Phenix

Model-building using cryo-EM and crystallographic maps:

Methods and techniques in integrated structural biology:
Beyond Black Boxes

Strasbourg, October 7, 2016

Tom Terwilliger, Li-Wei Hung
Los Alamos National Laboratory
Pavel Afonine, Oleg Sobolev, Paul Adams
Lawrence Berkeley National Laboratory

Outline

Are X-ray and cryo-EM maps the same?

Optimal sharpening of a map

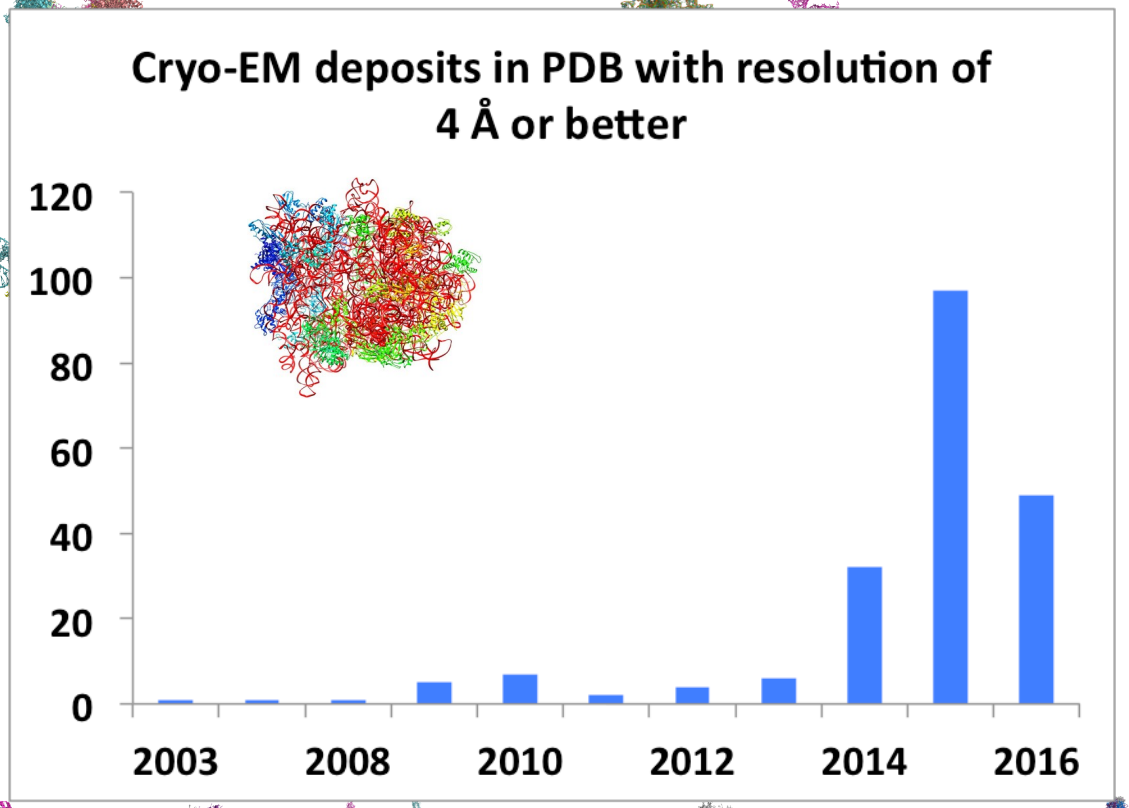
Finding the unique part of a cryo-EM map

Model improvement by iterative secondary-structure assignment and real-space refinement

Automated interpretation of cryo-EM maps

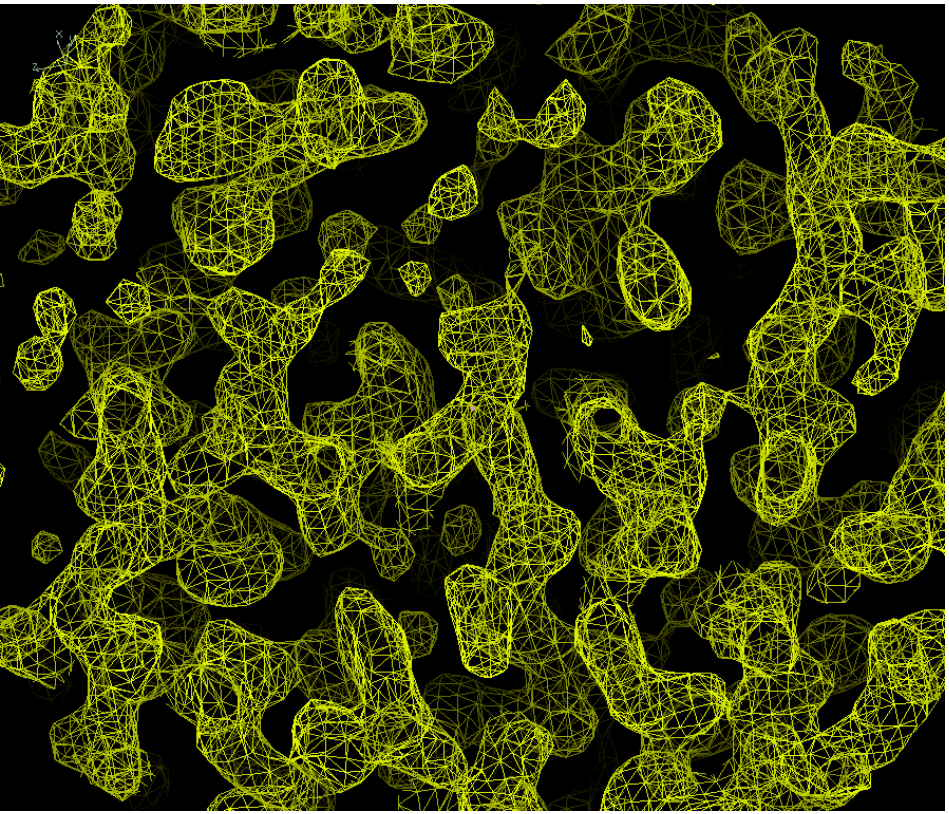
1147 Cryo-EM structures in PDB

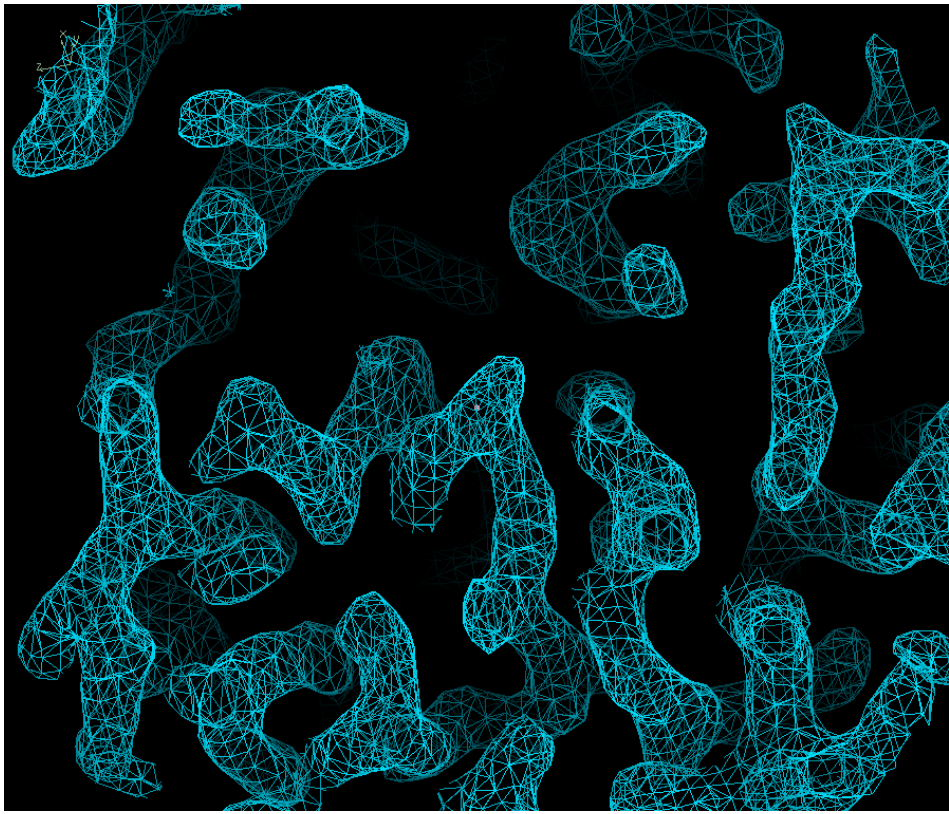
4 Å or better: 220, 3.5 Å or better: 92



X-ray vs cryo-EM

Beta galactosidase at 2.2 Å

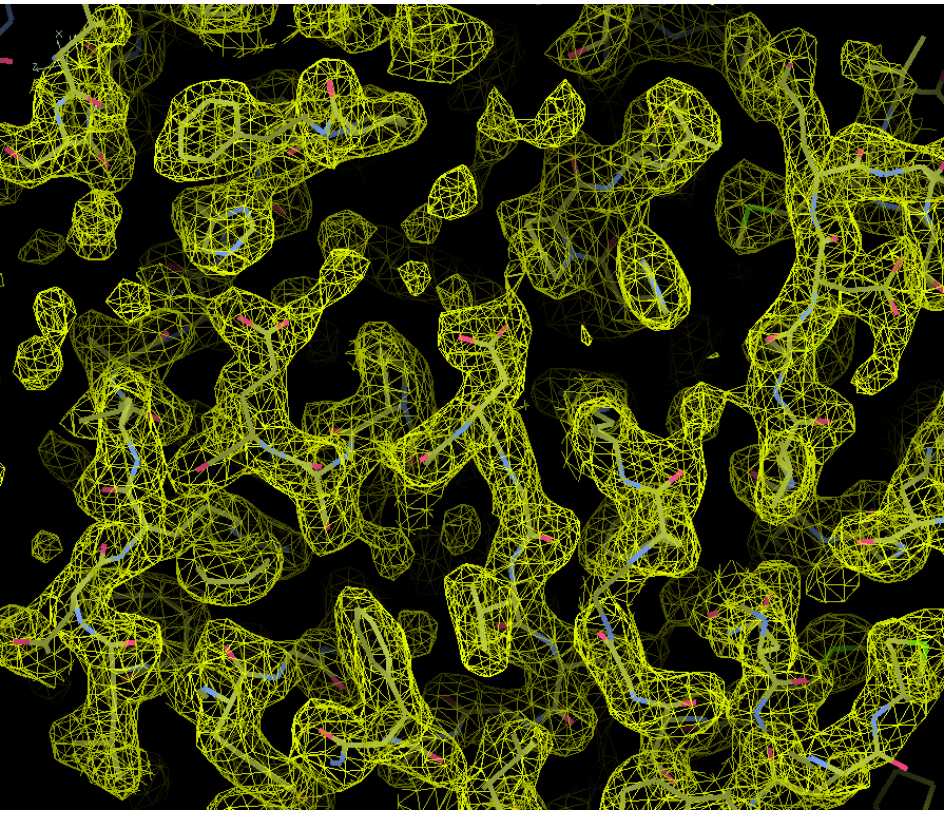


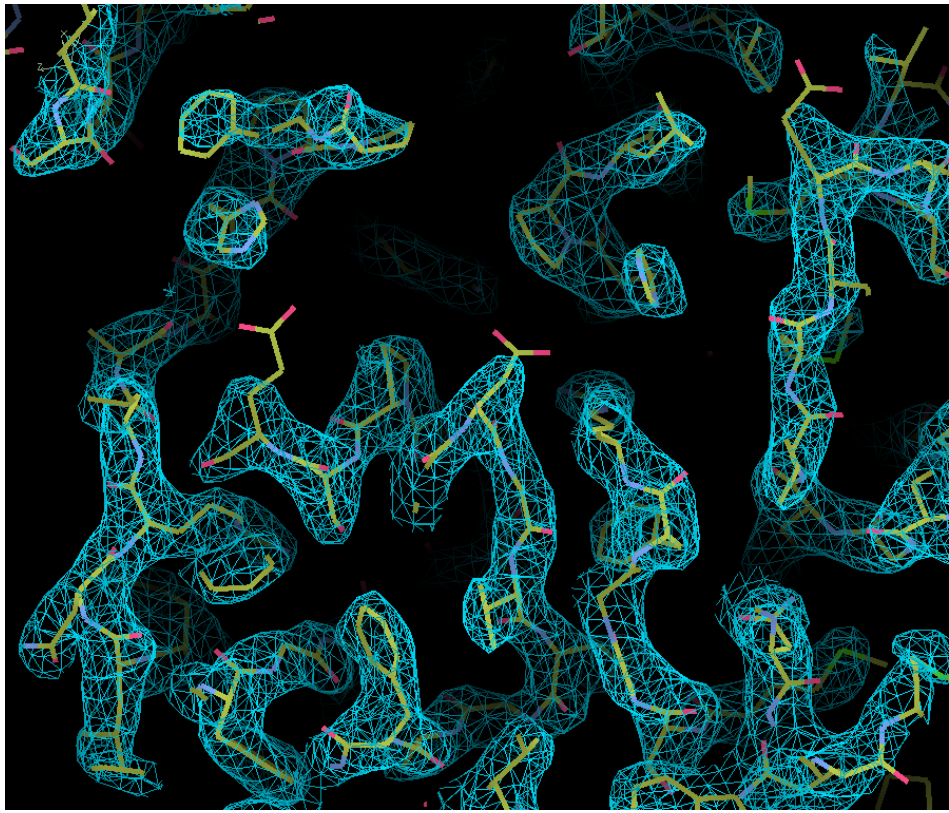


(which is the cryo-EM map?)

X-ray vs cryo-EM

Beta galactosidase at 2.2 Å

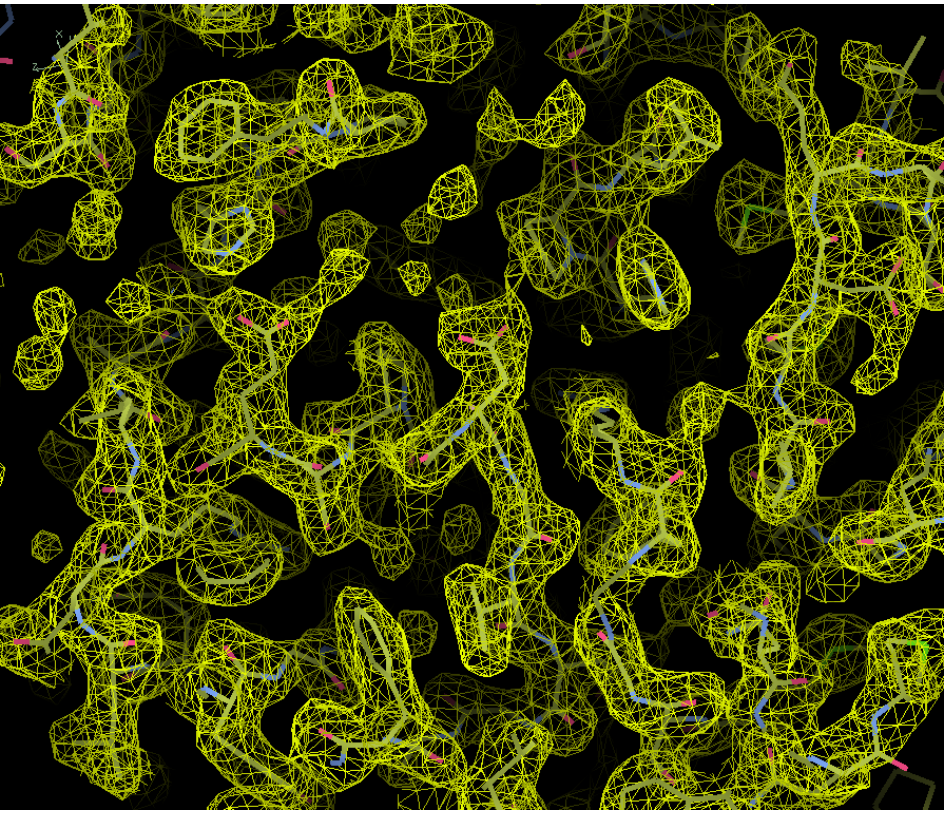




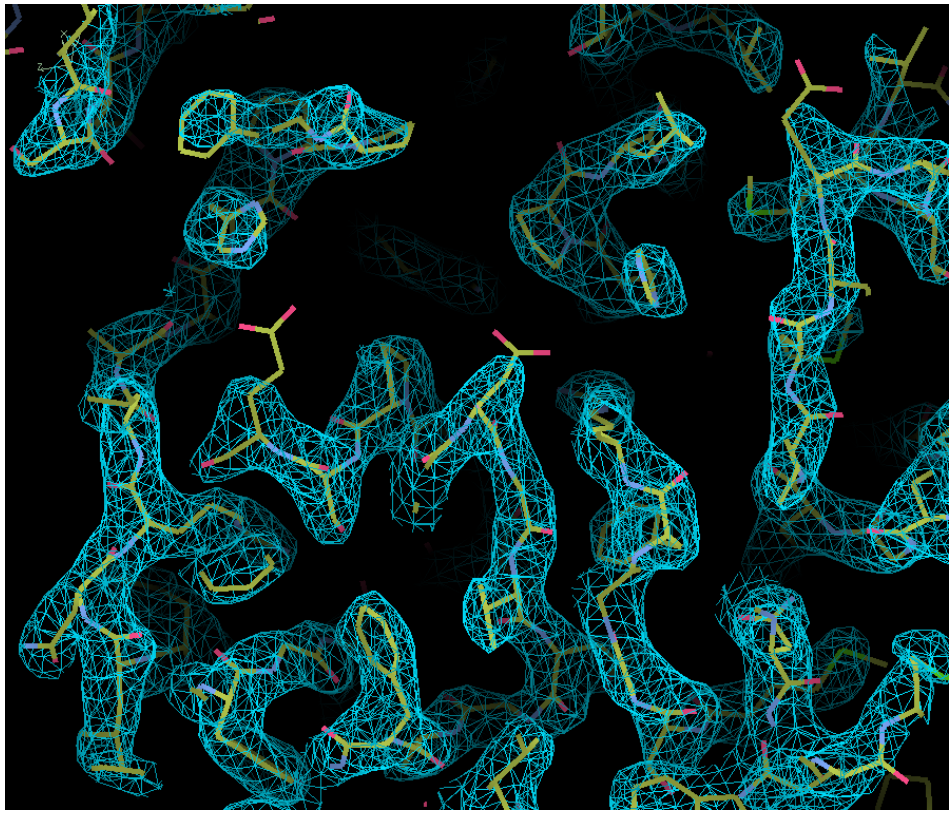
(which is the cryo-EM map?)

X-ray vs cryo-EM

Beta galactosidase at 2.2 Å

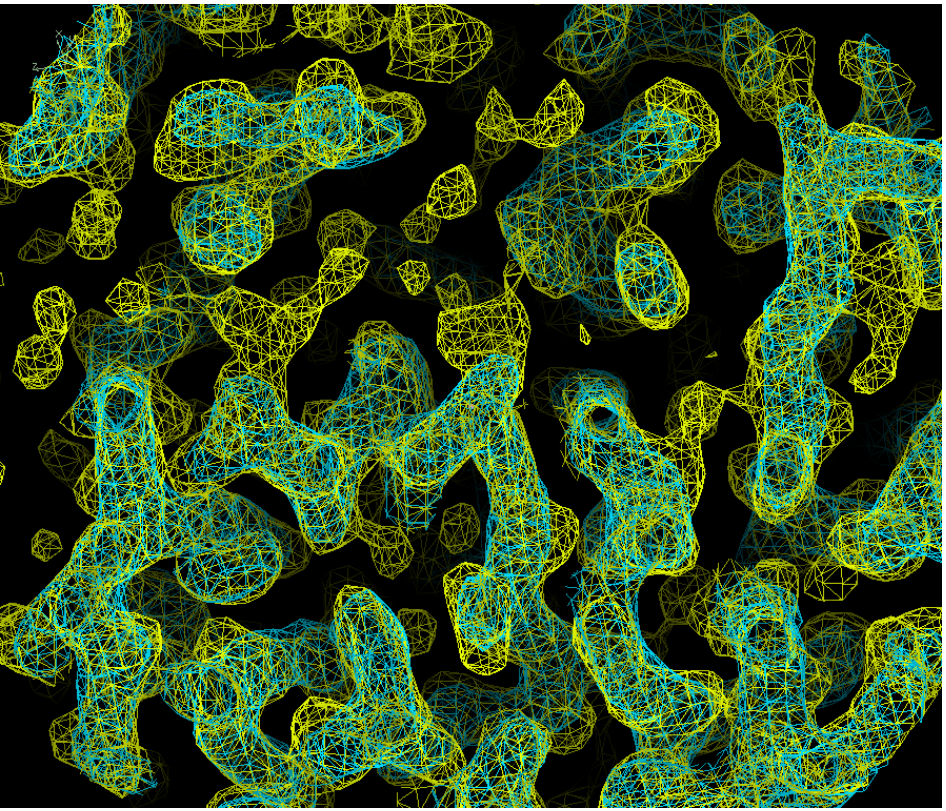


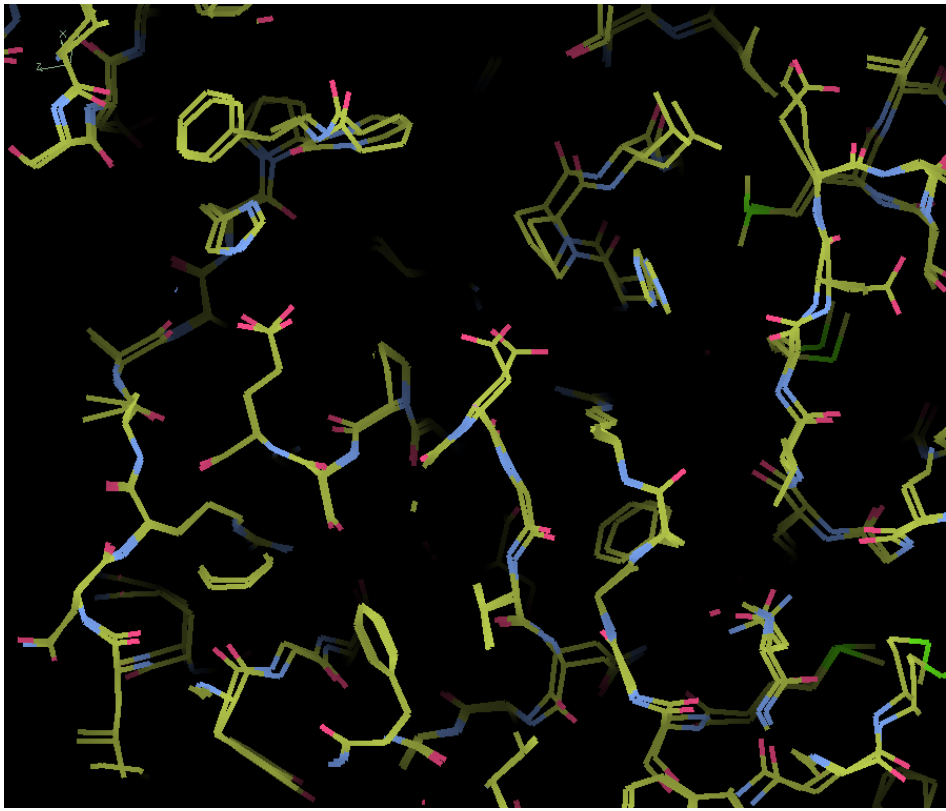
X-ray (PDB 3i3b)



Cryo-EM (PDB 5a1a)

X-ray and cryo-EM maps can be very similar…

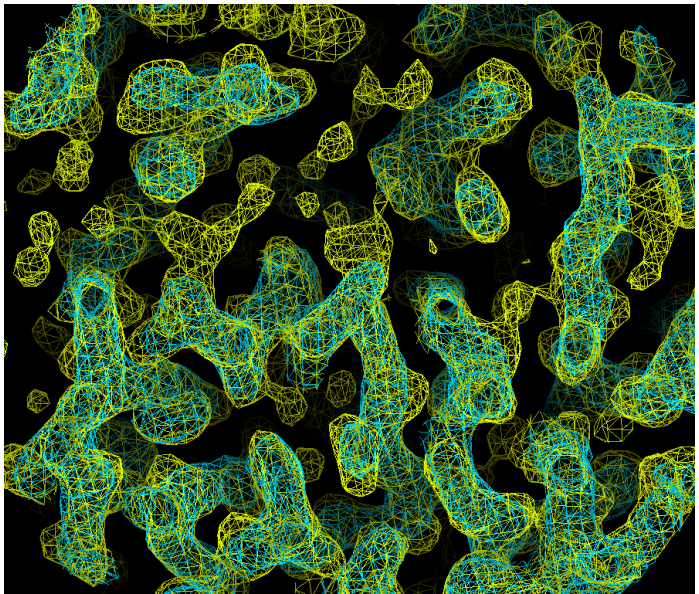


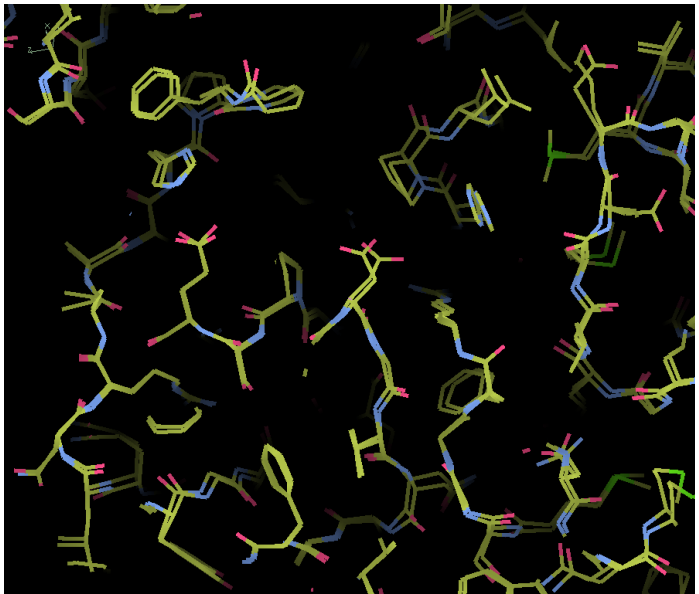


…but have different strengths

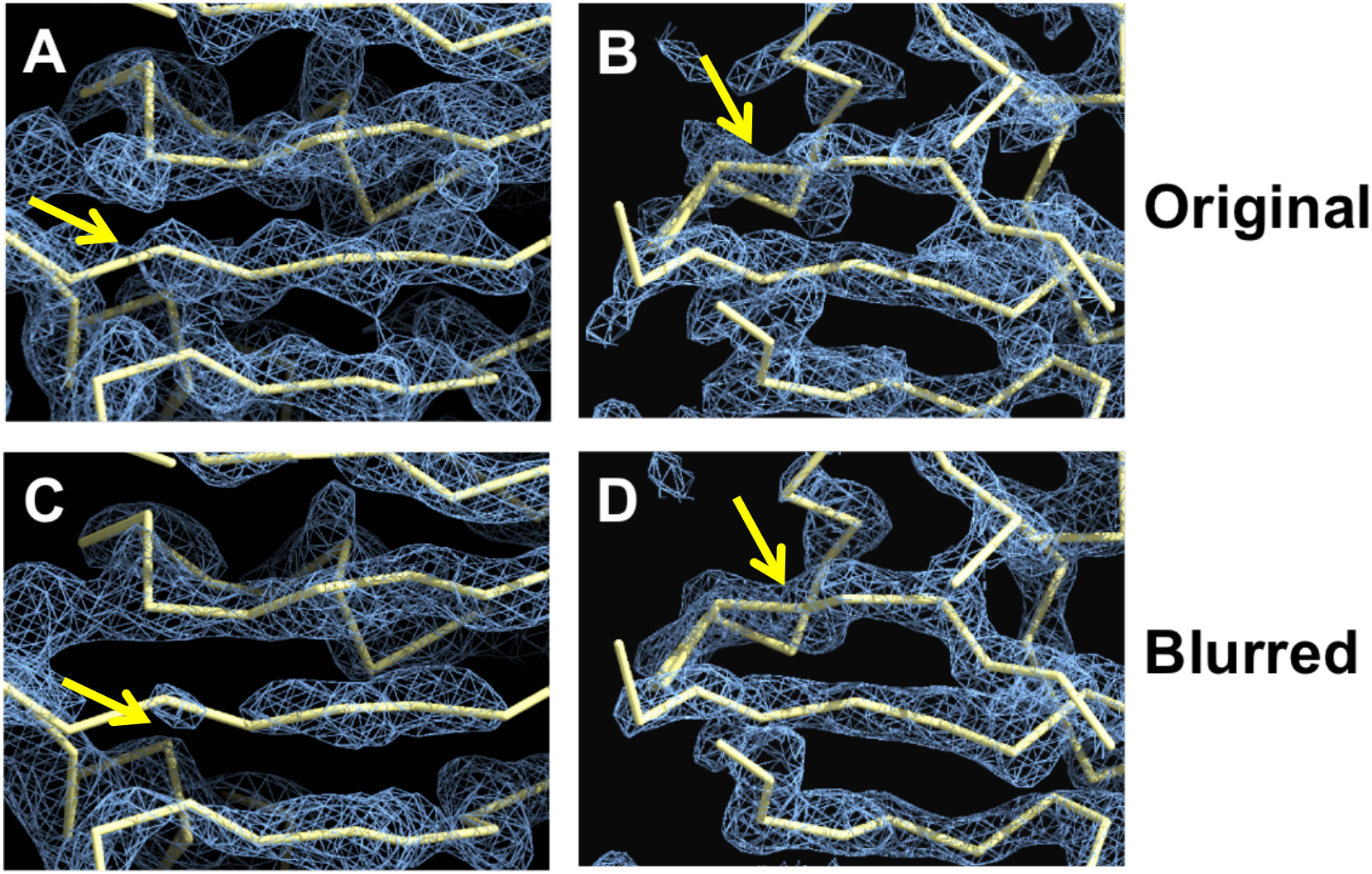
X-ray maps be improved by density modification—cryoEM maps are what you get

Cryo-EM maps may have more accurate low-resolution information

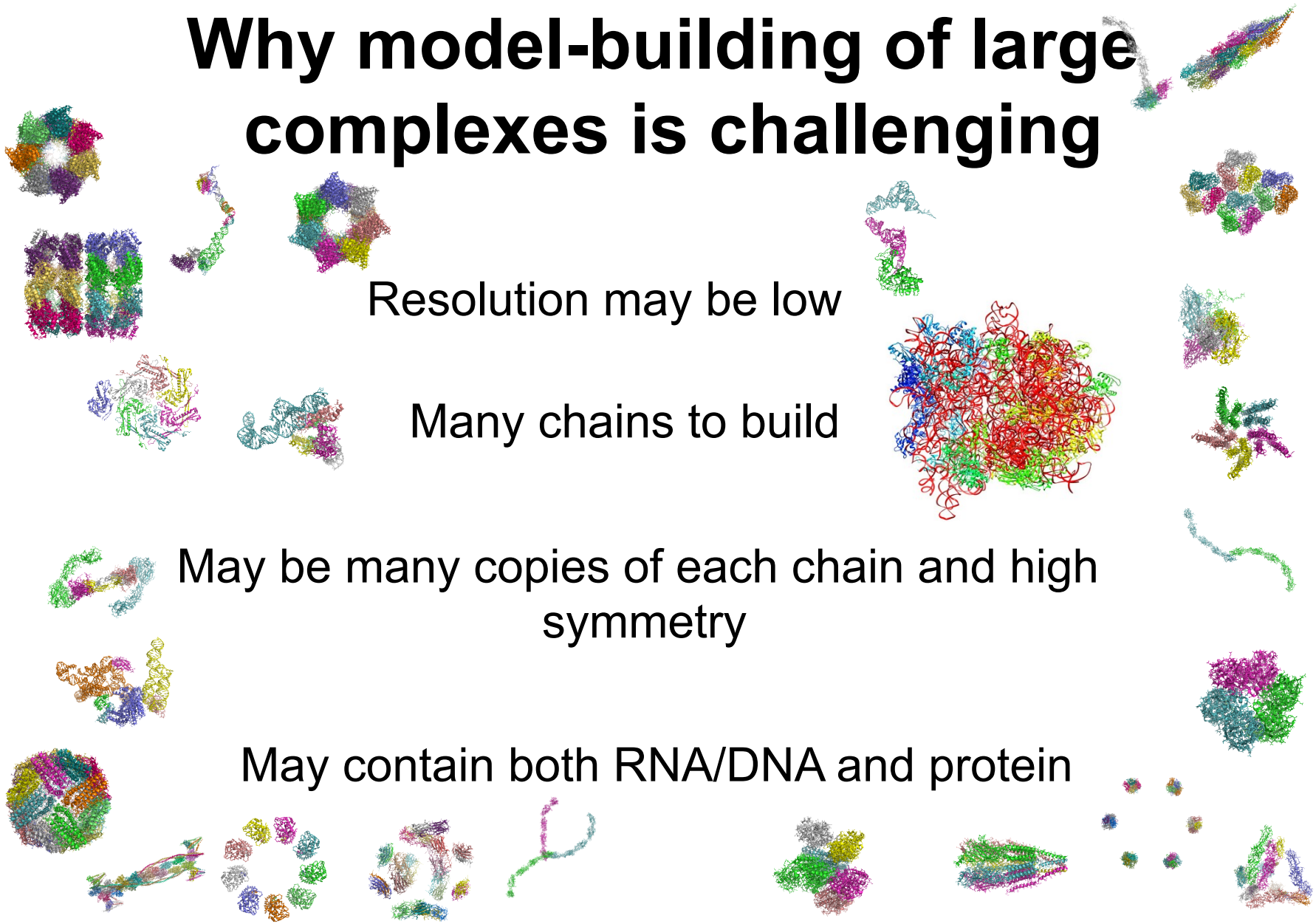




More accurate low-resolution information in cryo-EM



Why model-building of large complexes is challenging



Resolution may be low

Many chains to build

May be many copies of each chain and high symmetry

May contain both RNA/DNA and protein

Additional challenges for cryo-EM maps

What is the magnification of the map?
(as much as 10% uncertainty in scale factor)

What is the optimal sharpening of the map?
(X-ray maps too)

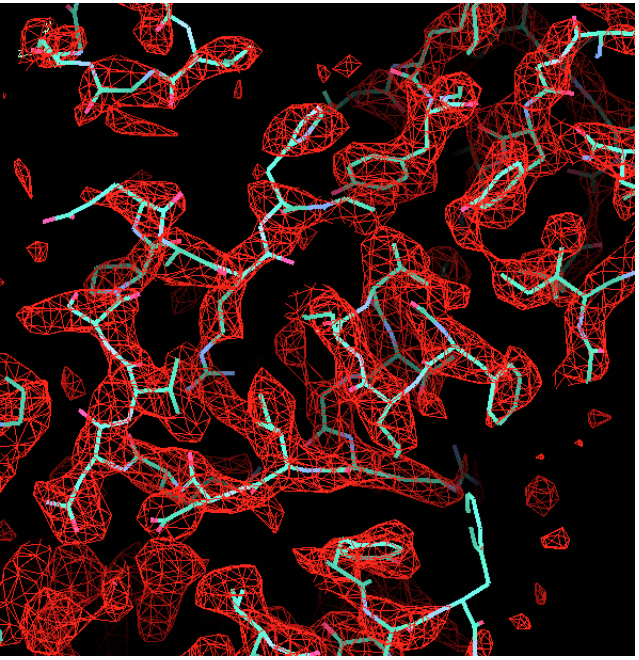
What is the region containing the molecule?

Anthrax toxin protective antigen pore at 2.9 Å

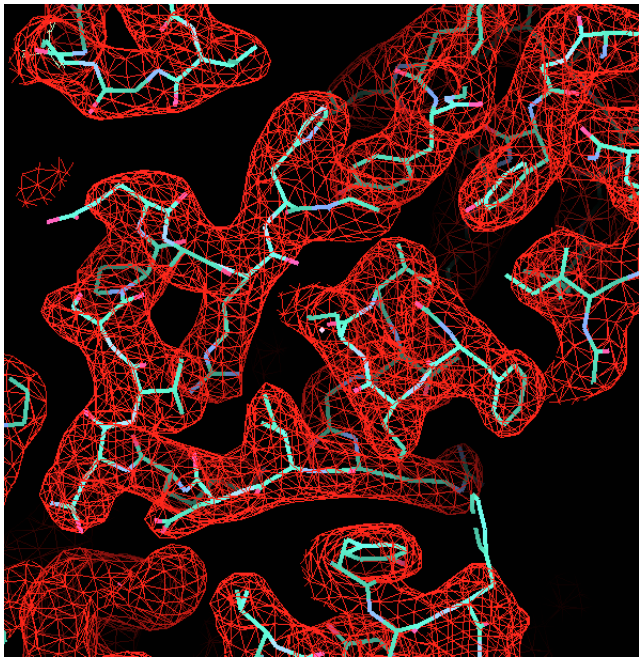
7-fold symmetry

Jiang et al., 2015

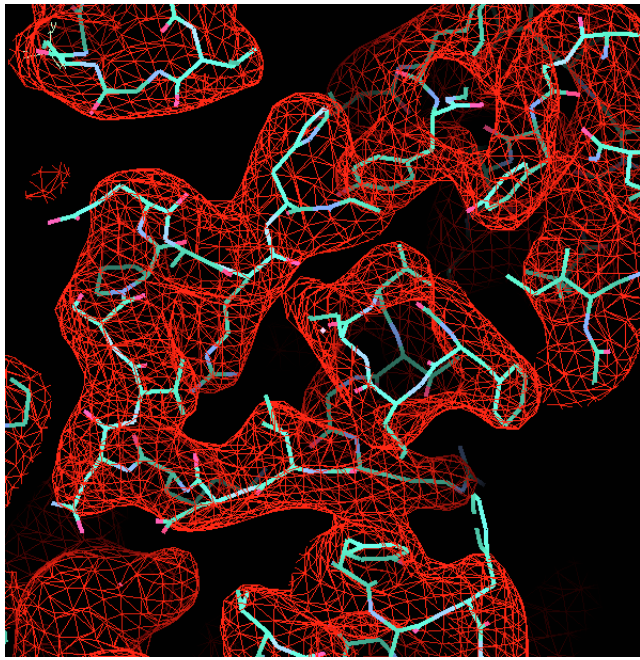
Automatic map sharpening



B_iso =-100
(density broken)

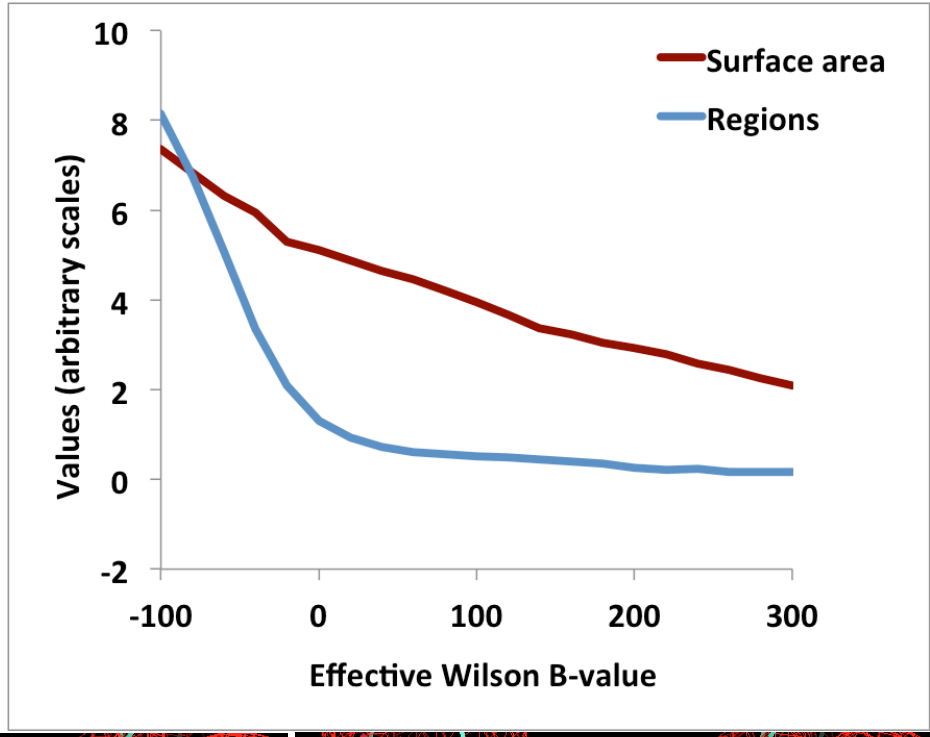


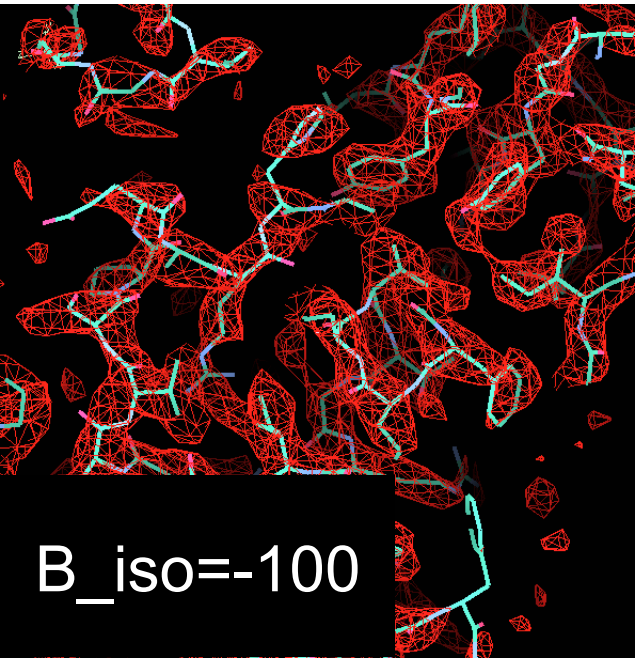
B_iso =60
(clear density)

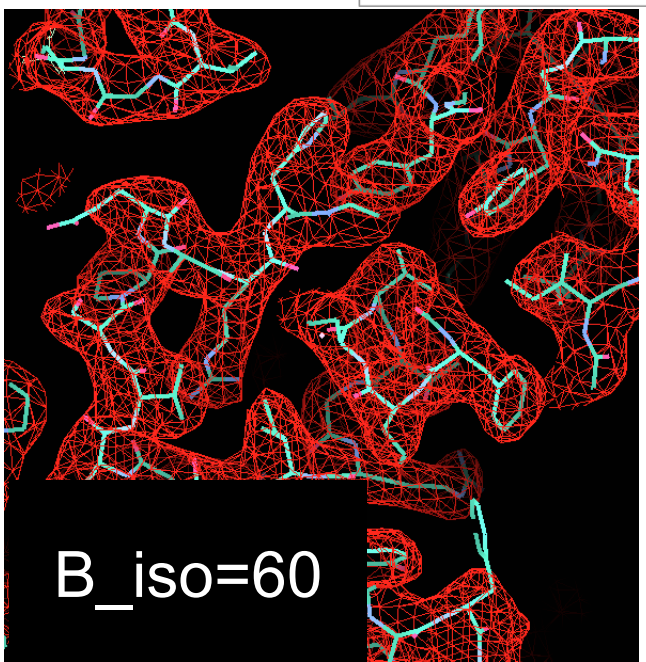


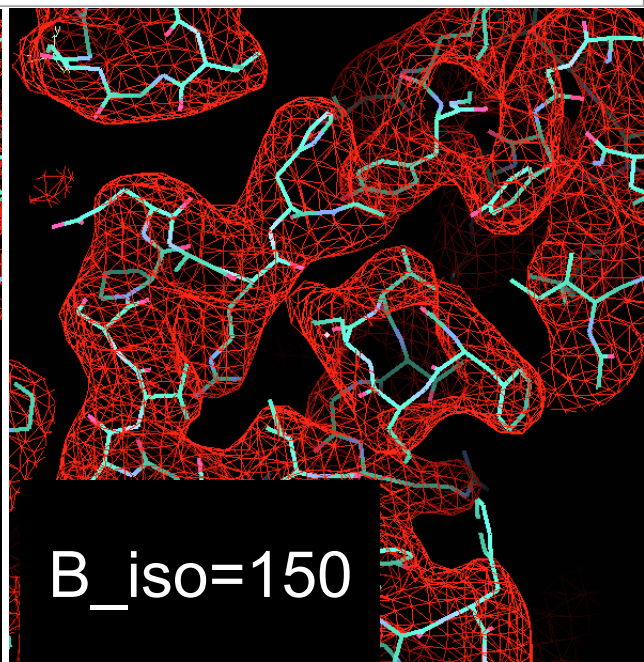
B_iso =150
(blurred density)

Sharpening based on contiguous regions and surface area

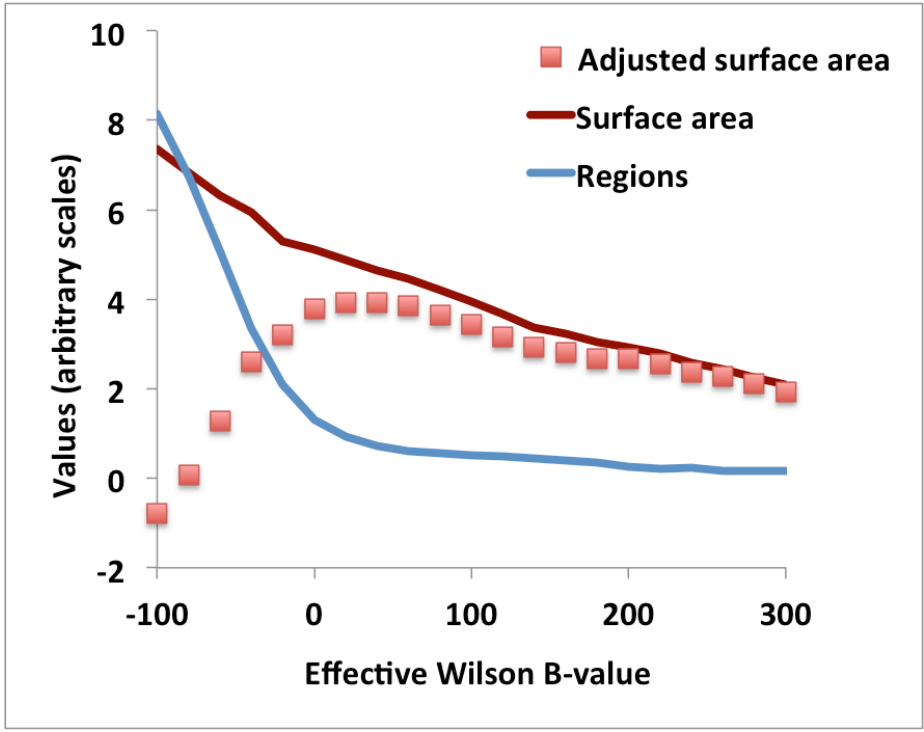


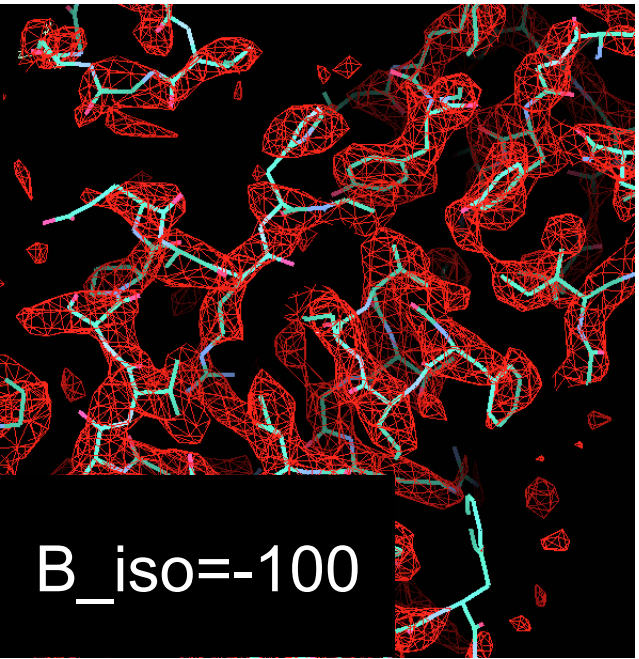


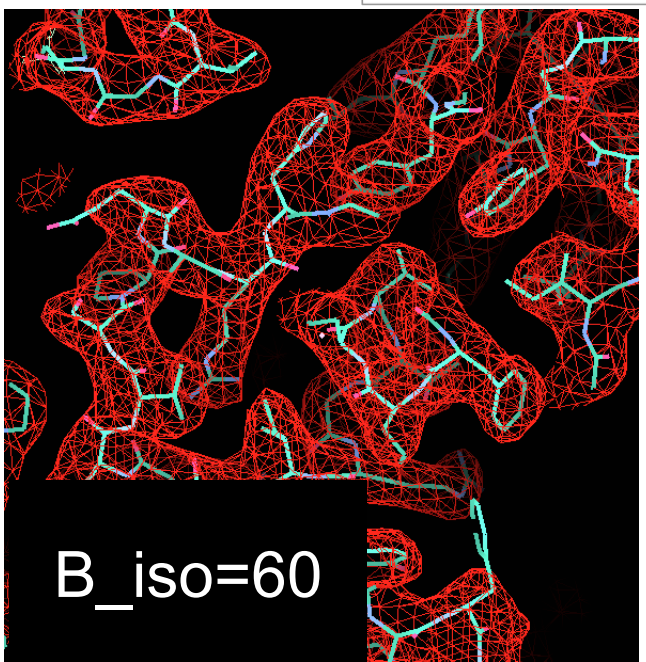


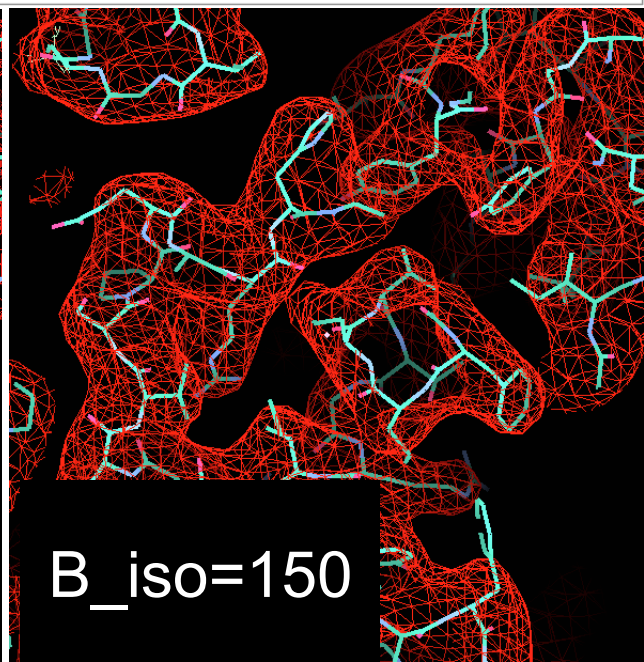


Adjusted surface area: surface area – weight * number of regions

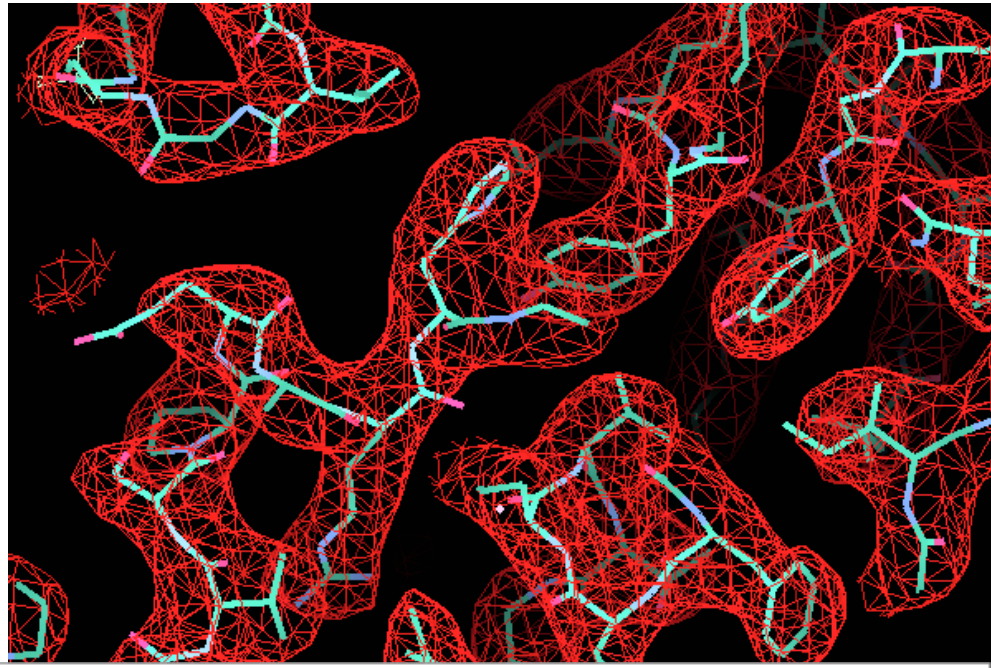




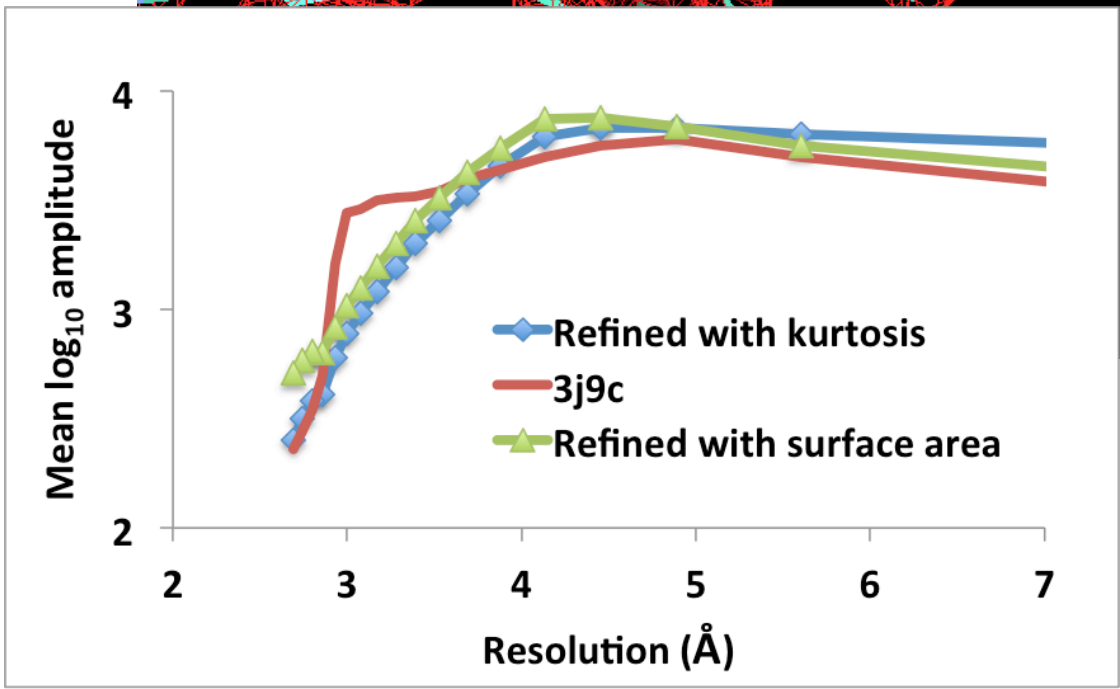




Adjusted surface area can be used to refine resolution-dependent normalization of map coefficients

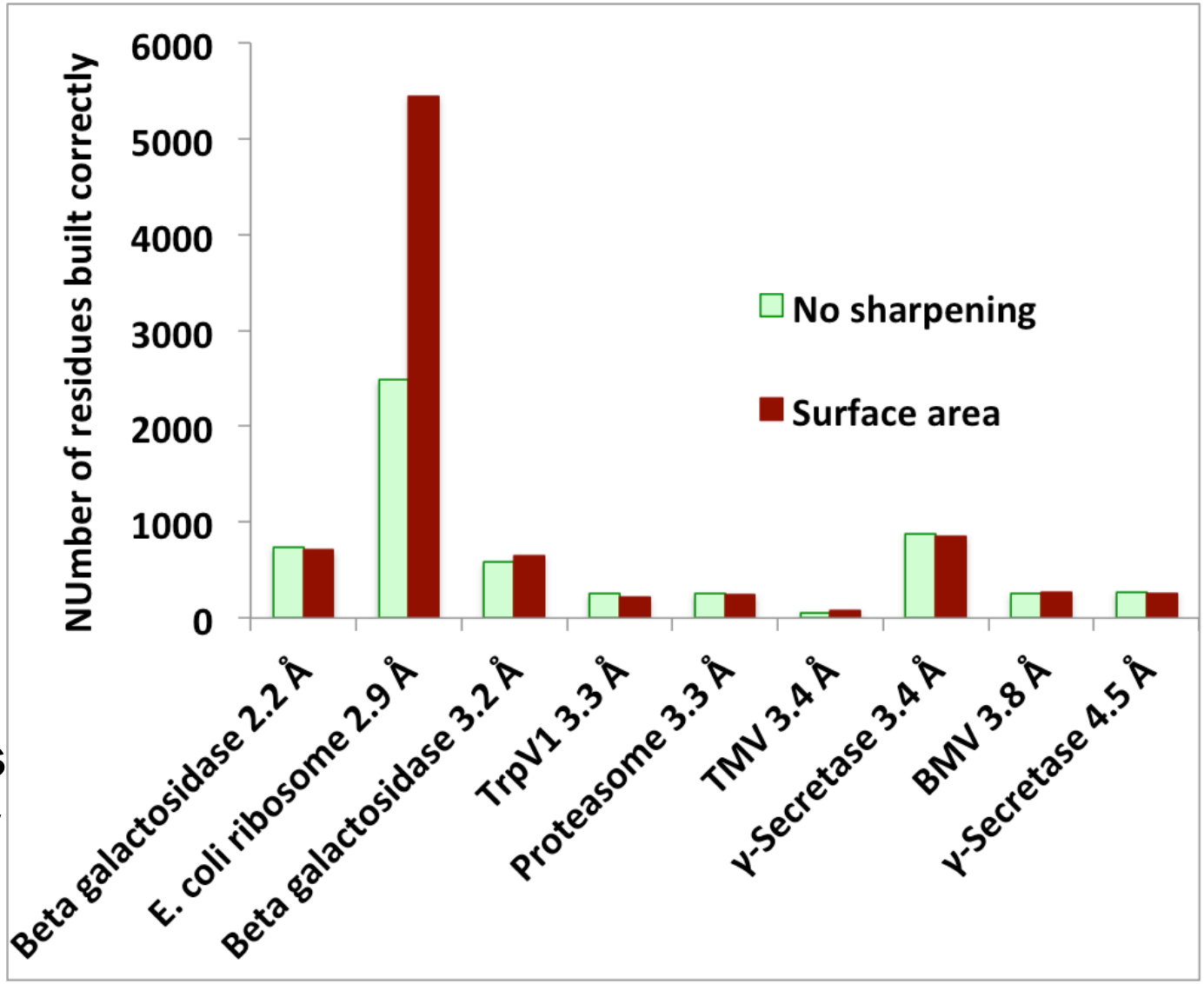


- Amplitudes normalized (B-iso=0)
- 3-parameter resolution-dependent weights applied to normalized amplitudes
- Log(<F>) varies linearly with $\sin^2\theta/\lambda^2$ in 3 ranges of resolution



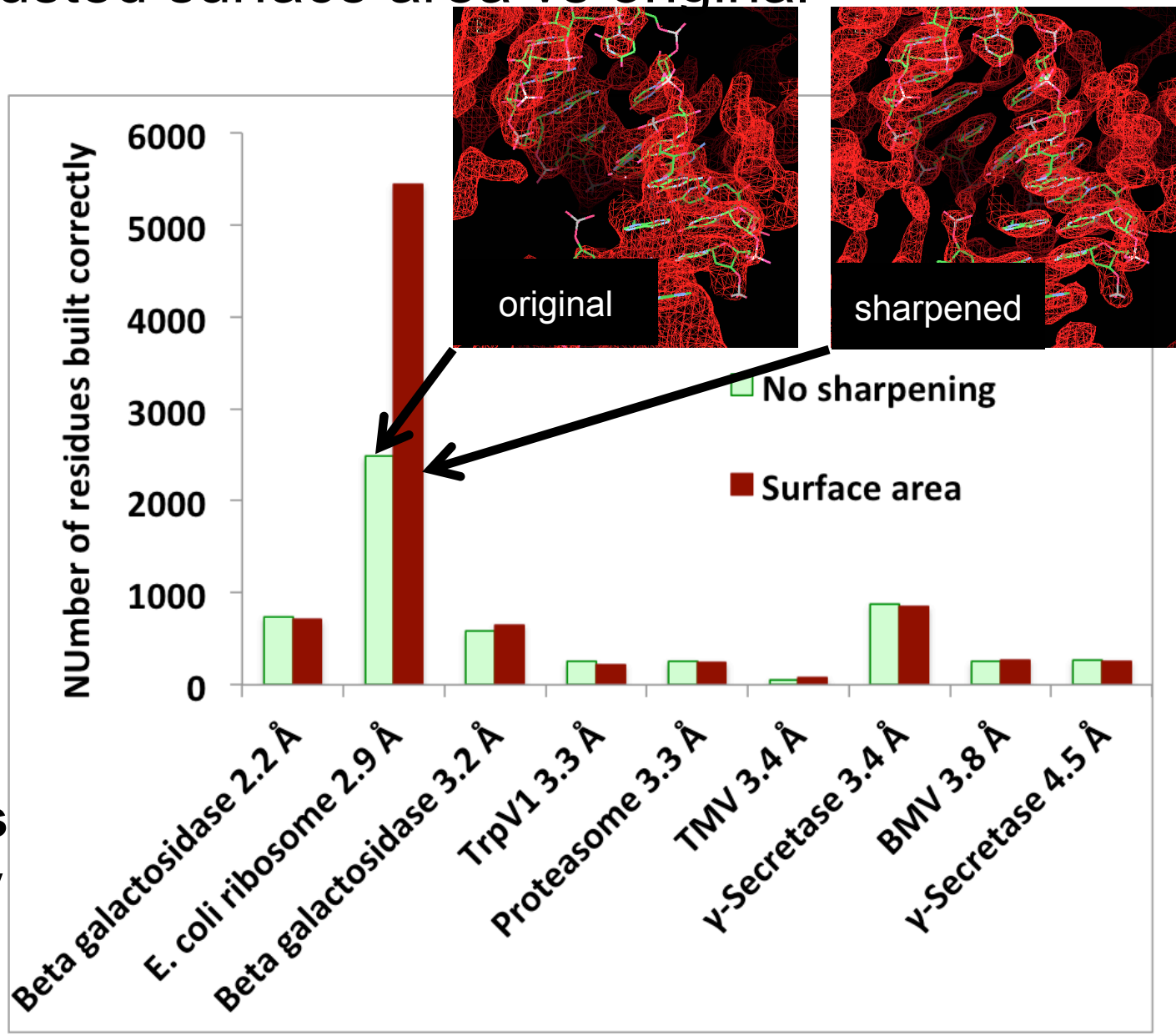
Map optimization:
Adjusted surface area vs original

- **7 cryo-EM maps**
- **2.2-4.5 Å**
- **Total residues built correctly**



Map optimization:
Adjusted surface area vs original

- 7 cryo-EM maps
- 2.2-4.5 Å
- Total residues built correctly



Automatic map segmentation

Use symmetry of the map

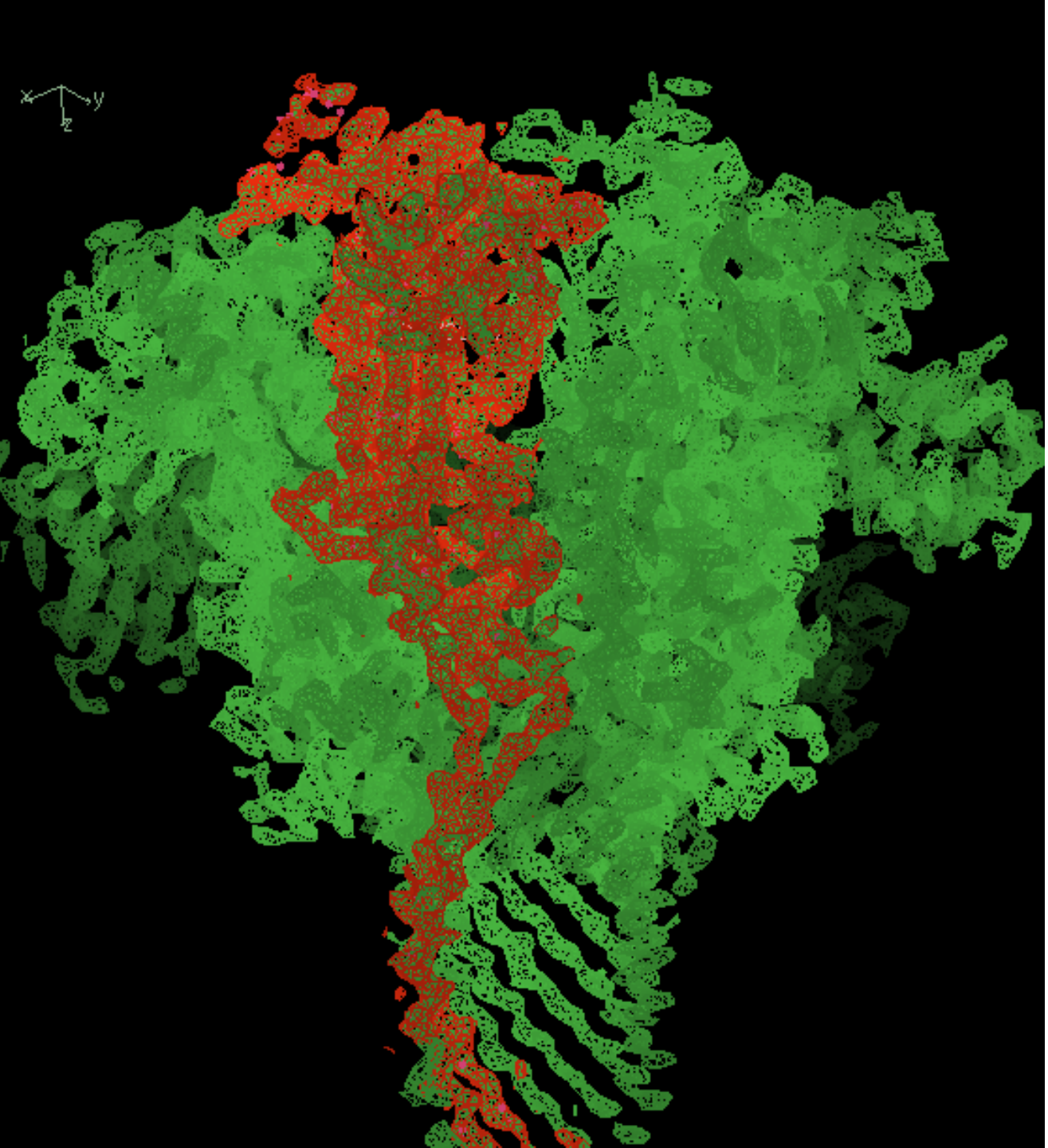
Identify contiguous regions representing asymmetric unit of the map

Choose symmetry-copies that make compact molecule

Anthrax toxin protective antigen pore at 2.9 Å

7-fold symmetry

Jiang et al., 2015



Anthrax toxin protective antigen pore at 2.9 Å

7-fold symmetry

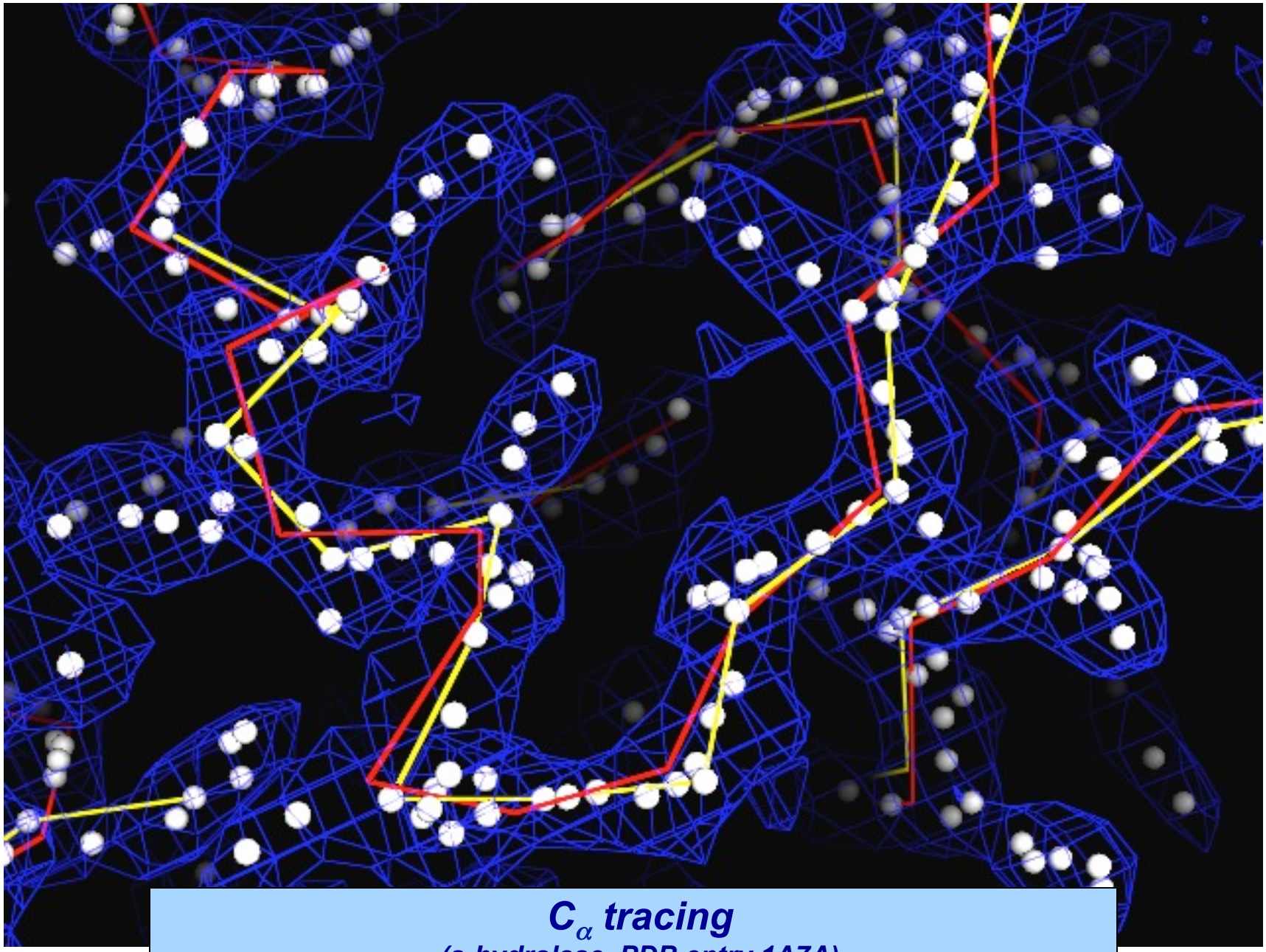
Jiang et al., 2015

Automated interpretation of Low-resolution maps

- Cut out asymmetric unit of the map
- Trace chain and build model
- Idealize secondary structure and refine
- Assemble and refine (protein/RNA/DNA)
- Apply molecular symmetry and re-refine

Low-resolution backbone chain-tracing for proteins

- Variable map sharpening
- Trace protein main chain
- Identify direction of main chain by fit to density



C_α tracing

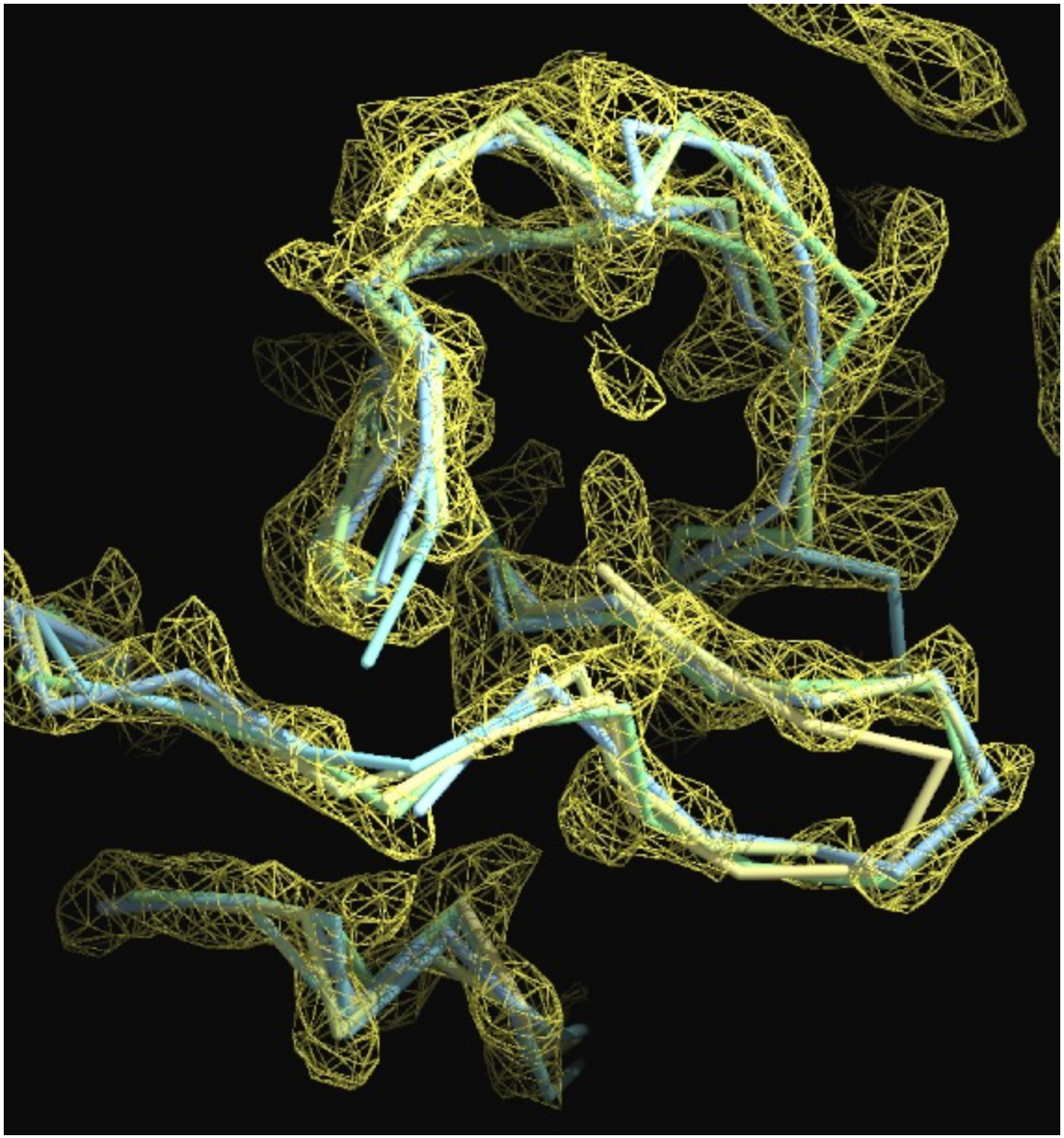
(s-hydrolase, PDB entry 1A7A)

Model improvement by iterative secondary-structure assignment and real-space refinement

- Find the secondary structure (helices/strands)
- Identify idealized atom-atom distances
- Refine including the secondary-structure restraints
- Score based on map correlation and number of suitable H-bonds in models

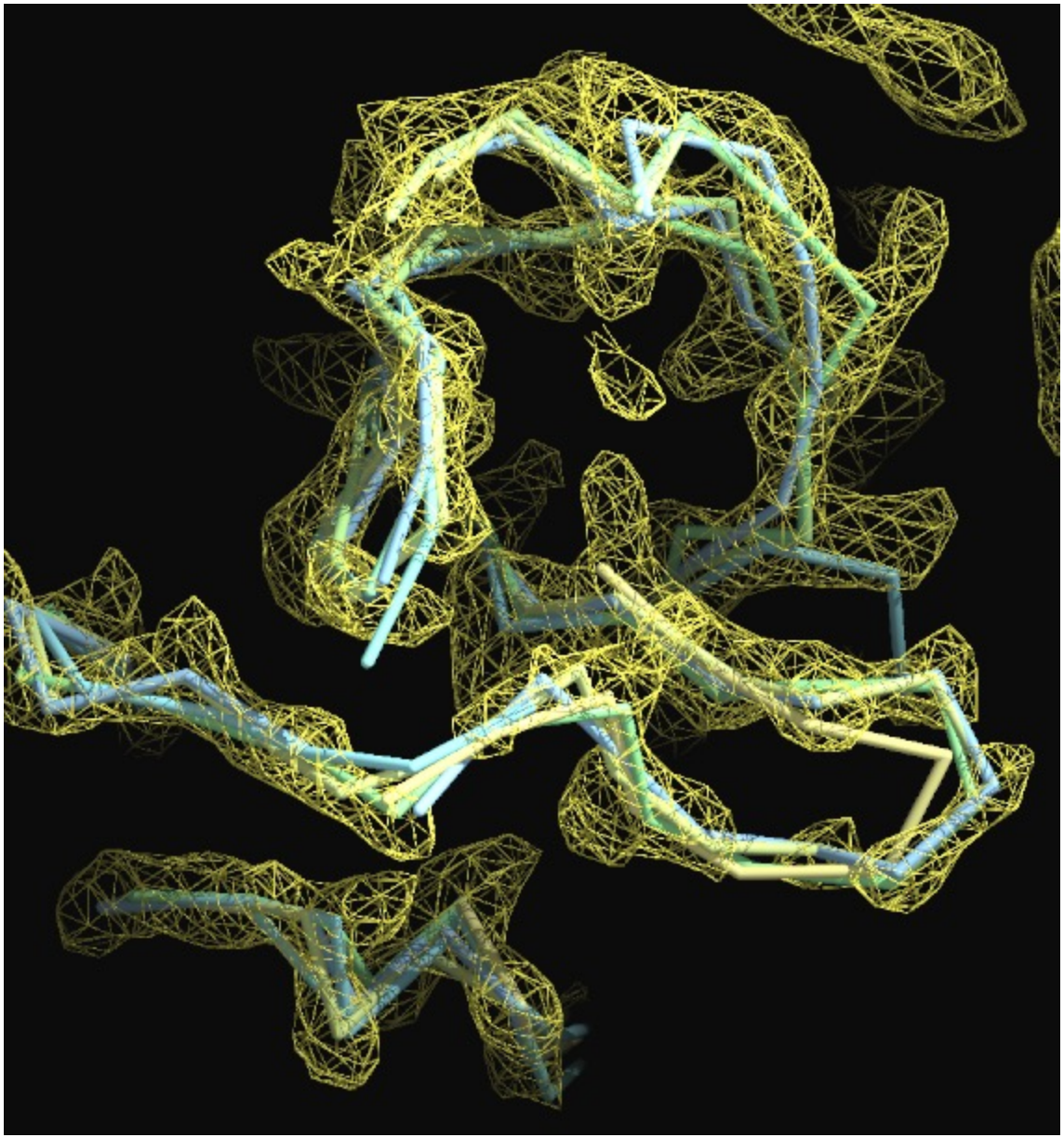
Chain tracings of cryo-EM map

(Chain I, yeast mitochondrial ribosome large subunit, 3.2 Å, 3j6b)



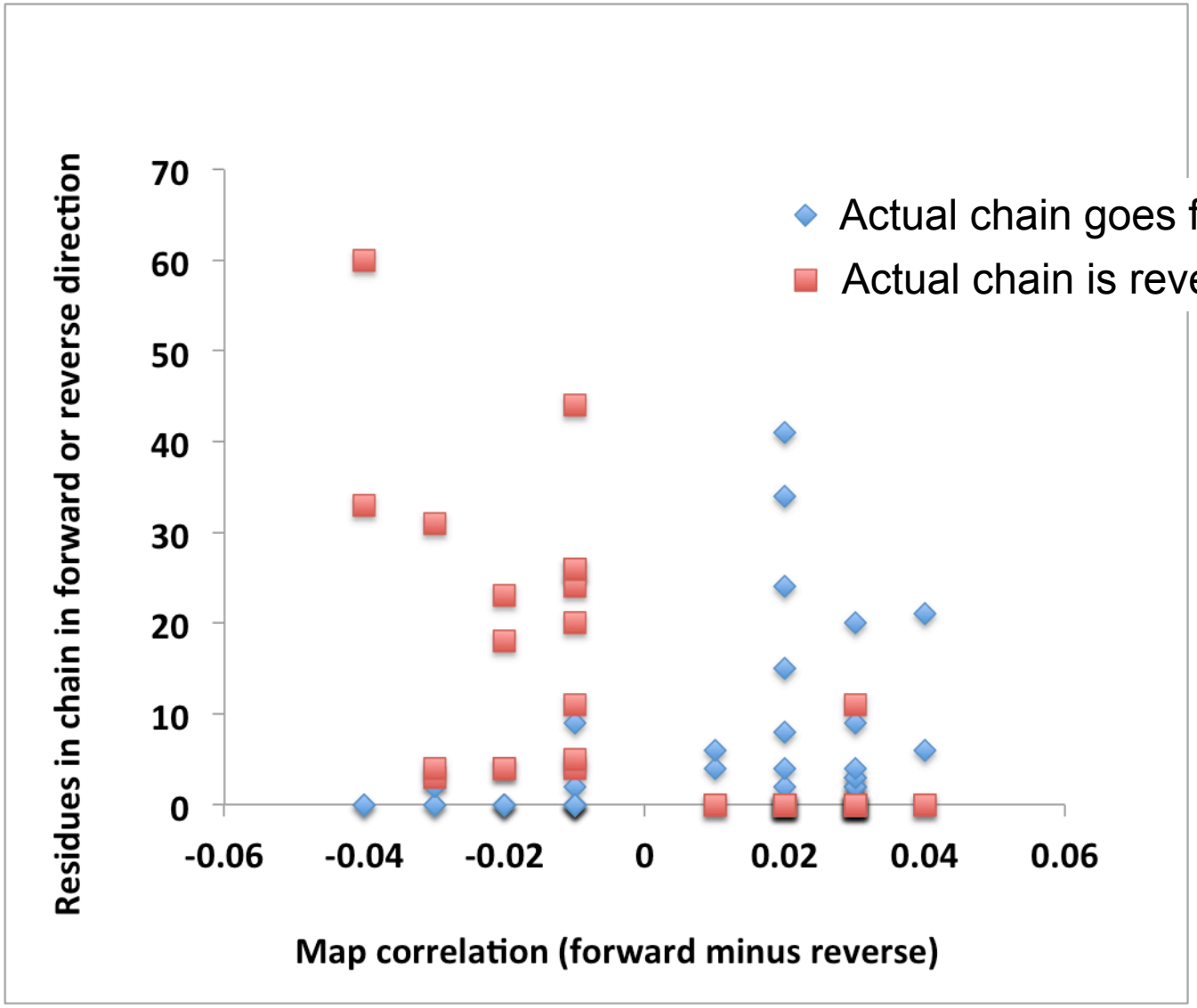
Chain tracings of cryo-EM map

(Chain I, yeast mitochondrial ribosome large subunit, 3.2 Å, 3j6b)



Which direction does the chain go?

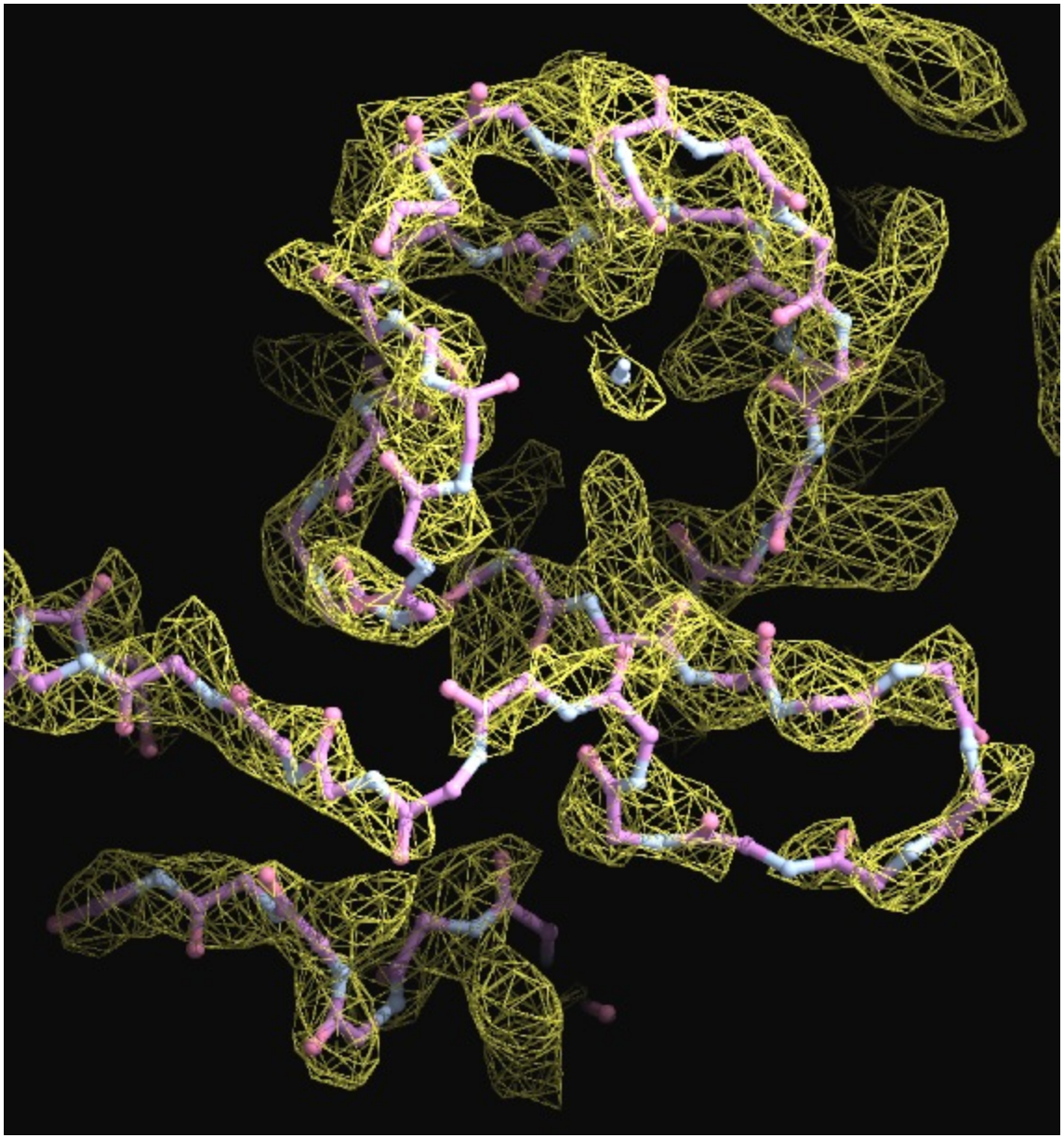
Identifying chain direction by map correlation



Optimizing model

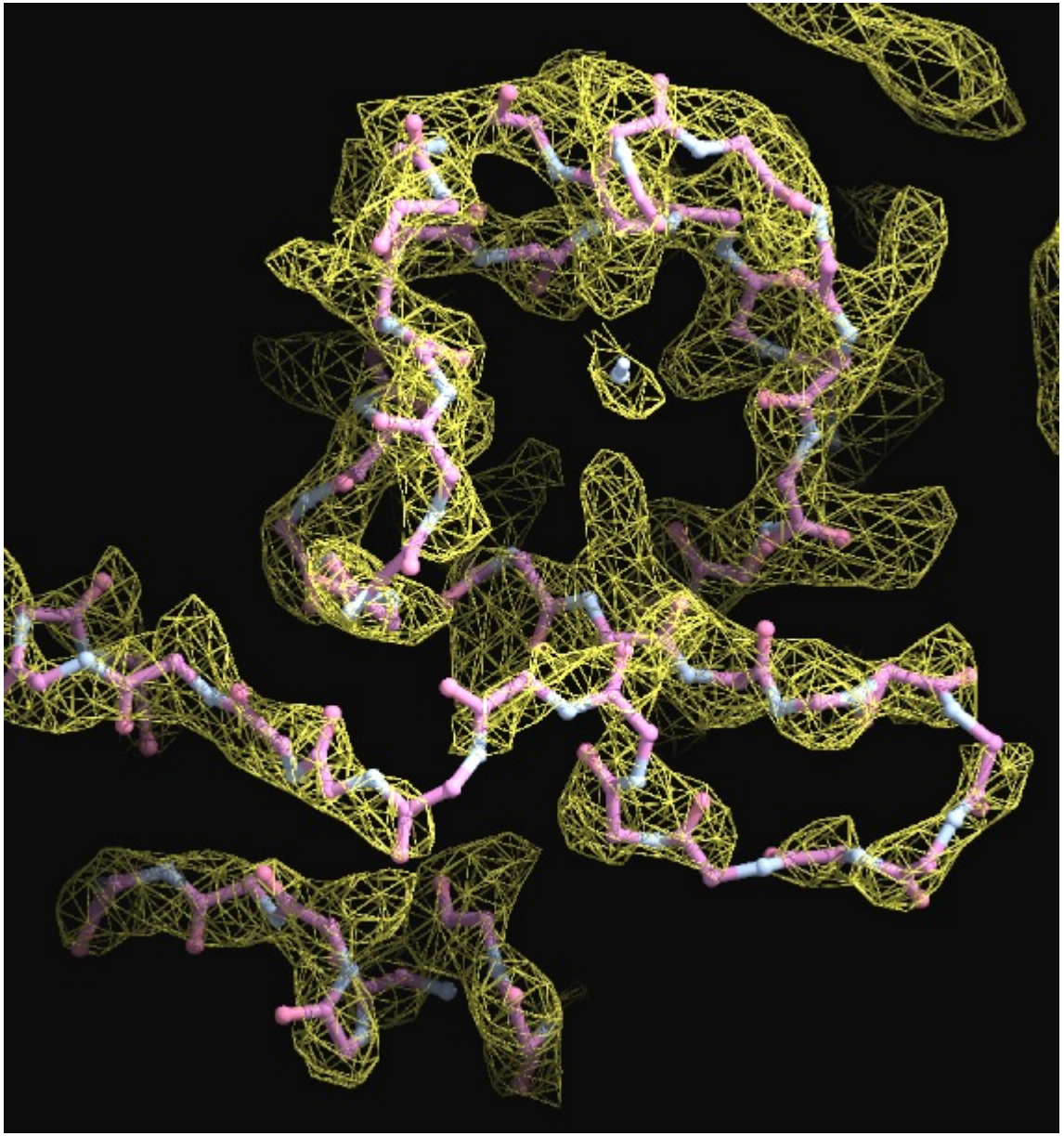
- Refine and rebuild model (simulated annealing, rebuilding and combination of best parts of each model)
- Replace segments with idealized structure
- Identify hydrogen-bonding (β-sheets, α-helices) and use them as restraints in real-space refinement

Simulated annealing refinement and recombination
(Chain I, yeast mitochondrial ribosome large subunit, 3.2 Å, 3j6b)



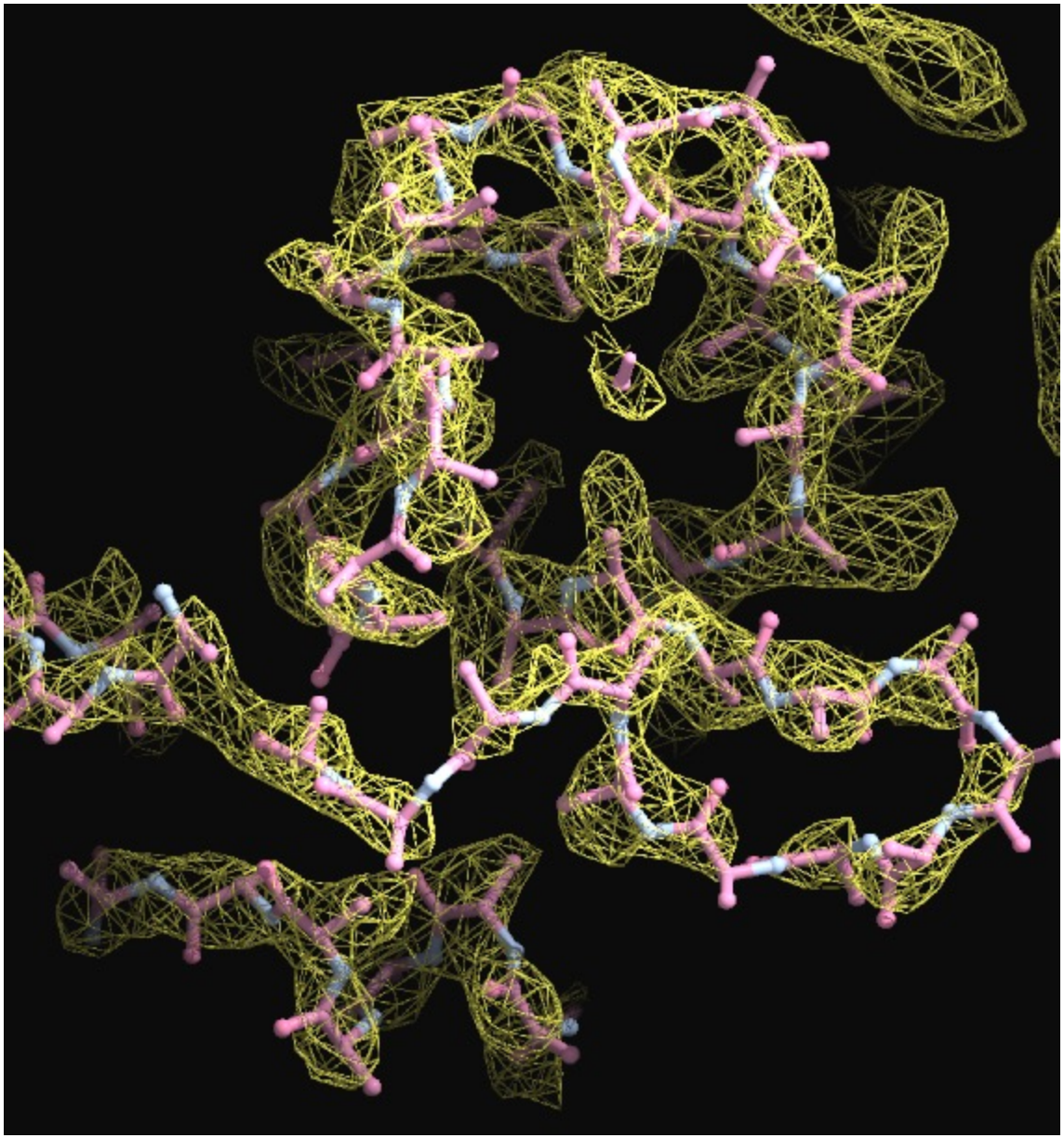
Rebuilding

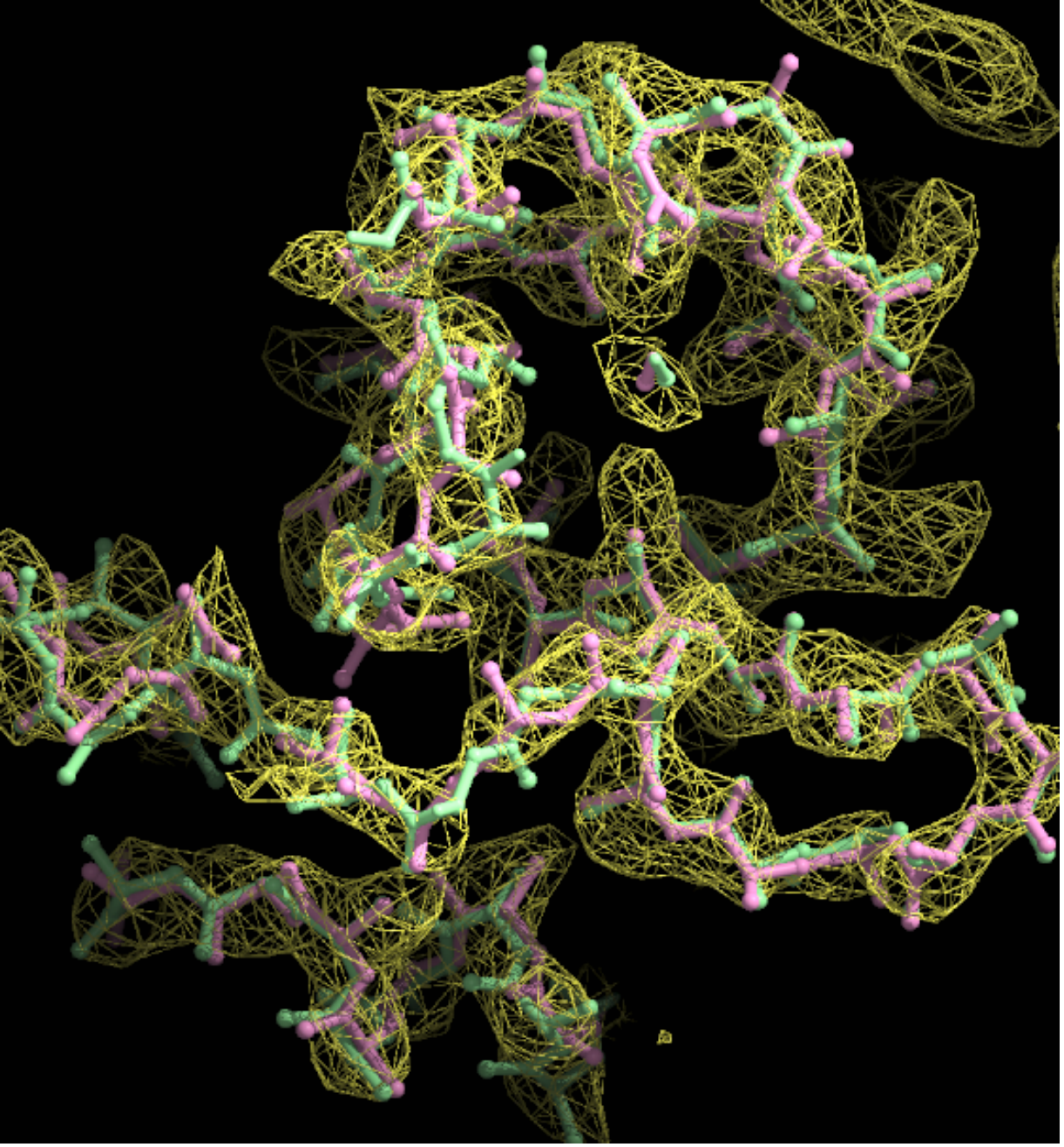
(Chain I, yeast mitochondrial ribosome large subunit, 3.2 Å, 3j6b)



Idealization and refinement

(Chain I, yeast mitochondrial ribosome large subunit, 3.2 Å, 3j6b)





Cryo-EM map from yeast mitochondrial ribosome (chain I of large subunit, 3.2 Å, Amunts et al., 2014)

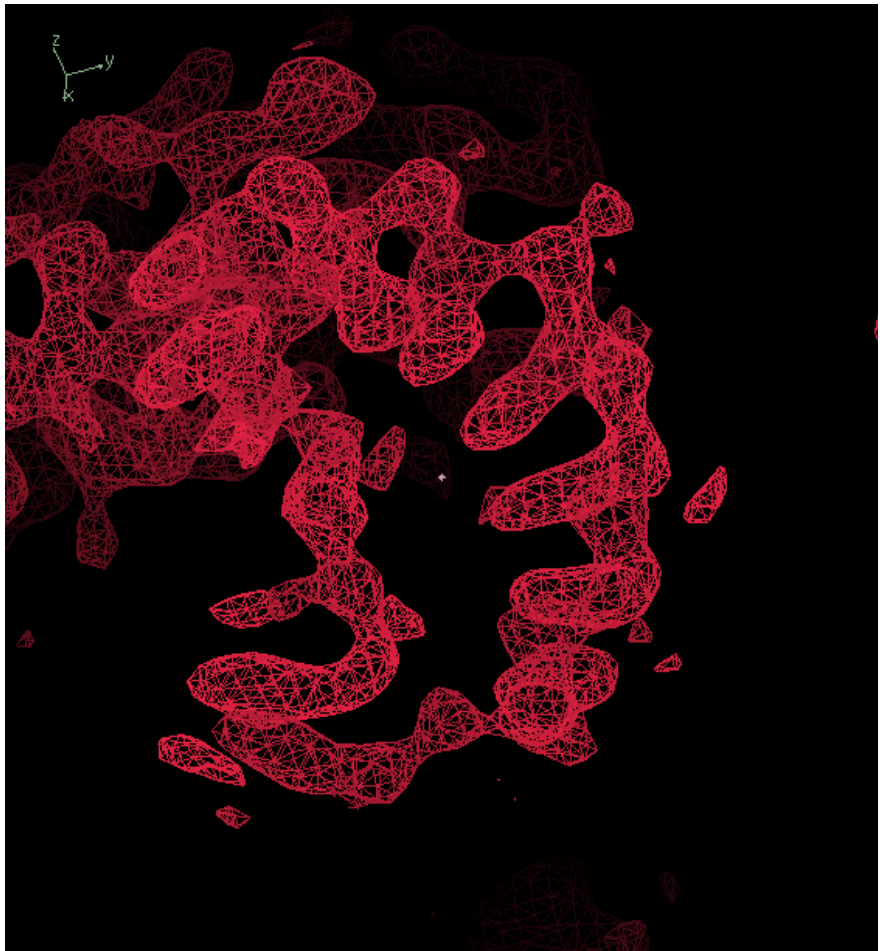
Autobuilt model (pink)
Deposited model (green)
(main-chain and C^{β} atoms)

Automated interpretation of cryo-EM maps

- Cut out molecule
- Identify optimal
- Try building protein/RNA/DNA (whatever may be there)
- Choose segment type by map correlation
- Assemble and refine
- Apply molecular symmetry and refine again

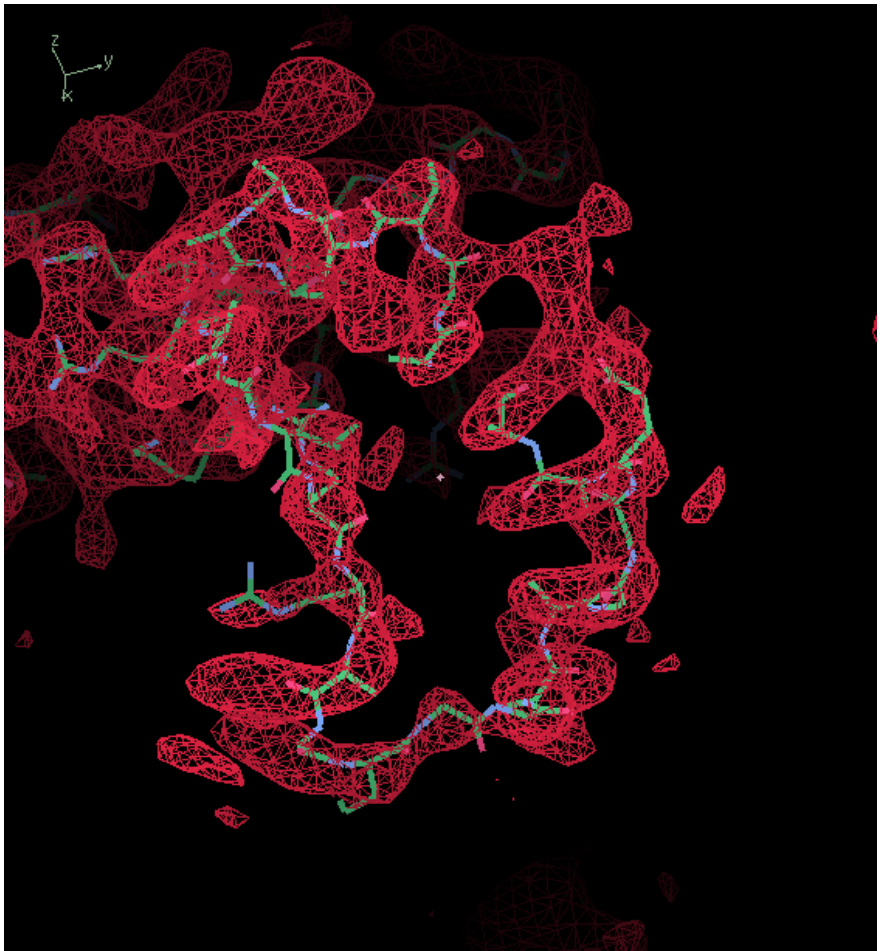
70S ribosome at 2.9 Å
RNA/Protein building into segmented map

Segmented density



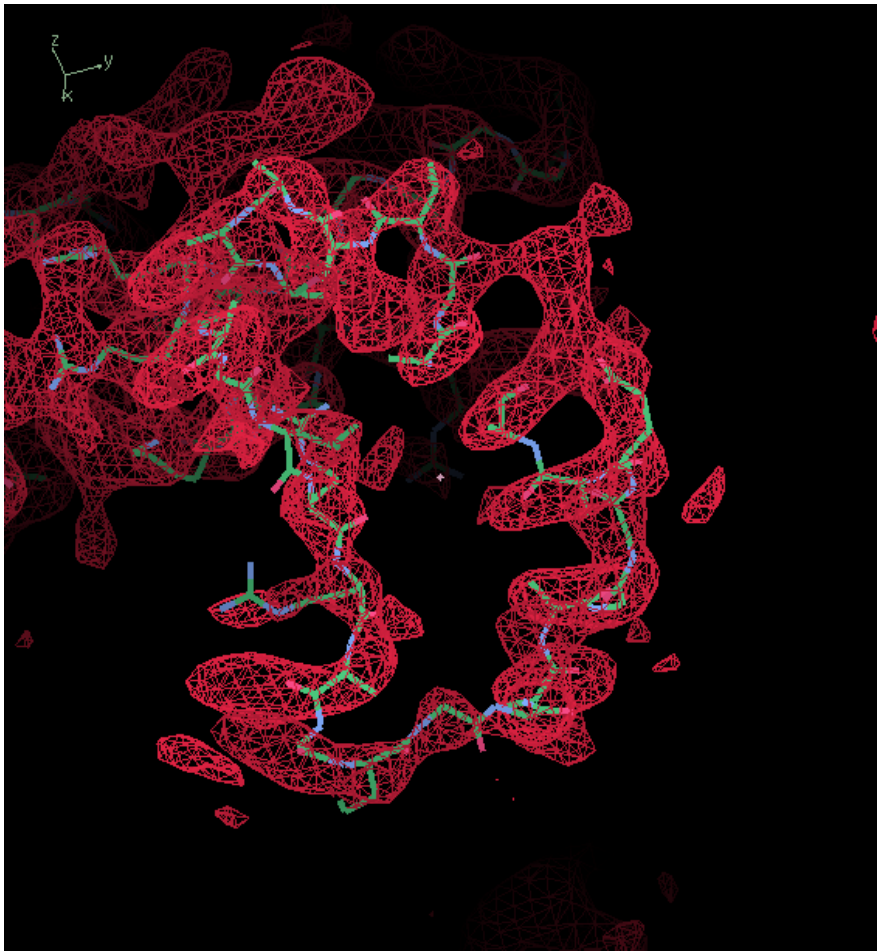
70S ribosome at 2.9 Å RNA/Protein building into segmented map

…as protein

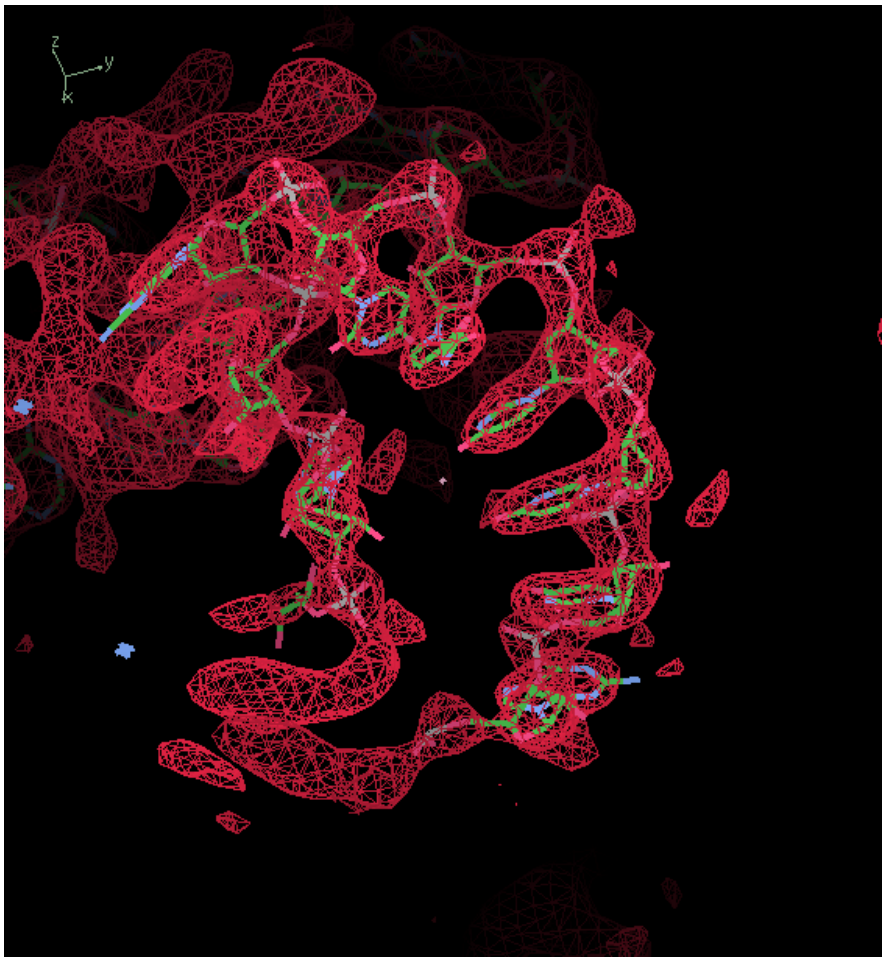


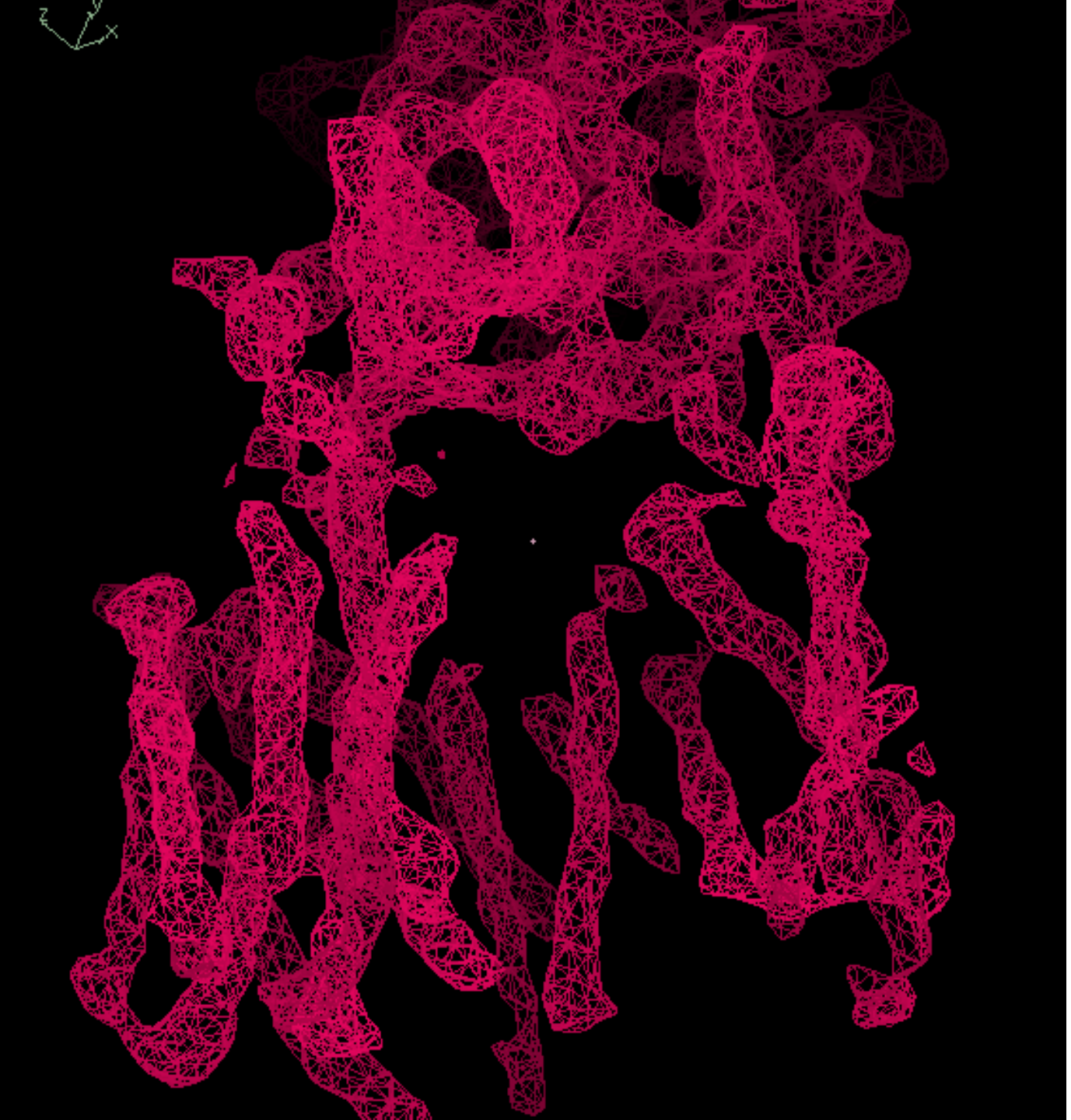
70S ribosome at 2.9 Å RNA/Protein building into segmented map

…as protein

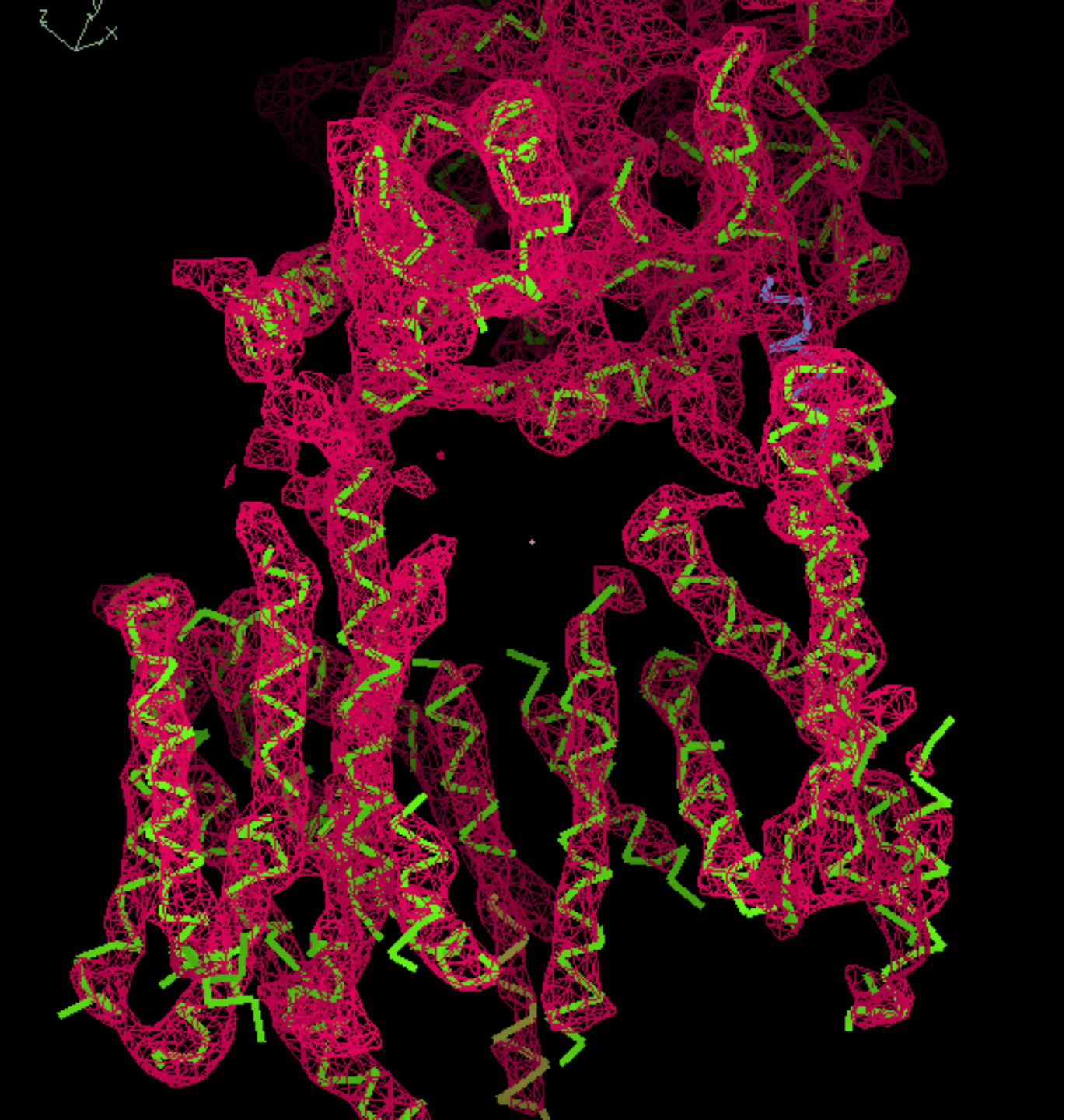


…as RNA



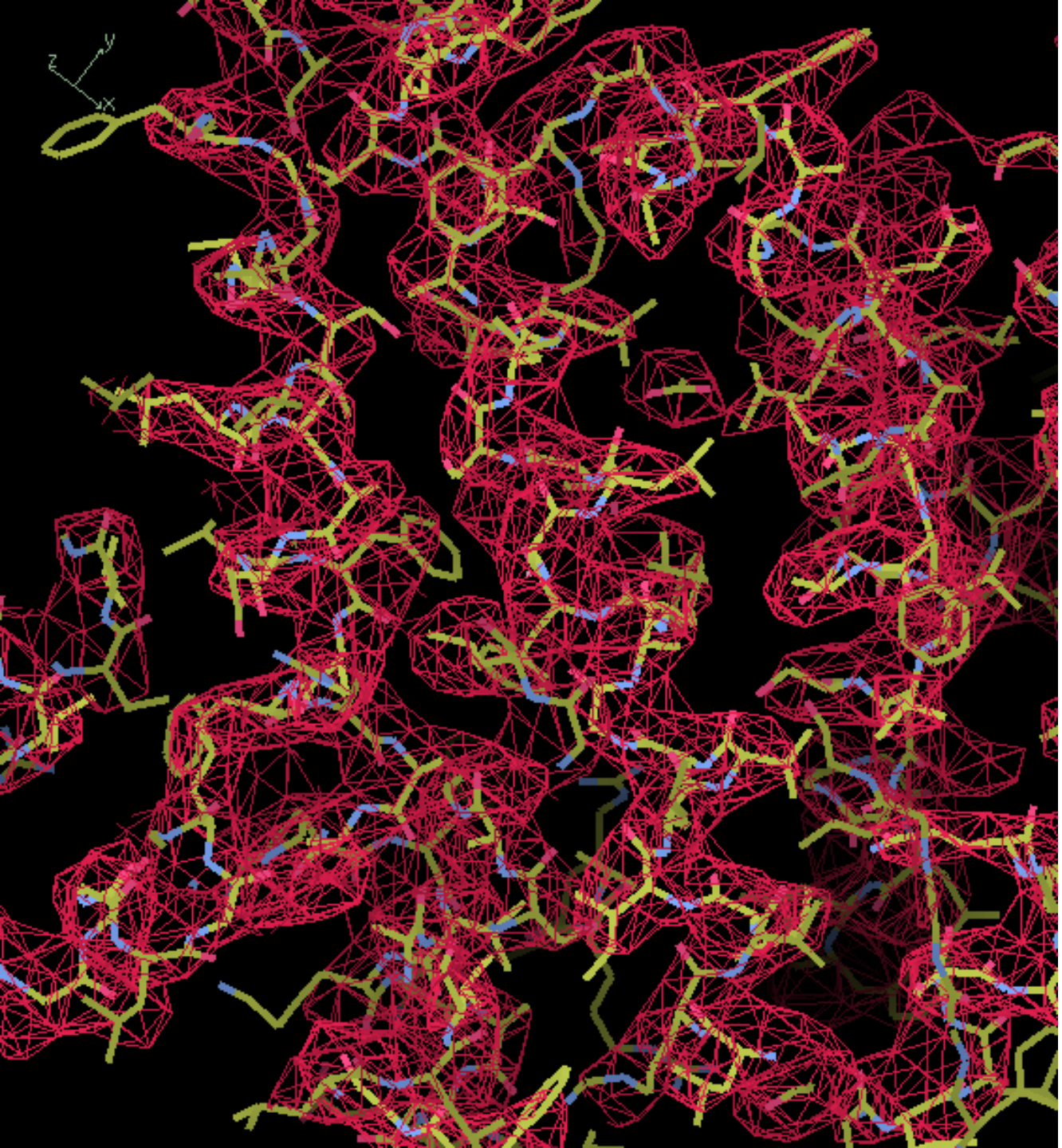


Gamma-secretase at 4.5 Å (emd_2677)



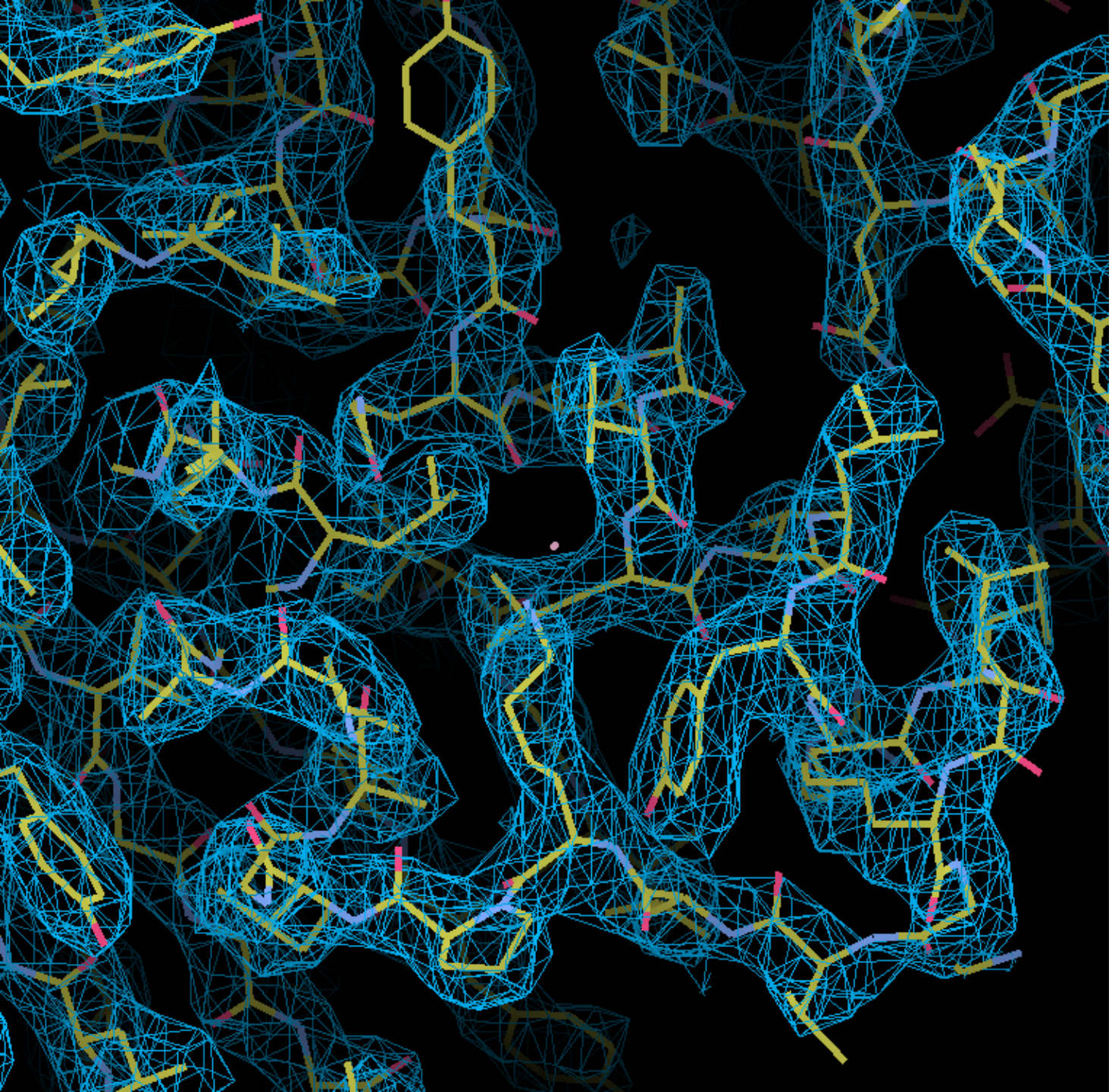
Gamma-secretase at 4.5 Å

(autobuilt model; emd_2677)



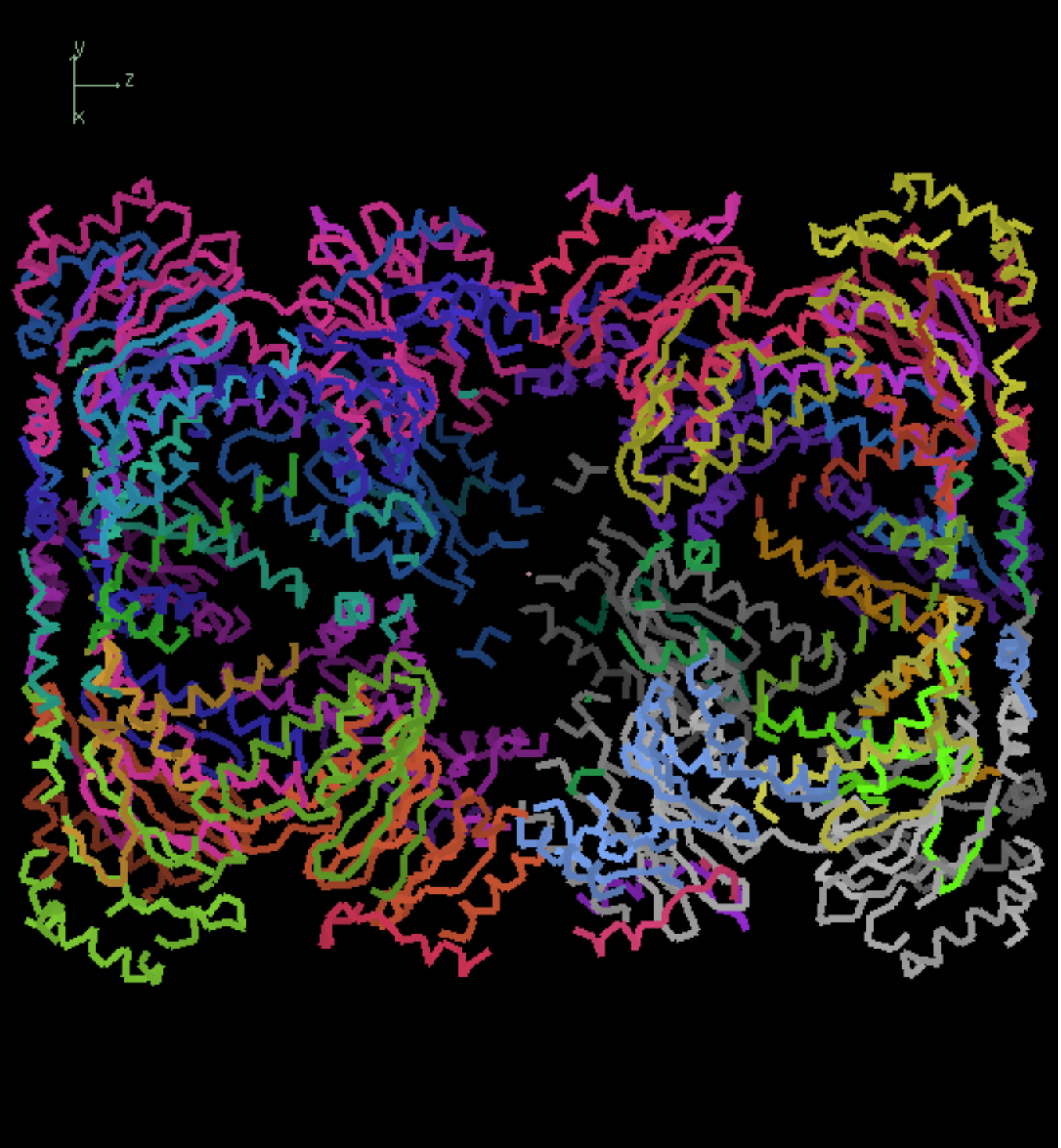
..and another Gamma-secretase structure at 3.4 Å

(autobuilt model; emd_3061)



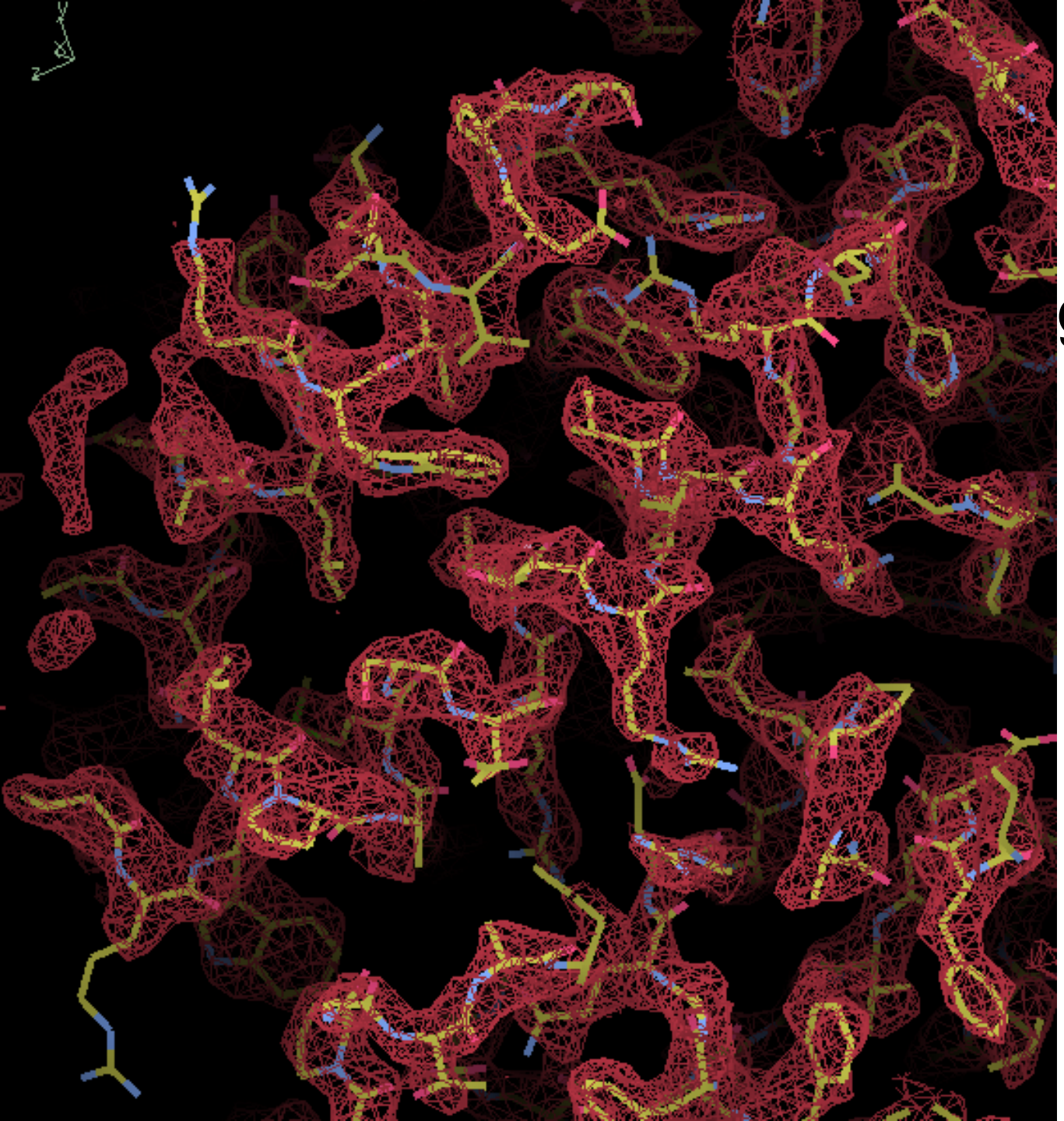
Proteasome at 2.8 Å

(autobuilt model; emd_6287)



Proteasome at 2.8 Å

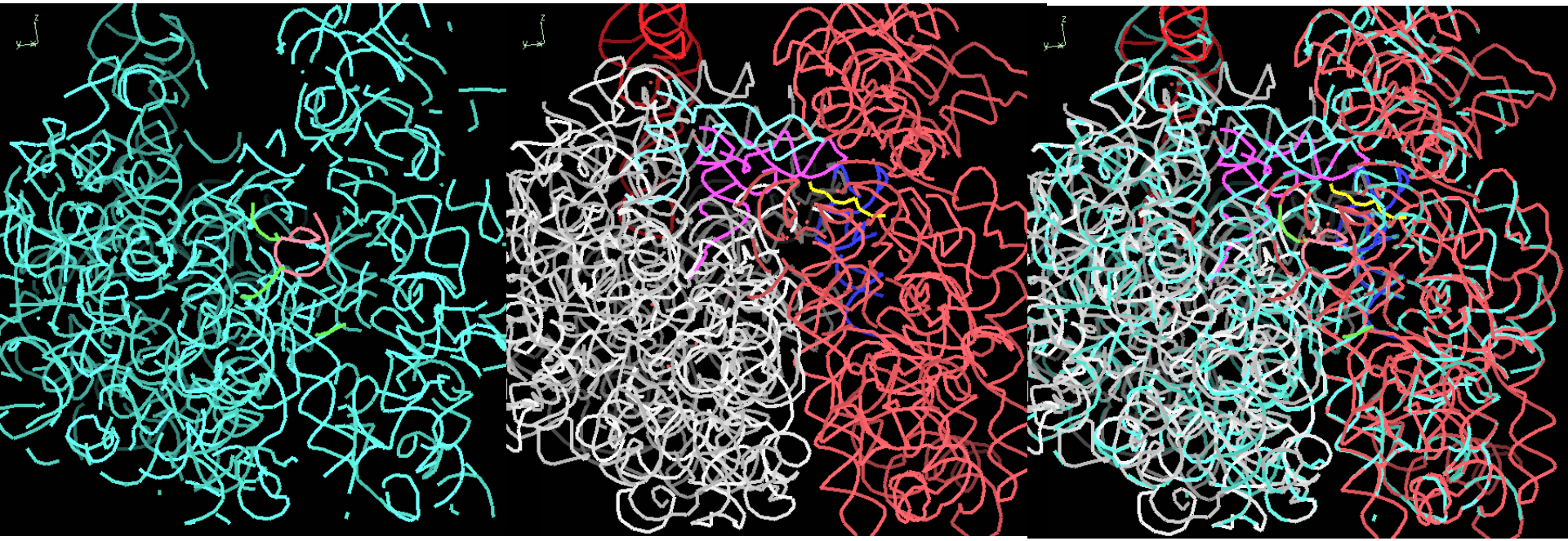
(autobuilt model; emd_6287)



Beta-galactosidase at 2.2 Å

(autobuilt model; emd_2984)

70S *E. coli* ribosome (5afi, 3.2 Å)

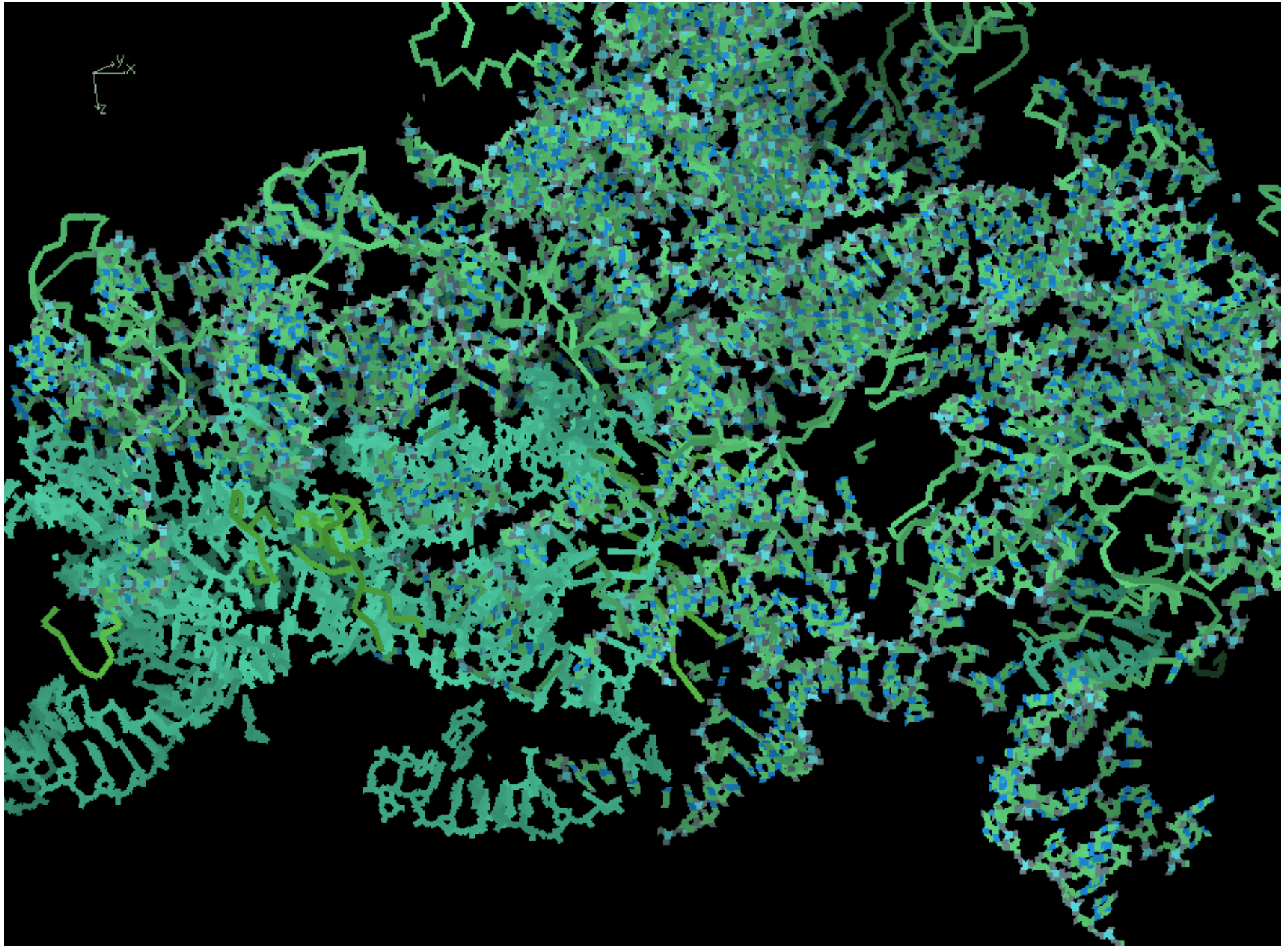


Autobuilt 5afi both

Total residues autobuilt correctly:

RNA: 2588 of 4763 (rmsd 0.63 Å)
Protein: 3212 of 6323 (rmsd 0.76Å)

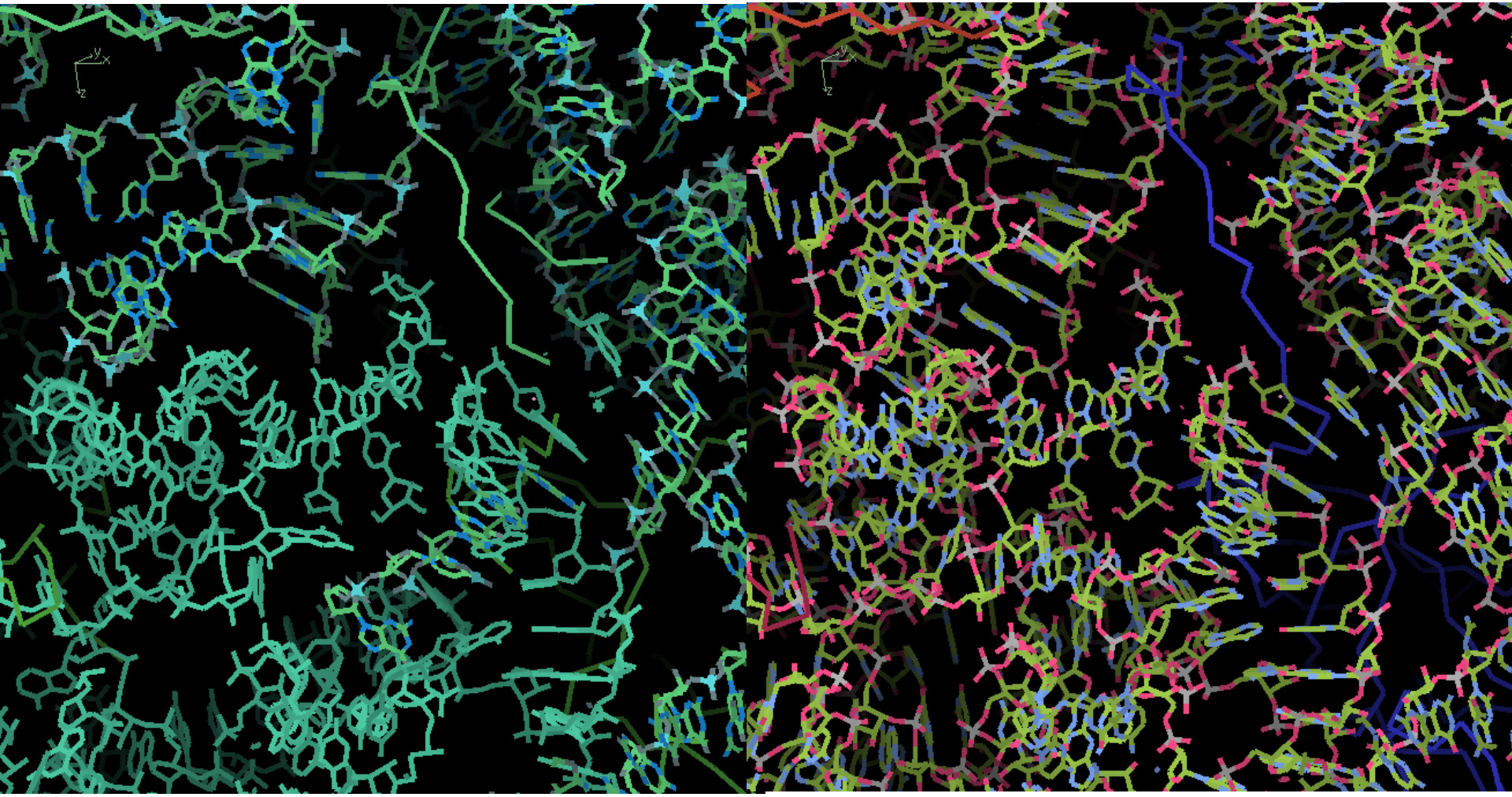
30S Ribosome (X-rray map autobuilt 1j5e, 2.9 Å)



30S Ribosome (autobuilt 1j5e, 2.9 Å)

autobuilt

1j5e



Perspectives…

- Local automatic map optimization could improve model-building
- Incorporation of validation (idealization) at model-building stage improves low-resolution models
- Approach may be enhanced by combining structure-modeling tools (Rosetta) with Phenix model-building
- Distance restraints from residue co-evolution could increase information about model
- Secondary structure prediction could be used in sequence assignment
- Partial model information from PDB could be used

The Phenix Project

Lawrence Berkeley Laboratory

Paul Adams, Pavel Afonine, Dorothee Leibschner, Nigel Moriarty, Nicholas Sauter, Oleg Sobolev, Billy Poon

Los Alamos National Laboratory

Tom Terwilliger, Li-Wei Hung

Cambridge University

Randy Read, Airlie McCoy, Gabor Bunkoczi, Rob Oeffner

Duke University

Jane & David Richardson, Chris Williams, Bryan Arendall, Bradley Hintze

An NIH/NIGMS funded Program Project

Phenix