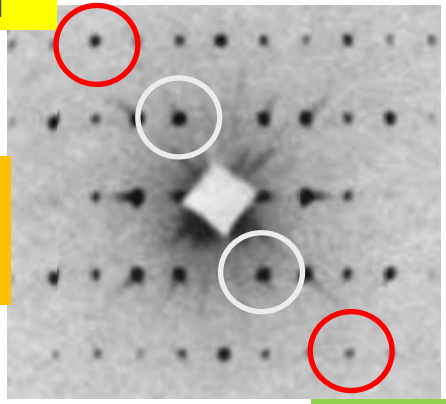


Length of F_{hkl} is different than F_{-h-k-l}

SAD phasing strategy

Key facts about anomalous differences:

F_{hkl}



F_{-h-k-l}

Due to sub-structure of anomalously-scattering atoms

Depend on phase differences between structure factors for sub-structure and all other atoms

Getting phases from anomalous differences:

Anomalous differences → sub-structure (Se atom positions)

Likelihood scoring: "How likely is it that I would measure F_{hkl} , $F_{-h,-k,-l}$ if this set of Se positions were correct?

Sub-structure and anomalous differences → phases for complete structure

Likelihood scoring: "Given this set of Se positions, how likely is it that I would measure F_{hkl} , $F_{-h,-k,-l}$ if this phase were correct?"

Will I find the anomalous substructure?

How many sites?

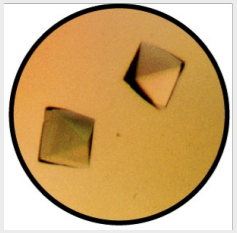
Are sites ordered?

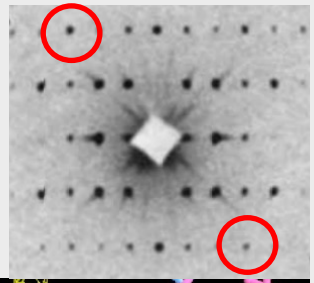
Anomalous atom?

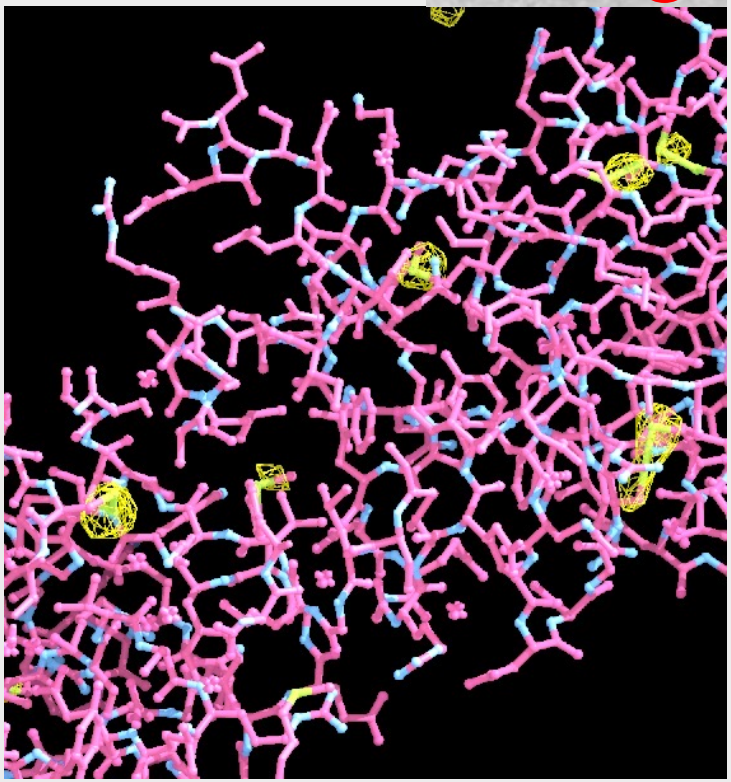
Wavelength?

Accurate data?

How many reflections?



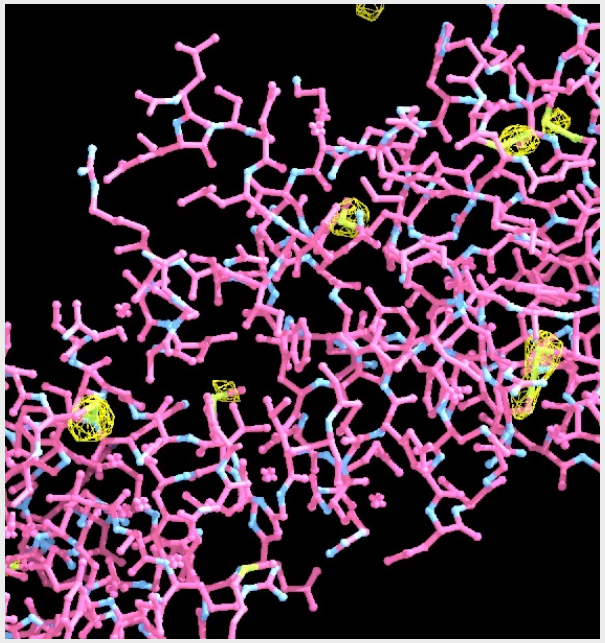


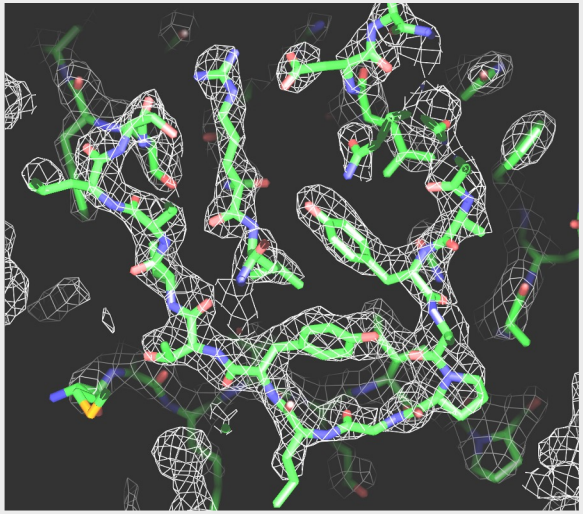


Key steps in SAD structure determination

1. Find the substructure

Anomalous signal S_{ano}



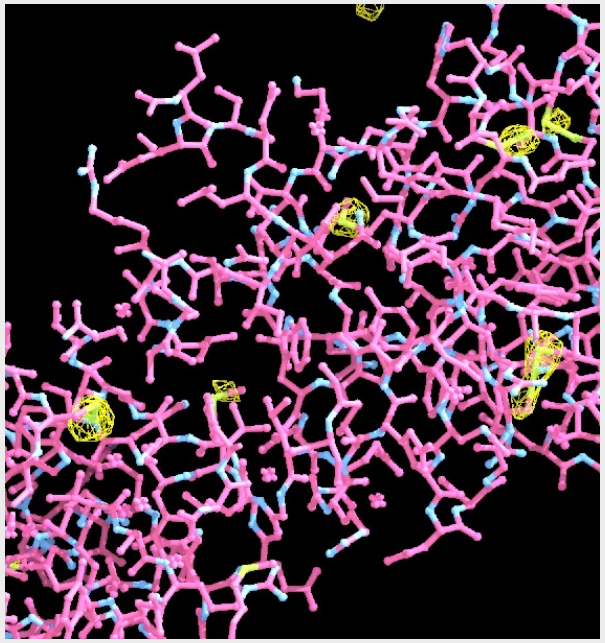


2. Calculate an interpretable map

Anomalous correlation CC^*_{ano}

Anomalous signal

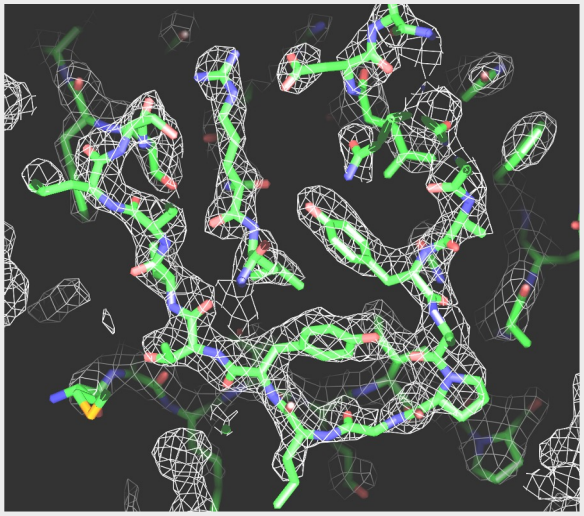
1. Find the substructure



Anomalous signal S_{ano}

- 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

Accuracy of the data

Anomalous correlation

Number of reflections

Anomalous signal S_{ano}

$$\langle S_{ano} \rangle = CC^{*}_{ano} \cdot \frac{\sqrt{N_{refl}}}{\sqrt{n_{sites}}} \cdot \frac{1}{f^{1/2}}$$

Will I find sites?

Number of sites

B-value for anomalous sub-structure

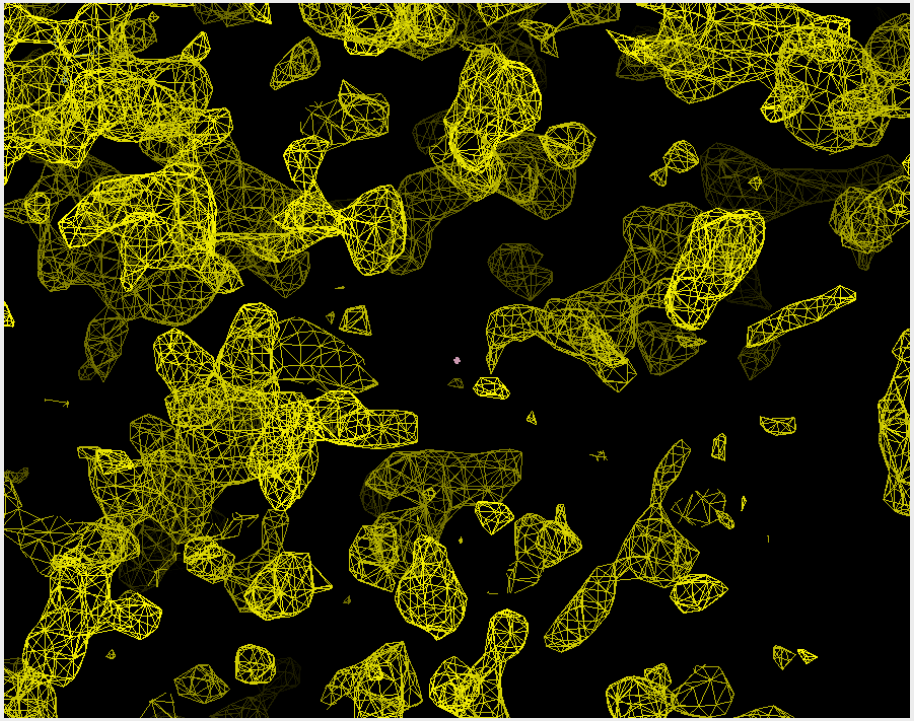
Map evaluation and improvement

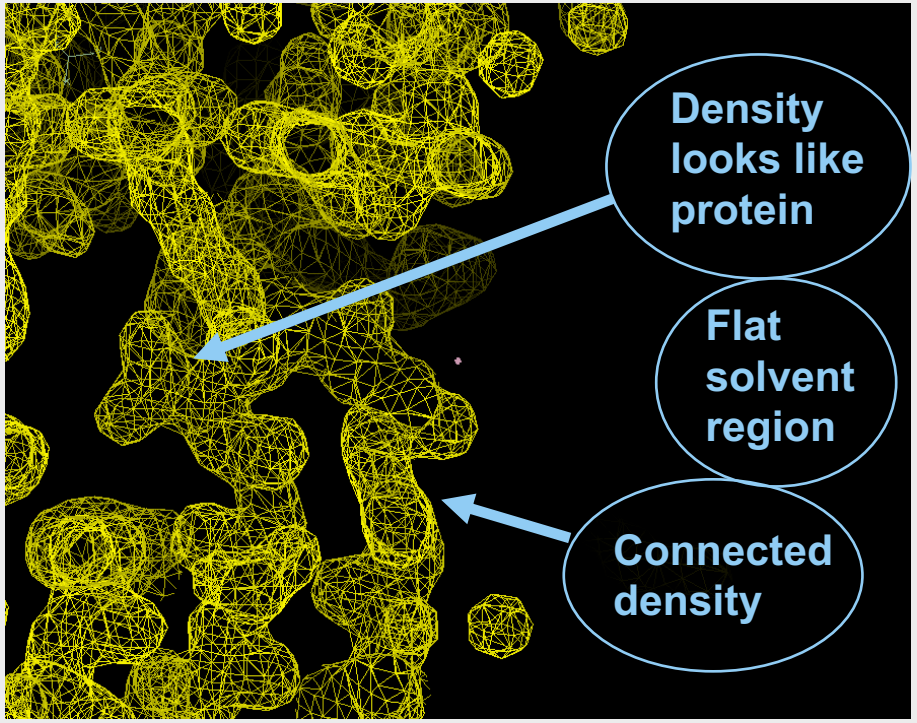
What does a good electron density map look like?

Using expected features of maps to make decisions and to improve maps

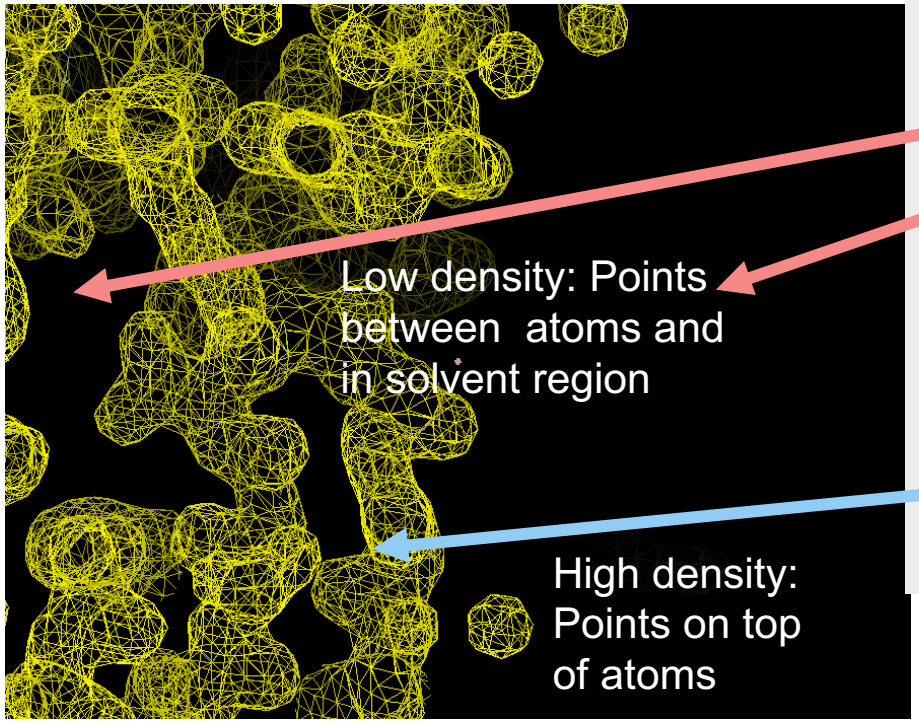
Map evaluation and improvement

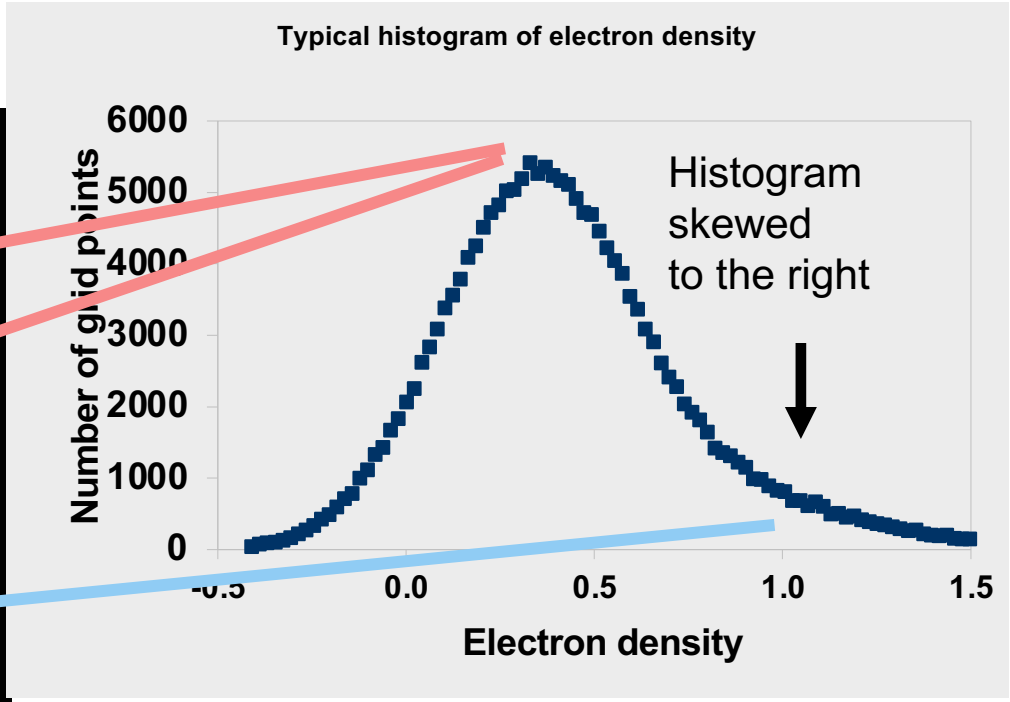
Which map is better?



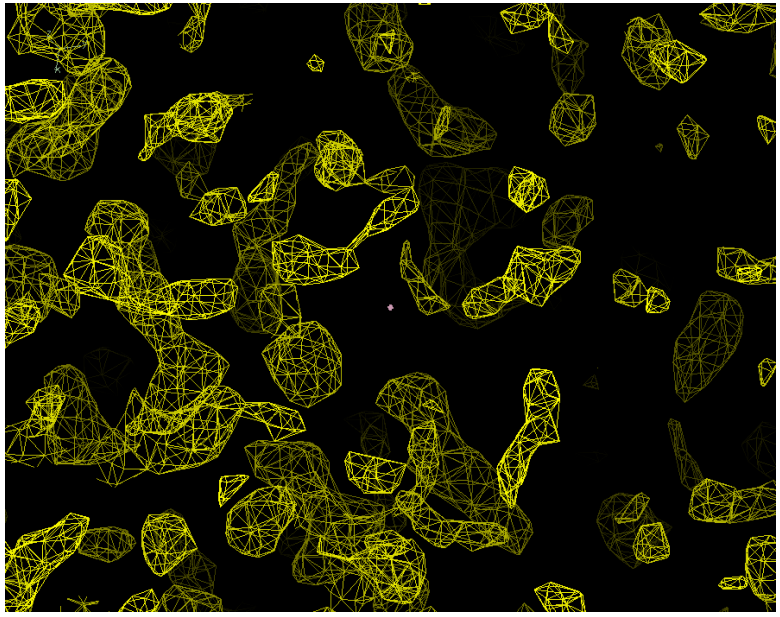


Histograms of density have positive skew

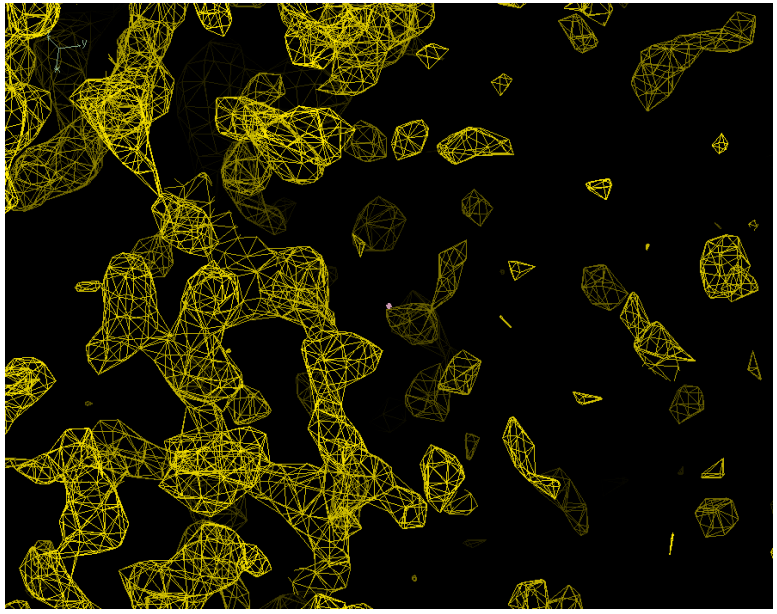




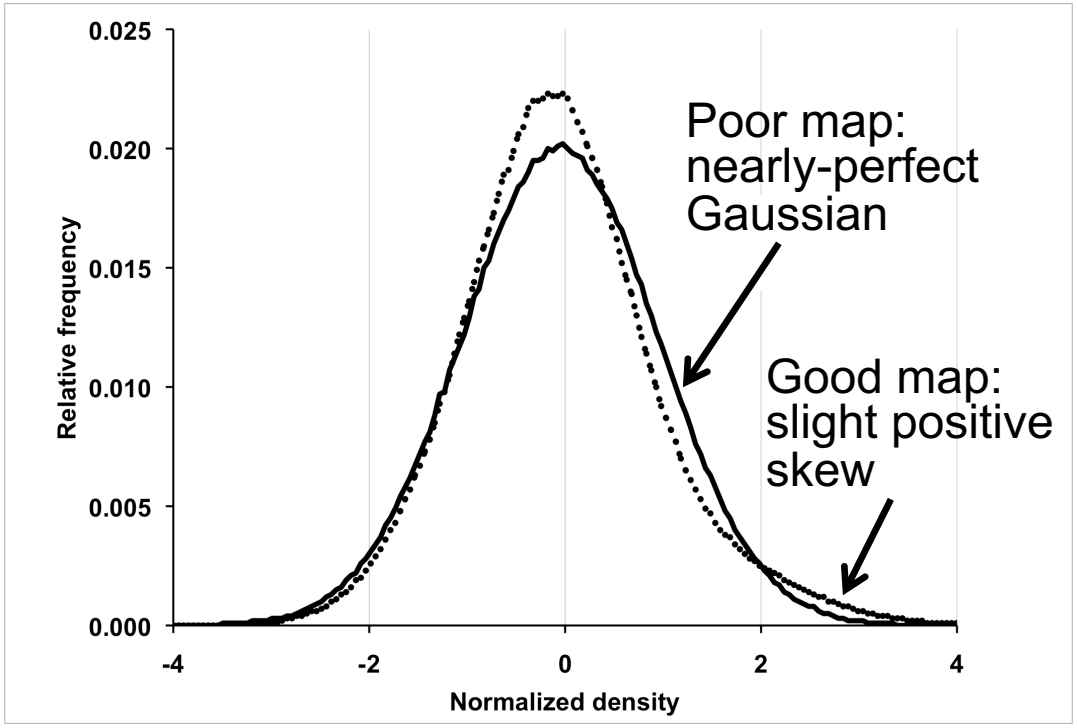
Histograms of density have positive skew



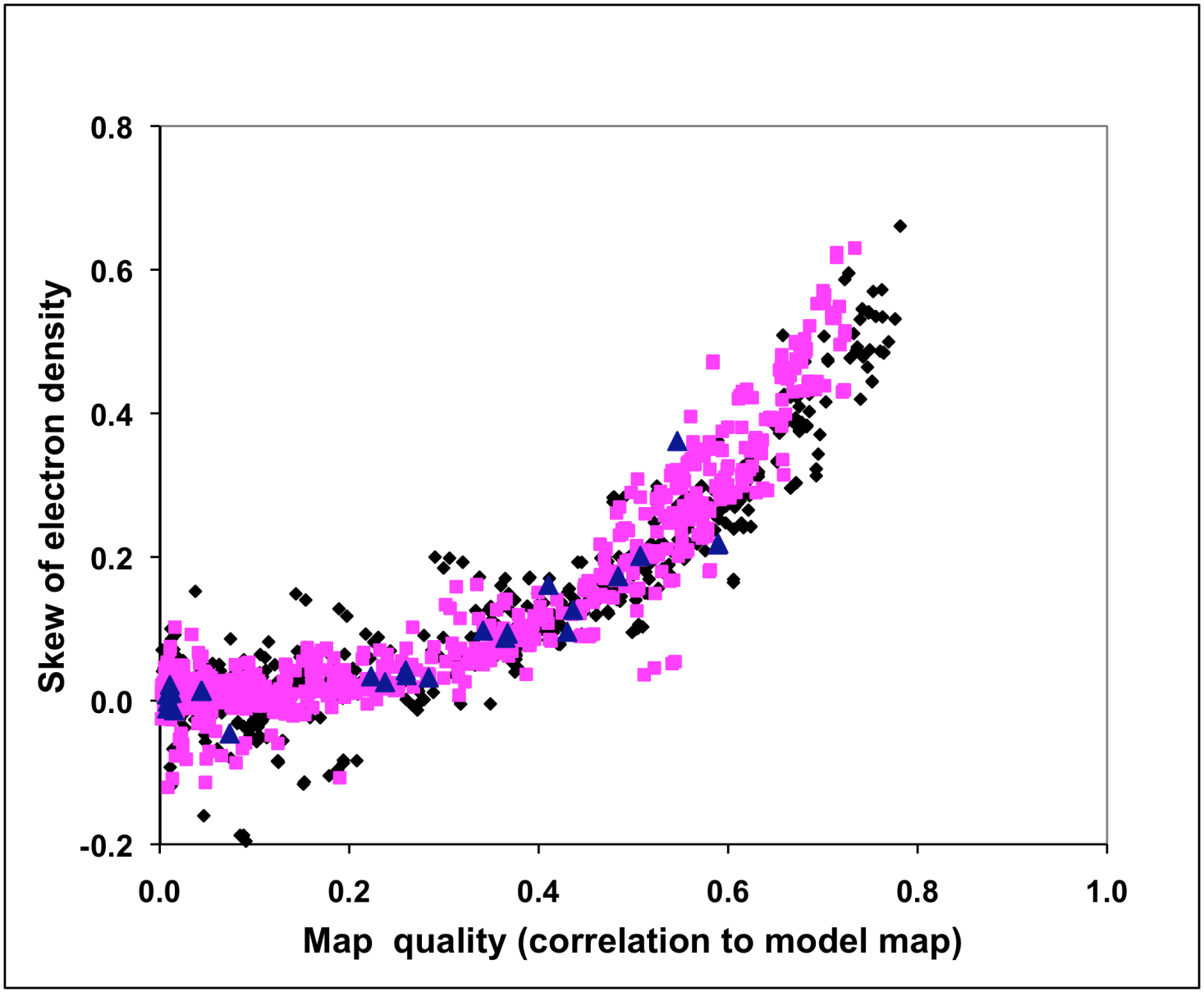
Poor map
(inverse hand)



Good map



Positive skew in good maps



Map improvement by density modification

What does a good electron density map look like?

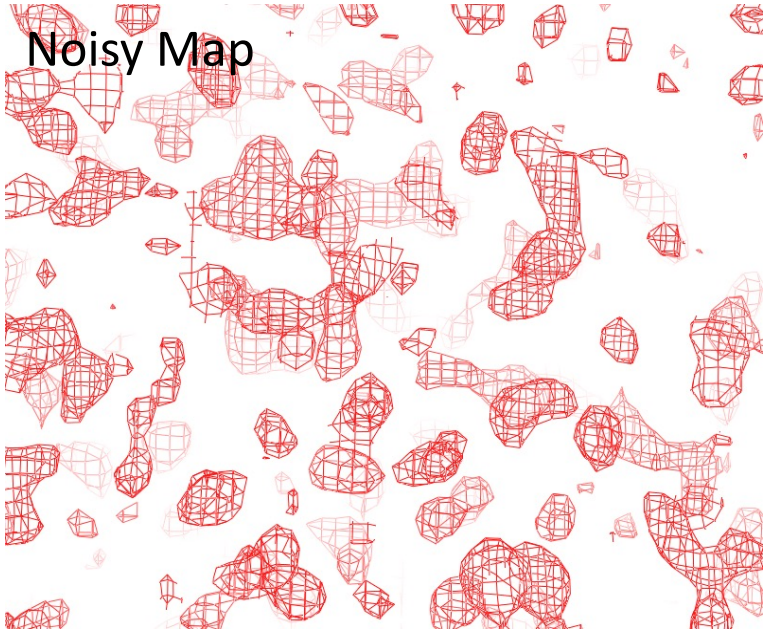
Using expected features to improve maps (X-ray or cryo-EM)

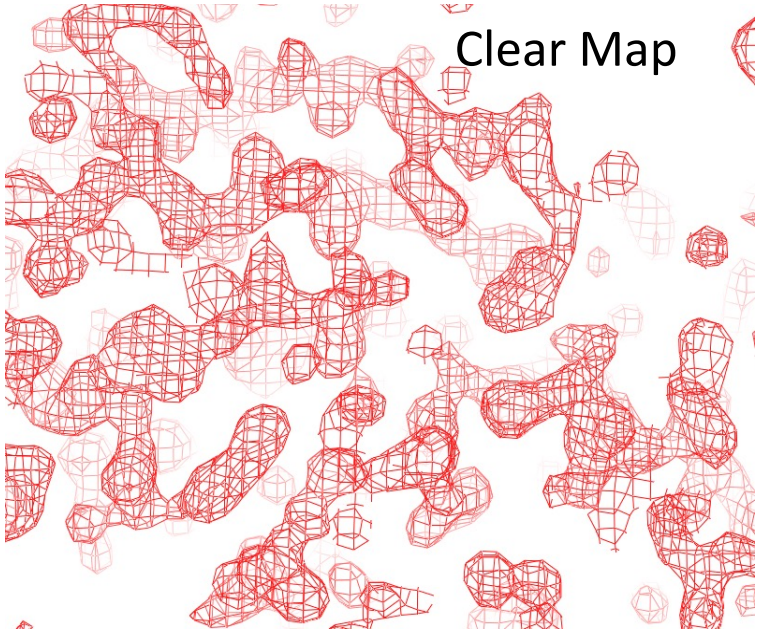
Density modification = “phase improvement”

Experimental Data

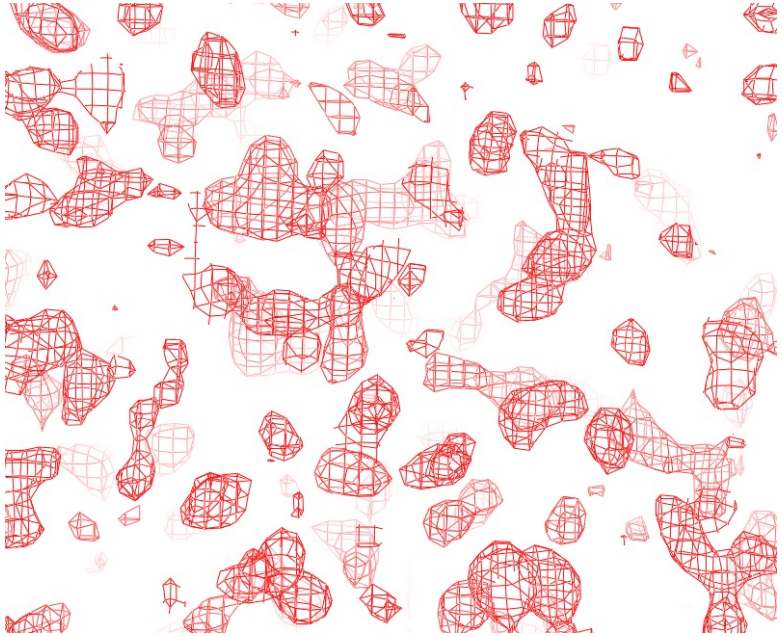
Initial phases

Improved phases

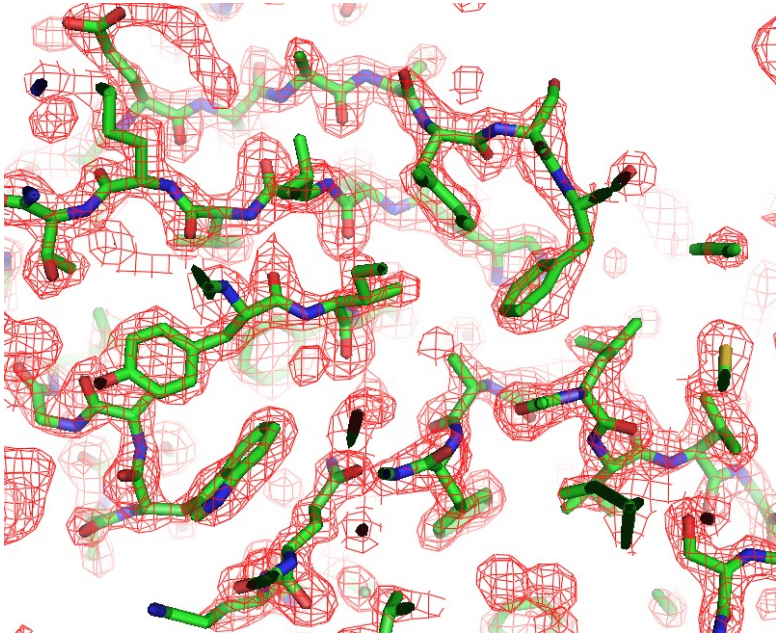




We know a good map when we see it

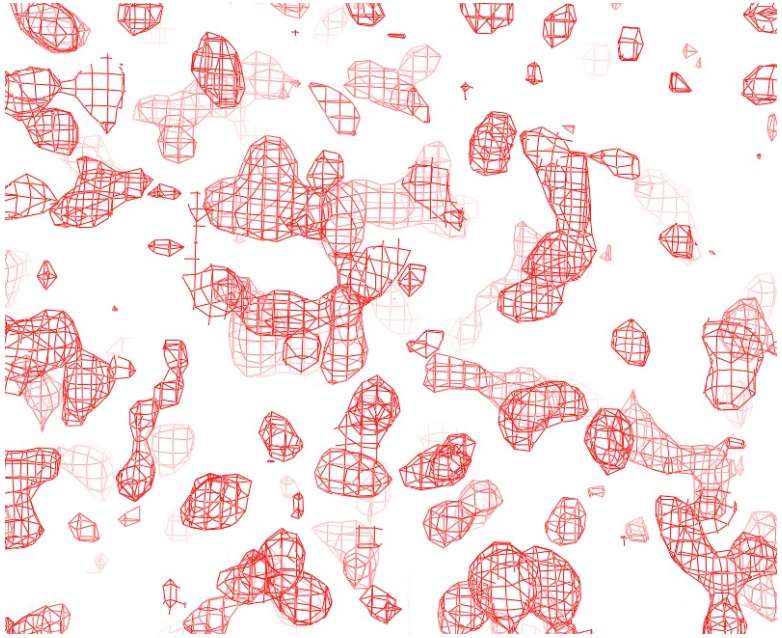


Noisy map

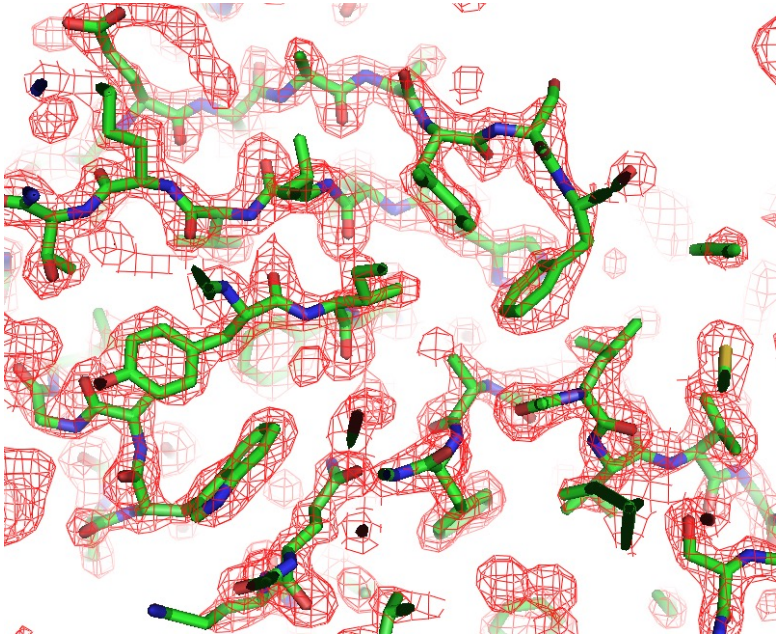


Clear map

Basis of density modification



Noisy map



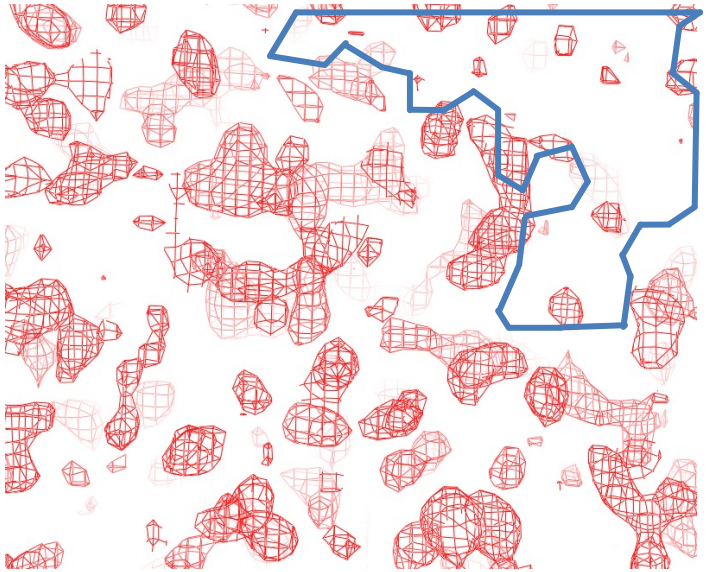
Clear map

1. We know a good map when we see it

2. Improvement anywhere means improvement everywhere

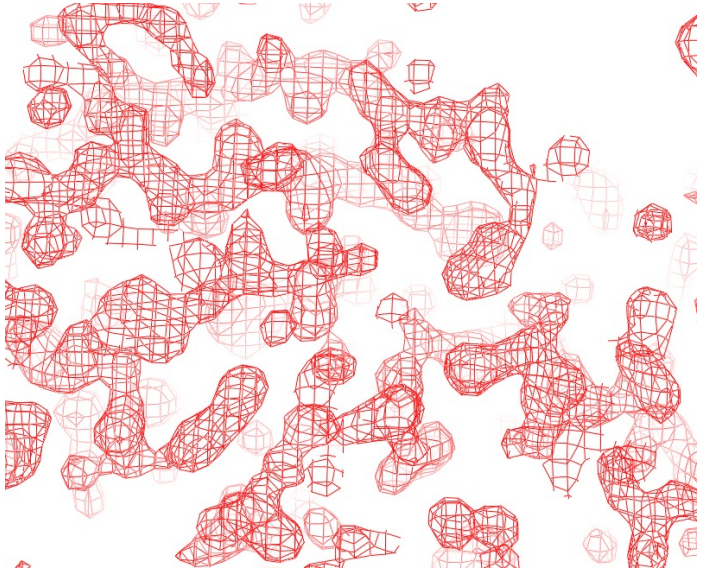
Density modification

Identify local expected density



Noisy map

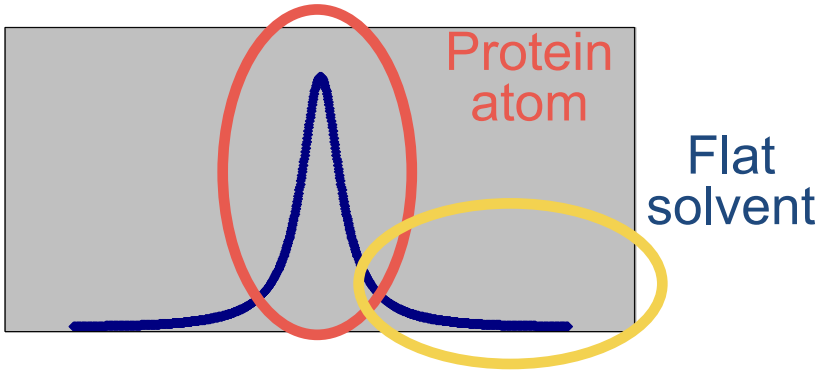
Find phases consistent with **experiment** and **expected density**



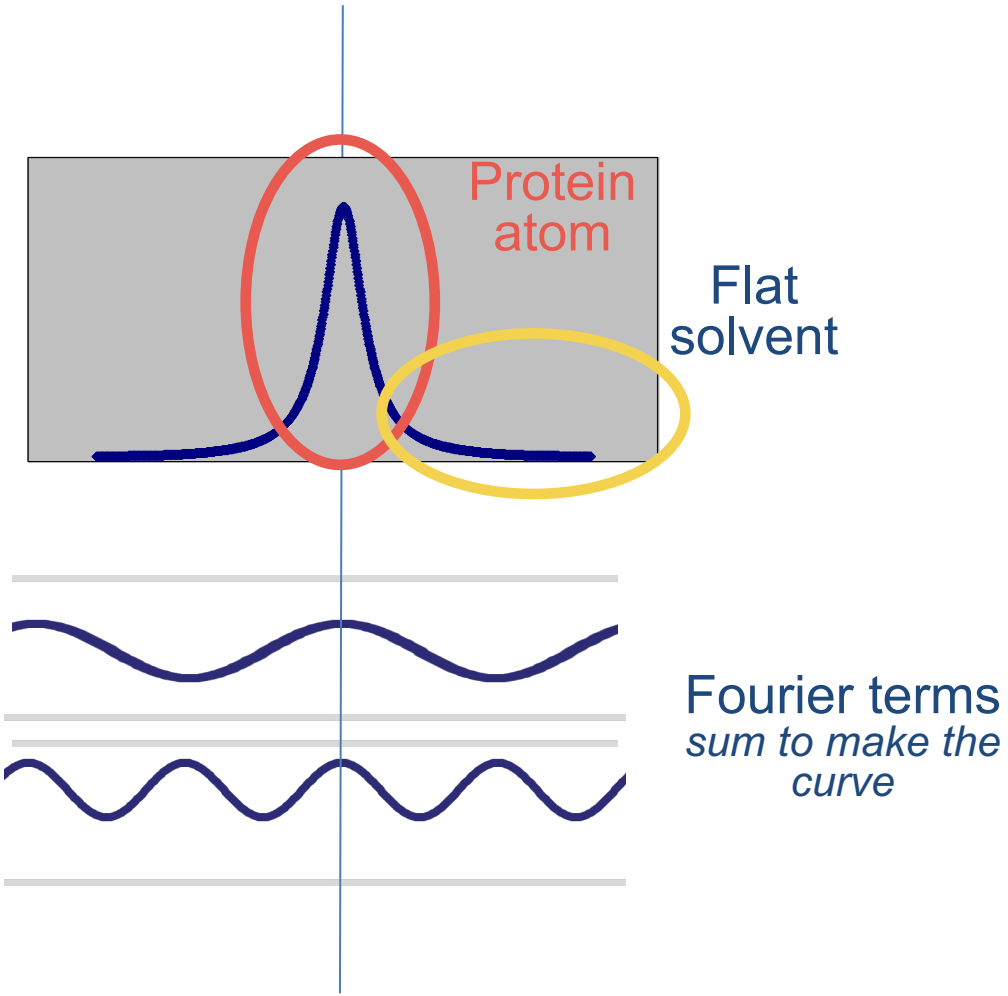
Clear map

Density everywhere is improved

One atom and a flat solvent region

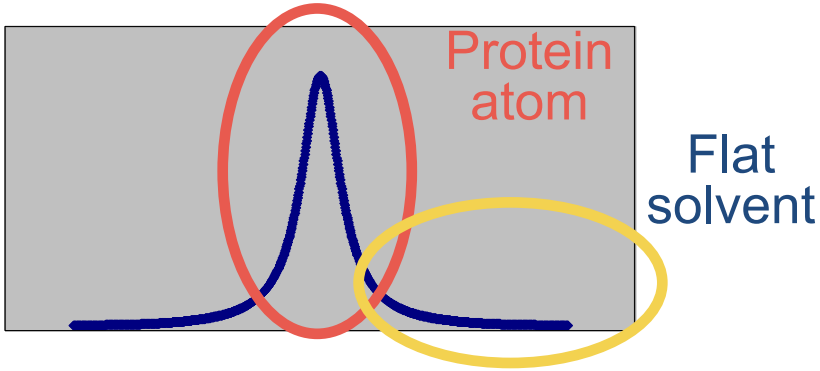


A Fourier sum of sines and cosines

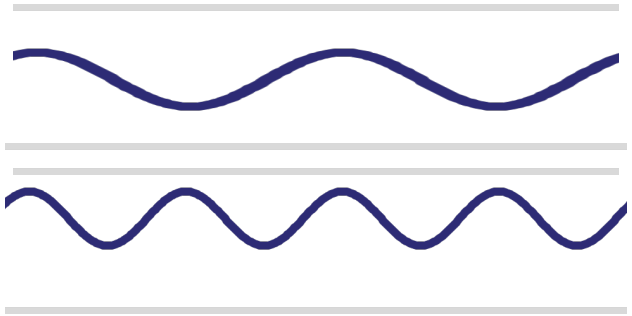


A Fourier sum of sines and cosines

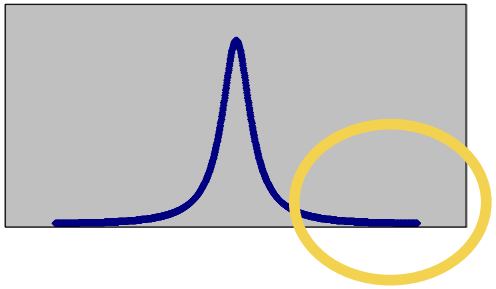
Find out the phase of one Fourier term using:



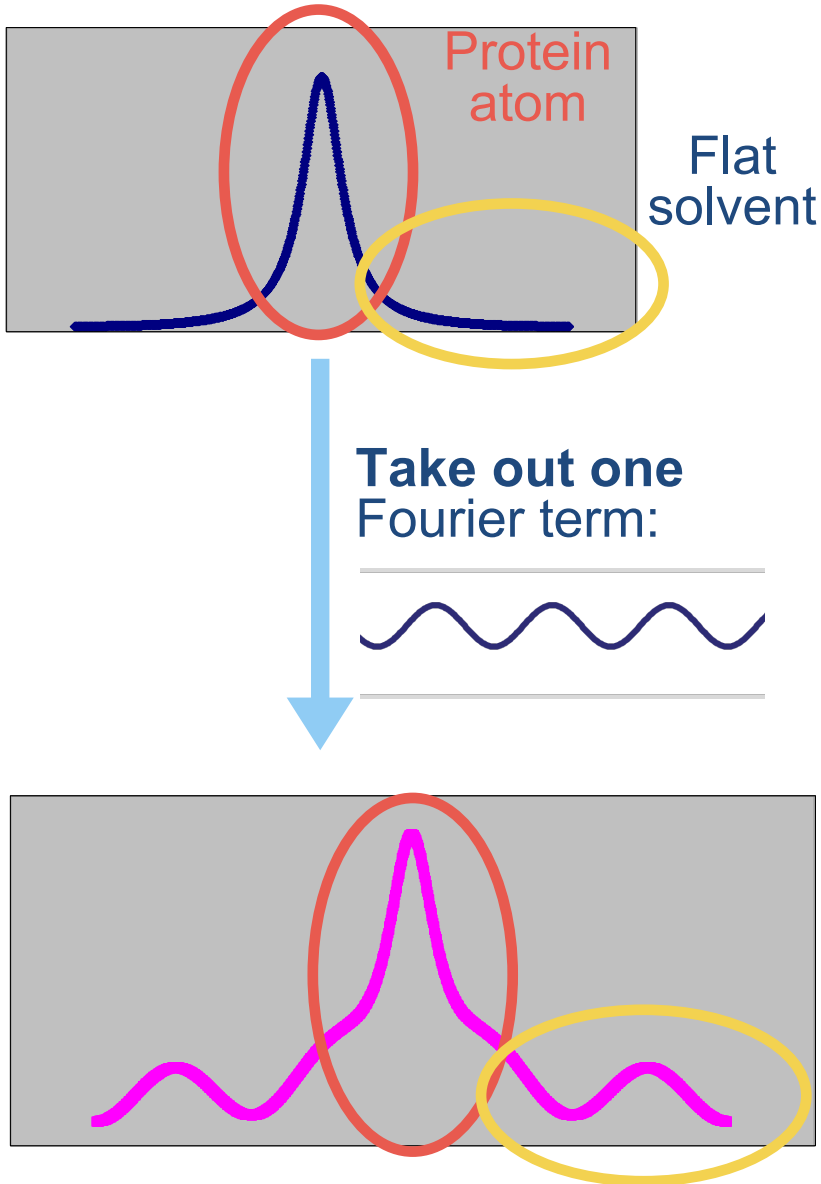
1) All other Fourier terms



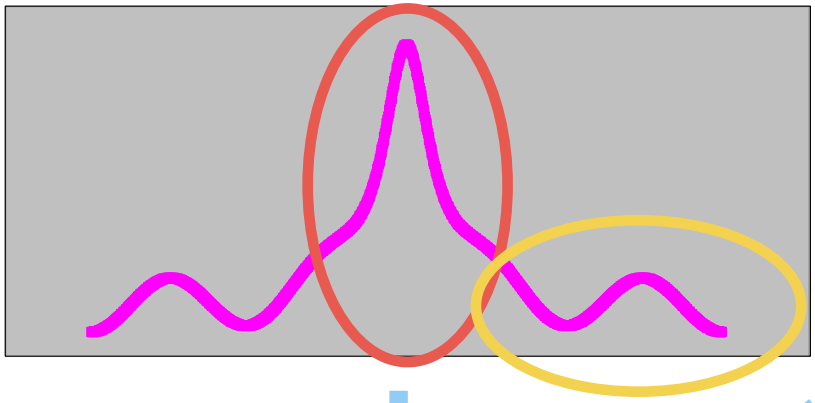
2) Flat solvent



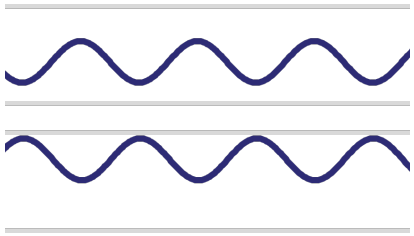
A Fourier sum of sines and cosines



Using flat solvent to identify phase of one term



One Fourier term:

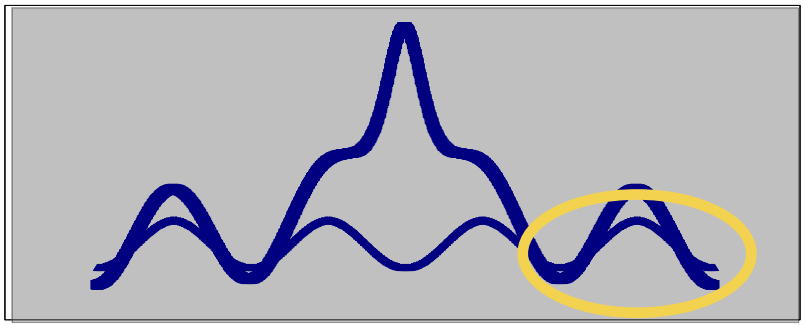


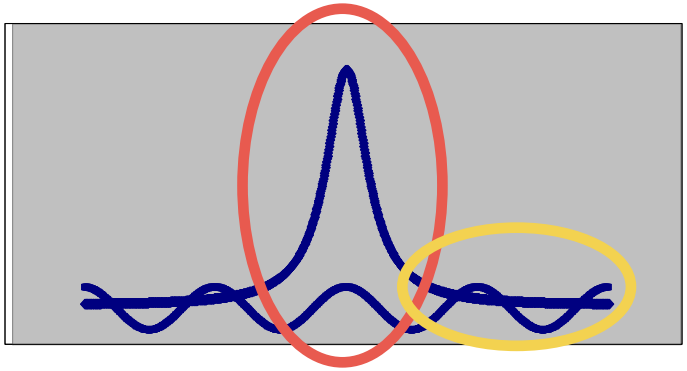
Correct phase

Phase 180° off

Adding the incorrect phase

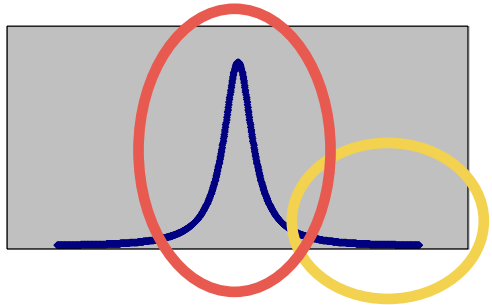
Adding the correct phase

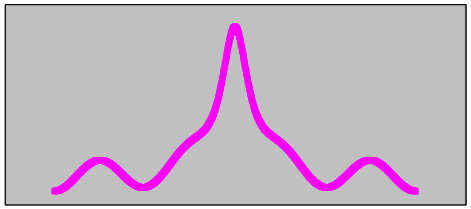


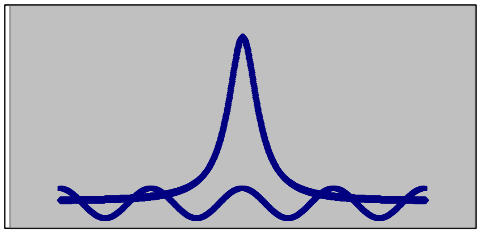


Solvent gets worse

Density modification of real maps







Incorrect phase

Correct phase

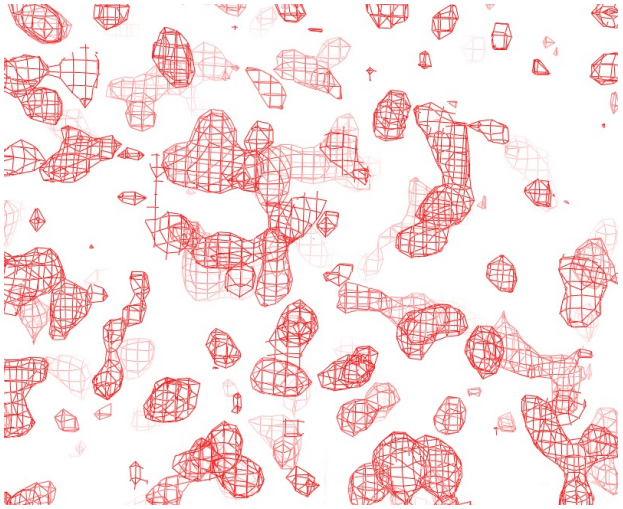
Real world:
Correct phase → $p_{map}(\varphi)$

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

Density modification phase probability:

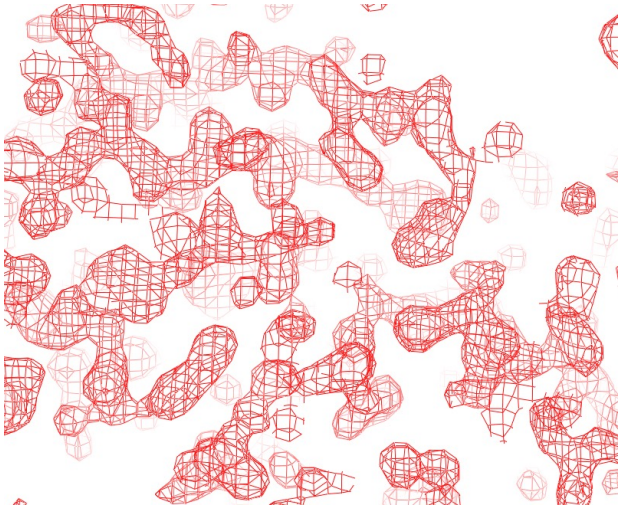
$$p(\varphi) = p_{exp}(\varphi)\, p_{map}(\varphi)$$

Key elements of density modification



Improved phases

$$p(\varphi) = p_{exp}(\varphi)\, p_{map}(\varphi)$$



We know a good map when we see it

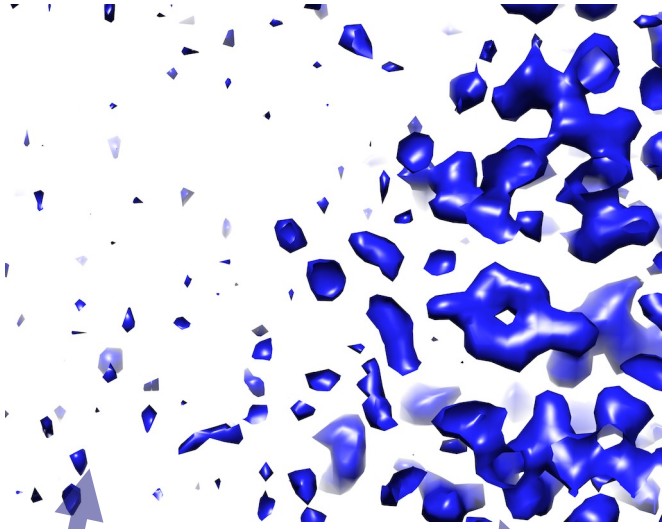
Improvement anywhere means improvement everywhere

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

Density modification with cryo-EM maps

Using expectations about one part of a map to improve another part of the map

Original map



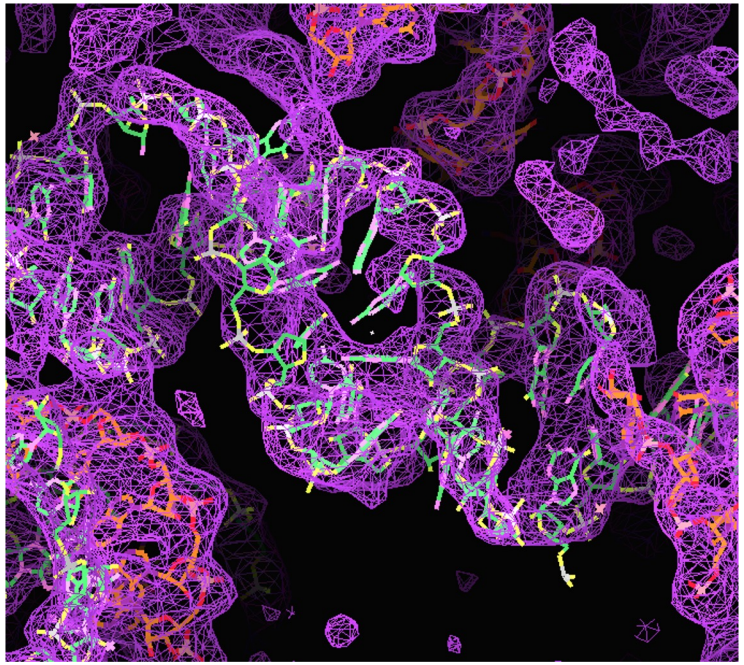
Density-modified

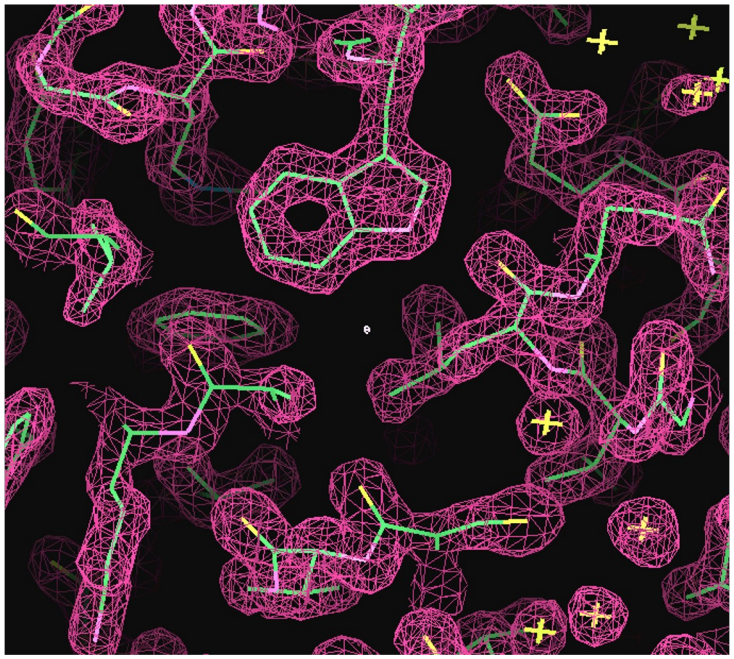
Solvent should be flat

Distribution of density values (histograms) should match typical protein

Automated model-building

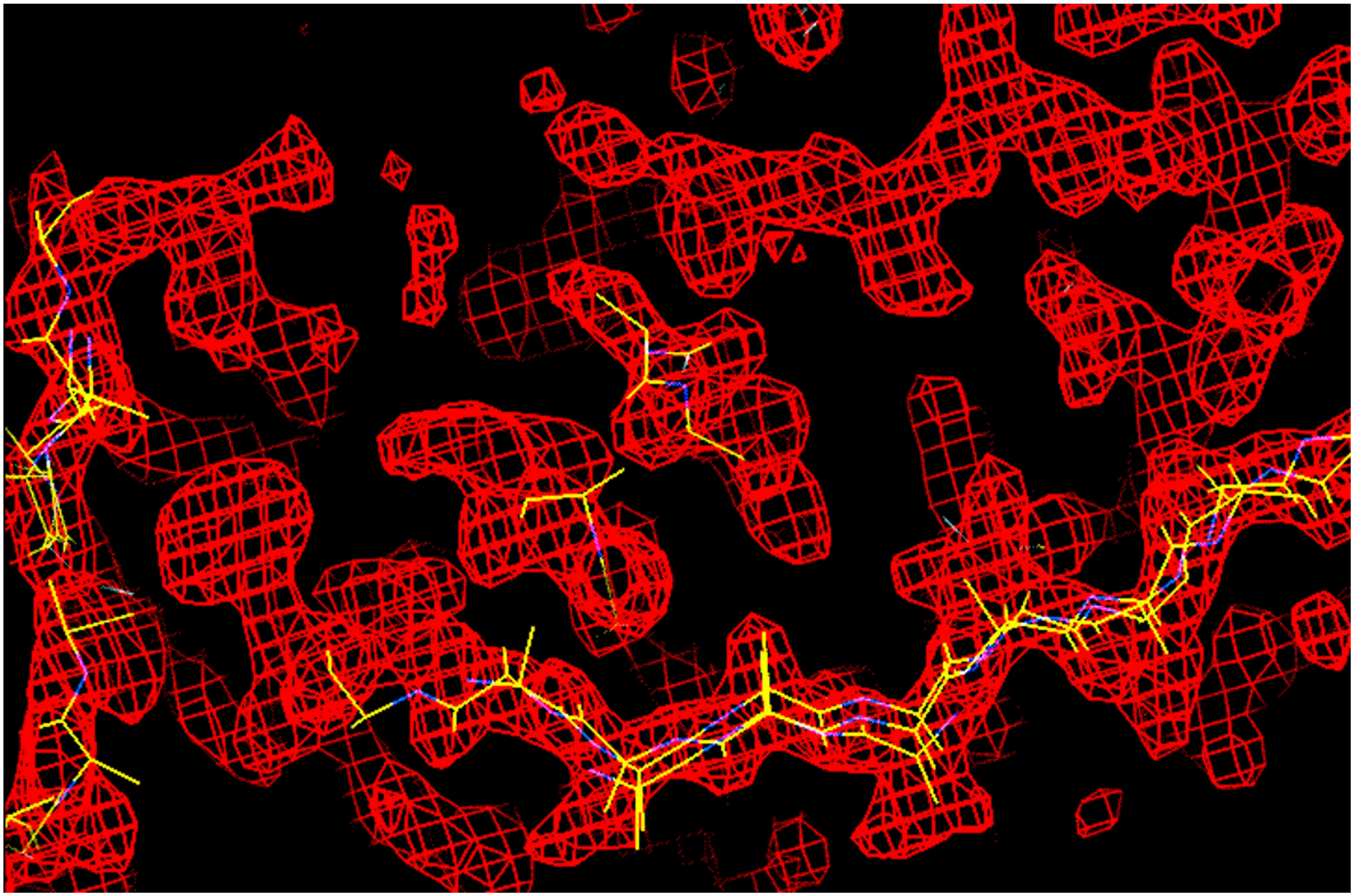
Examples



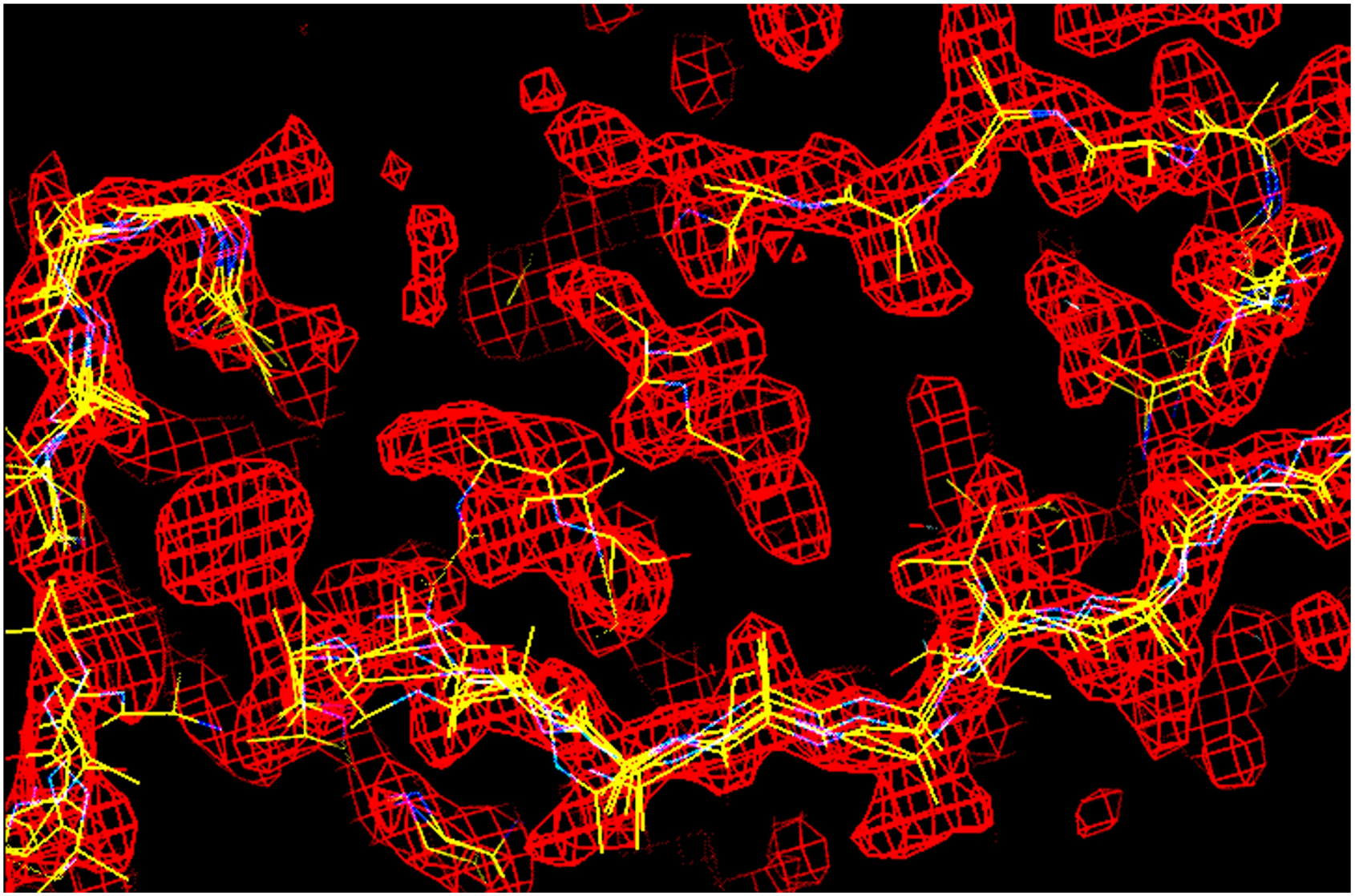


- Shape-based identification of regular secondary structure
- Extension with short fragments from high-resolution structures
- Probabilistic sequence alignment

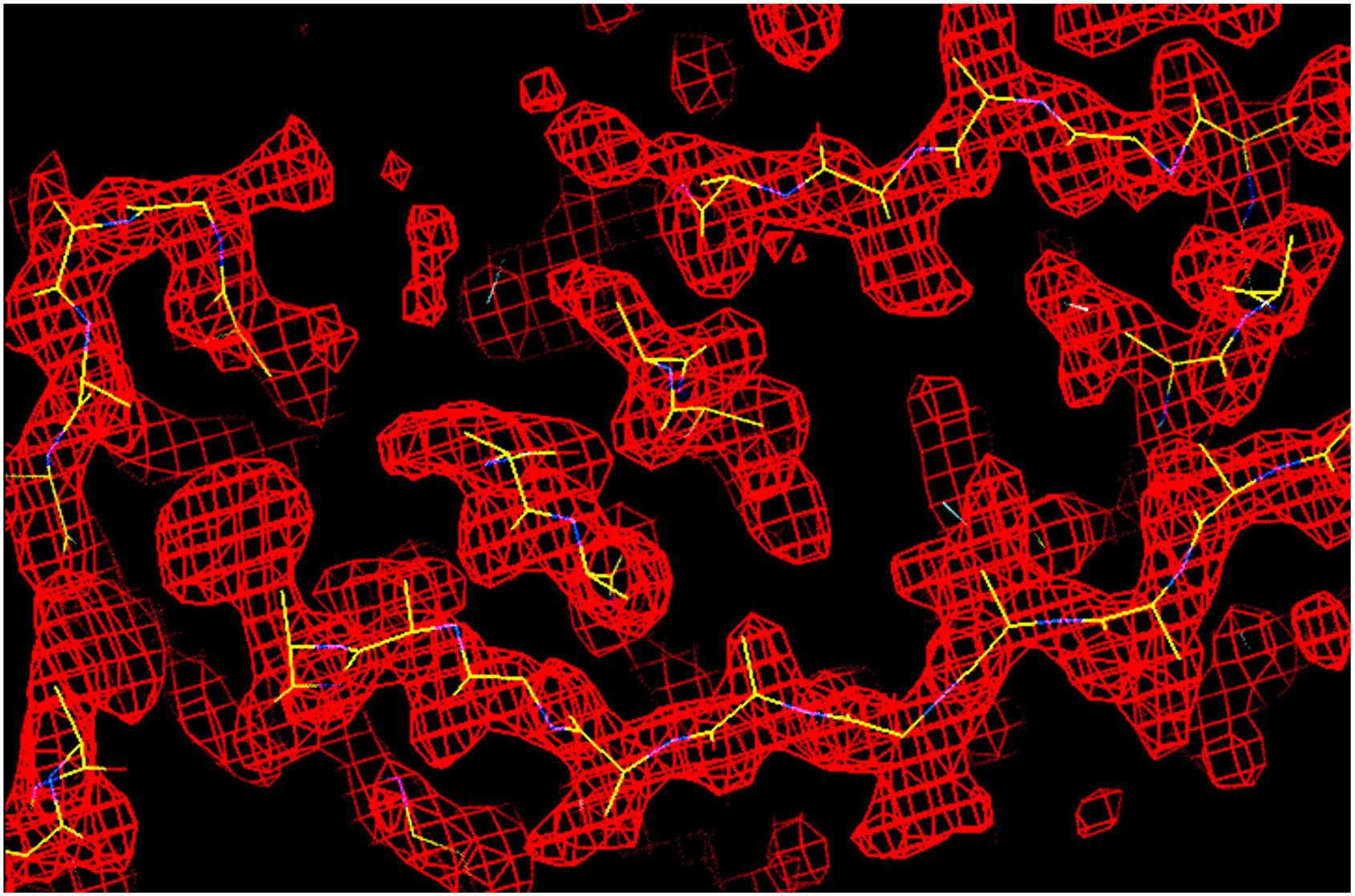
Finding regular protein structure



Extending with short fragments from PDB



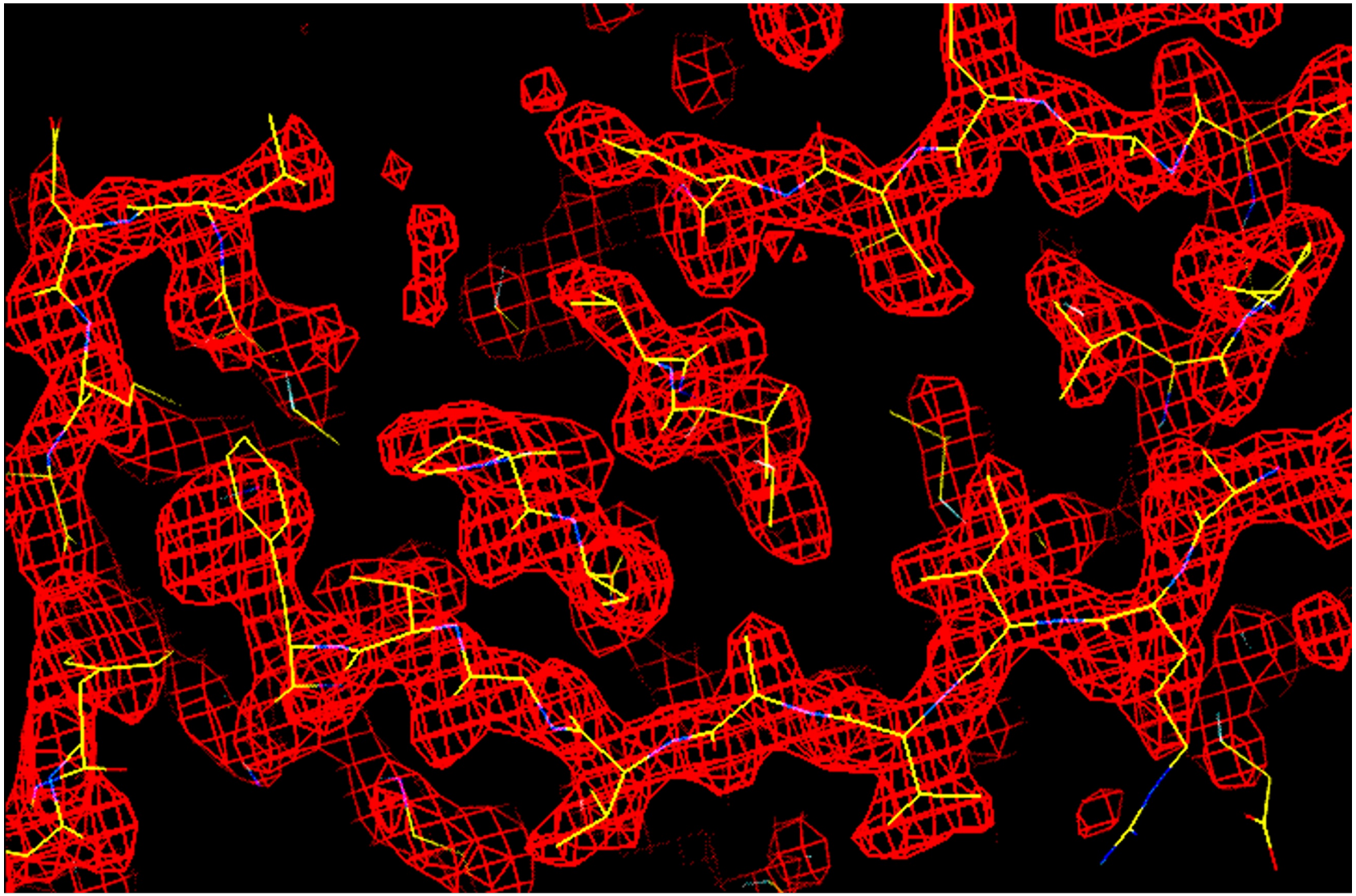
Assembling best model



Identifying residue type at each position

#	G	A	S	V	I	L	M	C	F	Y	K	R	W	H	E	D	Q	N	P	T
1	6	5	4	18	18	6	1	1	1	2	6	2	2	1	9	6	1	0	1	4
2	4	11	14	37	5	2	0	2	0	0	2	3	0	0	1	2	0	0	0	6
3	11	23	5	12	5	3	2	0	1	3	7	3	1	0	5	3	2	0	2	2
4	7	9	6	16	8	5	2	0	1	3	8	4	1	0	7	6	2	0	3	4
5	31	7	3	7	4	2	1	0	1	3	5	4	1	0	6	2	2	0	11	1
6	1	3	3	41	14	8	0	0	0	0	2	1	0	0	2	4	0	0	1	9
7	0	0	0	0	0	0	0	0	15	63	1	0	17	1	0	0	0	0	0	0
8	2	3	6	23	10	6	2	1	0	1	4	3	0	0	5	16	1	0	1	6
9	96	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Inserting side chains based on sequence



Automated structure solution

`phenix.autosol`

Experimental data, sequence, anomalously-scattering atom, wavelength(s)

Find heavy-atom sites with direct methods or likelihood (HYSS)

Calculate phases (Phaser/Solve)

Improve phases, find NCS, build model (phase_and_build)

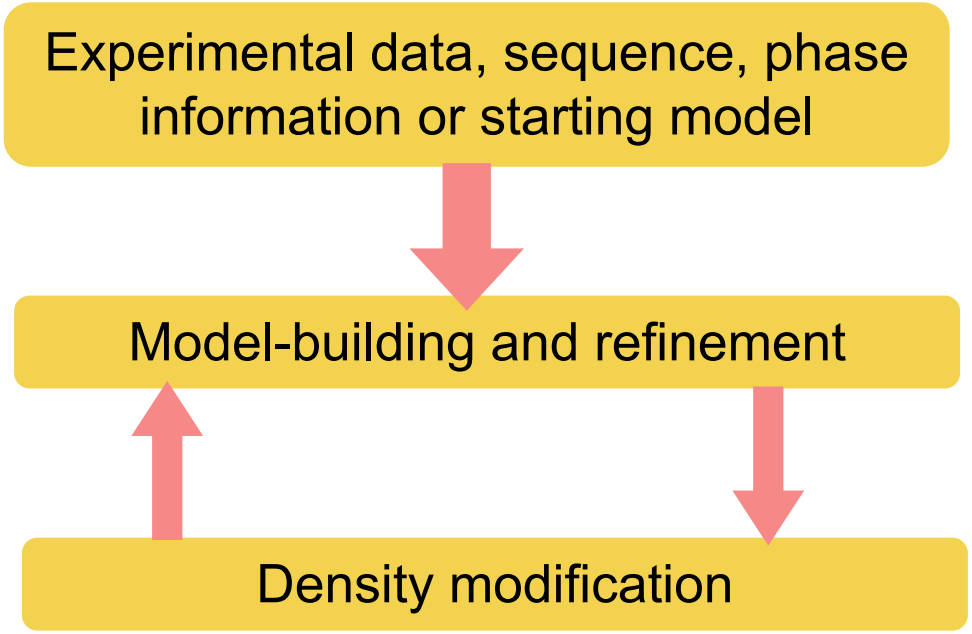
Decision to be made:

Multiple solutions, different derivatives or wavelengths

Alternative hands of space-group and substructure

Iterative map and model improvement

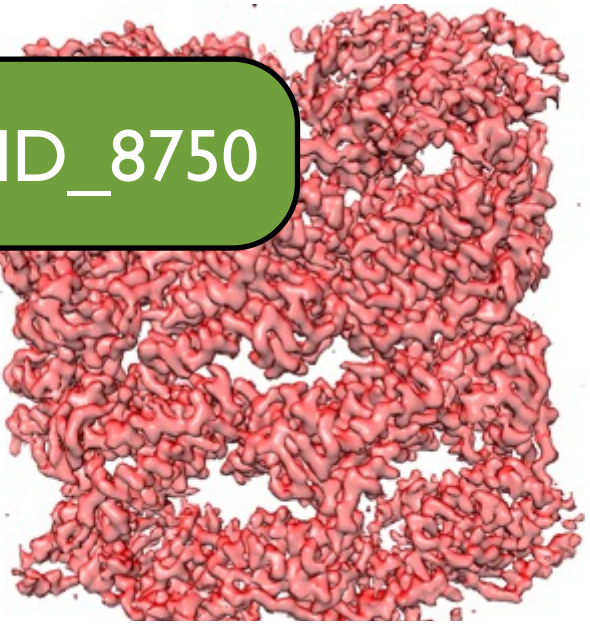
`phenix.autobuild`

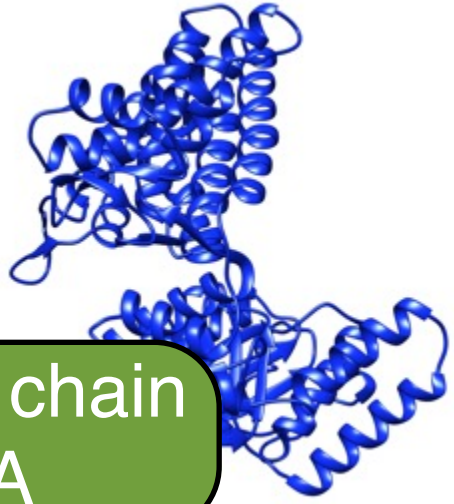


- Resolve building
- Secondary-structure only
- Connect chains
- Fit loops
- Build outside model

Cryo-EM: Docking models

EMD_8750





1ss8 chain A

Search procedure:

Pure translation

- low-res
- high-res

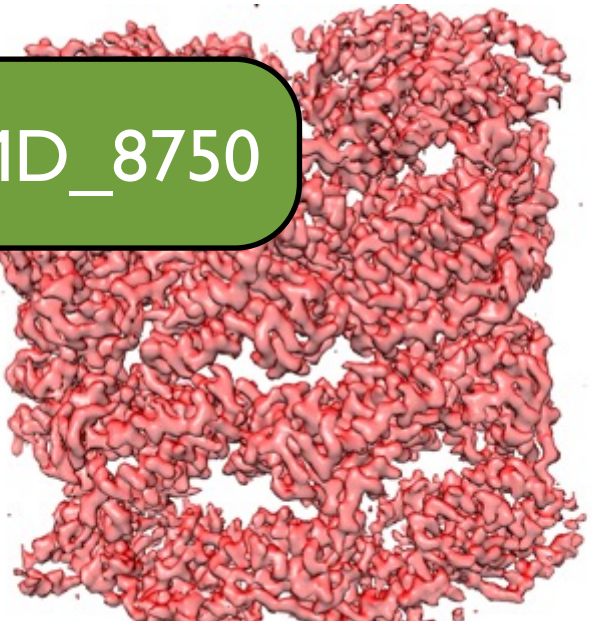
Rotation / translation

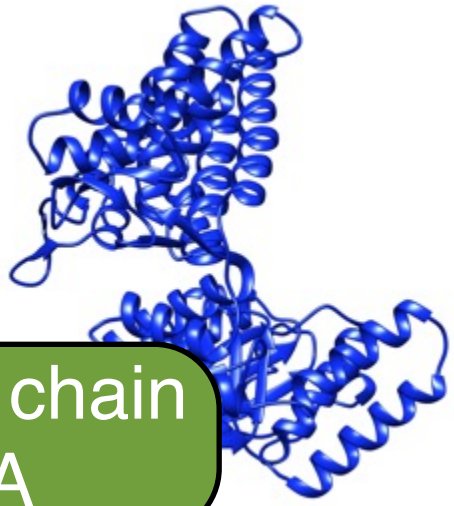
- low-res
- high-res

Score based on rigid-body refinement map-model correlation

Cryo-EM: Docking models

EMD_8750





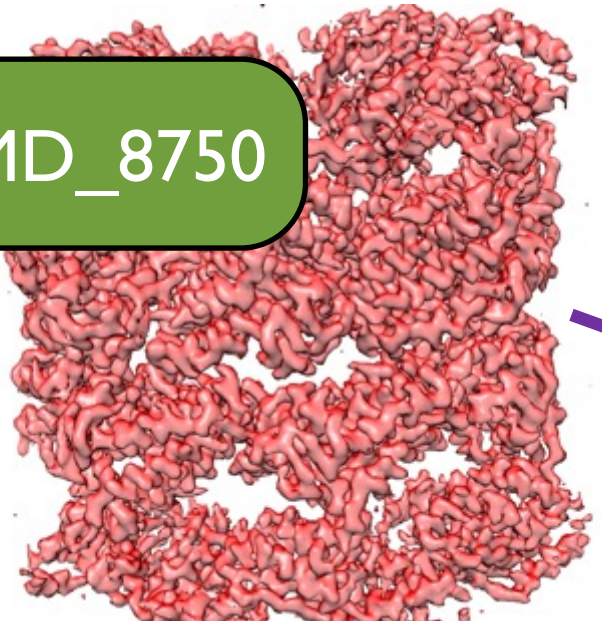
1ss8 chain A

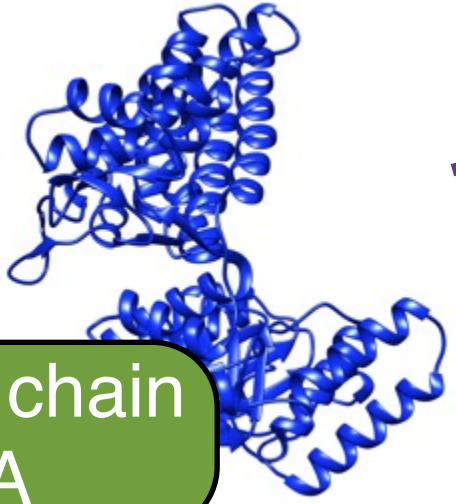
Features

- Multiple chains
- Density search
- Symmetry
- Multiprocessing

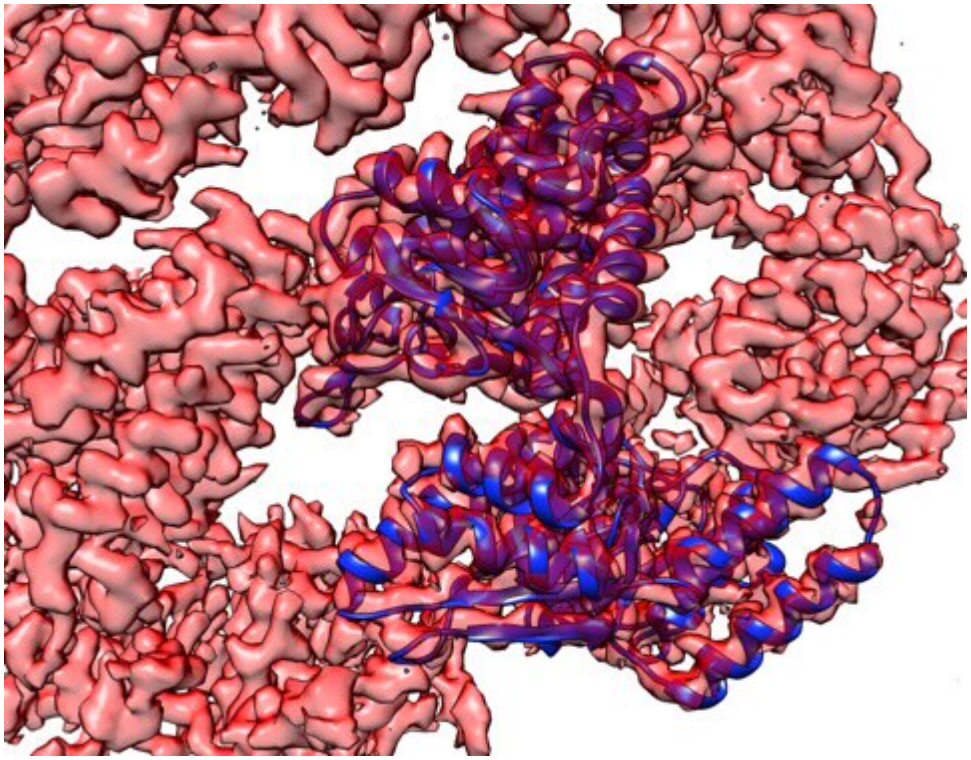
Cryo-EM: Docking models

EMD_8750





1ss8 chain A



1ss8 chain A docked in map

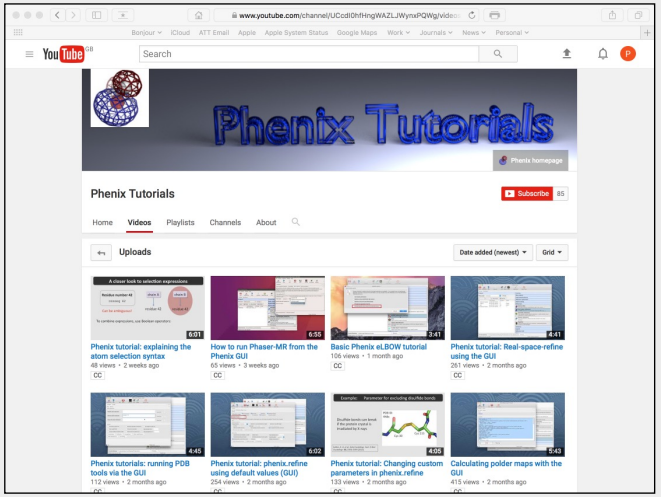
Resources

phenix-online.org

Phenix documentation

Tutorials with sample data

Video tutorials



The Phenix Project

Lawrence Berkeley Laboratory

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

University of Cambridge

Randy Read, Airlie McCoy, Rob Oeffner

Los Alamos National Laboratory
New Mexico Consortium

Tom Terwilliger, Li-Wei Hung

UTHealth

Matt Baker

Duke University

Jane Richardson, Vincent Chen, Michael Prisant, Christopher Williams,

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