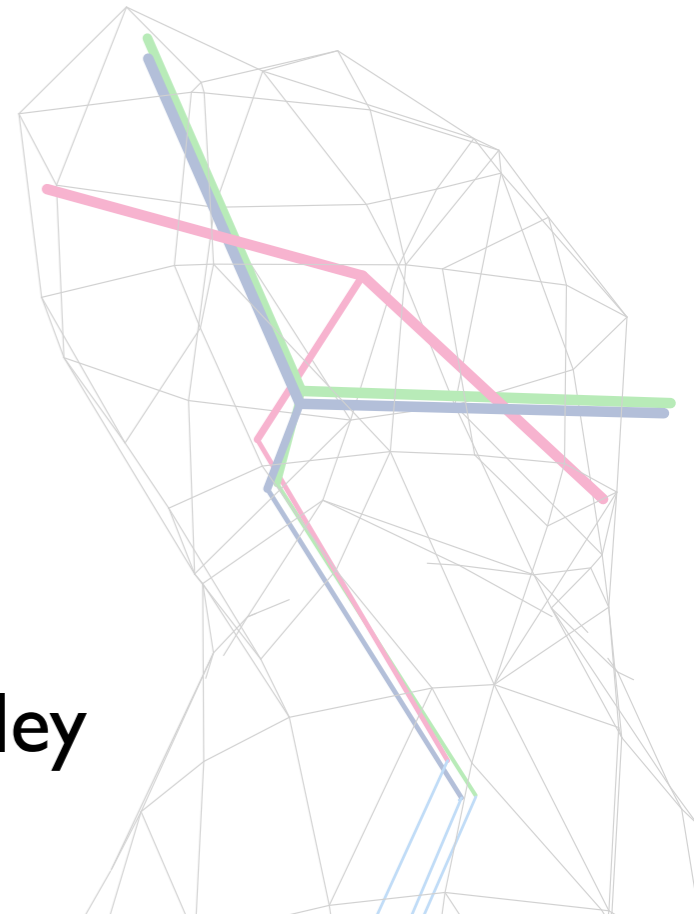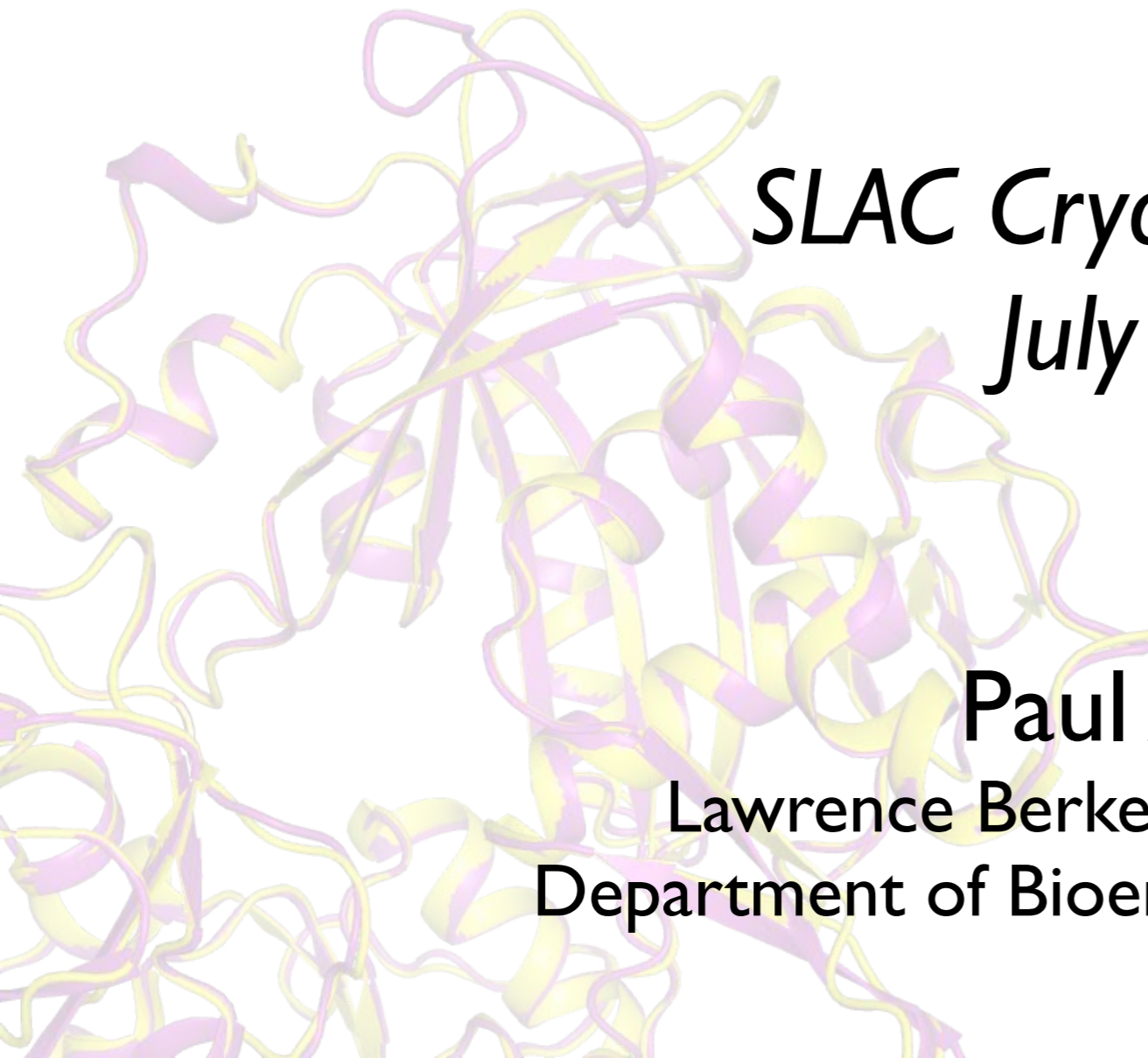# Atomic Models from Cryo-EM Data
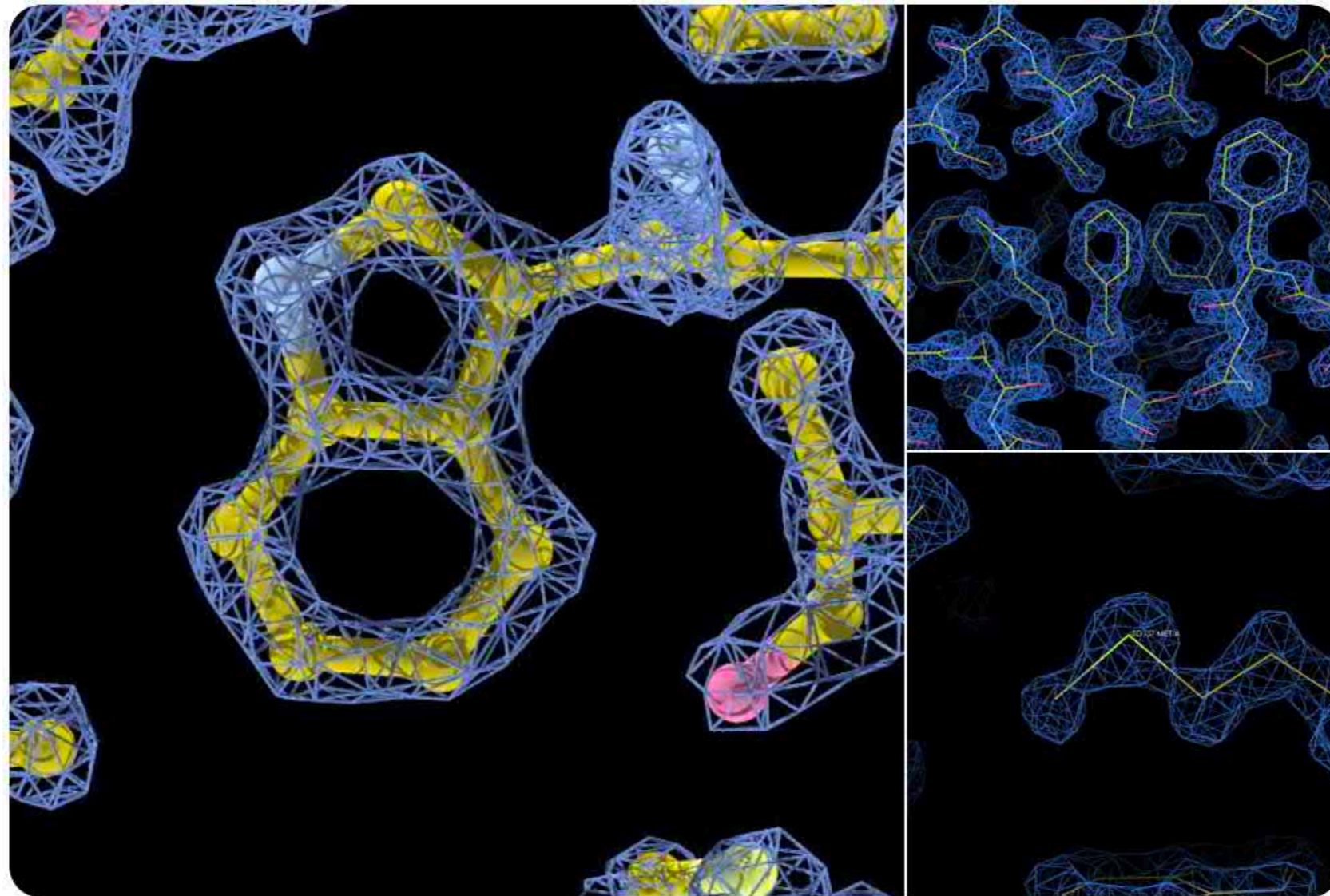
SLAC Cryo-EM School
July 2019

Paul Adams
Lawrence Berkeley Laboratory and
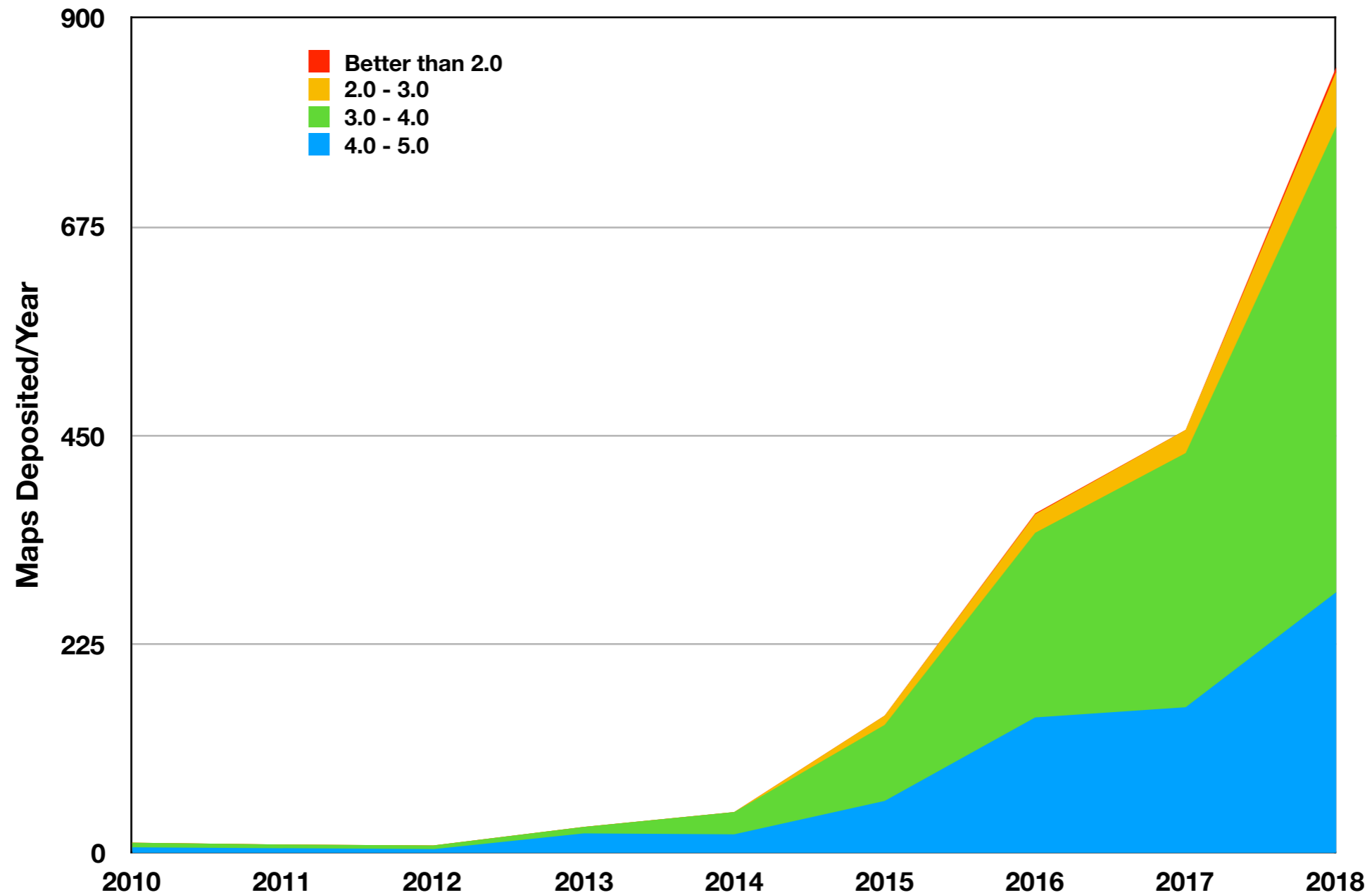Department of Bioengineering UC Berkeley

Impressive Cryo-EM Achievements

Namba Lab, Osaka

Map Resolution

- Biggest growth is in the 3-4Å range

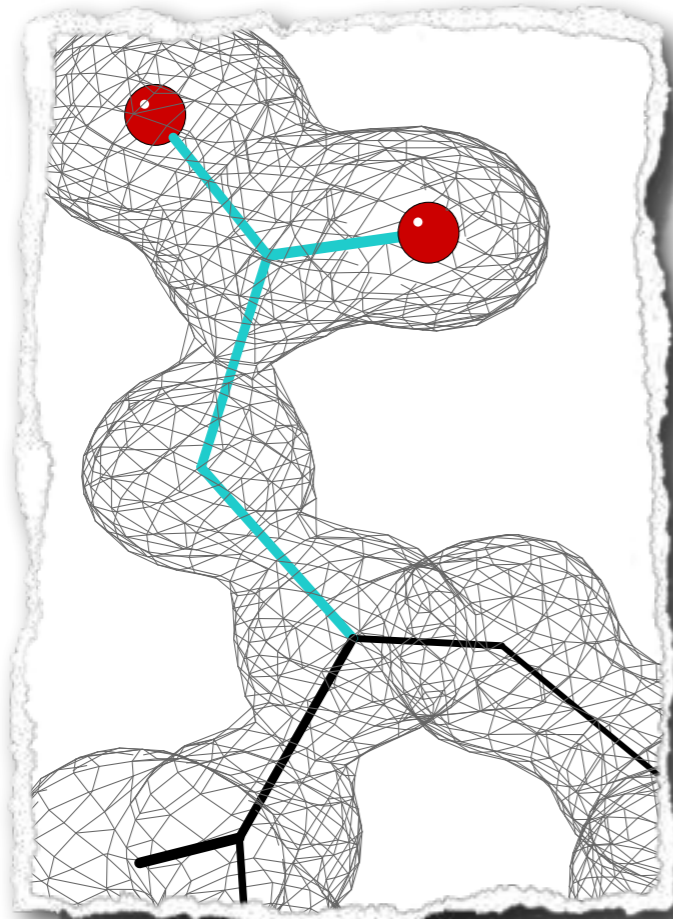- Substantial number of maps in 4-5Å range
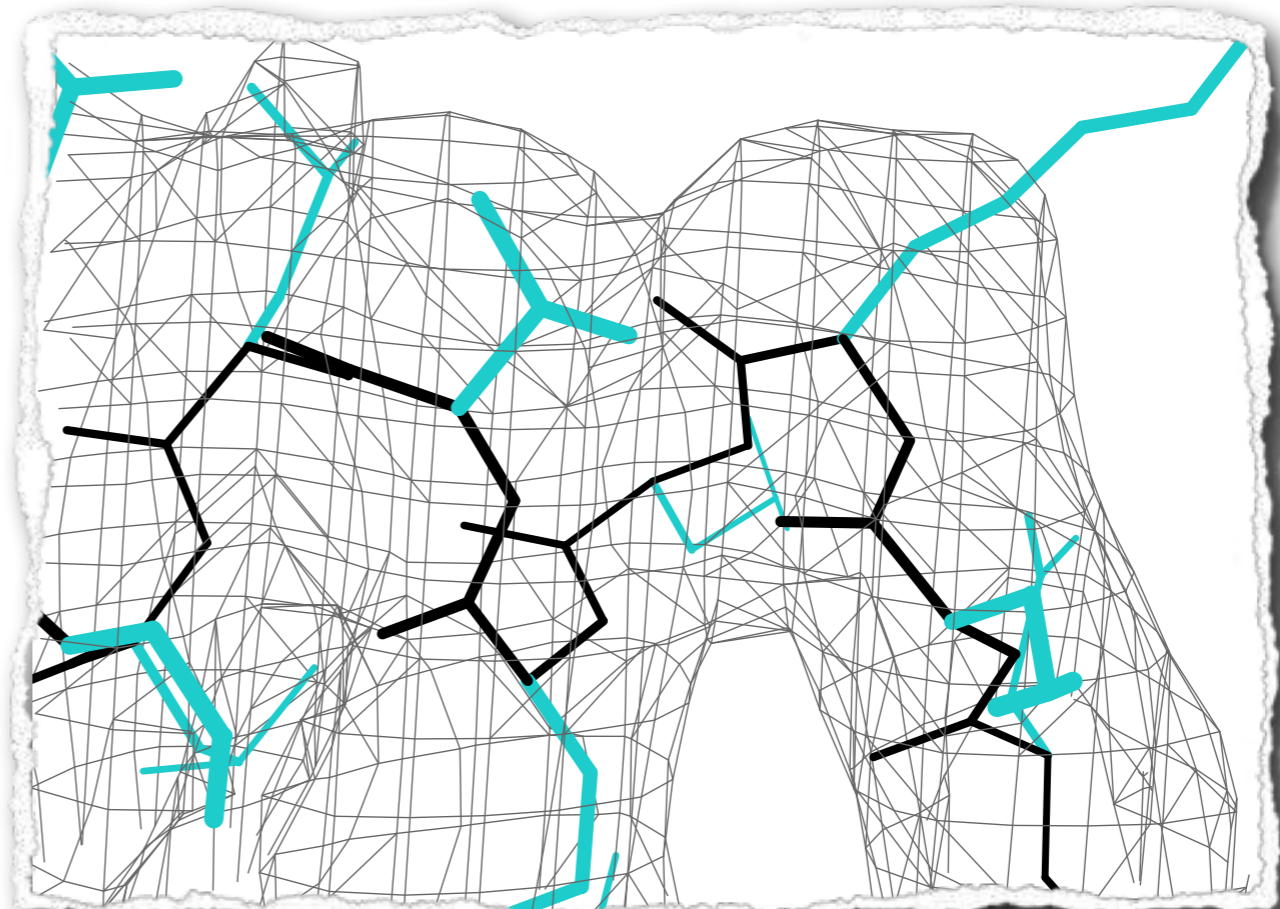

* Not all maps have an associated model

Low Resolution

PDBID: 2gkg
Resolution: 1.00Å
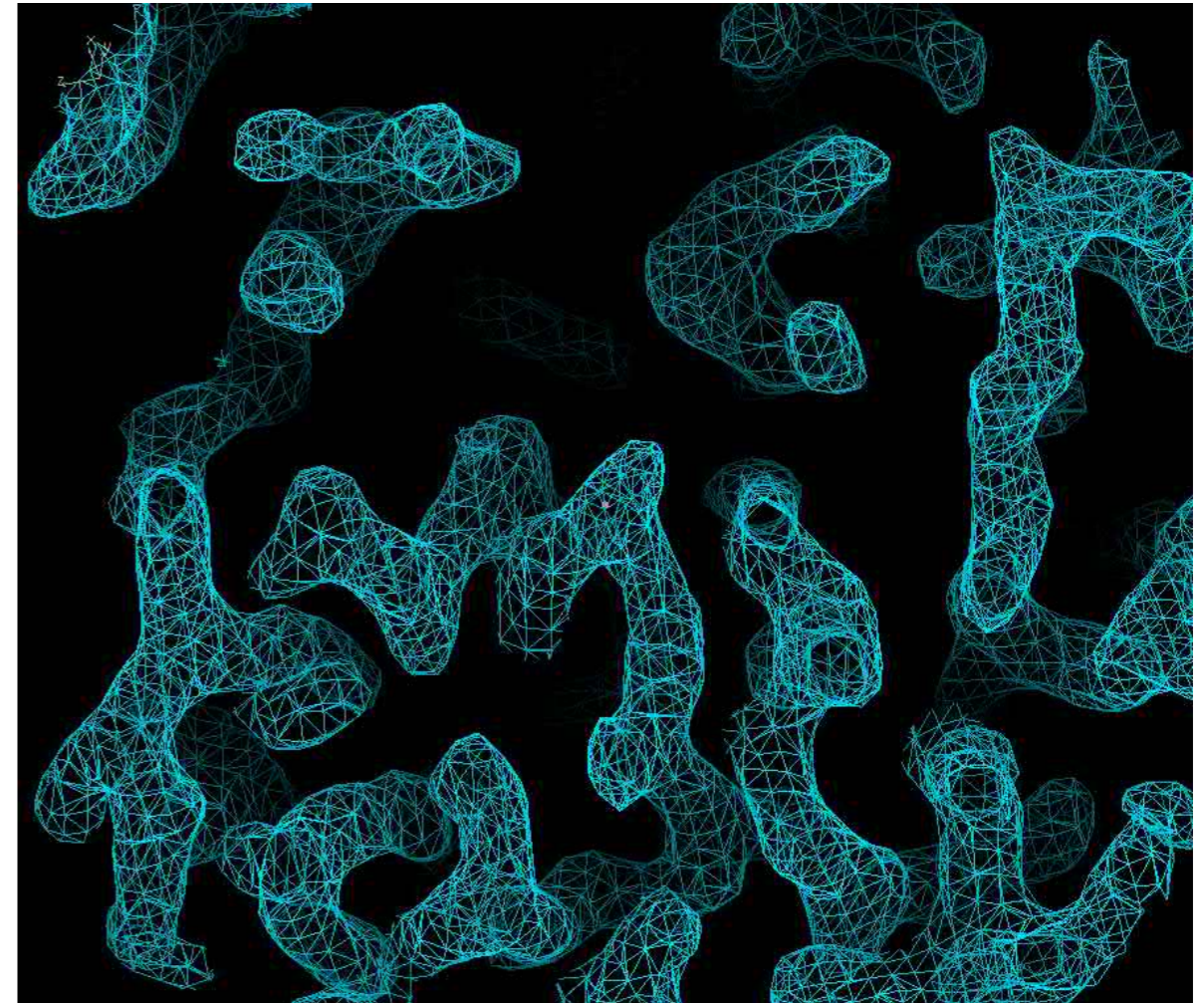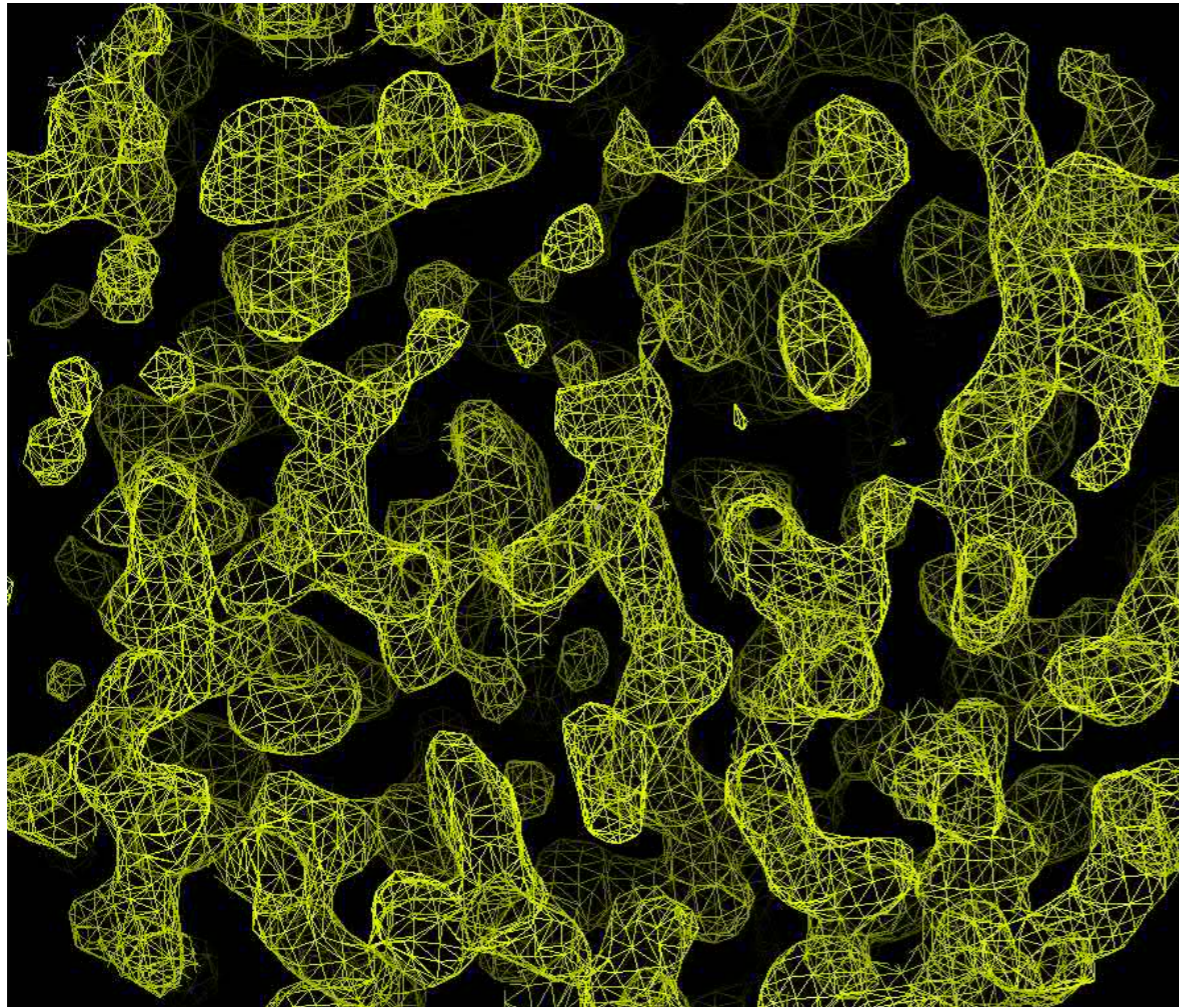
PDB ID: 3k7a
Resolution: 3.80Å

- Many challenges:

  - How to interpret "featureless" maps (pattern matching, chemical constraints)

  - How to optimize models with sparse data (prior information)

Phenix

Crystallographic vs. Cryo-EM Maps
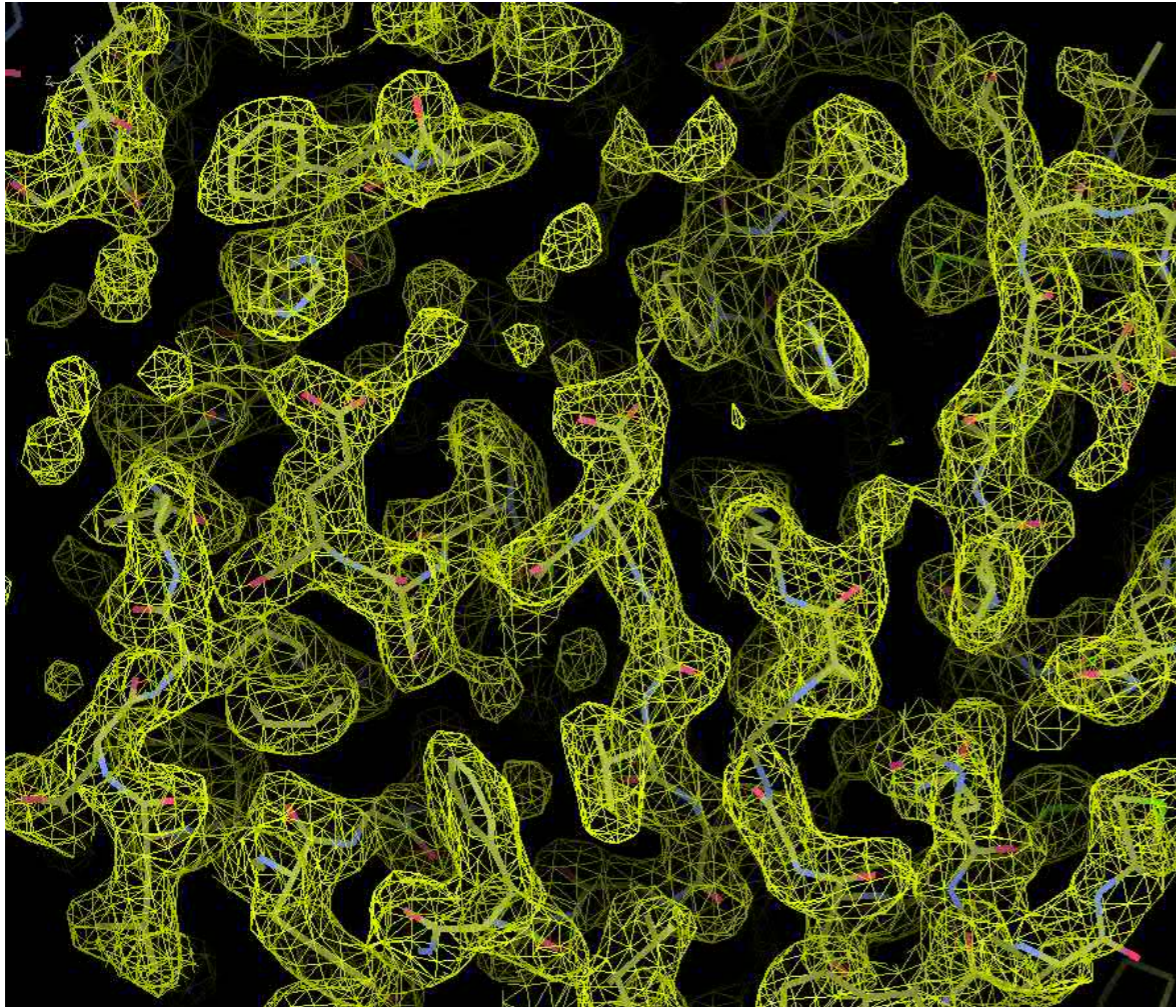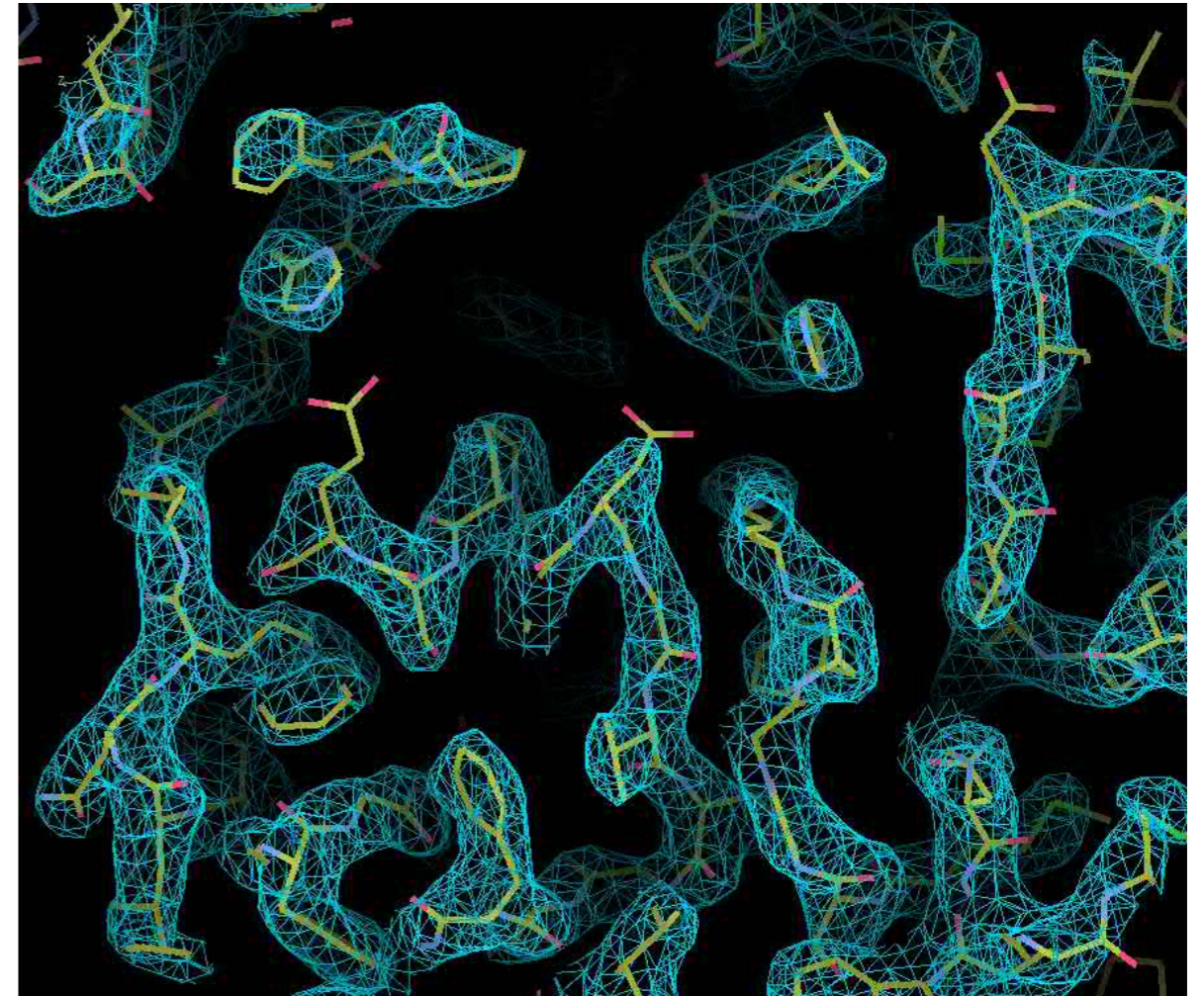
Beta galactosidase at 2.2 Å

Tom Terwilliger, Los Alamos National Lab

Crystallographic vs. Cryo-EM Maps

Beta galactosidase at 2.2 Å

X-ray (PDB 3i3b)

Cryo-EM (PDB 5a1a)

Tom Terwilliger, Los Alamos National Lab
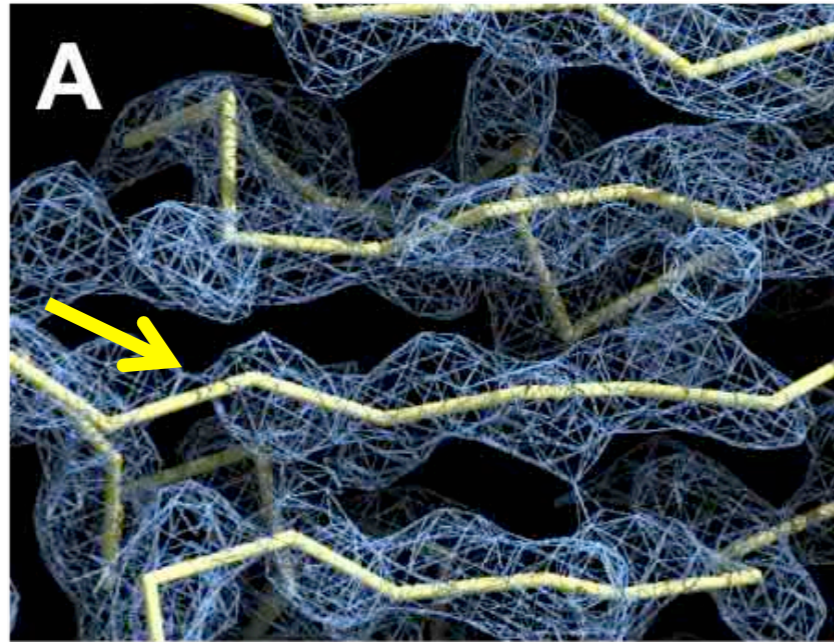
Crystallographic vs. Cryo-EM Maps

- The maps are very similar

Tom Terwilliger, Los Alamos National Lab
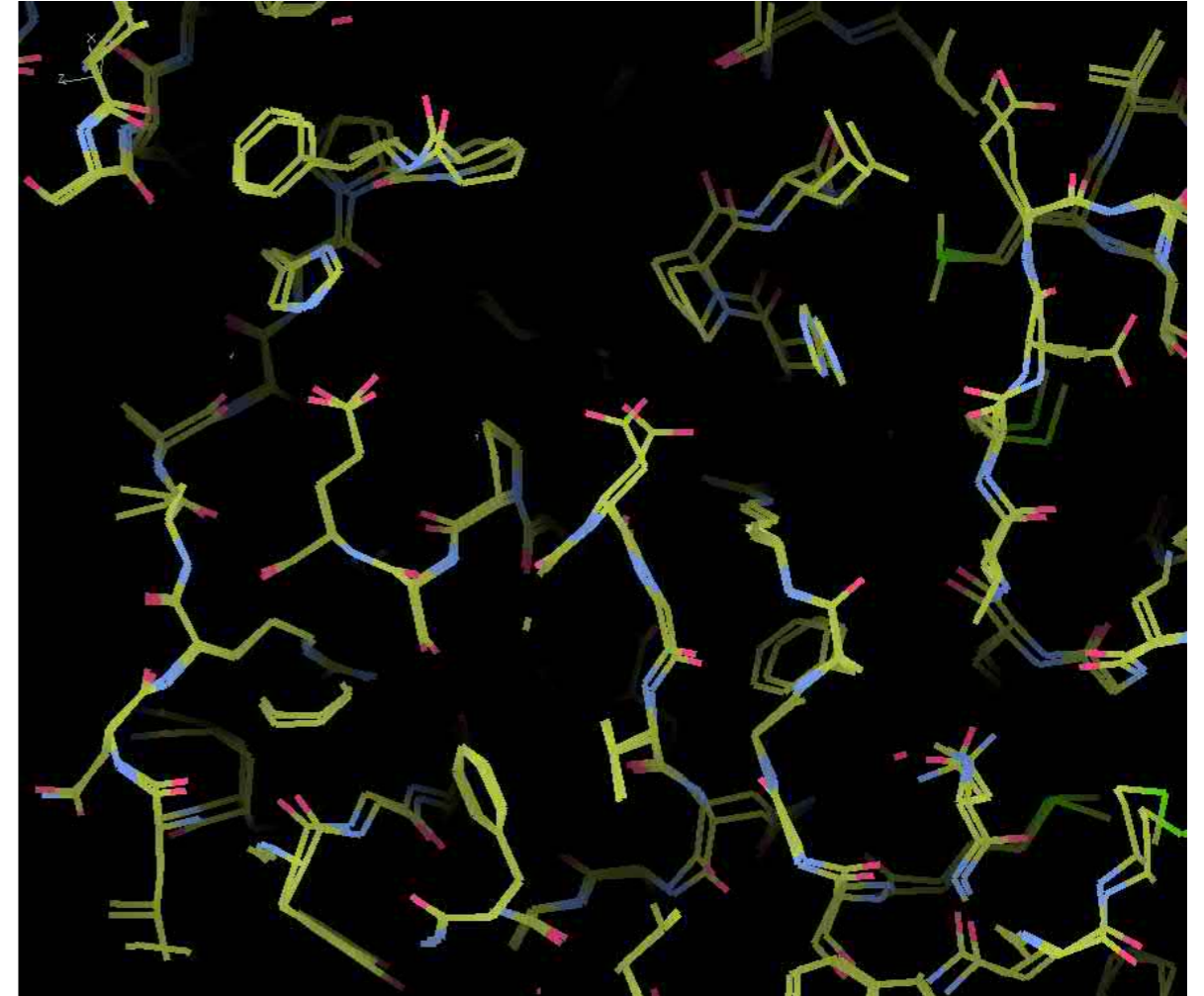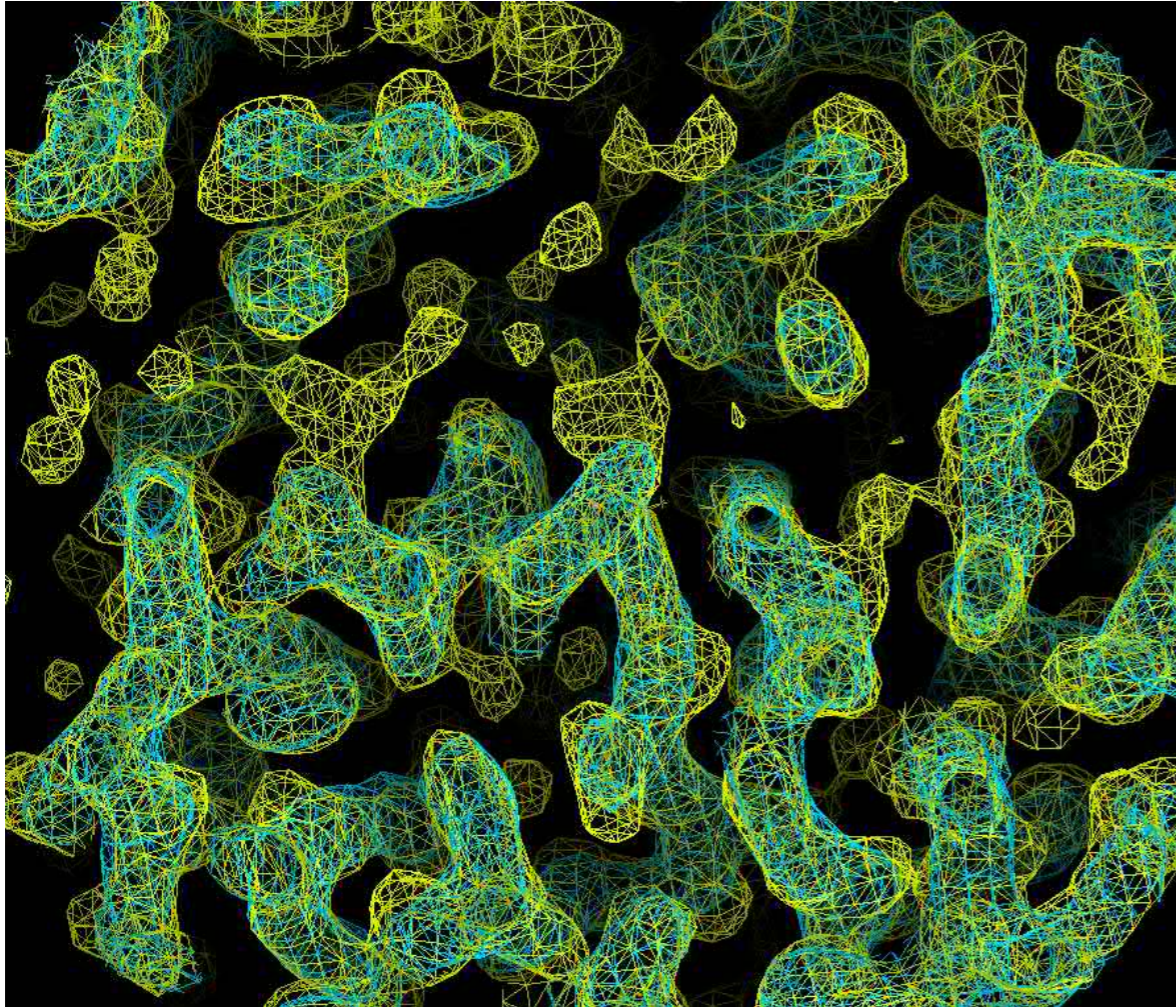
More Accurate Low Resolution Information in Cryo-EM Maps

Original

Phenix

Tom Terwilliger, Los Alamos National Lab
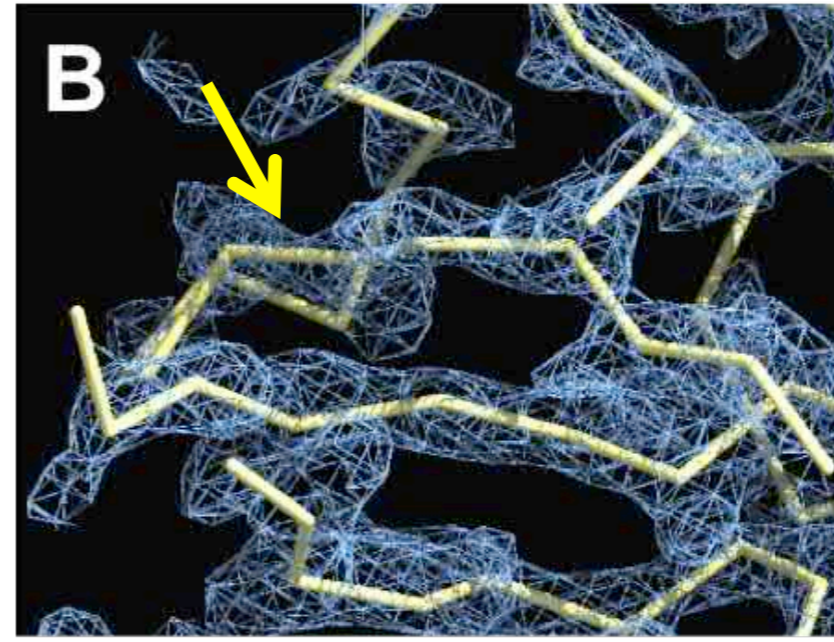
The Phenix Project

Lawrence Berkeley Laboratory

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

Los Alamos National Laboratory
New Mexico Consortium
Baylor College of Medicine

Tom Terwilliger, Li-Wei Hung, Matt Baker

Duke University

Jane & David Richardson, Chris Williams, Vincent Chen, Bradley Hintze

Randy Read, Airlie McCoy, Tristan Croll, Rob Oeffner, Kaushik Hatti, Massimo Sammito, Duncan Stockwell

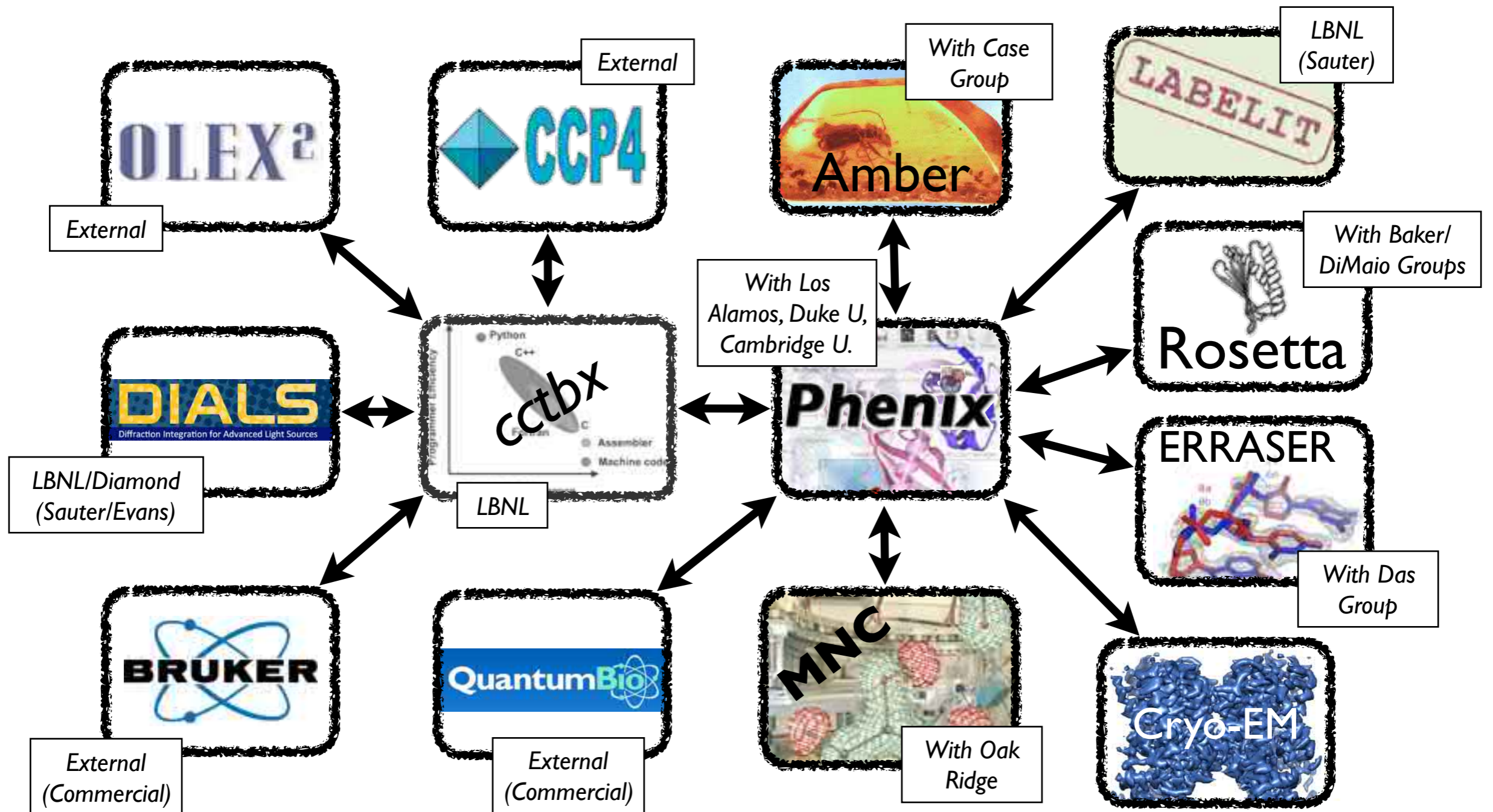**Cambridge University**

An NIH/NIGMS funded Program Project

Adams PD et al., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.* 2010, **D66**:213-221.
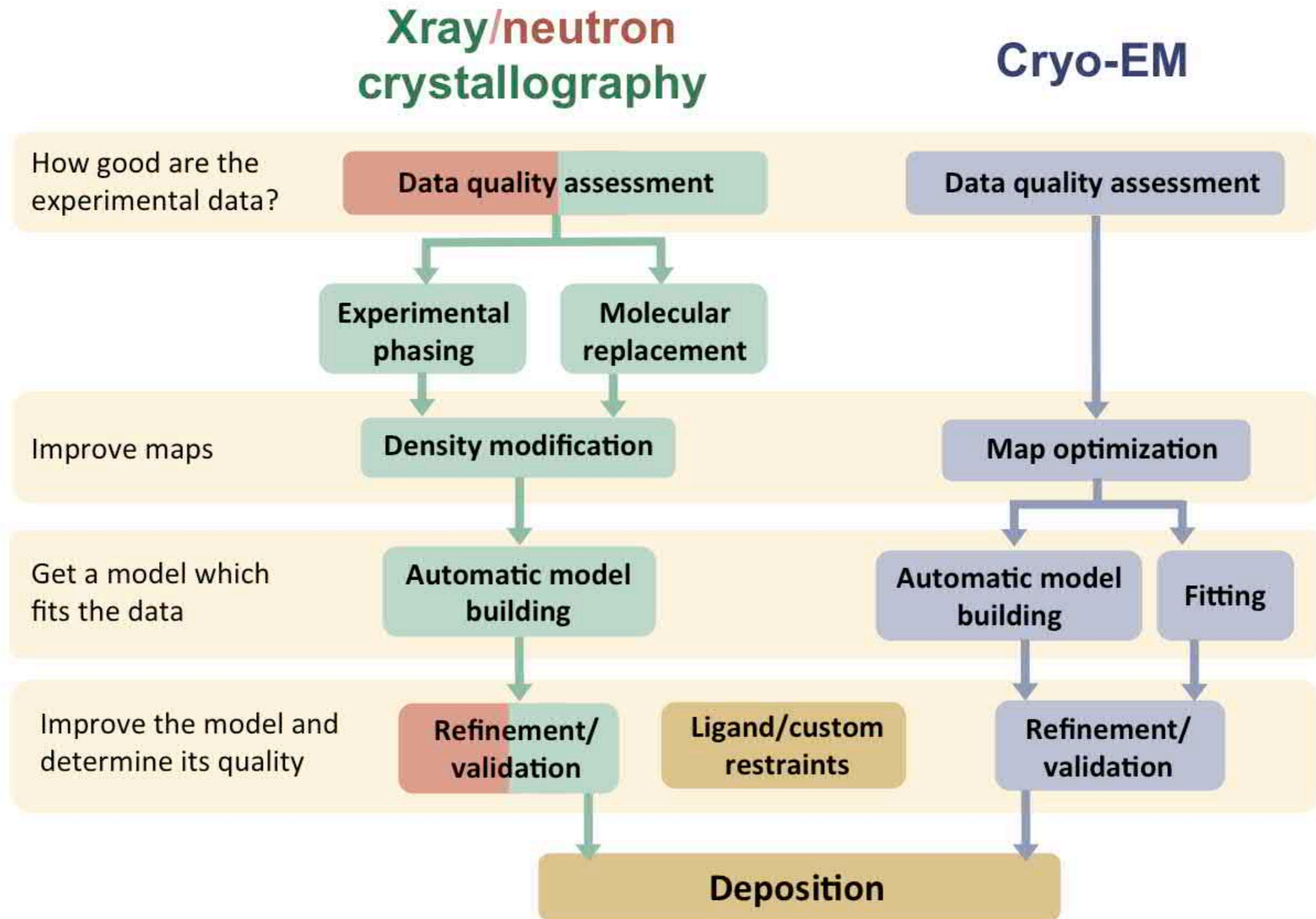
Phenix - a Structural Biology Hub

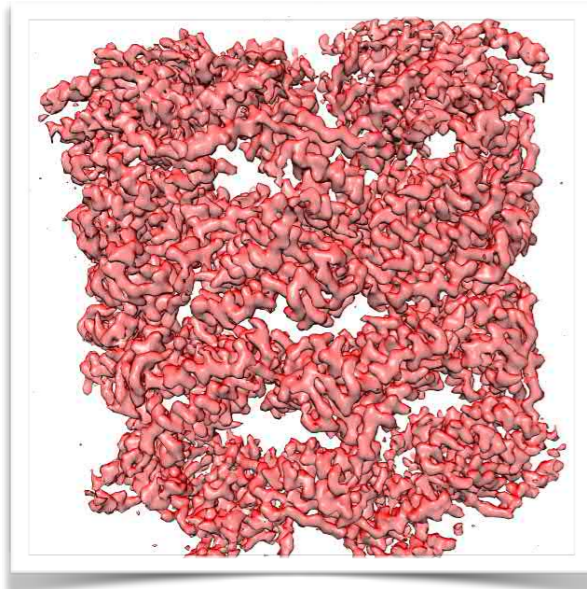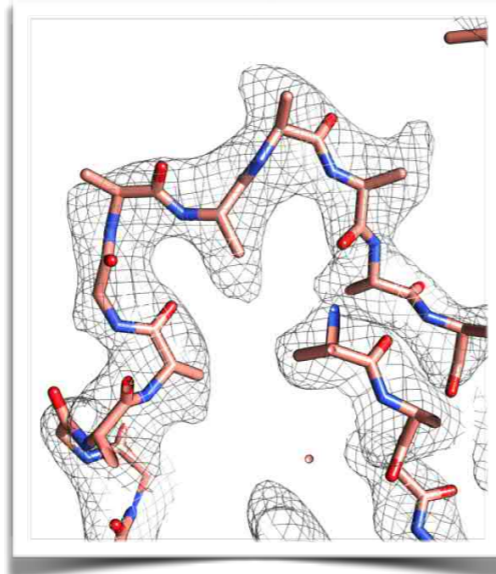# Structural Biology Workflows

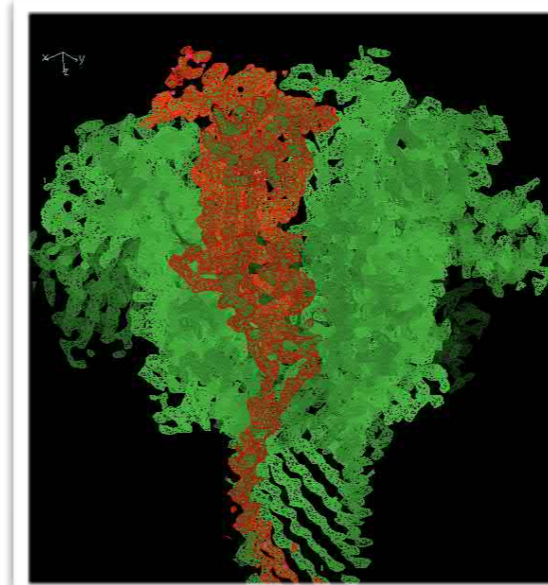*Dorothee Liebschner,*
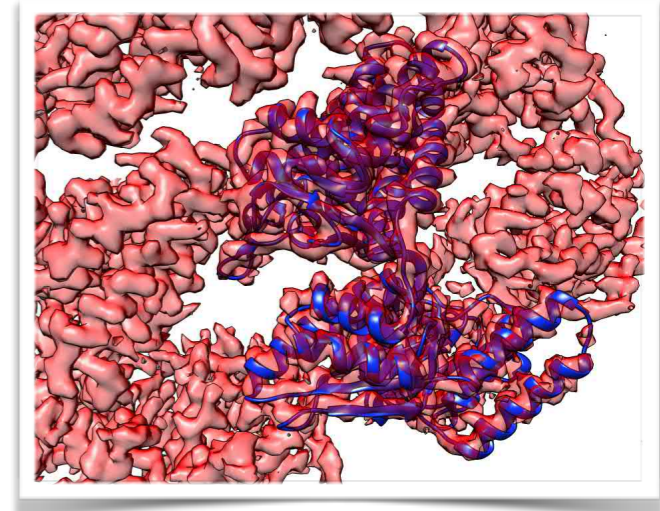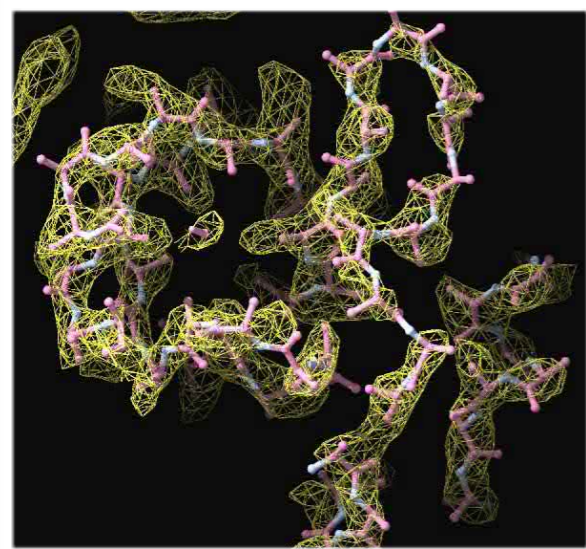*Lawrence Berkeley Lab*

New Tools for Cryo-EM in Phenix

Symmetry from a map

Automated map sharpening

Map segmentation

Rigid model docking

Automated model building

Real space refinement

Model and map validation

Tutorials

- Model placement and building

  - Symmetry determination

  - Rigid body model fitting

  - Map sharpening

  - Map segmentation

  - Automated model building

  - [Focused map/model combination]

- Atomic model optimization and validation

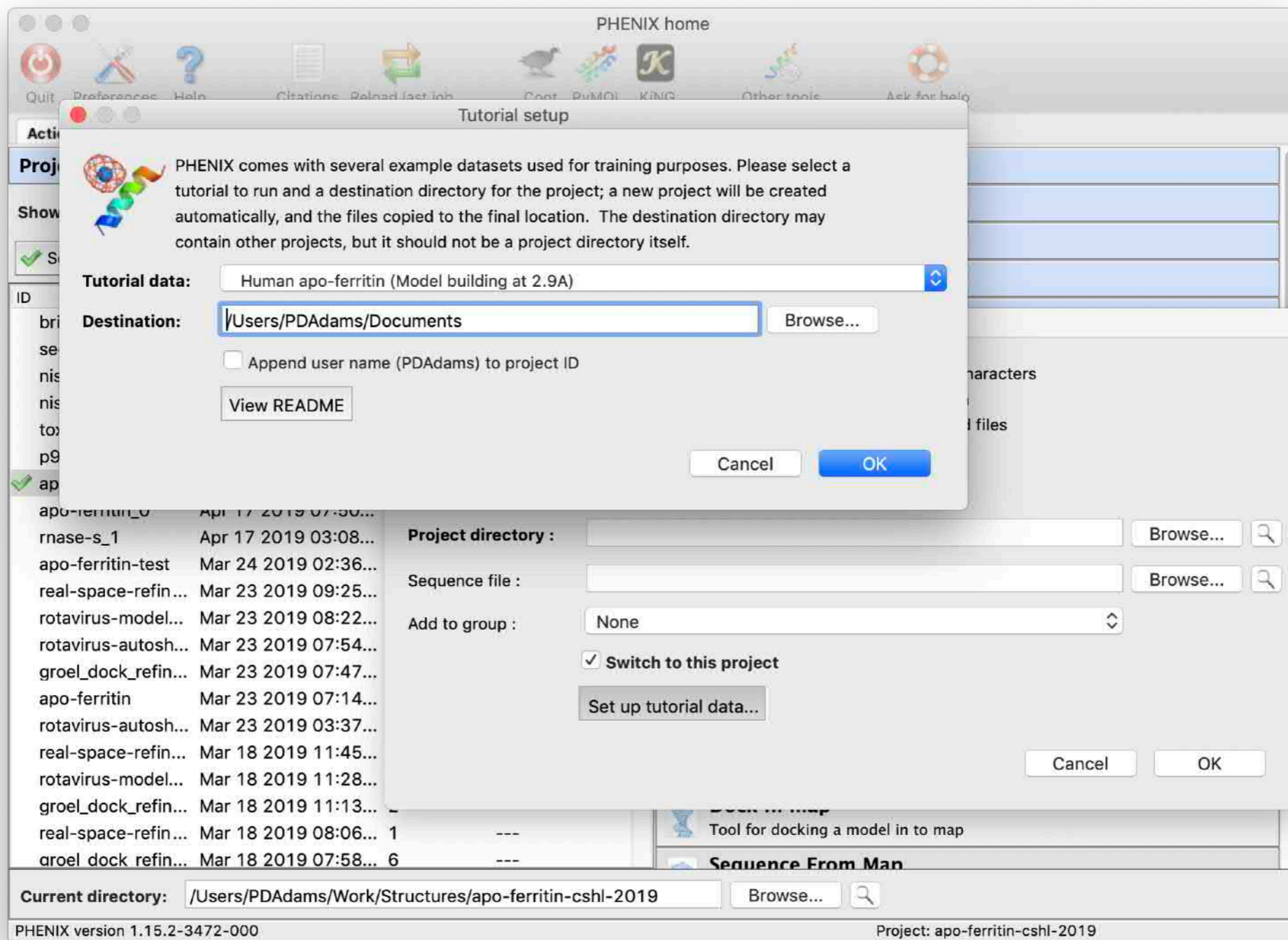 - Structure refinement

  - Validation

Tutorial Format

- Use graphical user interface

Tutorial Format

- Use tutorial datasets distributed with Phenix
- Should run on most laptops (2GB RAM, multiple CPUs better)

Challenges

- Automated model building
  - What is the magnification of the map? (can be 5% uncertainty)
  - What is the optimal sharpening of the map?
  - What is the region containing the molecule?
  - Low and variable resolution across maps

- Structure optimization
  - Variable resolution across maps
  - Large molecules
  - Poor initial models

- Validation
  - How to validate a model against moderate resolution maps

Phenix

Automated Model Docking

Tom Terwilliger
Los Alamos National Laboratory
Pavel Afonine, Oleg Sobolev
Lawrence Berkeley National Laboratory

Phenix

Automated Model Docking

- Systematic cross correlation search of rotations and translations
- Performed in reciprocal space using FFT (very fast)
- Rigid body optimization of position

EMD8750

1SS8 chain A

Phenix
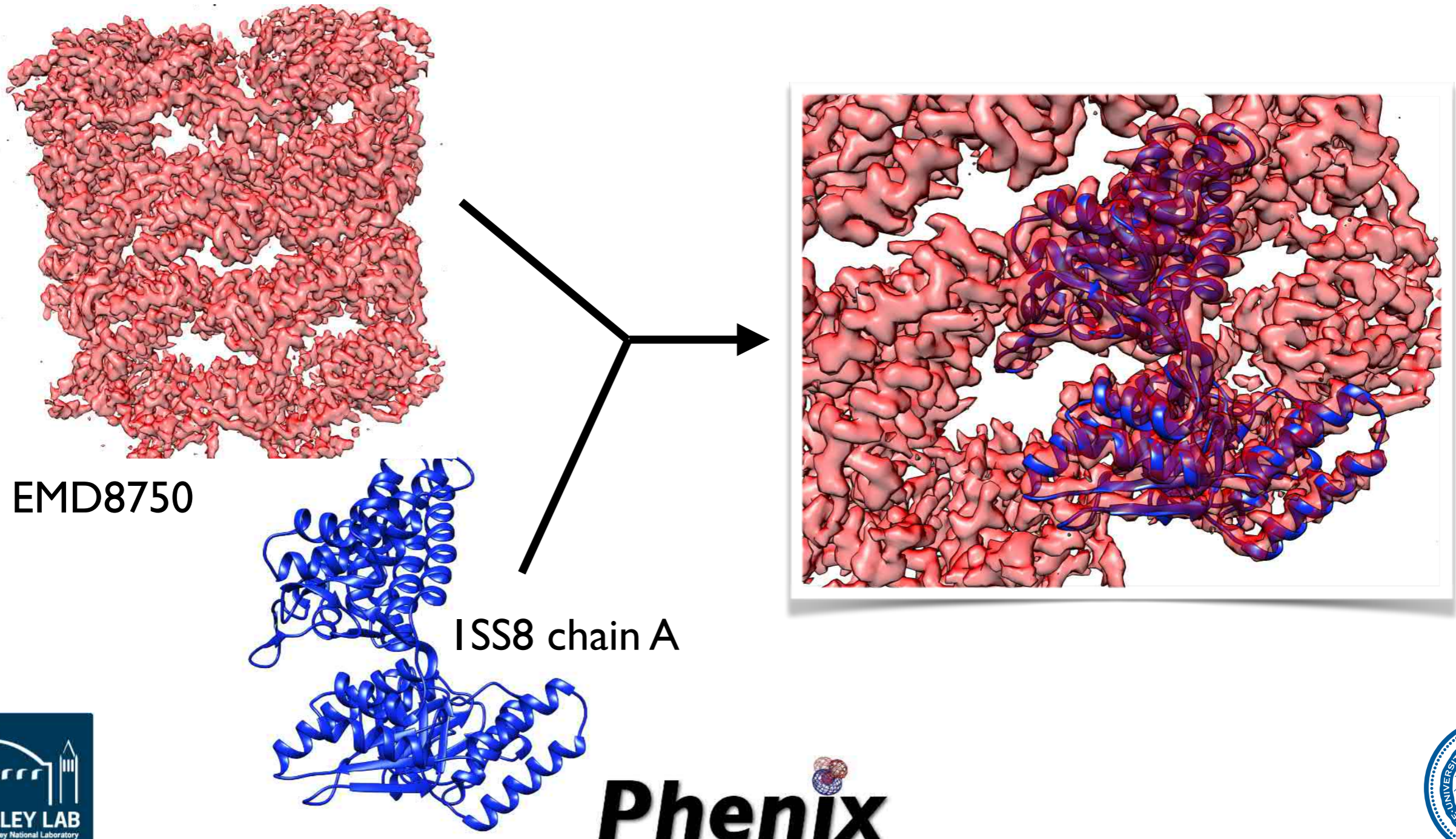
Automated Model Sharpening, Segmentation and Model Building

Tom Terwilliger
Los Alamos National Laboratory
Pavel Afonine, Oleg Sobolev
Lawrence Berkeley National Laboratory

Phenix

Automated Model Building Procedure

Determine optimal sharpening of the map

↓

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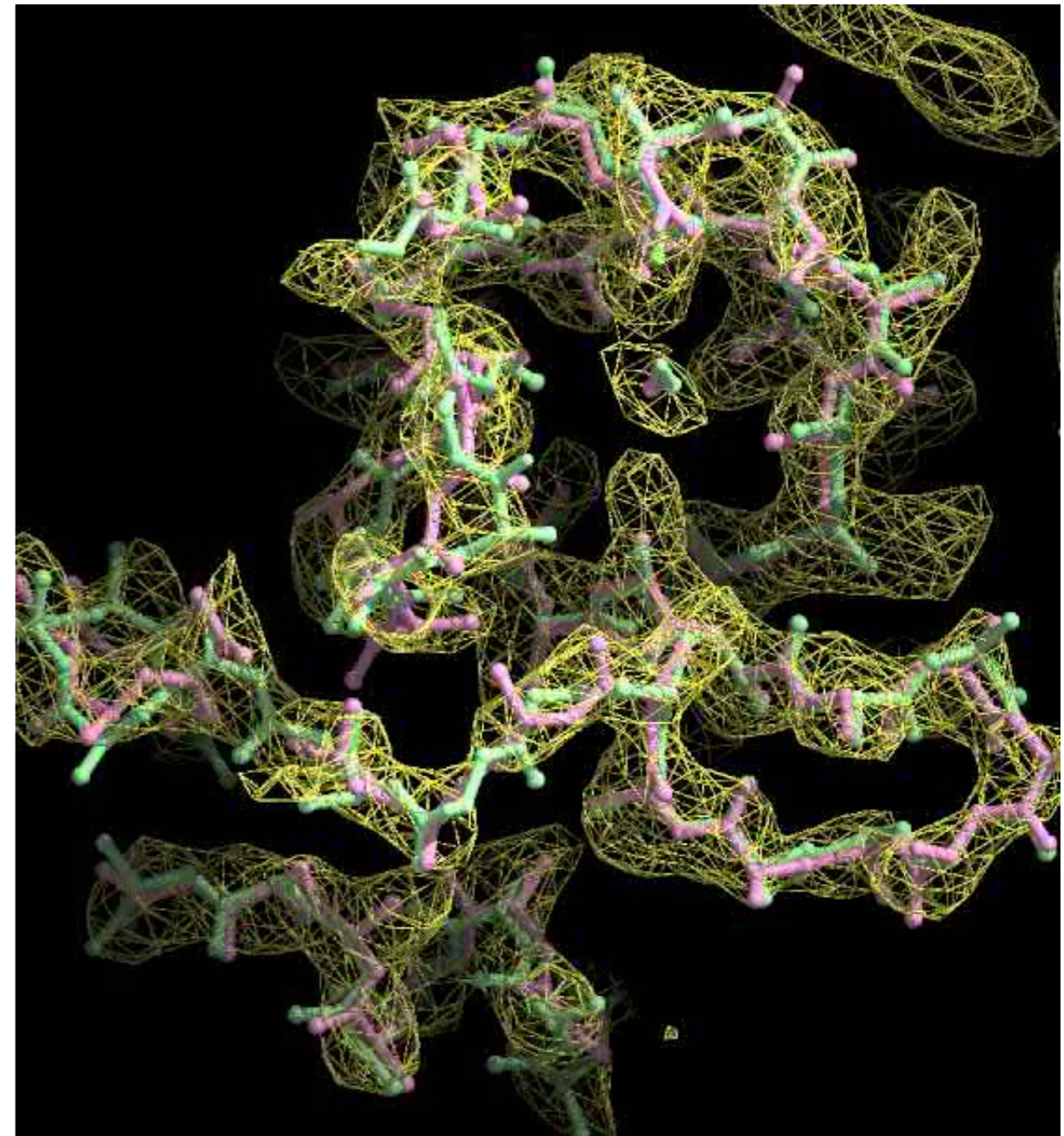
Cryo-EM map from the yeast mitochondrial ribosome (chain I of large subunit, 3.2Å, Amunts *et al.*, 2014)

Autobuilt model (pink)
Deposited model (green)

Terwilliger et al. A fully automatic method yielding initial models from high-resolution electron cryo-microscopy maps. *Nature Methods*, in press

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

Automated Map Sharpening

Create series of maps with variable overall B-values

Analyze maps for detail and connectivity
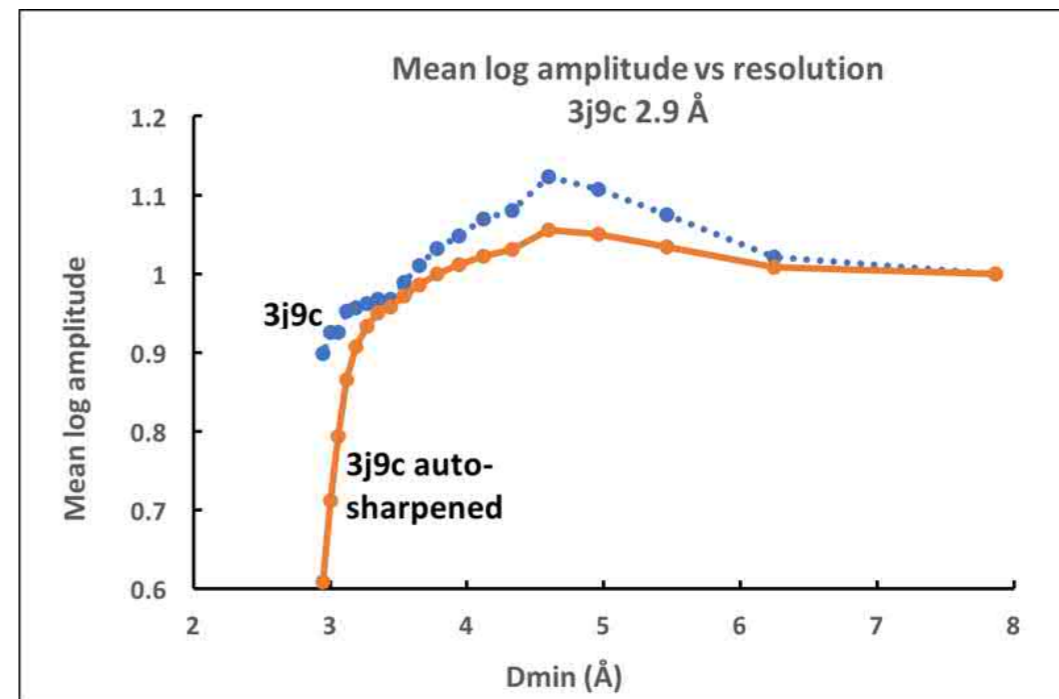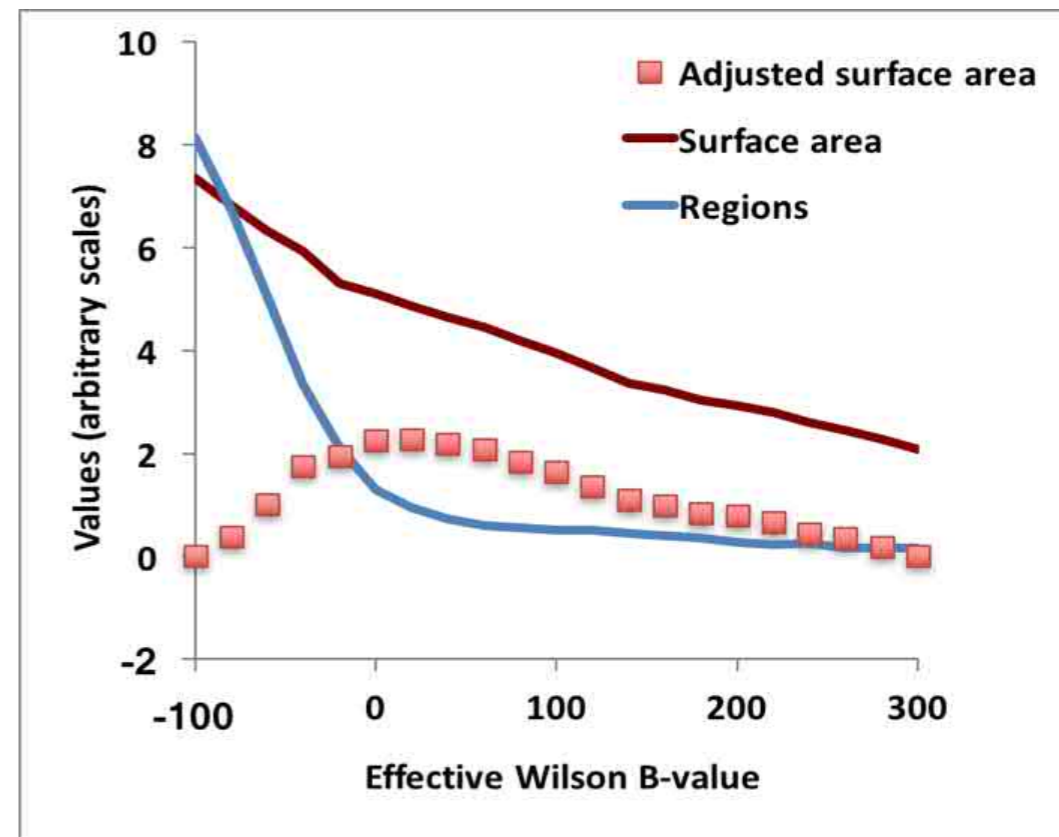
Set contour level enclosing 20% of molecular volume

Calculate surface area of contours

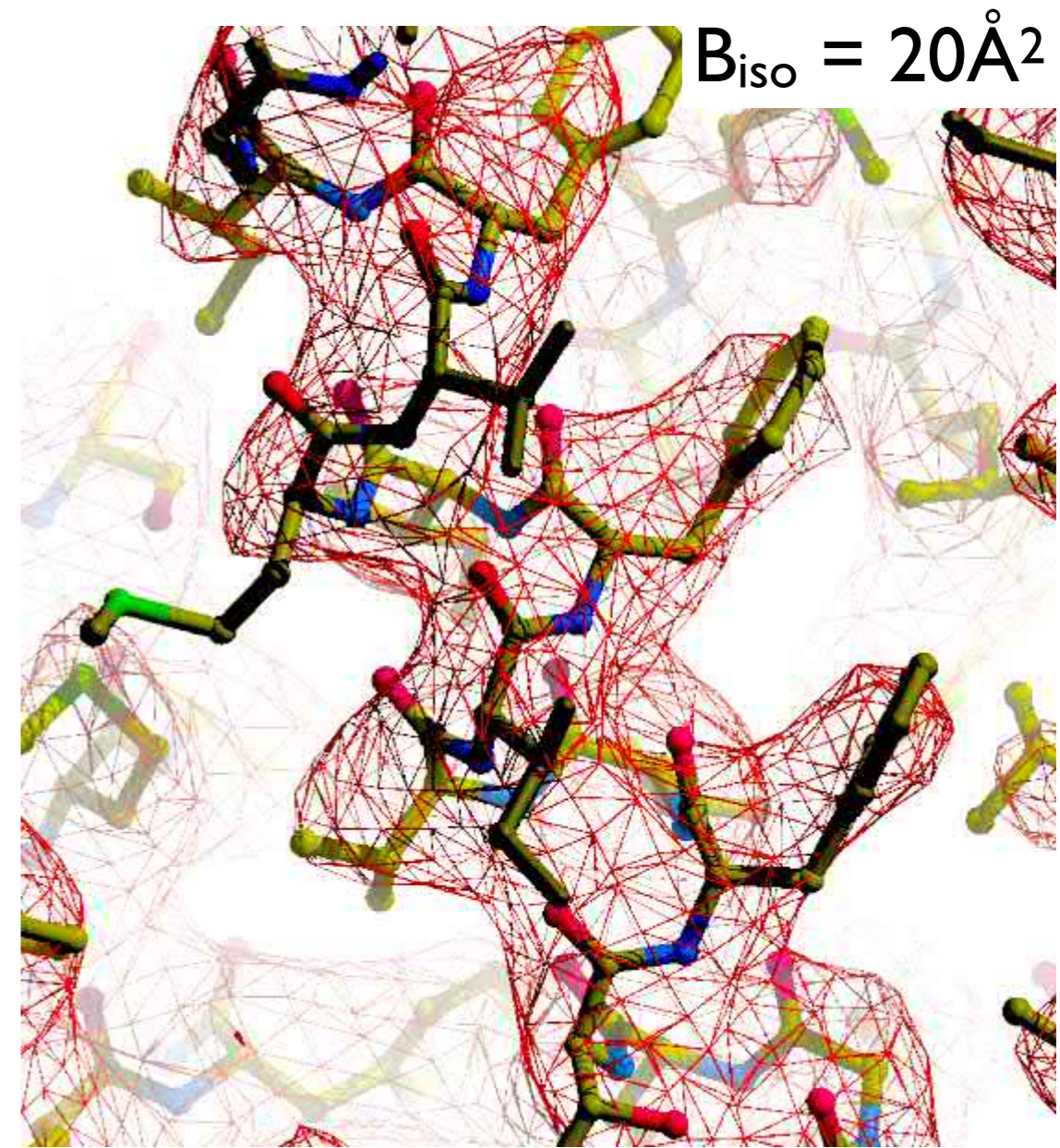Count number of distinct regions enclosed by contours

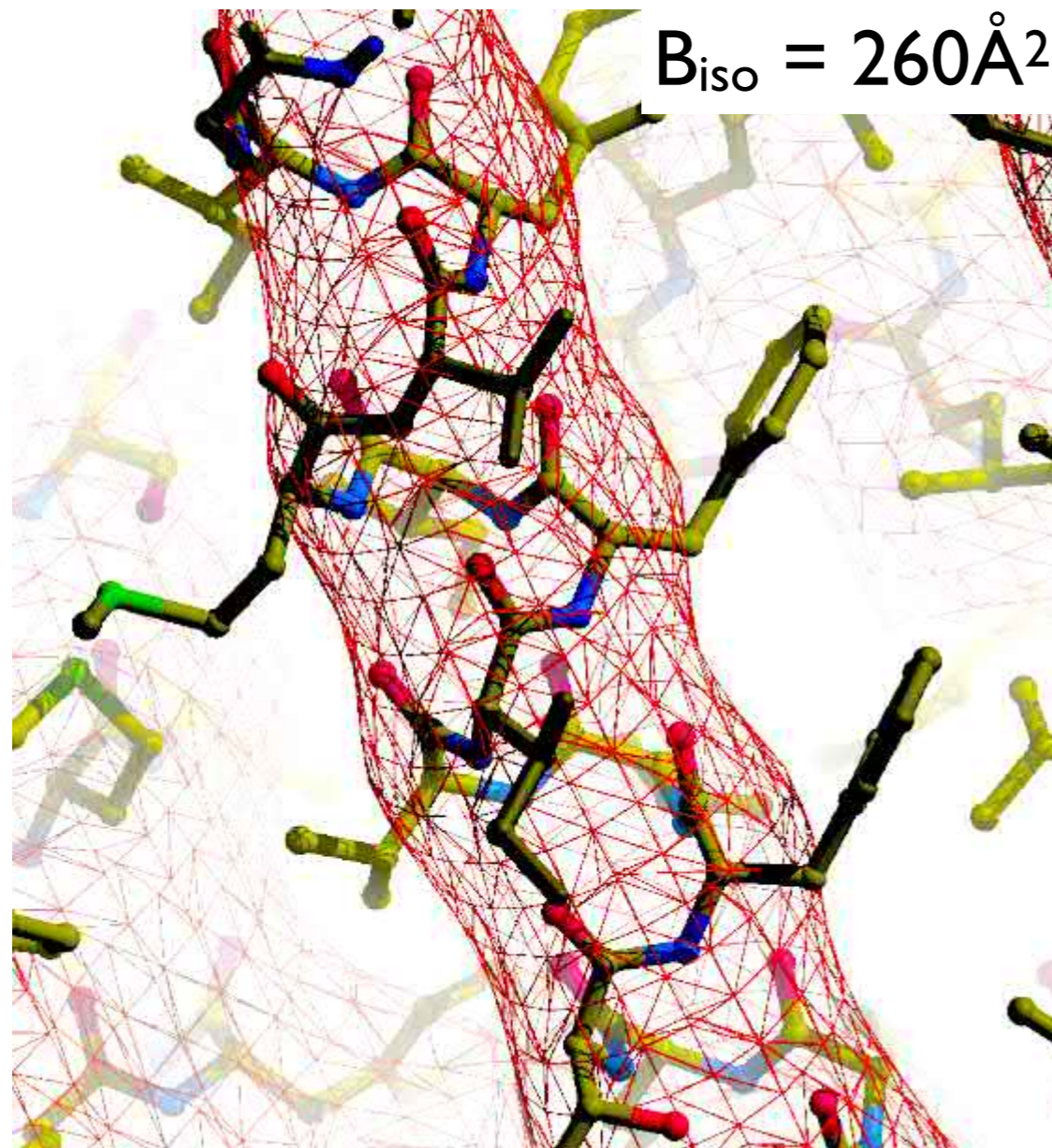Choose map with maximum of adjusted surface area

adjusted area = surface area – weight * number of regions

phenix.auto_sharpen

Automated Map Sharpening
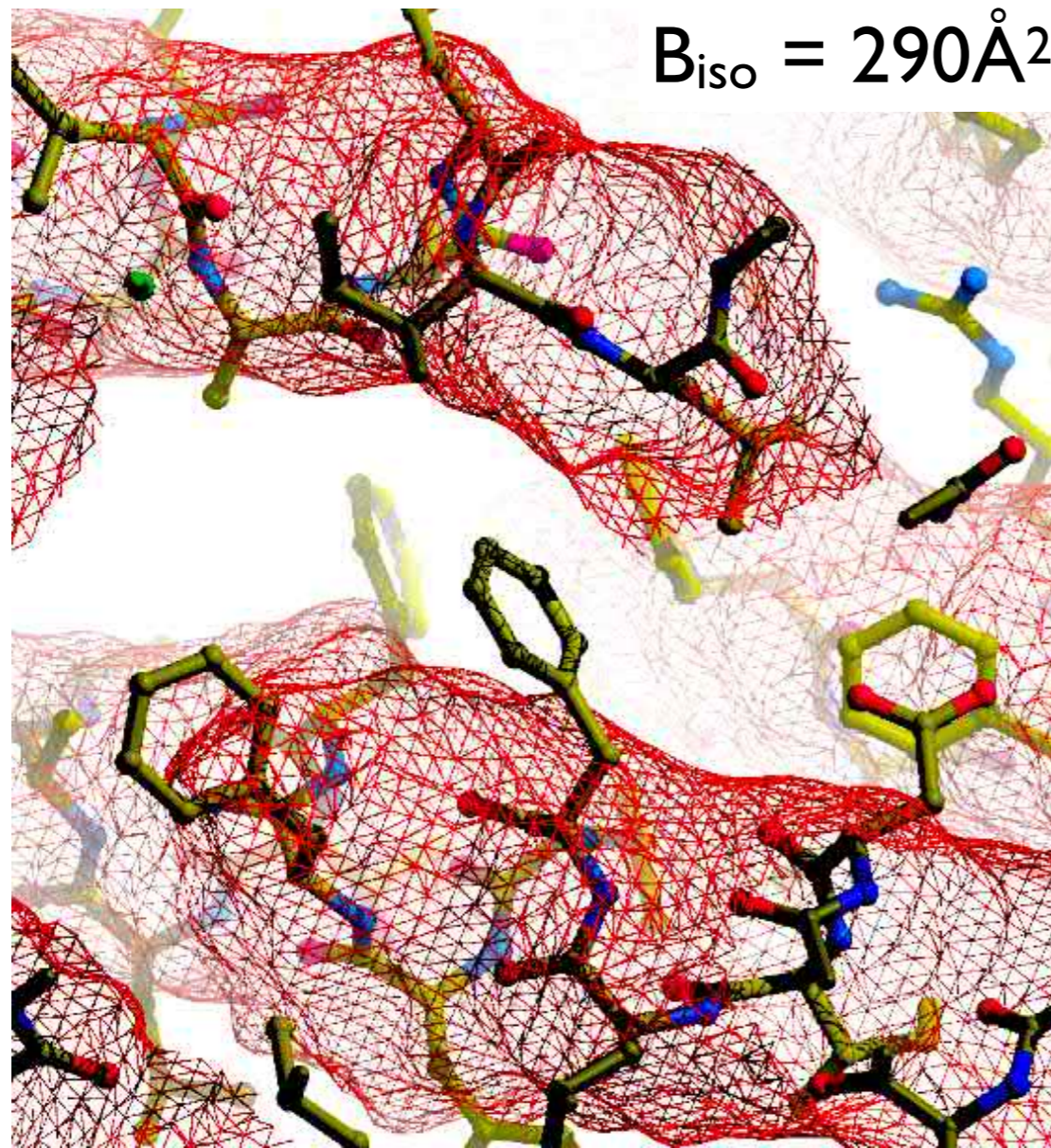

$B_{iso} = 260 Å^2$

$B_{iso} = 20 Å^2$

Deposited Map

Autosharpened Map

High-conductance Ca(2+)-activated K(+) channel (emd_8414 and PDB entry 5tji; Hite et al., 2017)

Phenix

BERKELEY LAB
Lawrence Berkeley National Laboratory
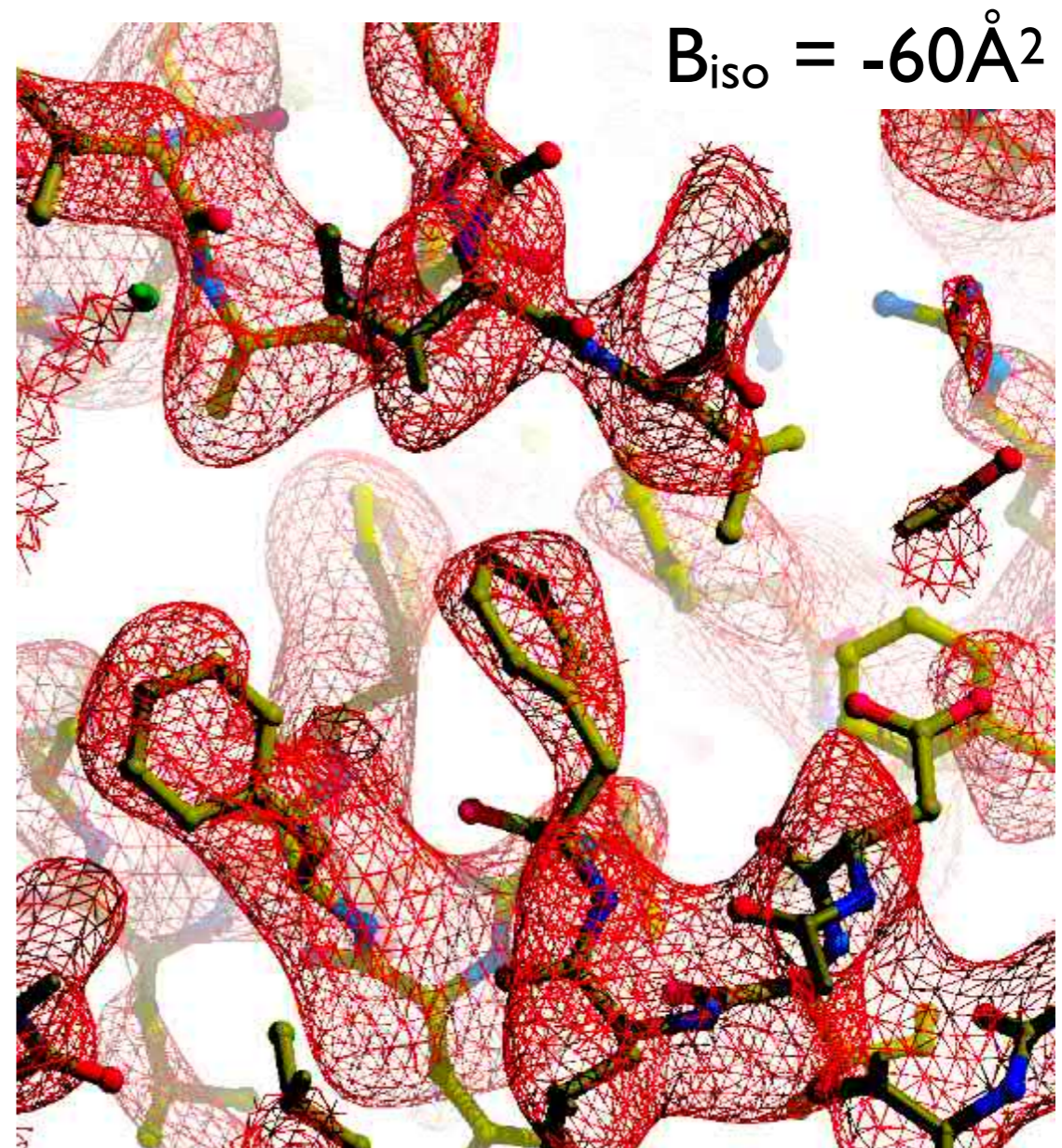
Automated Map Sharpening

$B_{iso} = 290\text{Å}^2$
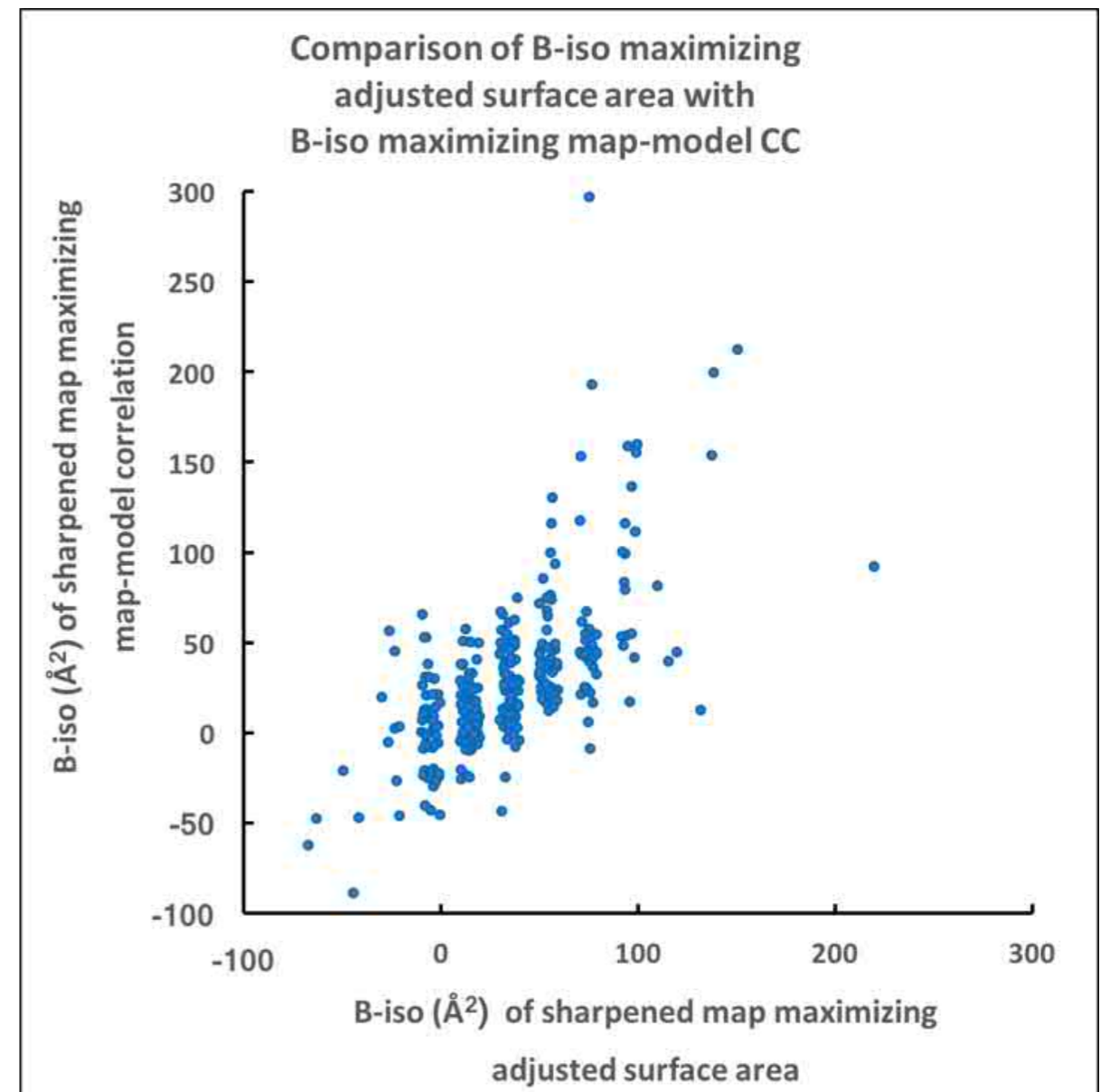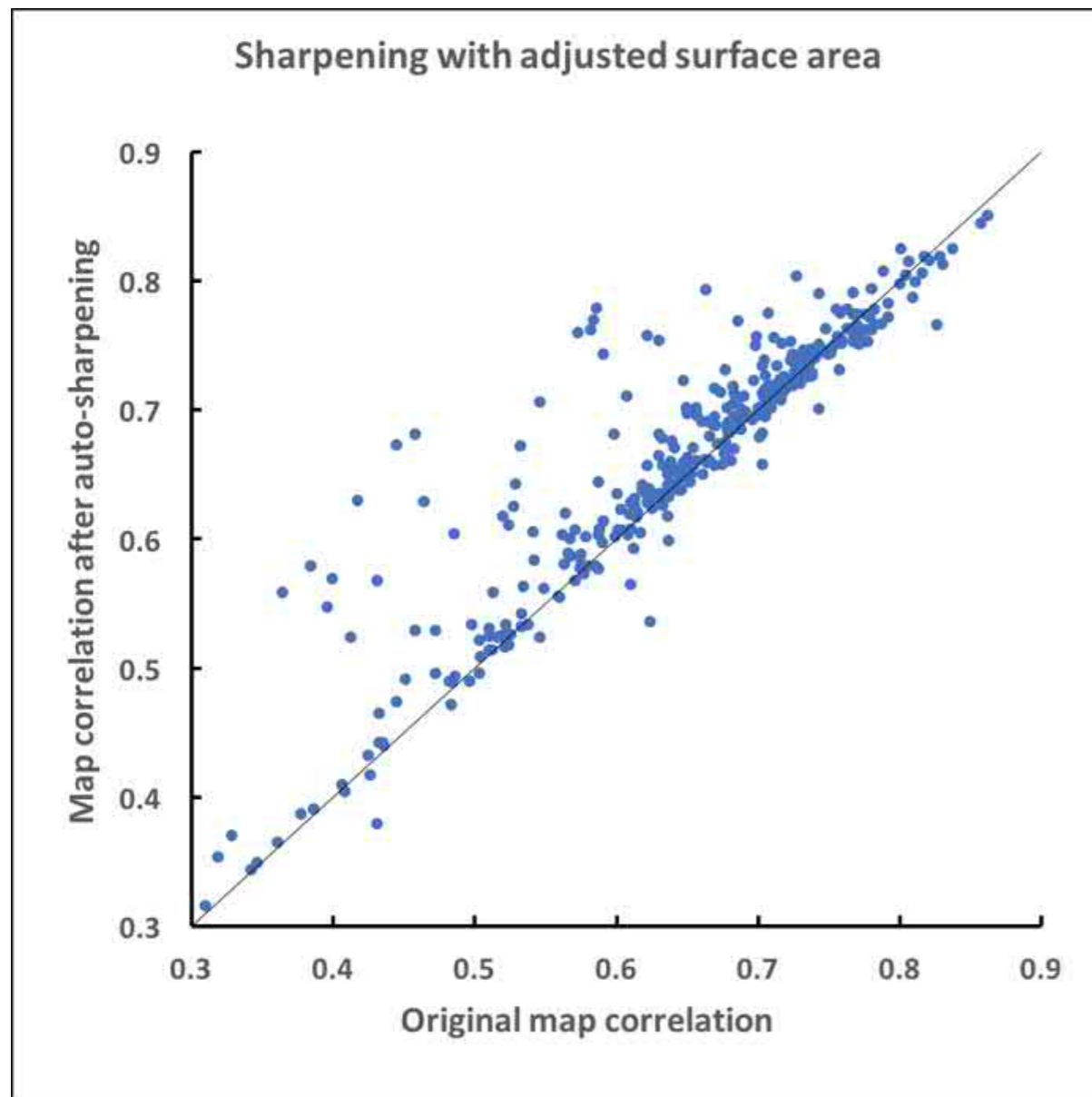
$B_{iso} = -60\text{Å}^2$

Deposited Map

Autosharpened Map

Cystic fibrosis transmembrane conductance regulator
(emd_8461 and PDB entry 5uar; Zhang and Chen, 2016)

Phenix

Automated Map Sharpening

Terwilliger et al. Automated map sharpening by maximization
of detail and connectivity. *Acta Cryst* 2018, **D74**:545-559

Automated Segmentation

Determine optimal sharpening of the map

⬇

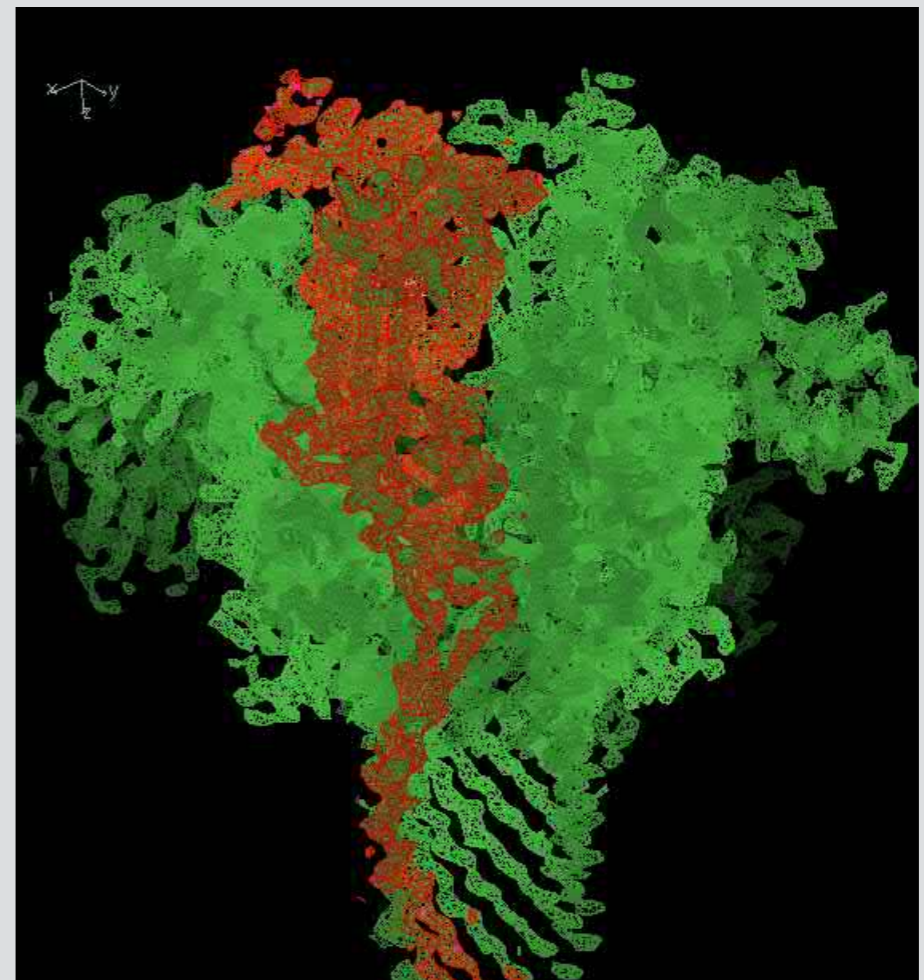Cut out asymmetric unit of the map

⬇

Trace chain and build model

⬇

Idealize secondary structure and refine

⬇

Assemble and refine (protein/RNA/DNA)

⬇

Apply molecular symmetry and re-refine

- Use the symmetry of the map
- Identify contiguous regions representing asymmetric unit of the map
- Choose symmetry-copies that make compact molecule

Terwilliger et al. Map segmentation, automated model-building and their application to the Cryo-EM Model Challenge. *J. Struct. Biol.* 2018, in press

emd_6224 (anthrax toxin protective antigen pore at 2.9 Å; Jiang *et al.* 2015)

Phenix

BERKELEY LAB
Lawrence Berkeley National Laboratory

Chain Tracing

Determine optimal sharpening of the map

↓

Cut out asymmetric unit of the map

↓

Trace chain and build model

↓

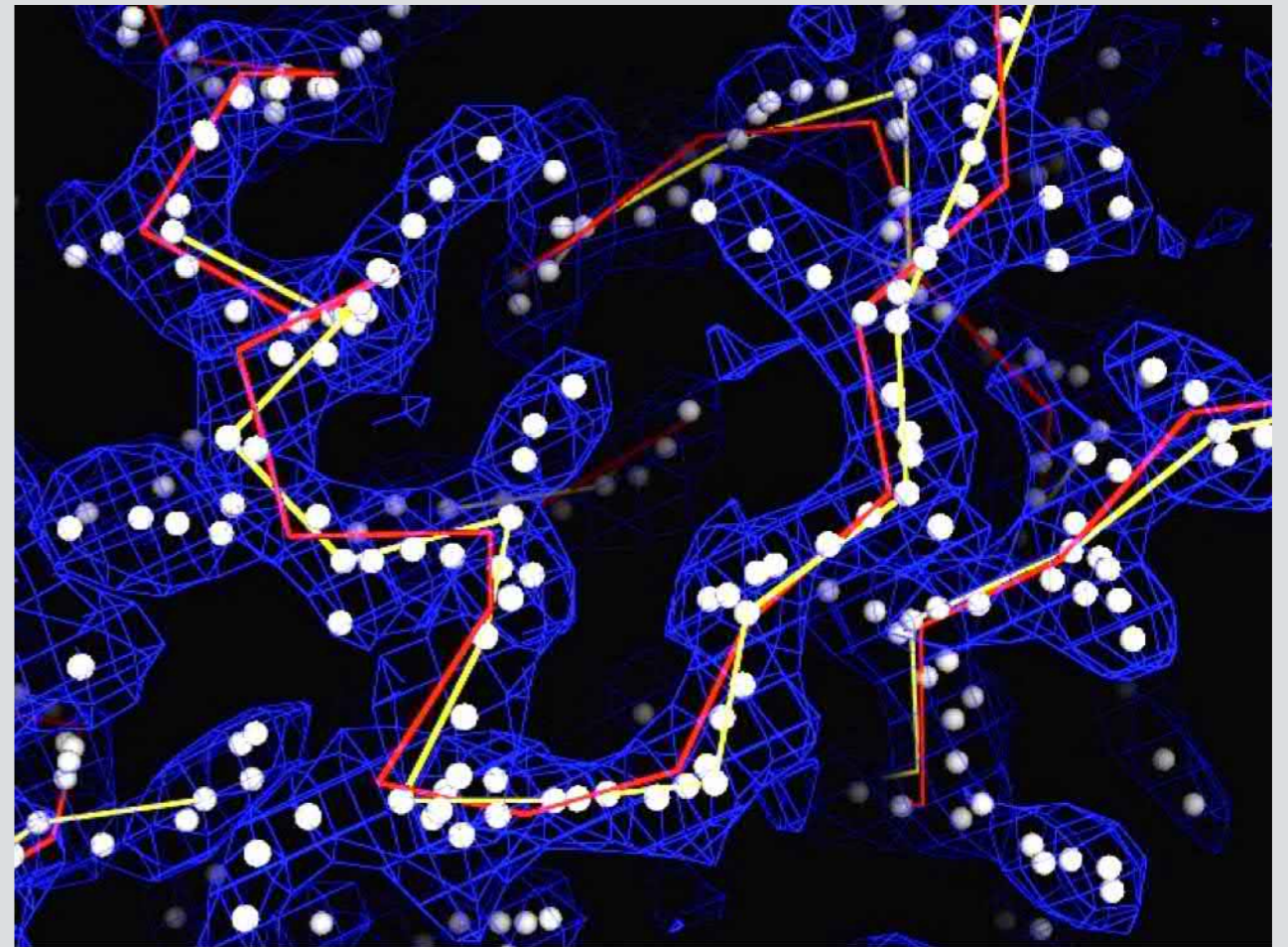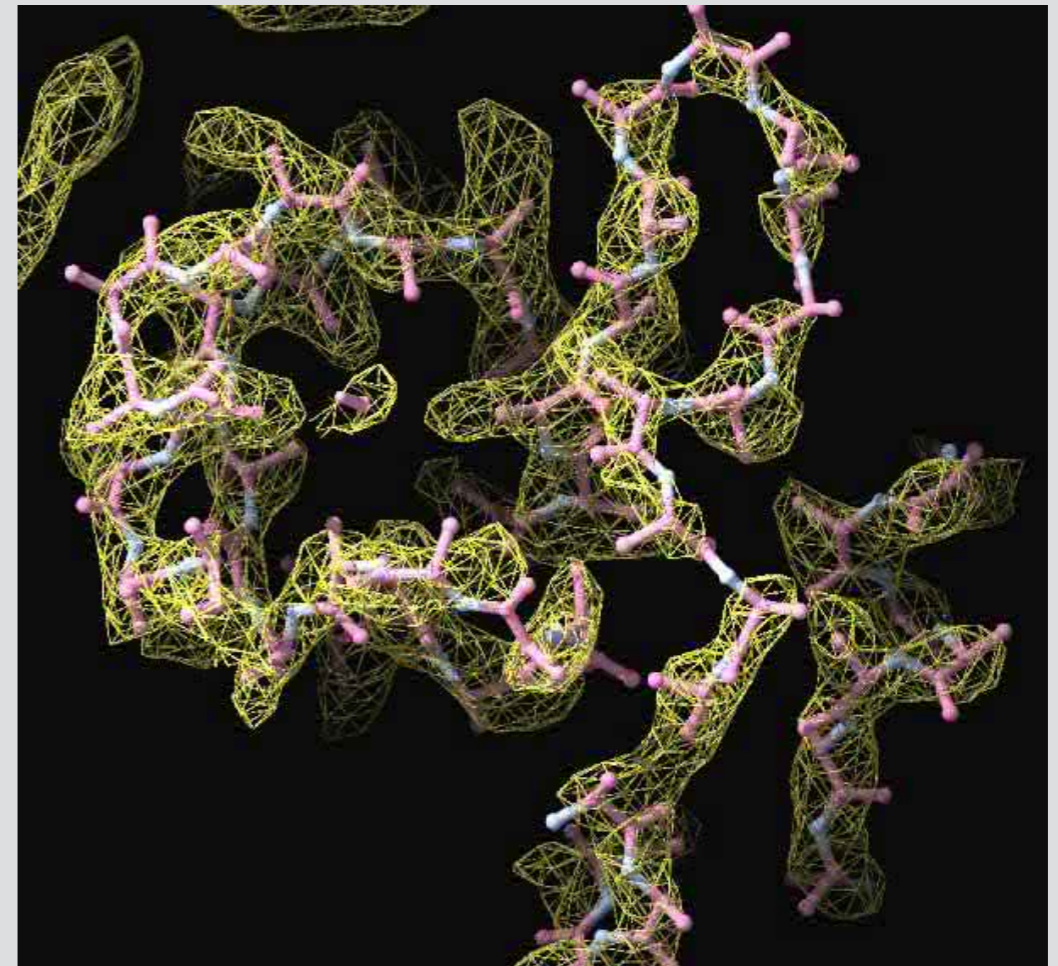Idealize secondary structure and refine

↓

Assemble and refine (protein/RNA/DNA)

↓

Apply molecular symmetry and re-refine

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

Phenix

Idealization and Refinement

```
Determine optimal sharpening of the map
                    ↓
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
```

- 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

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

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

Assembly and Polymer Recognition

Determine optimal sharpening of the map

↓

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

- Try building protein/RNA/DNA (whatever may be there)
- Choose segment type by map correlation
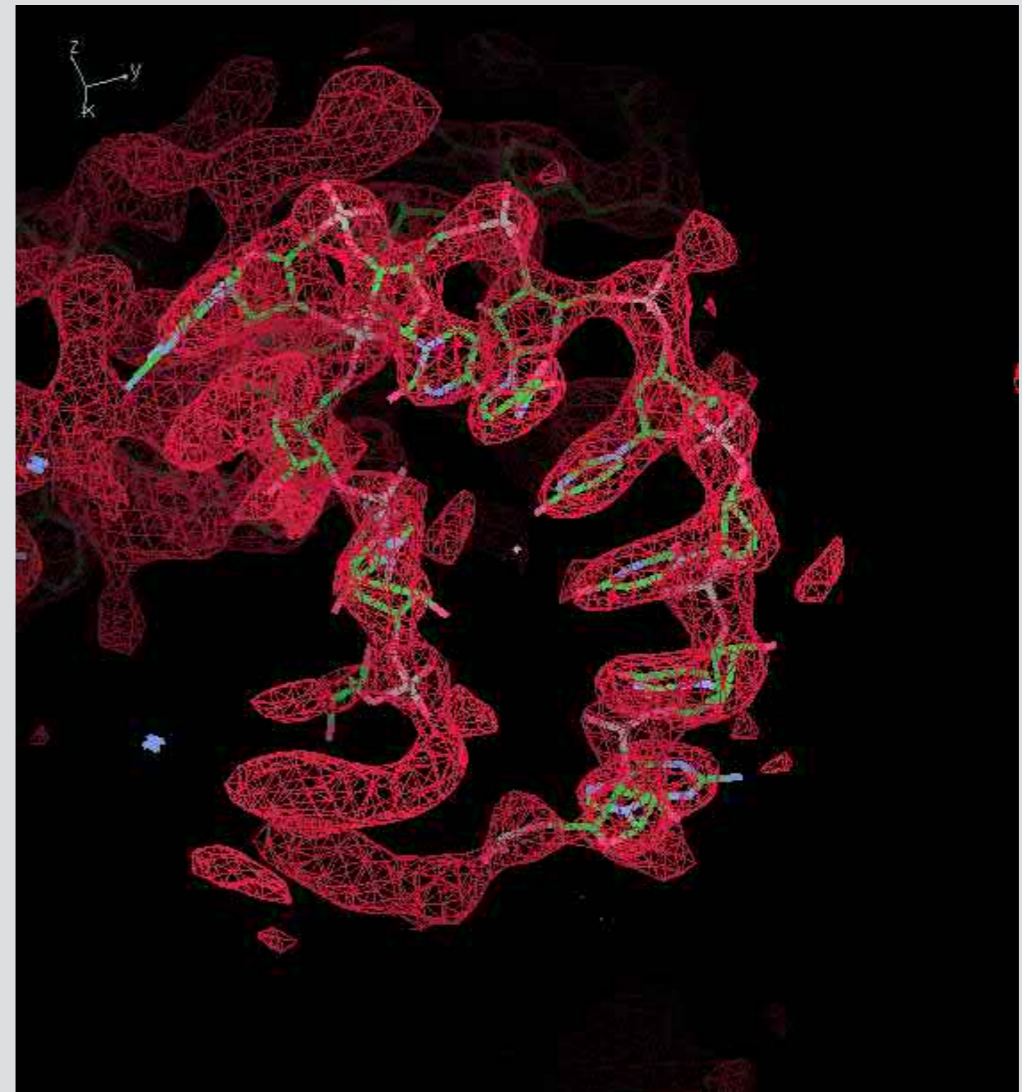
70S ribosome at 2.9 Å

Phenix

The Final Model

Determine optimal sharpening of the map

⬇

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

- `phenix.map_to_model`

30S Ribosome (1j5e, 2.9 Å)

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix
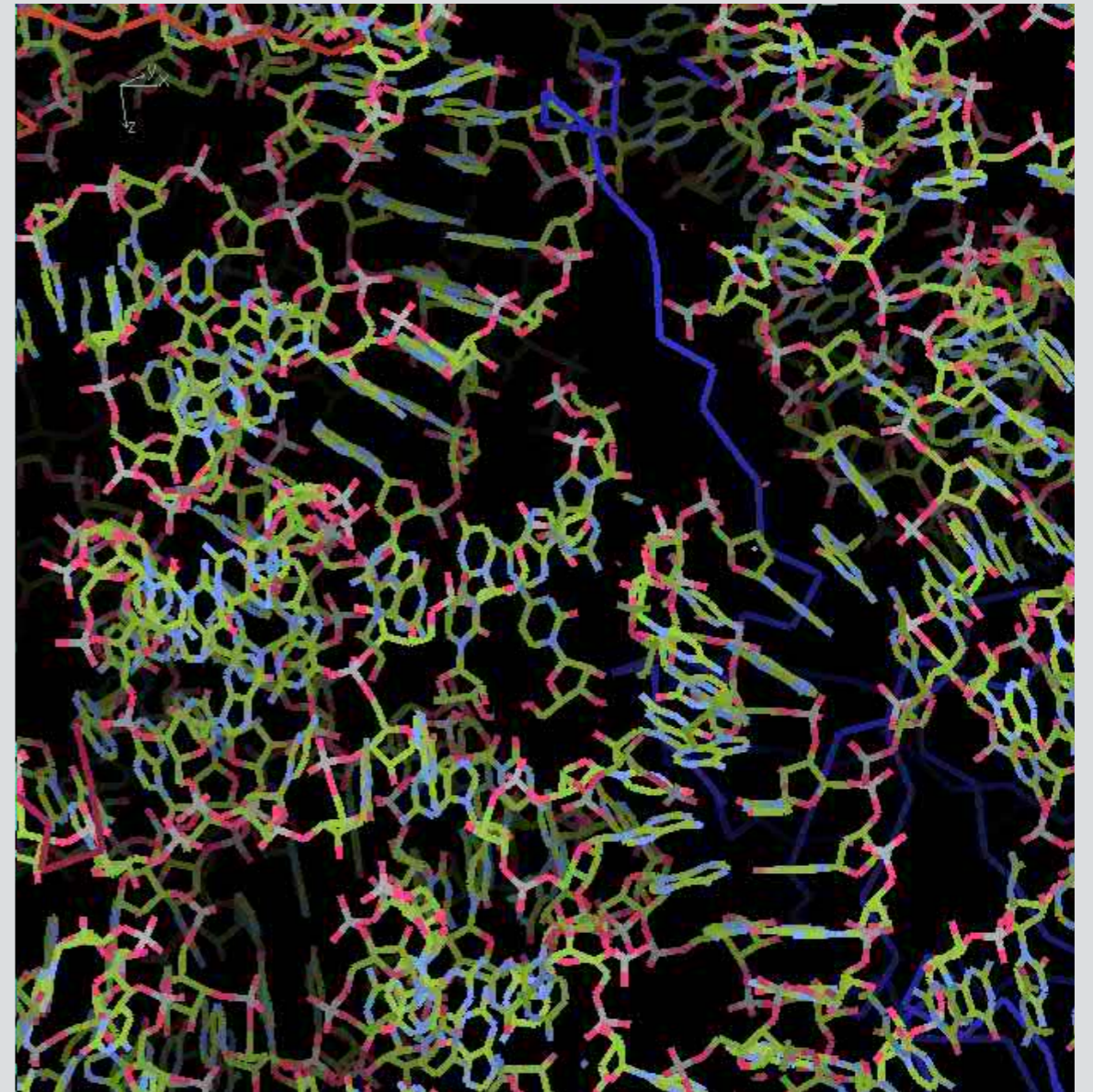
Automated Building - Sharpening
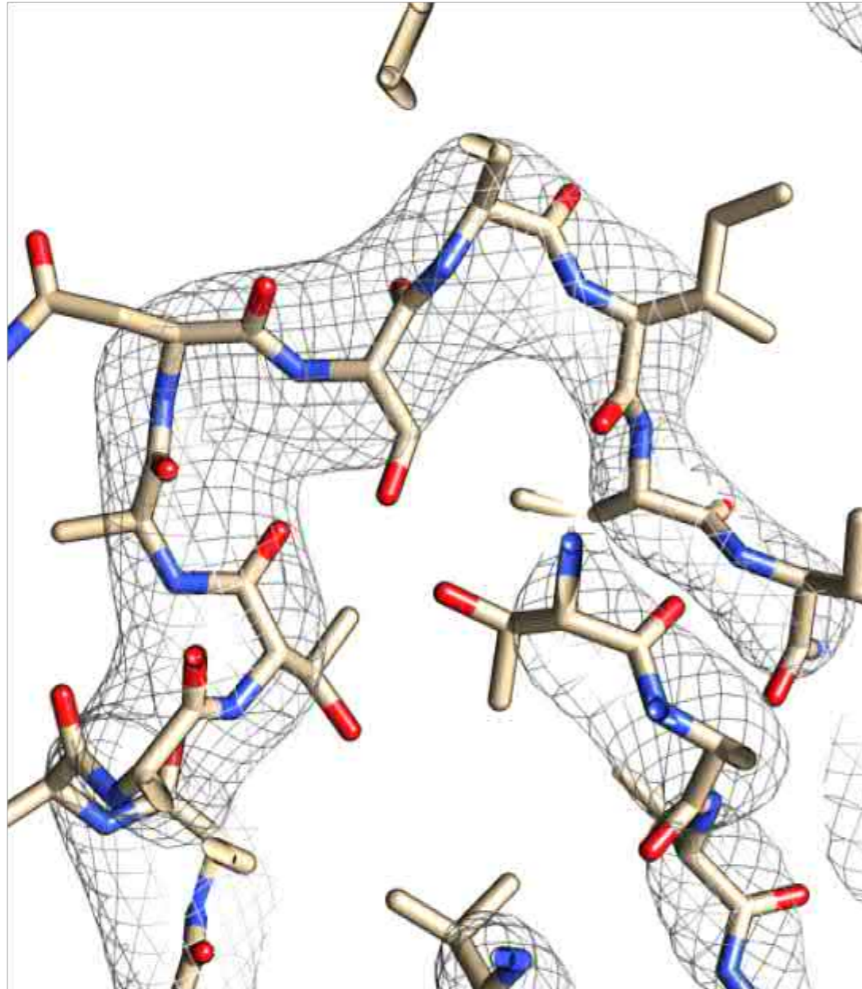
Original
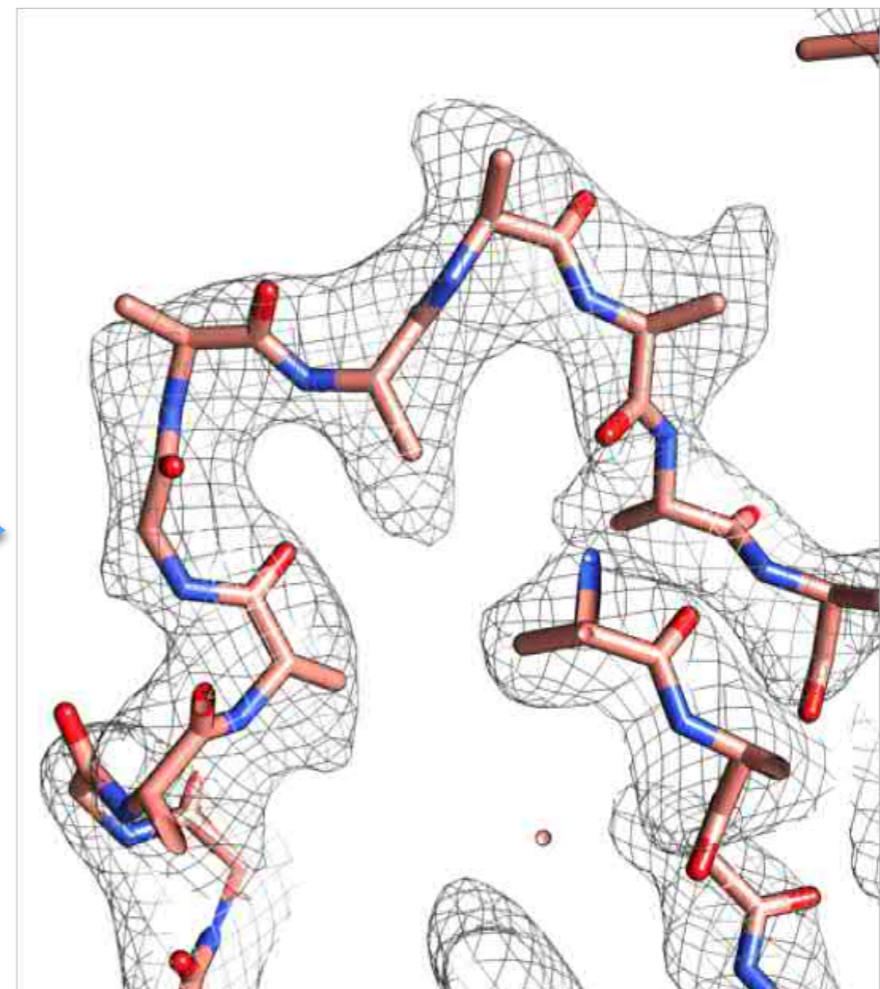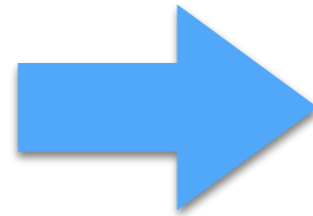
Automatically Sharpened

Automated Building - Combining Multiple Models
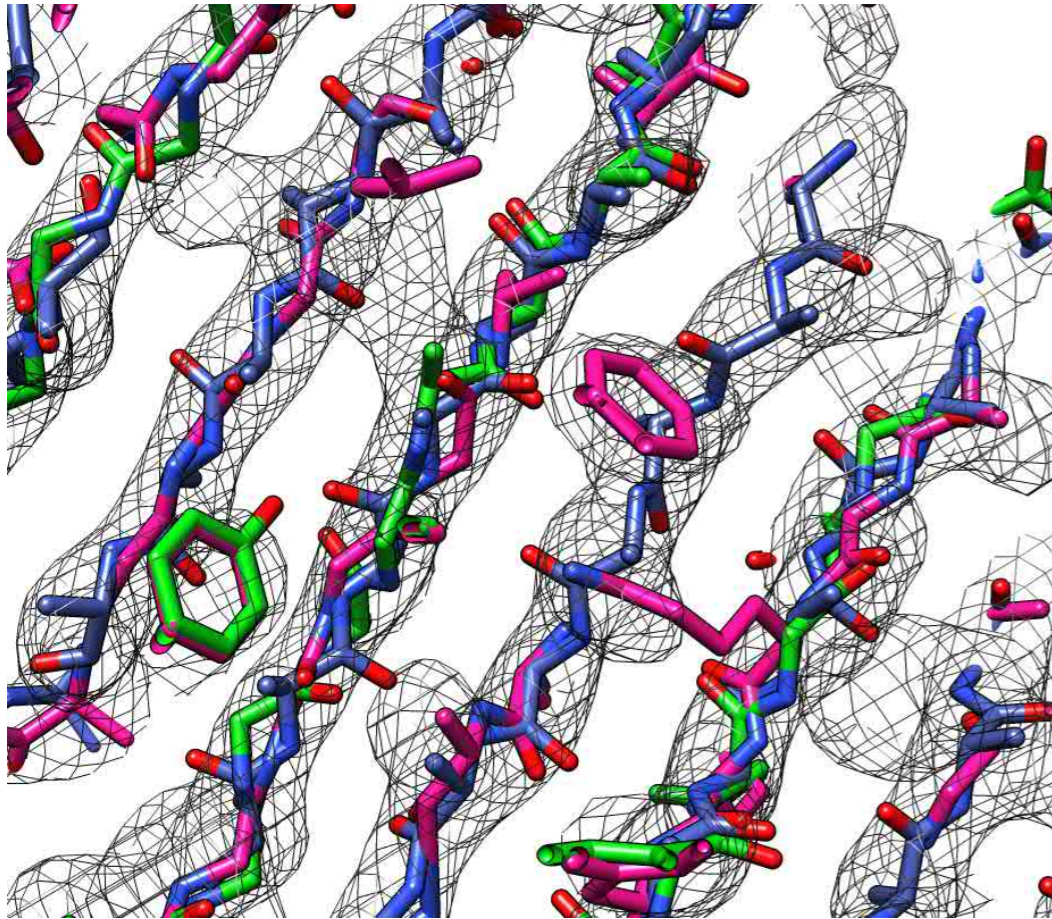
Three Independently Built Models
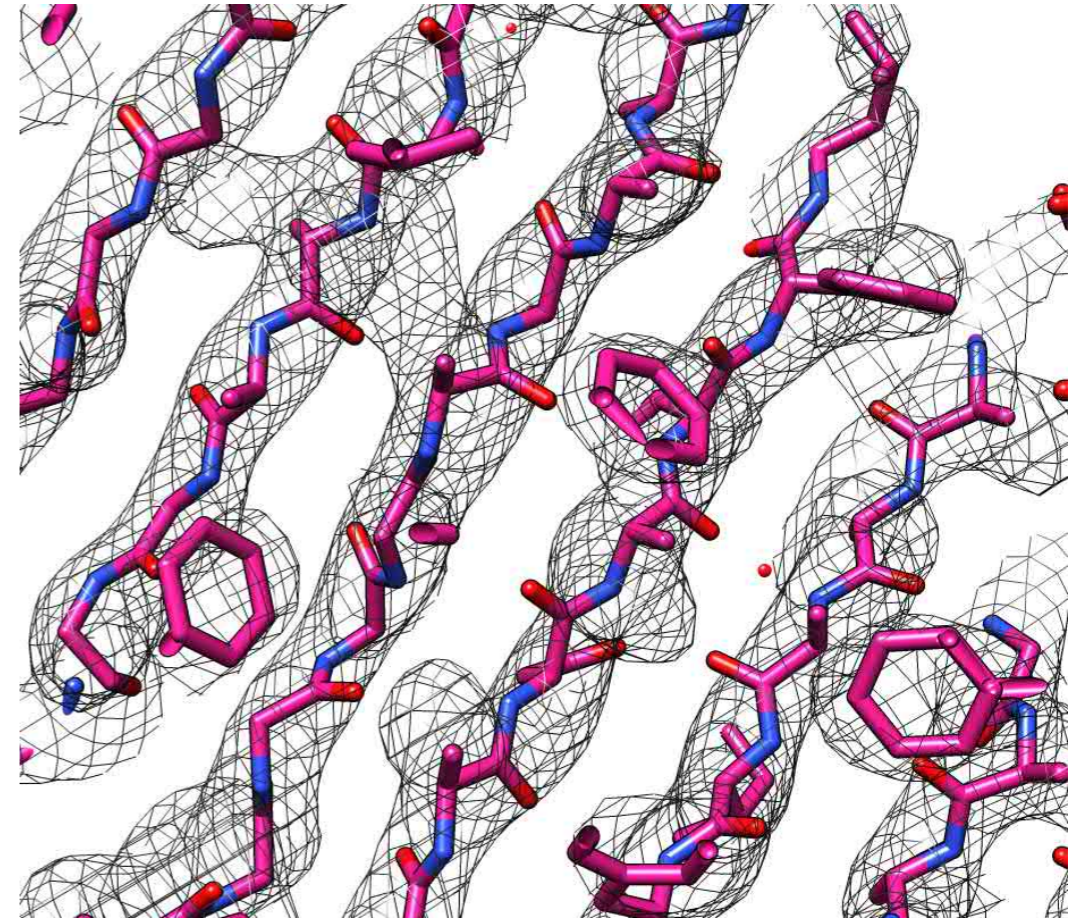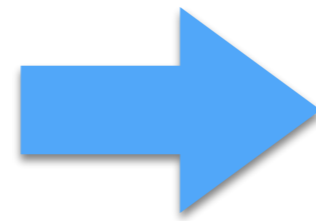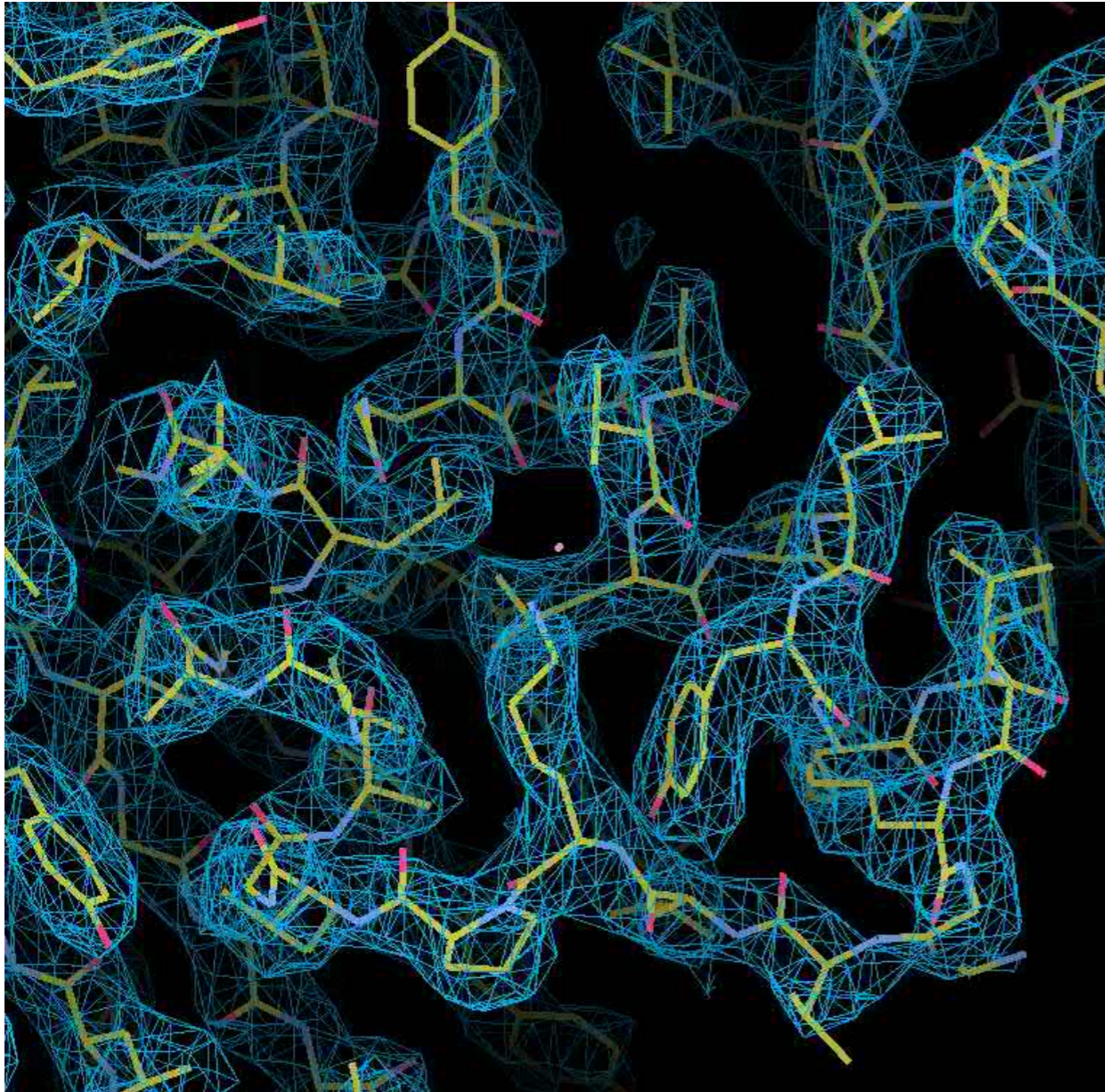
Composite Model

Building at Low Resolution
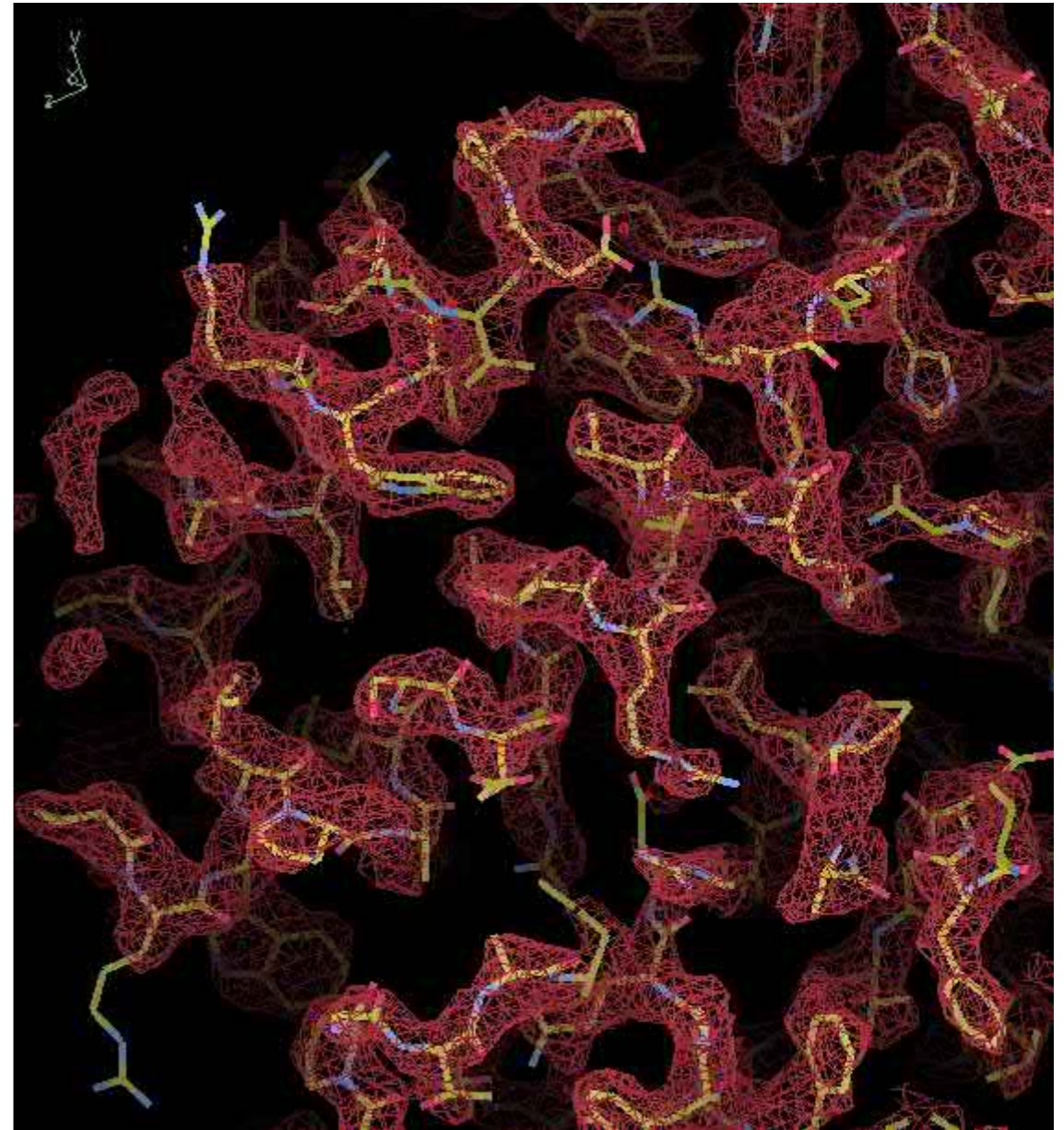

Gamma-secretase at 4.5 Å
(autobuilt model; emd_2677)
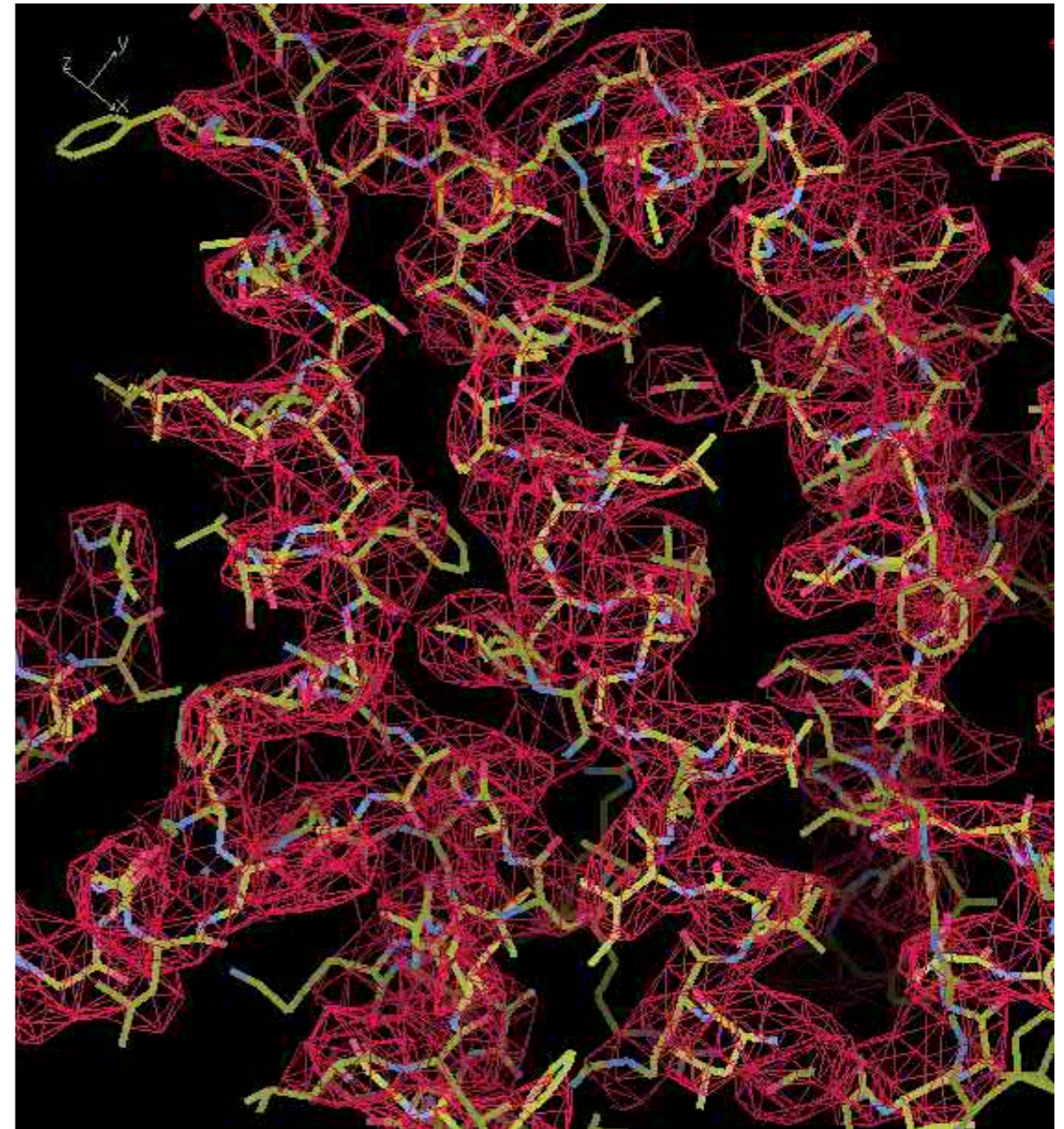
Gamma-secretase structure at 3.4 Å
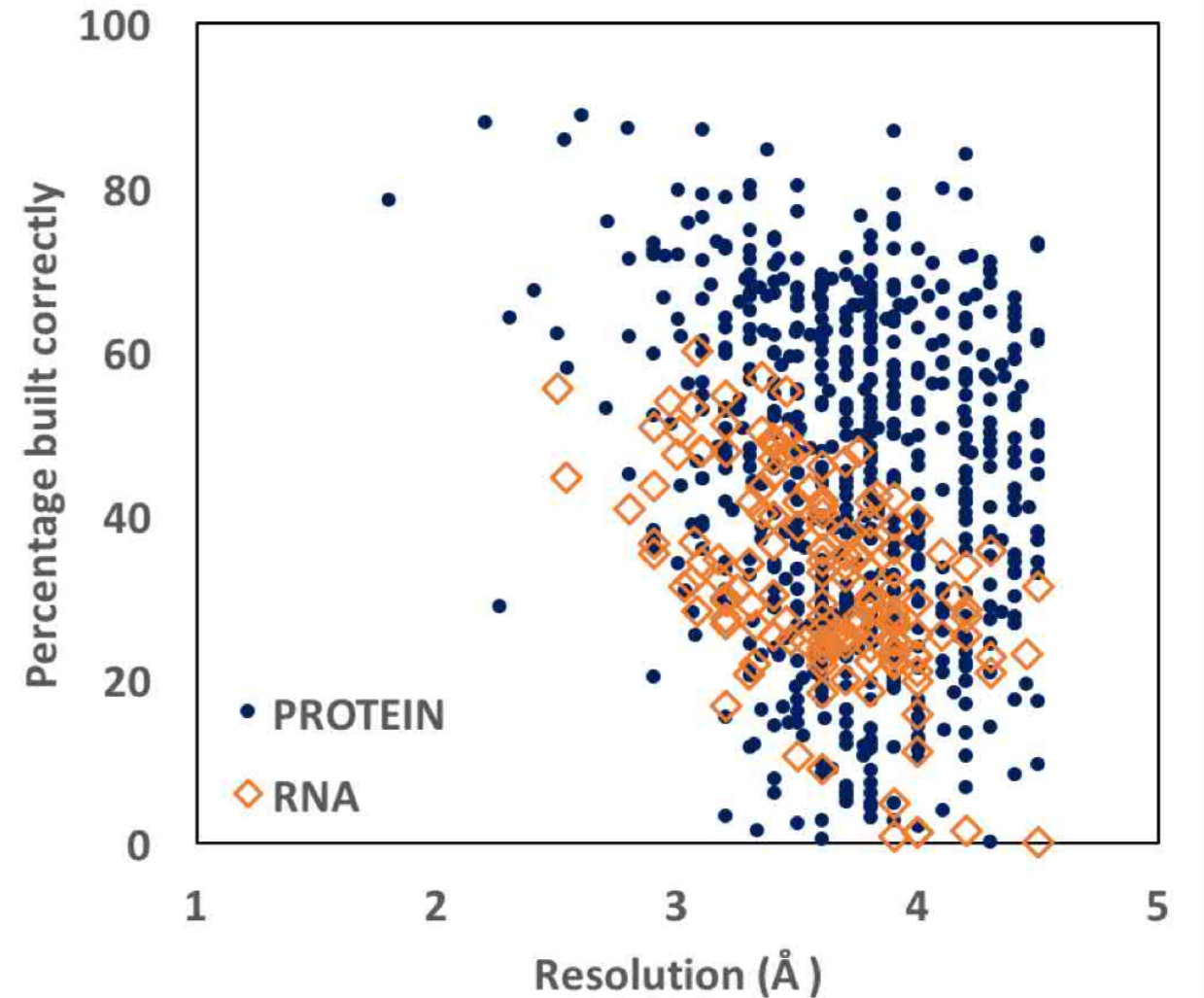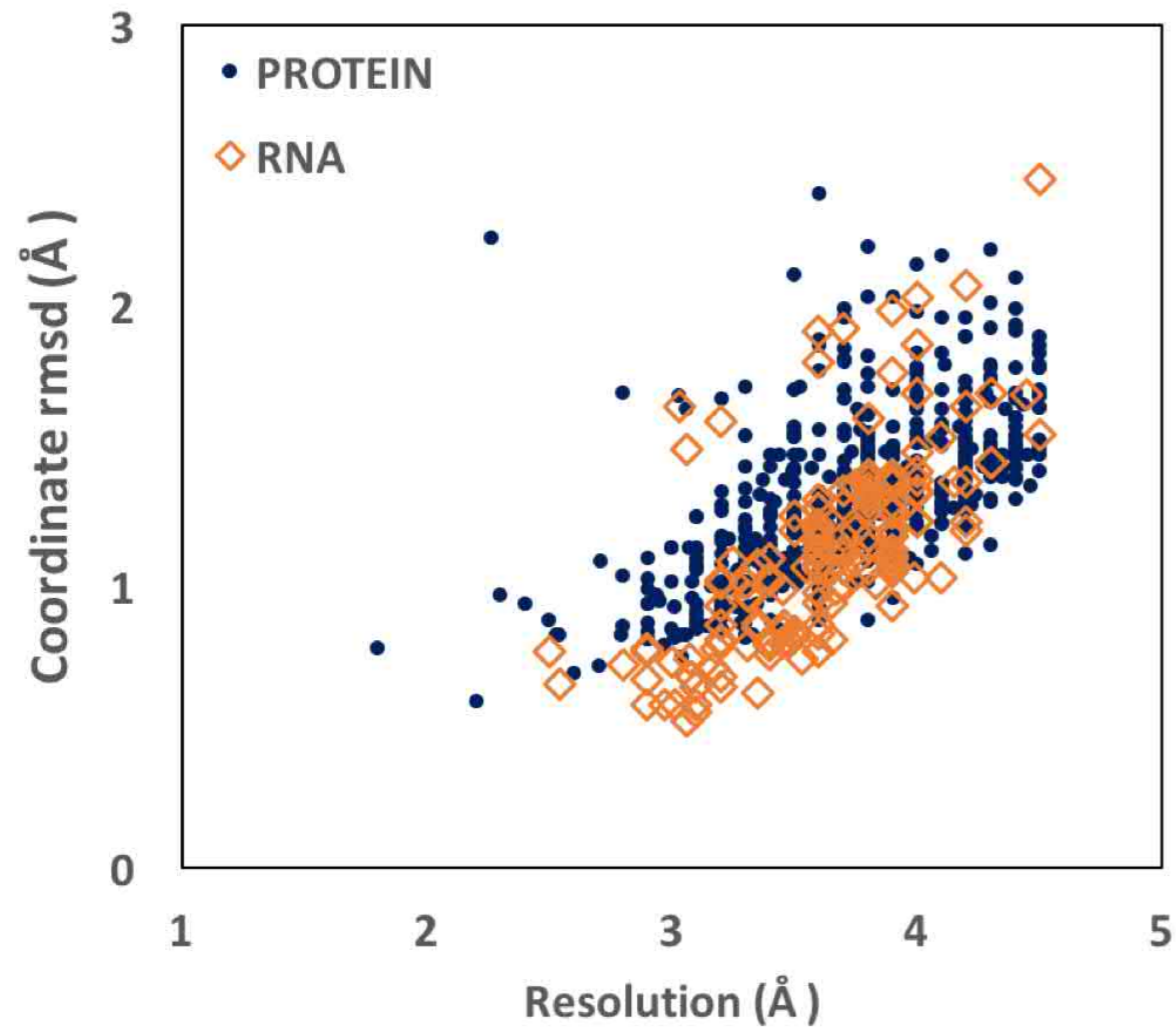(autobuilt model; emd_3061)

Building at Medium/High Resolution

Proteasome at 2.8 Å
(autobuilt model; emd_6287)

Beta-galactosidase at 2.2 Å
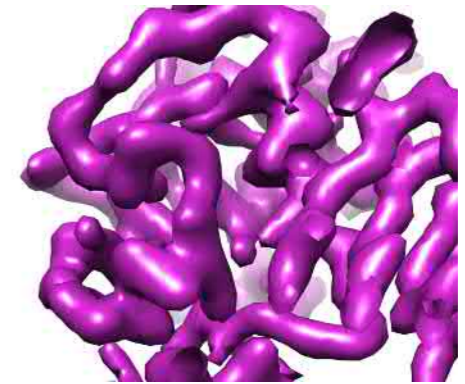(autobuilt model; emd_2984)
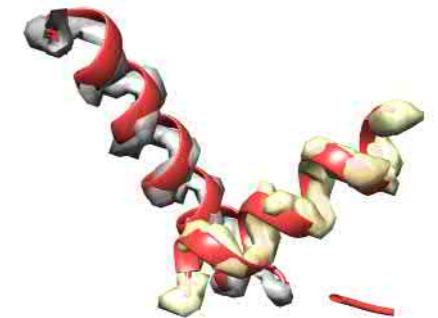
Phenix

Autobuilding Performance

Model Building Version 2
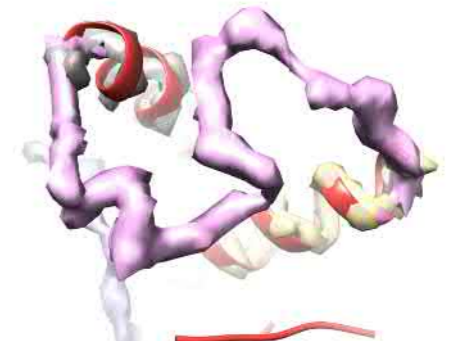
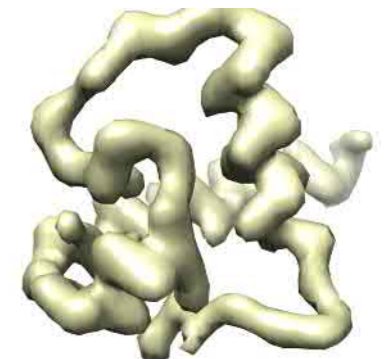Trace chain the way a person does

Find secondary structure

Find clear regions of density

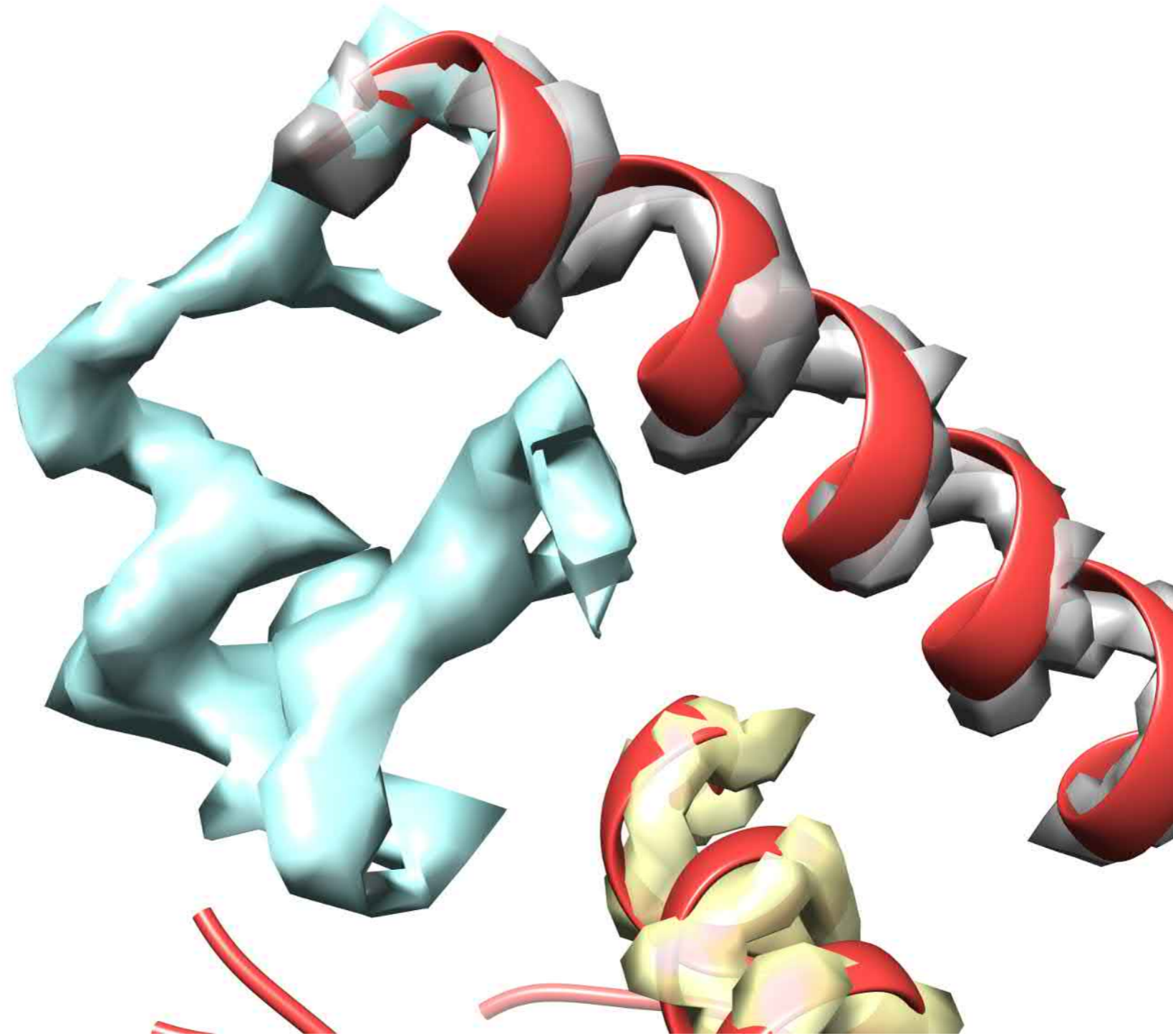Adjust contour level until a region just connects to another

Iterate to build up a connected chain
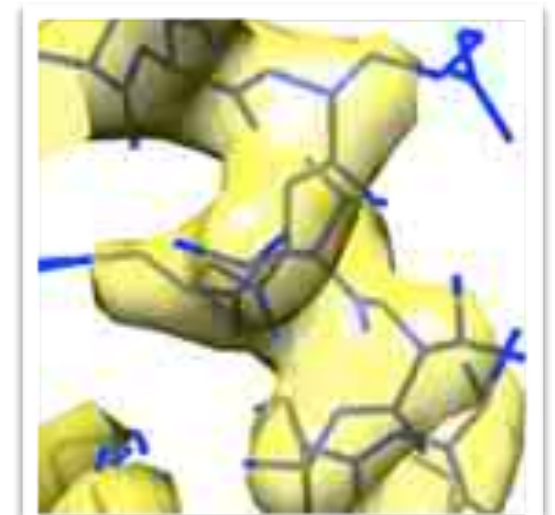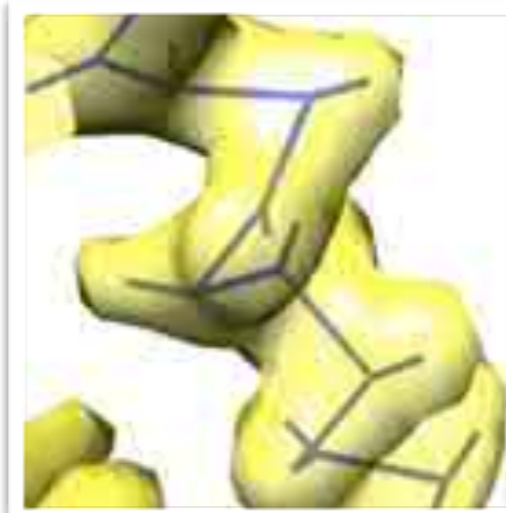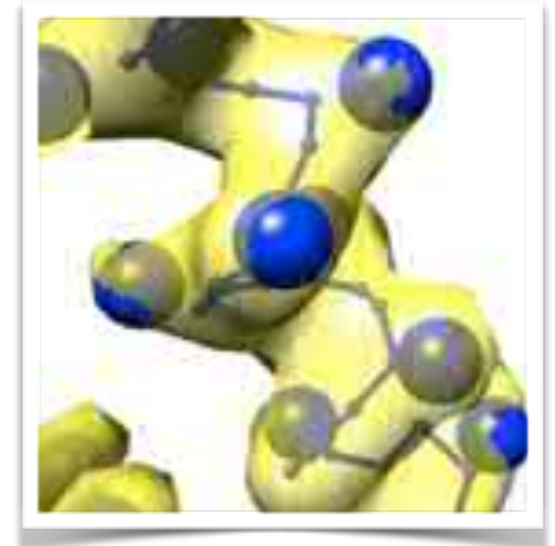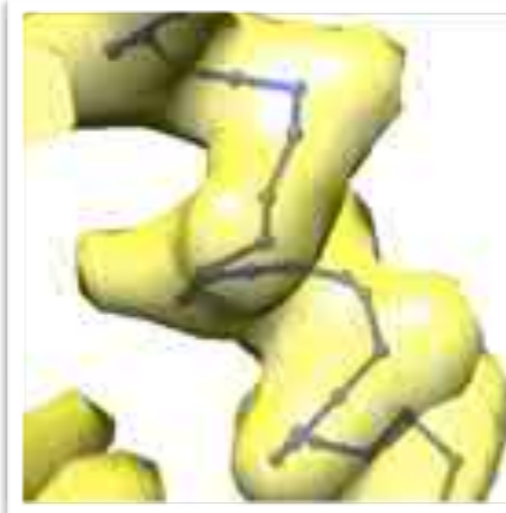
Phenix

Model Building Version 2

Finding C$_\alpha$ and C$_\beta$ positions

Trace chain path through high density

Find C$_\beta$ positions from side-chain density

Choose C$_\alpha$ positions 3.8 Å apart and next to C$_\beta$ positions
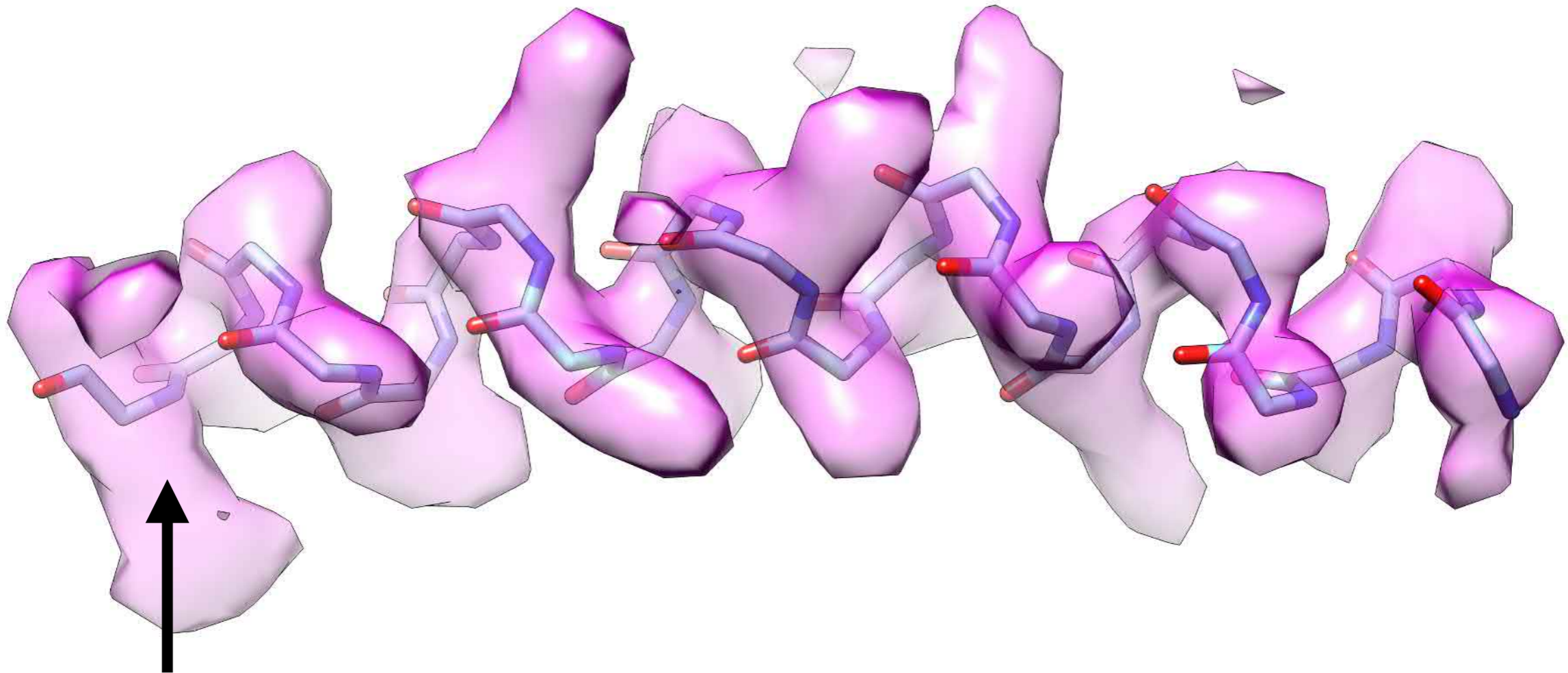
Construct all-atom model with Pulchra* and refine

*Rotkiewicz & Skolnick (2008). J. Comp. Chem. 29, 1460.

Phenix

Sequence Assignment
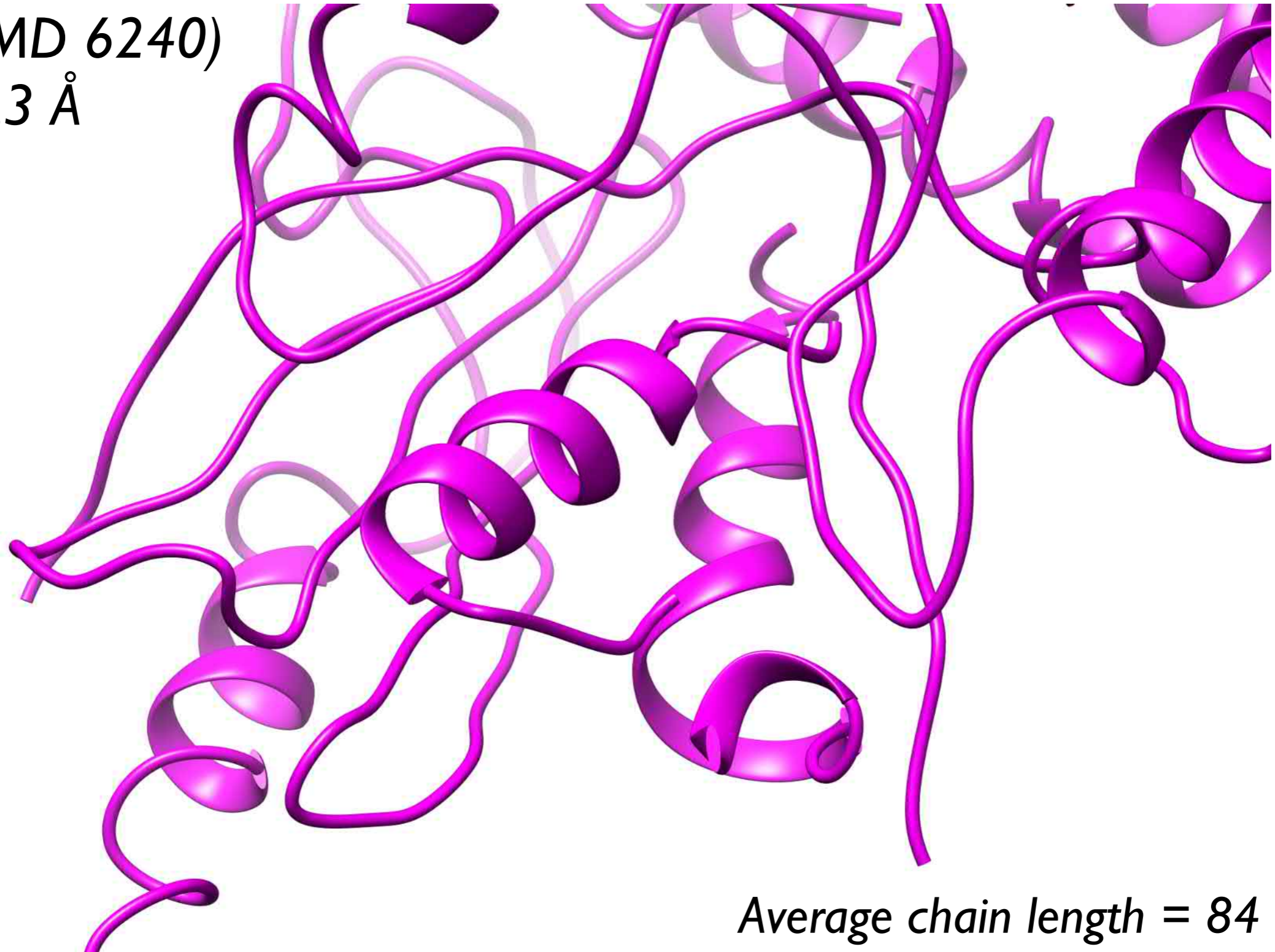

Residue	G	A	S	V	I	L	M	C	F	Y	K	R	W	H	E	D	Q	N	P	T
CC	0.30	0.50	0.53	0.47	0.58	0.62	0.68	0.59	0.83	0.77	0.71	0.69	0.70	0.82	0.65	0.64	0.60	0.60	0.35	0.47
Prob	3	0	0	0	0	0	1	0	40	23	5	5	4	9	2	2	1	0	2	0

- Determine probability of side chain at each C_α
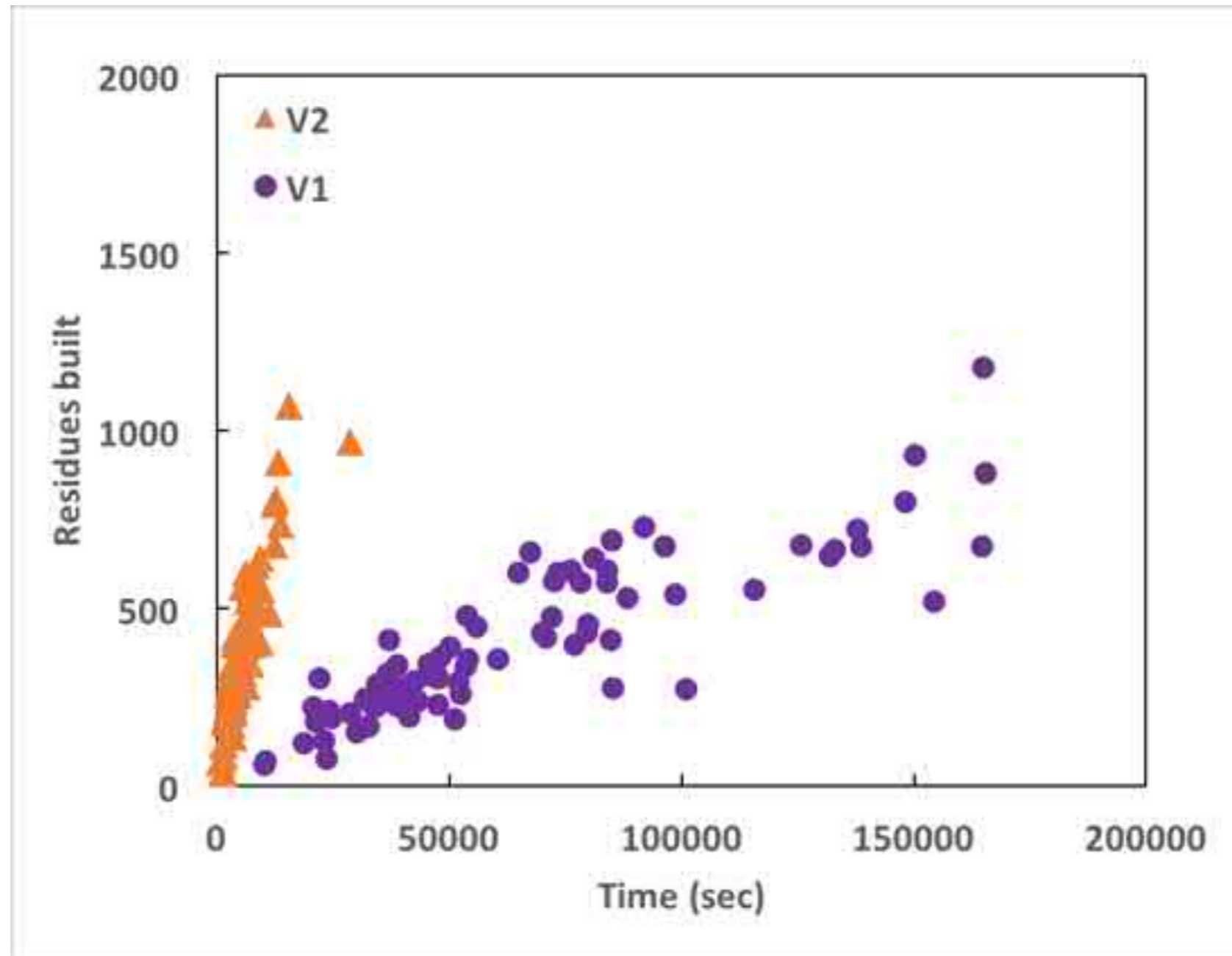- Align sequence to maximize total probability for the chain

Improved Connectivity

3j9e (EMD 6240)
3.3 Å

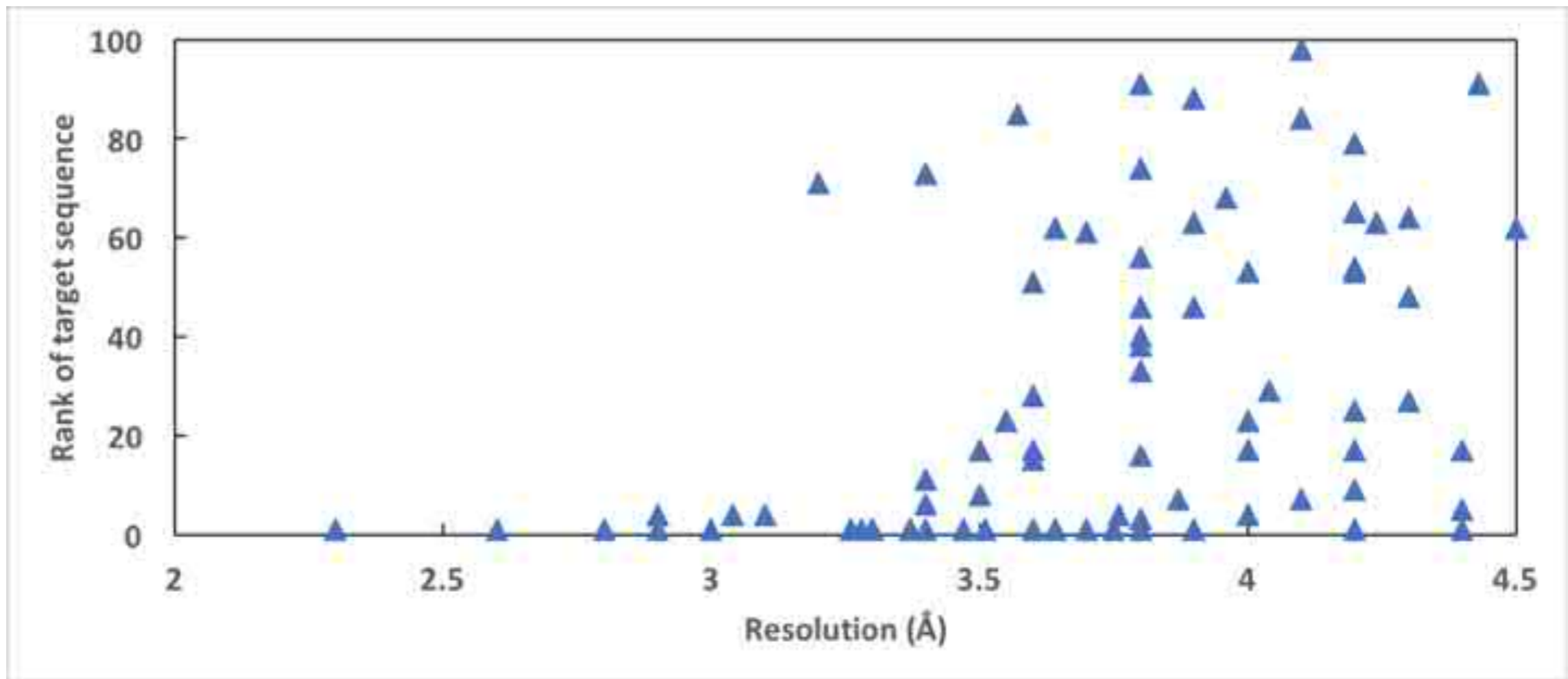*Average chain length = 84*

Improved Performance

What's The Molecule?

- Use the highest side chain probabilities to determine a sequence (from the map)

- Search the sequence database to identify the molecule

With Xiaorun Li, Chi-min Ho & Hong Zhou, UCLA

Conclusions

- Automated model building is possible, but can be improved

  - Include information from secondary structure prediction, evolution etc.

  - Combine structure-modeling tools (Rosetta) with Phenix model-building

- Many challenges remain:

  - Reliably accounting for uncertainty in magnification

  - Local variation in resolution leads to uncertainties in interpretation

Phenix

Acknowledgements

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix