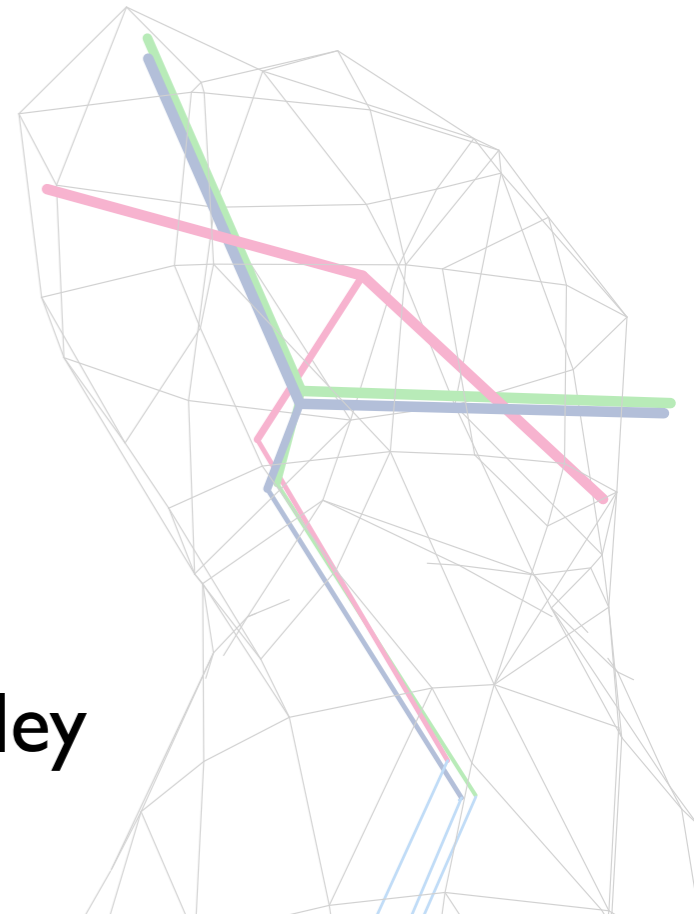
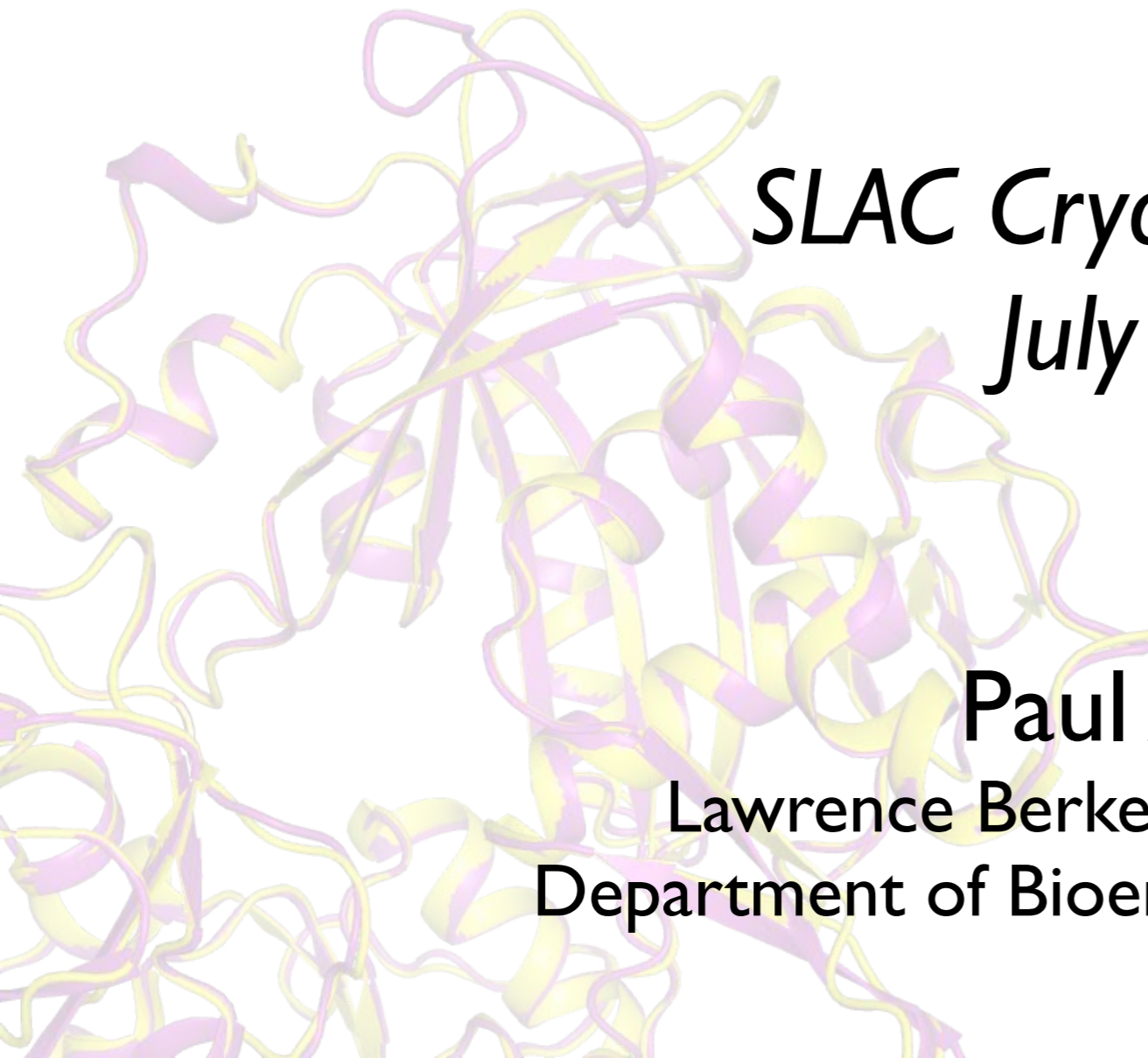


# Atomic Models from Cryo-EM Data

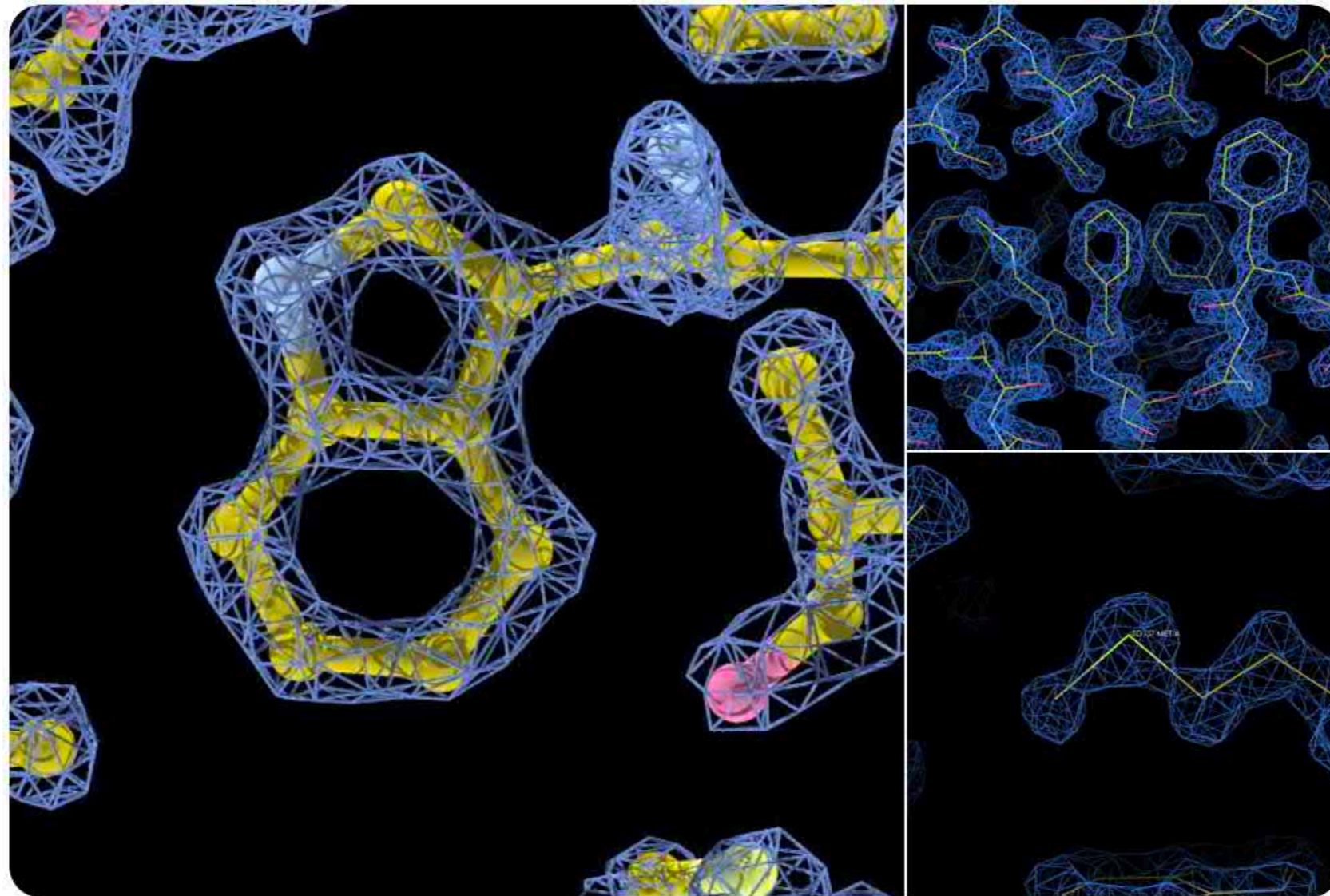
*SLAC Cryo-EM School  
July 2019*

**Paul Adams**

Lawrence Berkeley Laboratory and  
Department of Bioengineering UC Berkeley



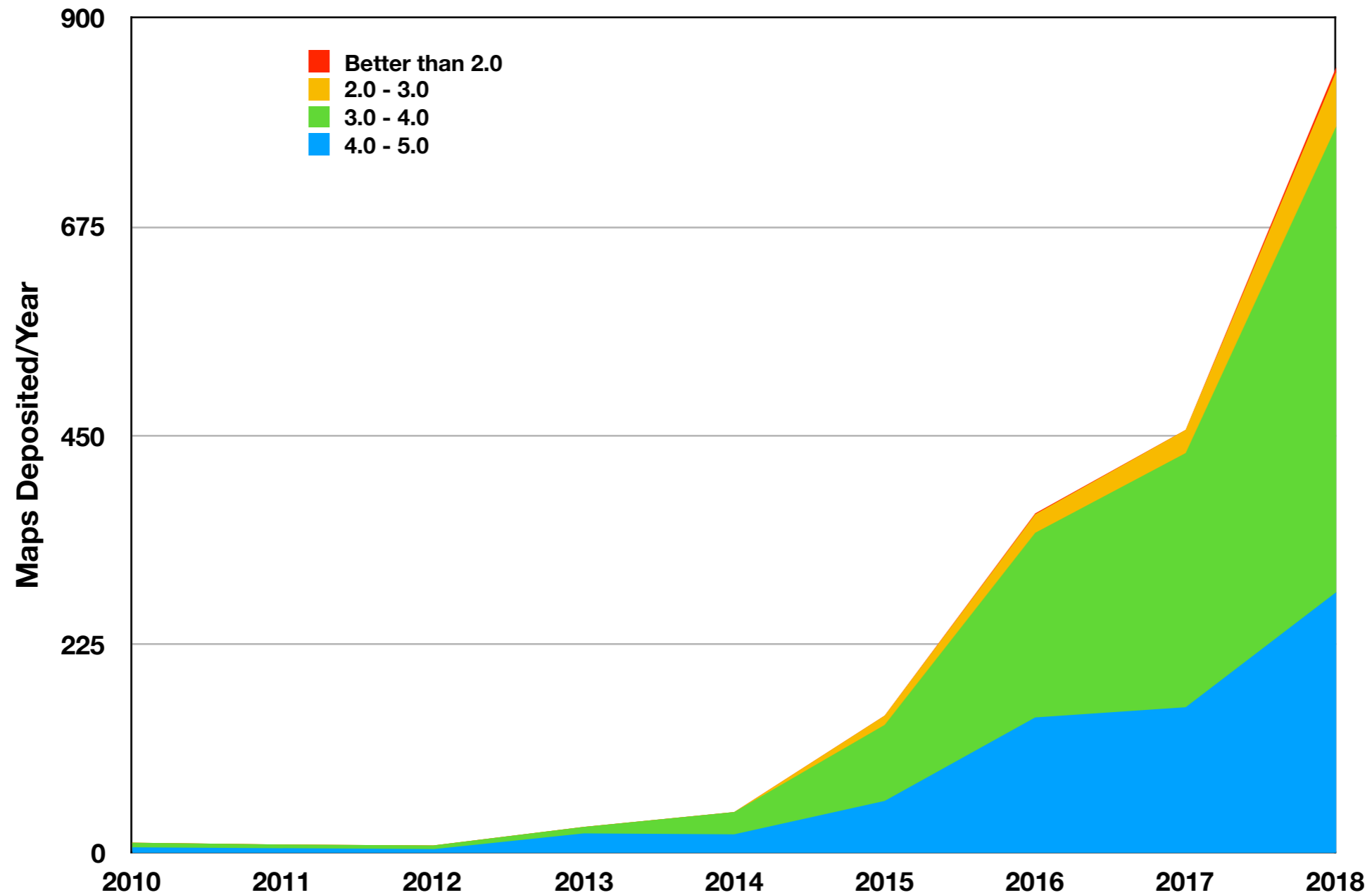
# Impressive Cryo-EM Achievements



*Namba Lab, Osaka*

# Map Resolution

- Biggest growth is in the 3-4Å range
- Substantial number of maps in 4-5Å range

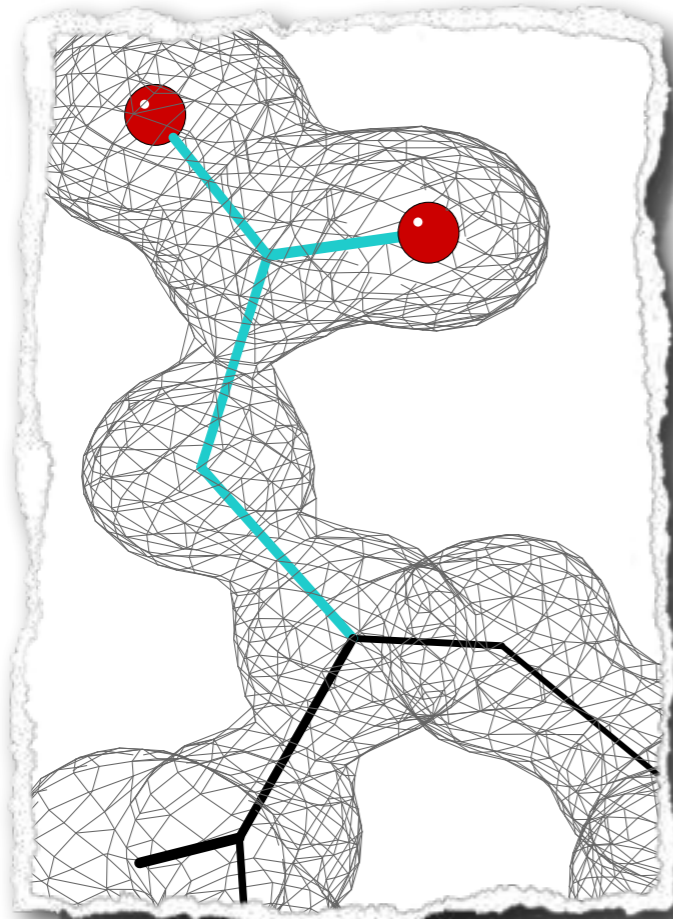


\* Not all maps have an associated model

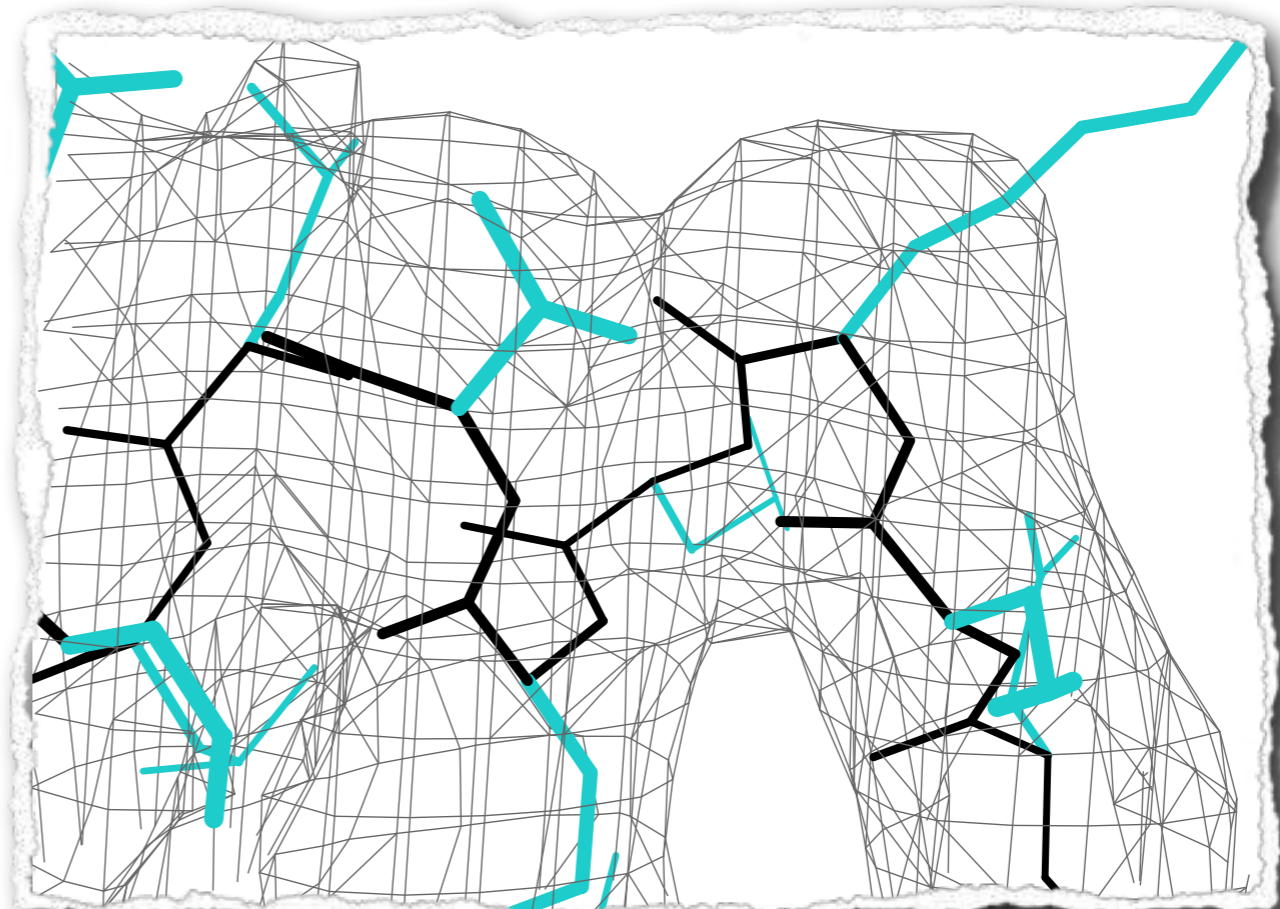


# Low Resolution

PDBID: 2gkg  
Resolution: 1.00Å



PDB ID: 3k7a  
Resolution: 3.80Å



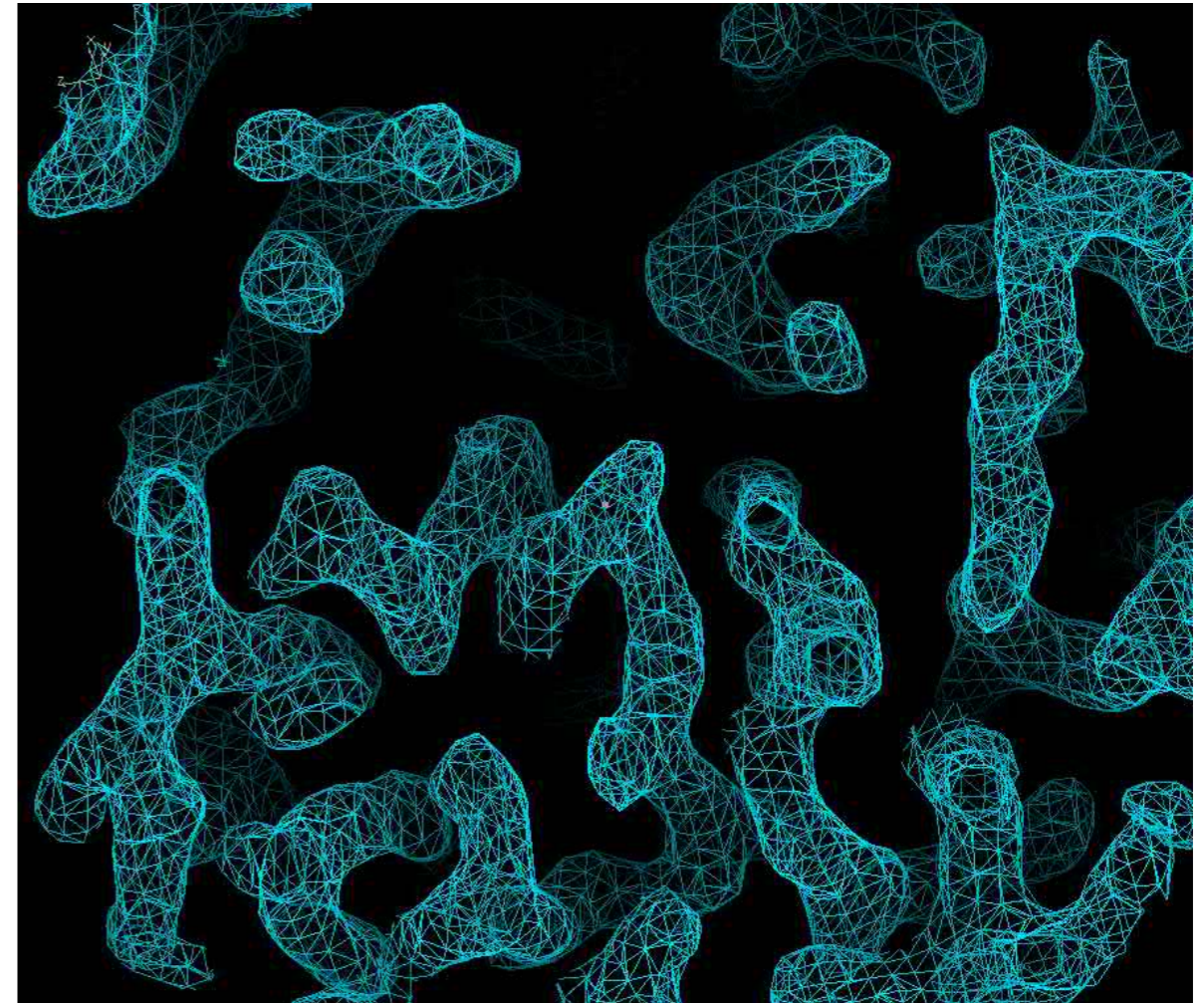
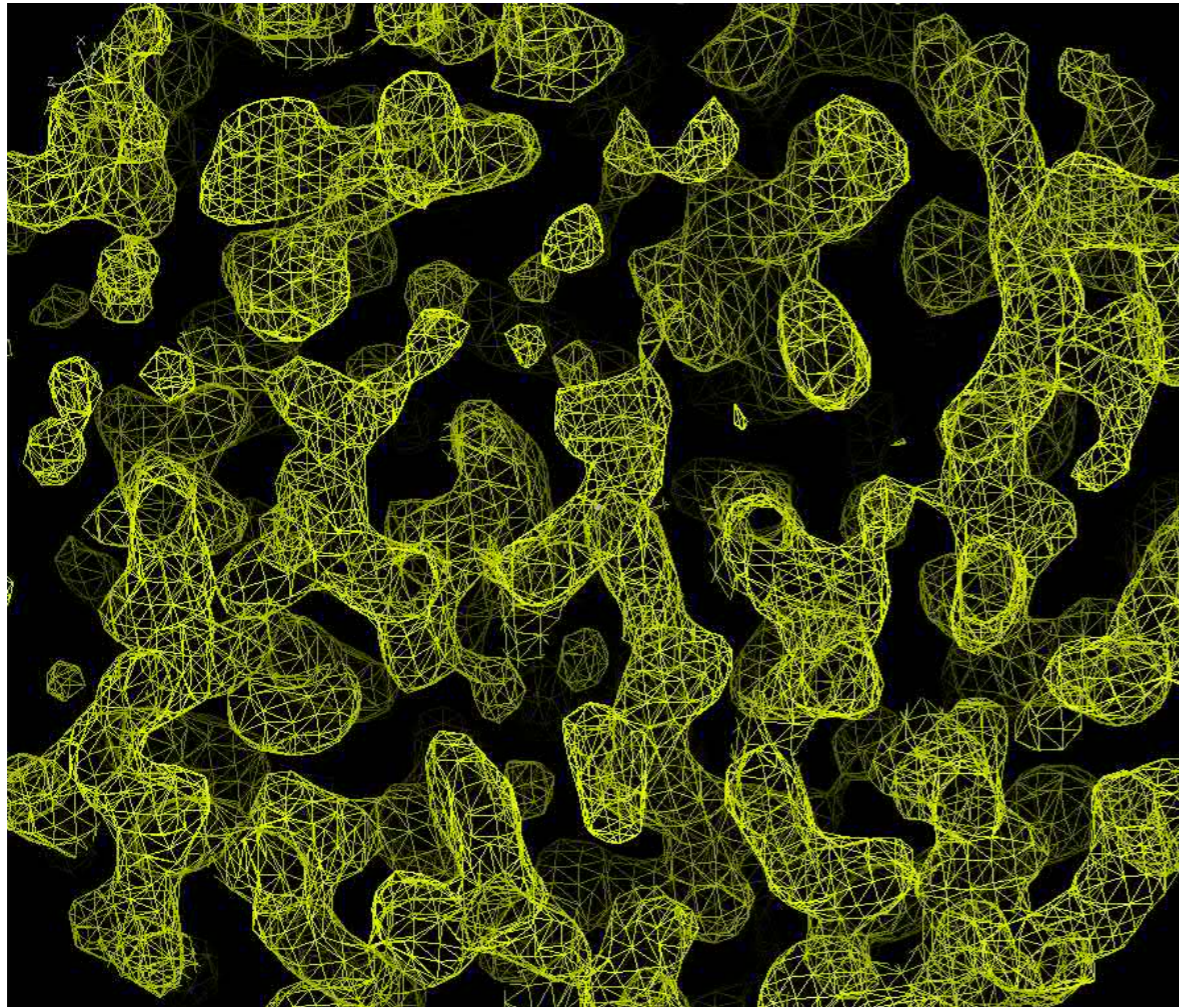
- Many challenges:
- How to interpret “featureless” maps (pattern matching, chemical constraints)
- How to optimize models with sparse data (prior information)

  
**Phenix**



# Crystallographic vs. Cryo-EM Maps

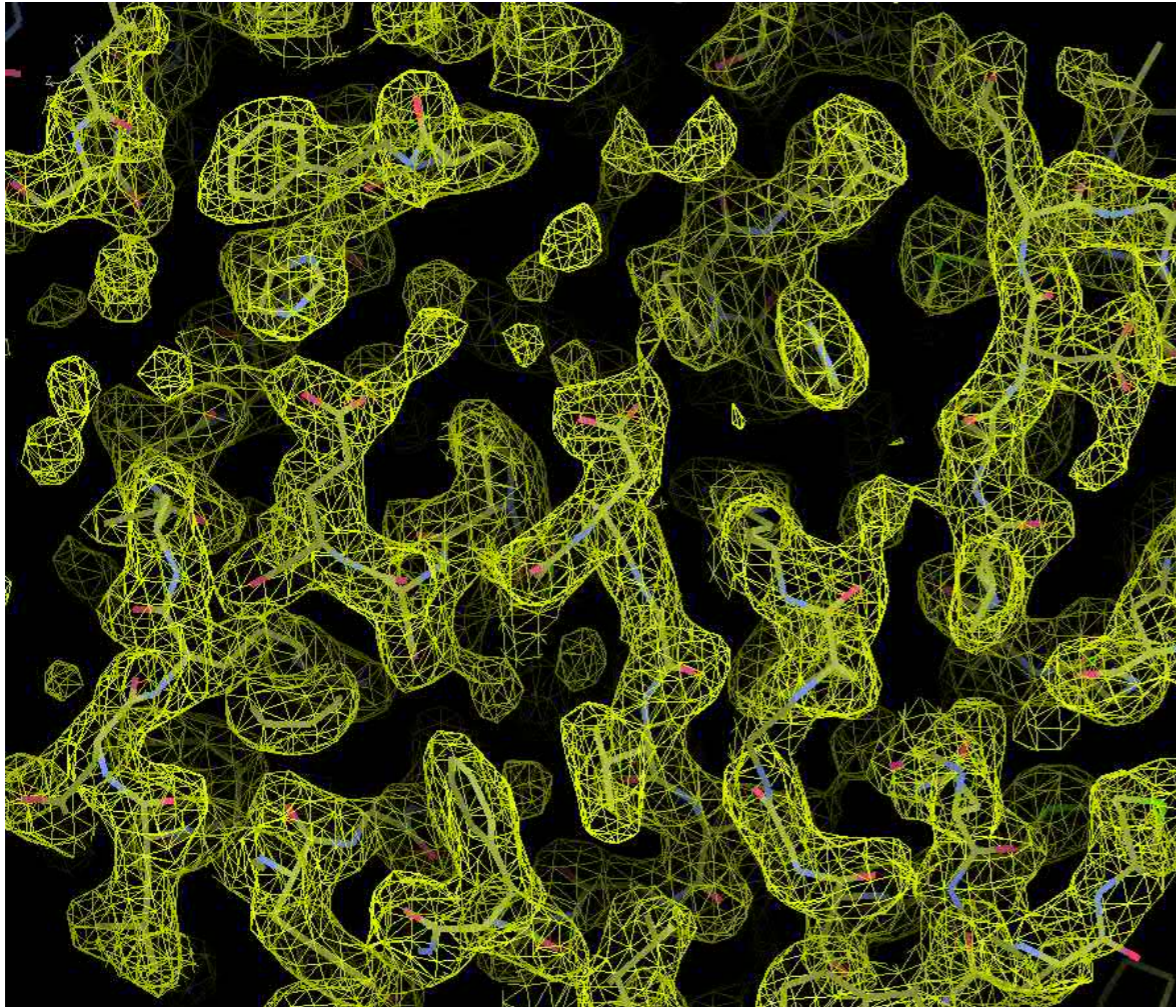
Beta galactosidase at 2.2 Å



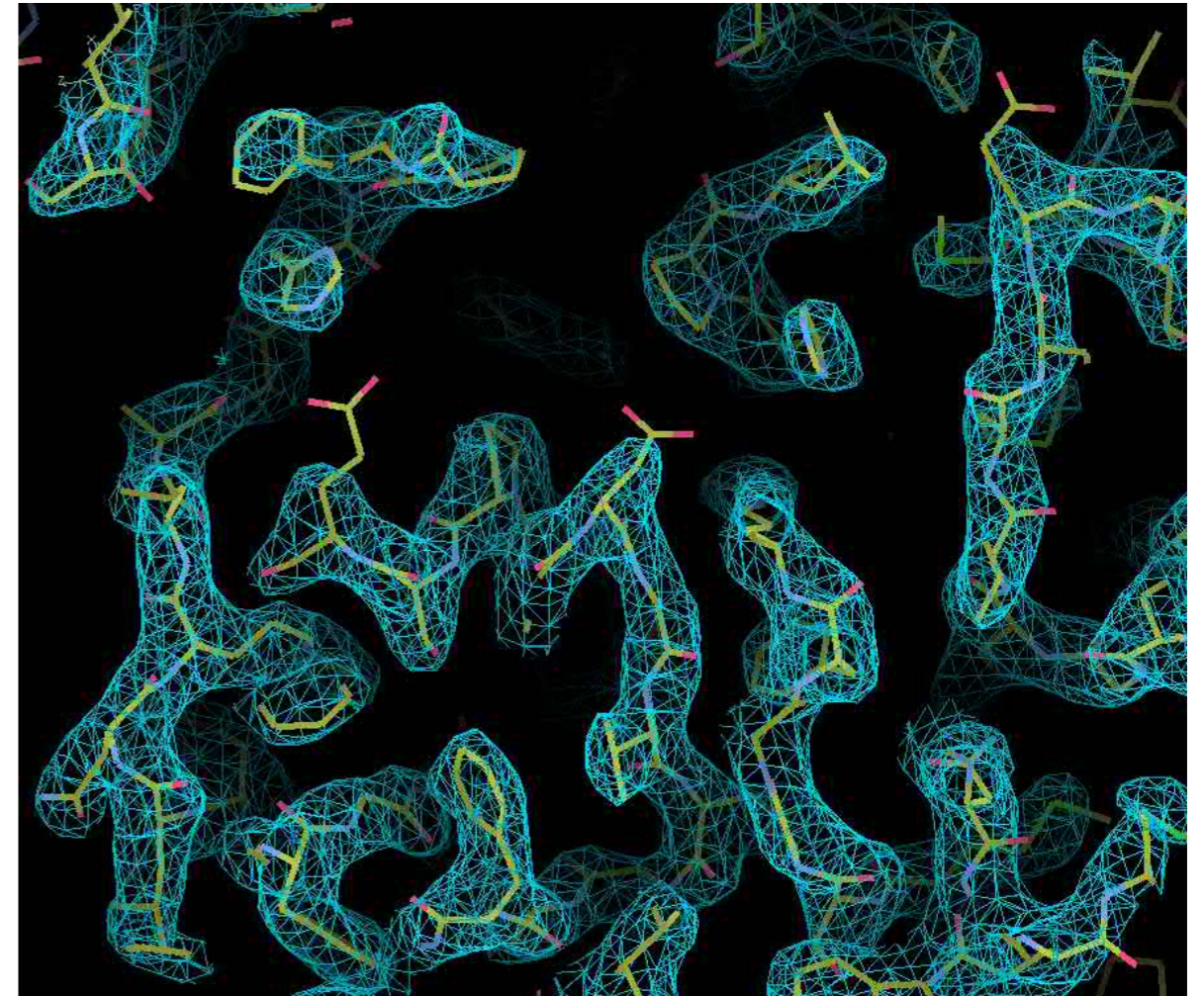


# Crystallographic vs. Cryo-EM Maps

Beta galactosidase at 2.2 Å



X-ray (PDB 3i3b)

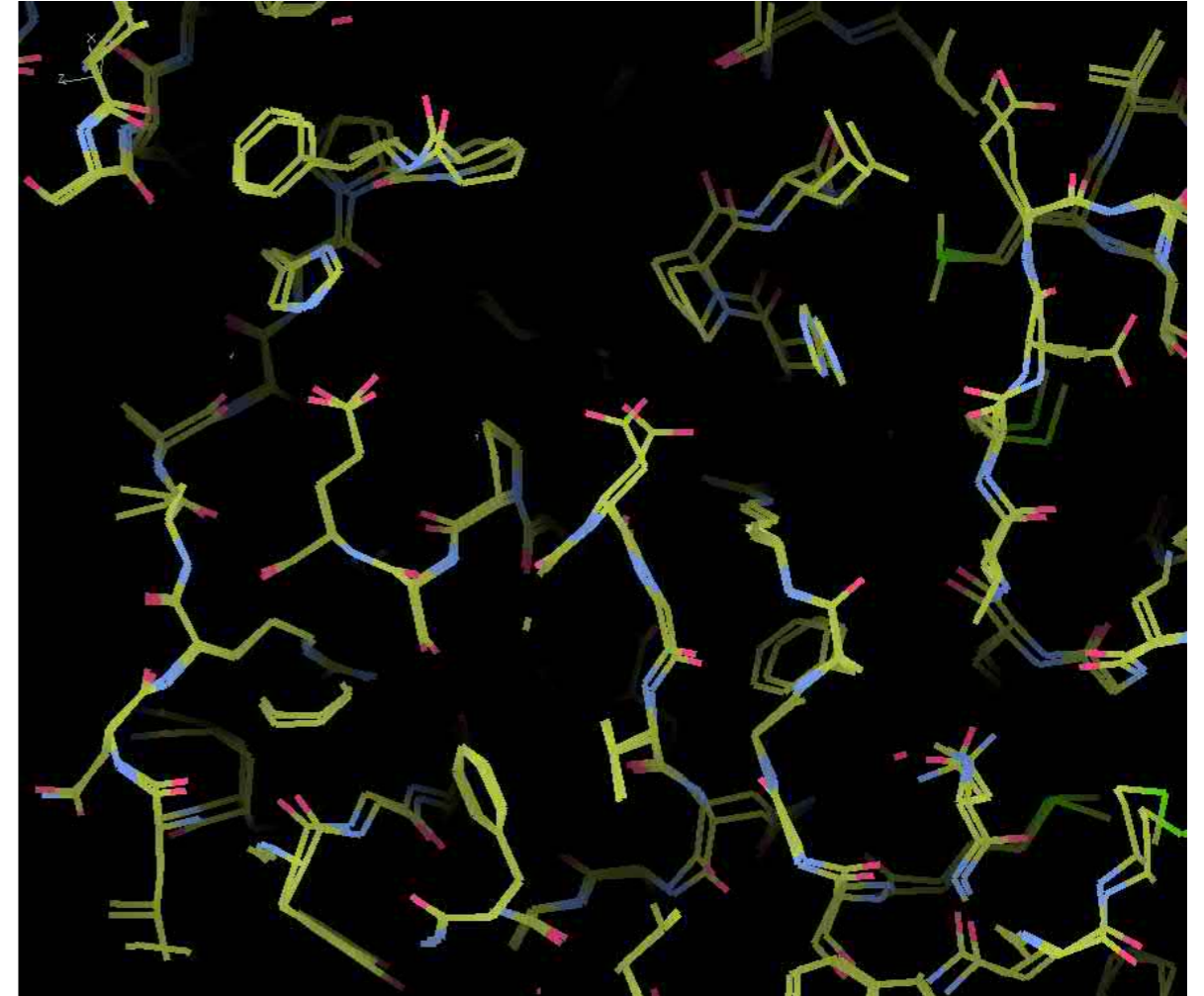
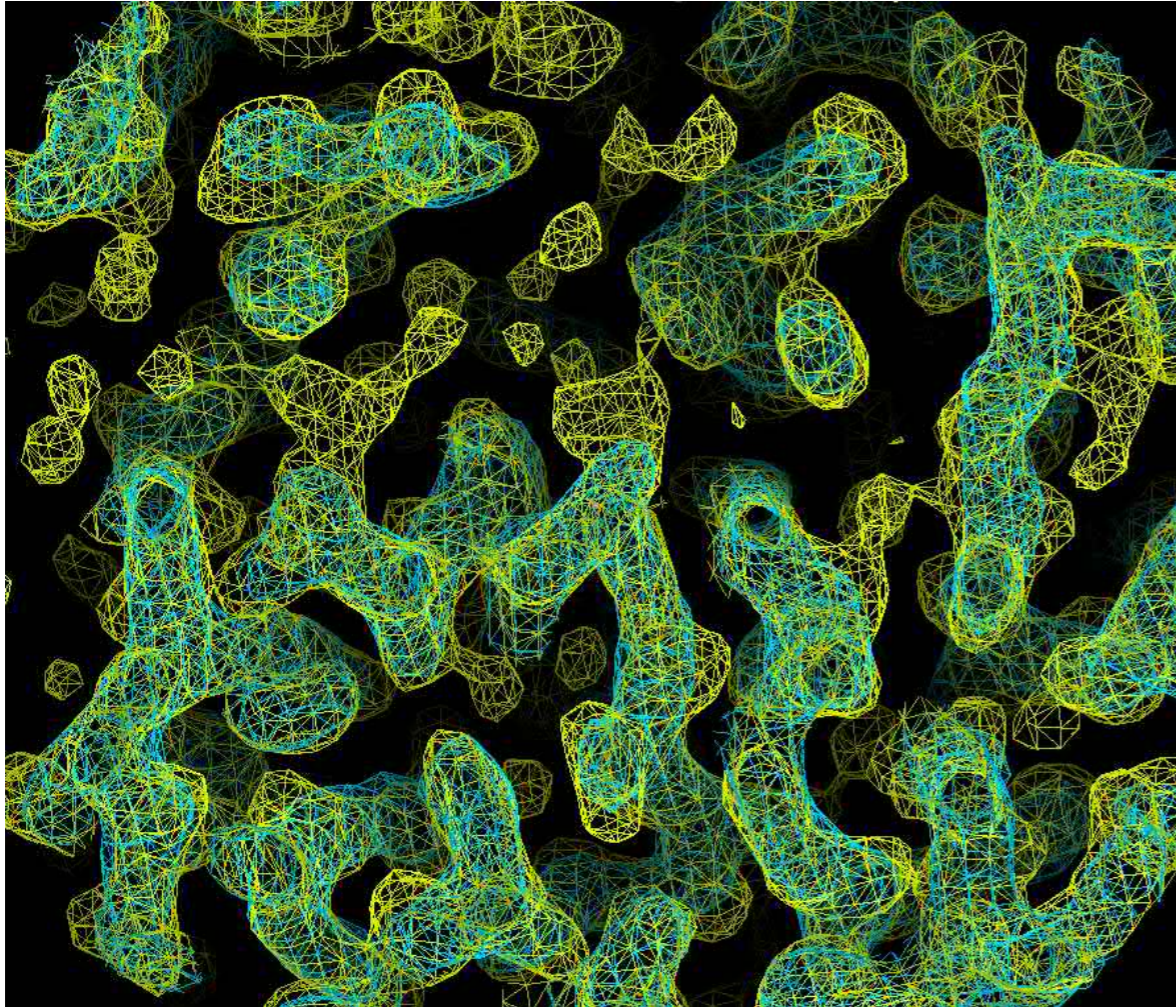


Cryo-EM (PDB 5a1a)



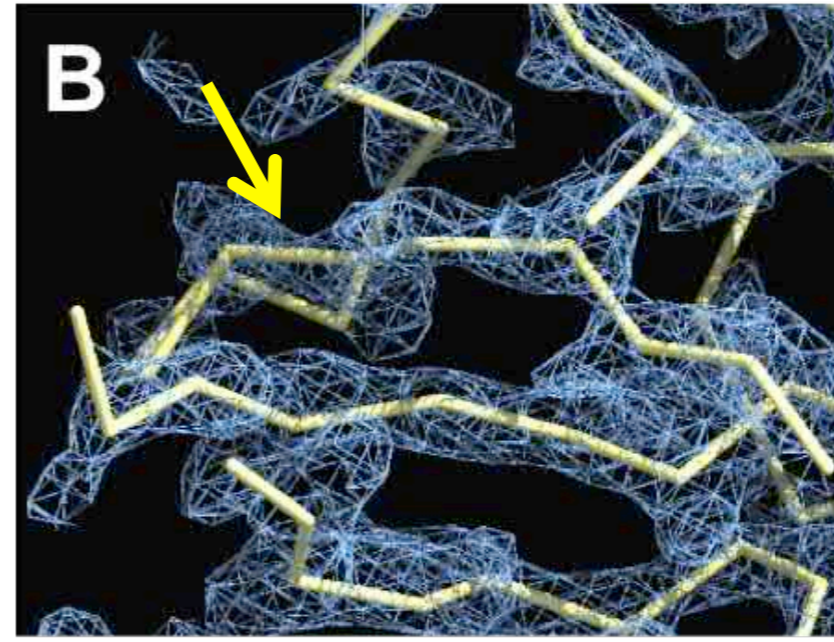
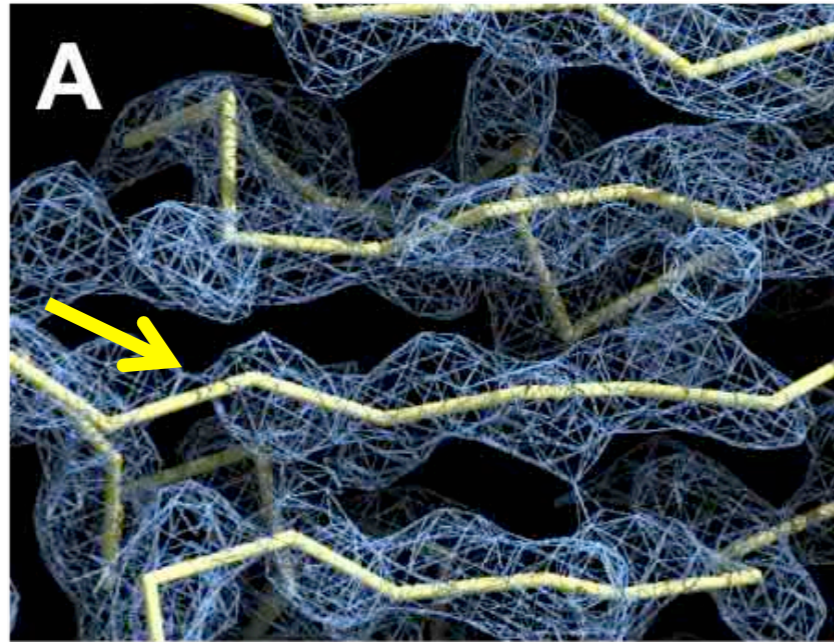
# Crystallographic vs. Cryo-EM Maps

- The maps are very similar





# More Accurate Low Resolution Information in Cryo-EM Maps



**Original**

# 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



UNIVERSITY OF  
CAMBRIDGE

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

## Cambridge University



## Duke University

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



An NIH/NIGMS funded  
Program Project

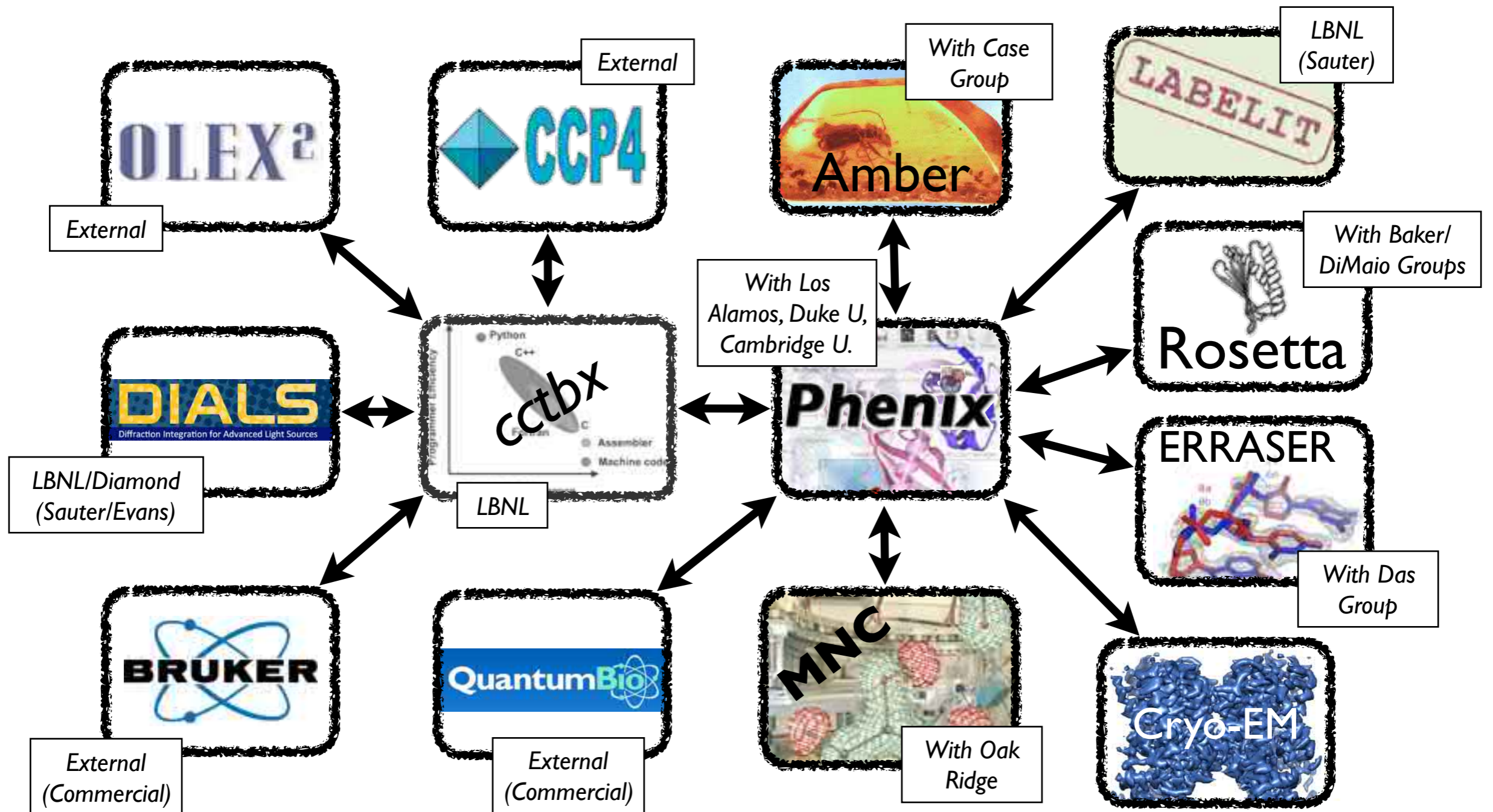
Adams PD et al., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.* 2010, **D66**:213-221.

**Phenix**



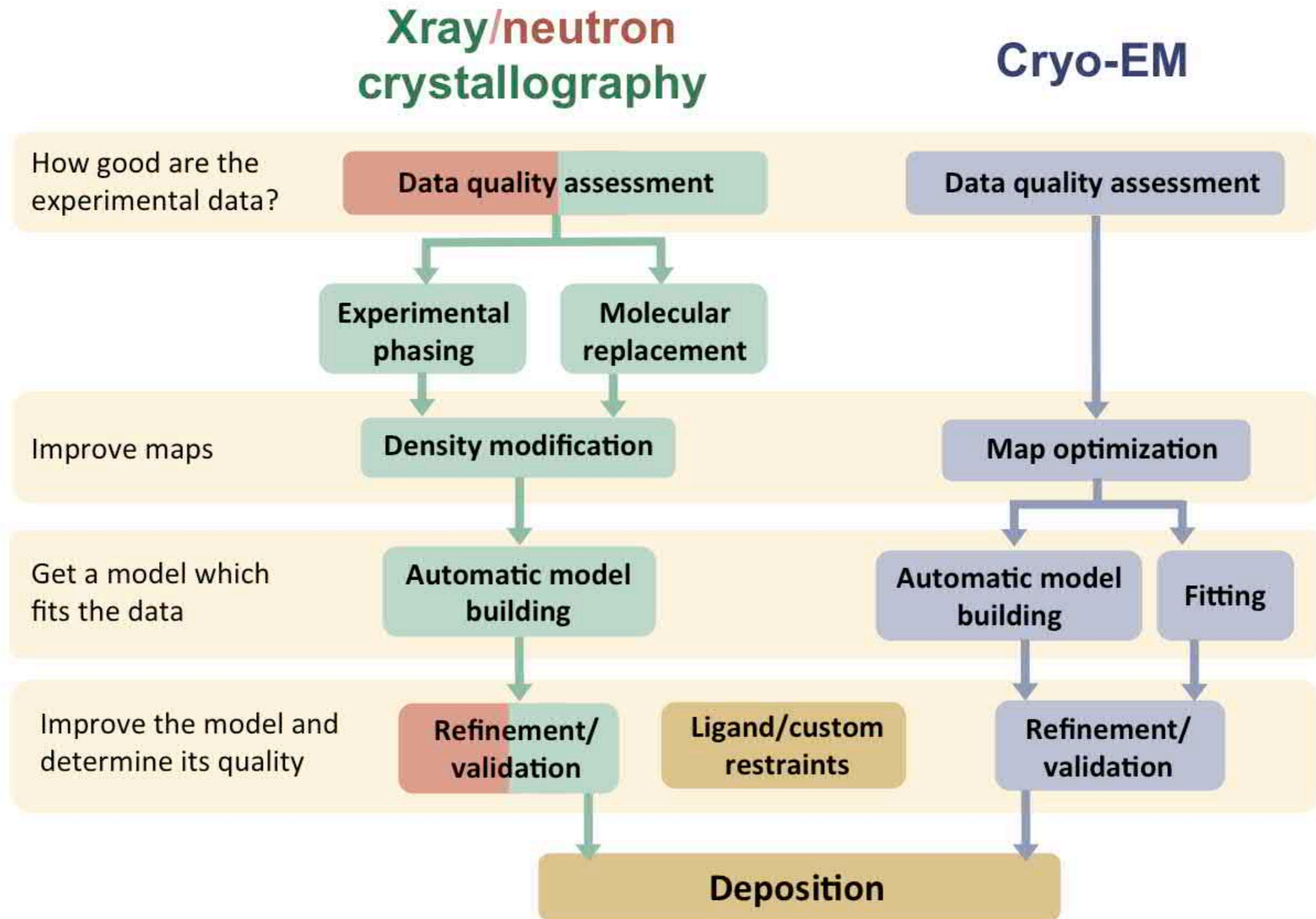


# Phenix - a Structural Biology Hub



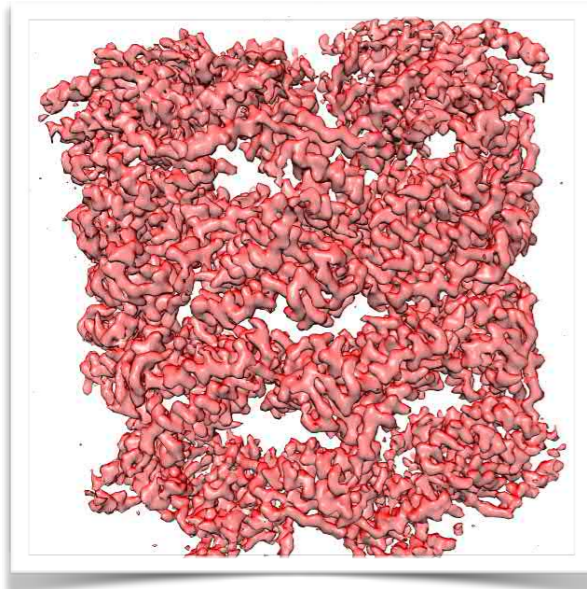


# Structural Biology Workflows

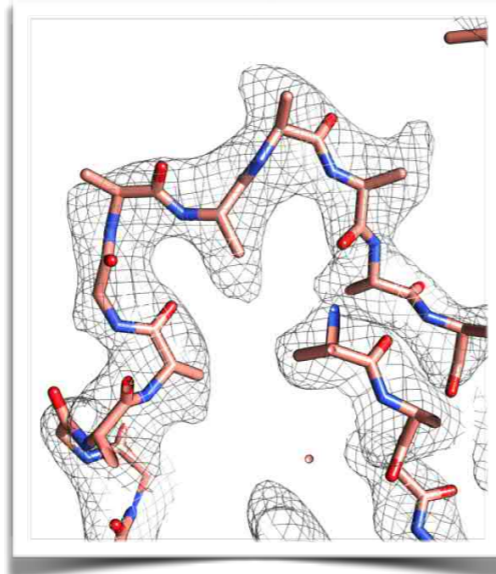




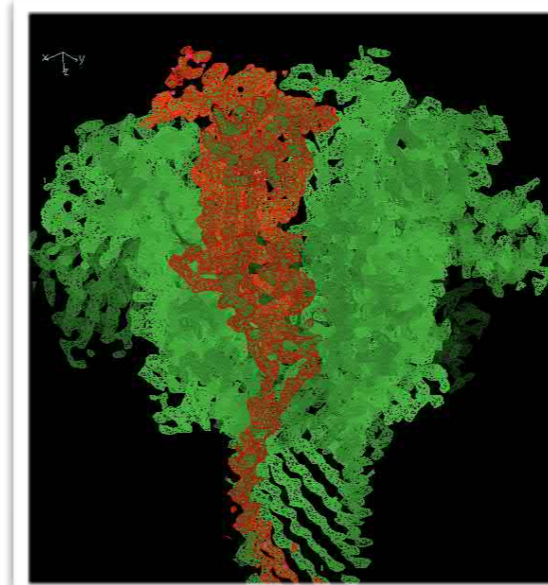
# New Tools for Cryo-EM in Phenix



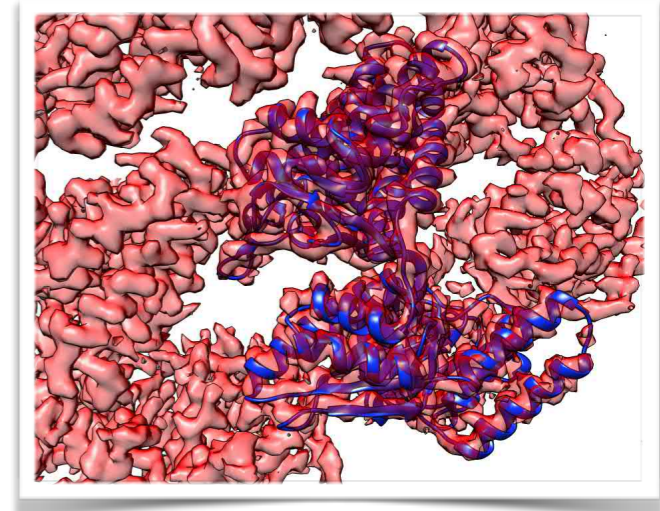
*Symmetry from a map*



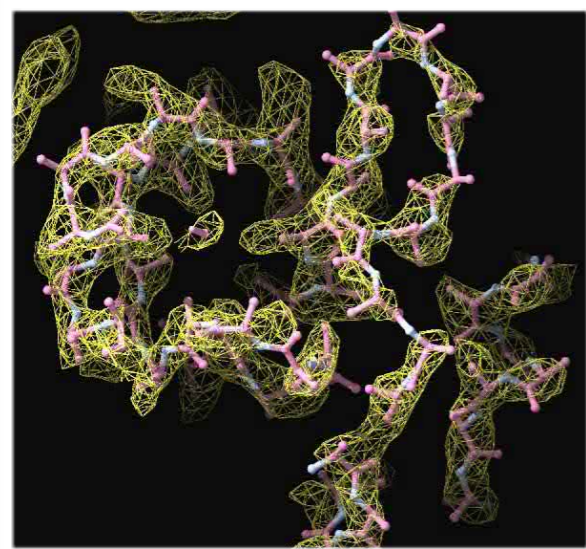
*Automated map sharpening*



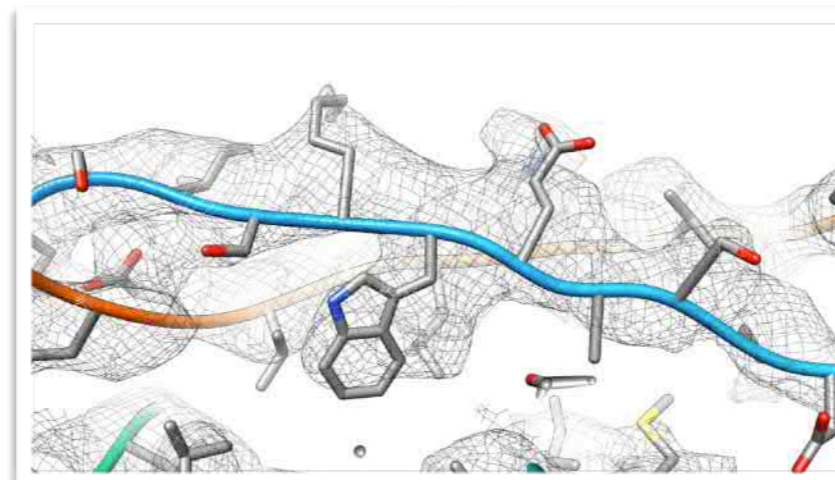
*Map segmentation*



*Rigid model docking*



*Automated model building*



*Real space refinement*

Model	
MolProbity score	1.72
Clash score	5.44
Rotamer outliers (%)	0.00 (Goal: < 1%)
CB outliers	0 (Goal: 0)
Ramachandran	0.00 (Goal: < 0.2%)
Allowed (%)	0.45
Favored (%)	93.55 (Goal: > 98%)

CαBLAM	
Outliers (%)	3.88 (Goal: <= 1%)
Disfavored (%)	0.96 (Goal: <= 1%)
Cα outliers (%)	1.12 (Goal: <= 0.5%)

Peptide Plane	
cis-proline (%)	0.00
twisted proline (%)	0.00
cis-general (%)	0.00
twisted general (%)	0.00

*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

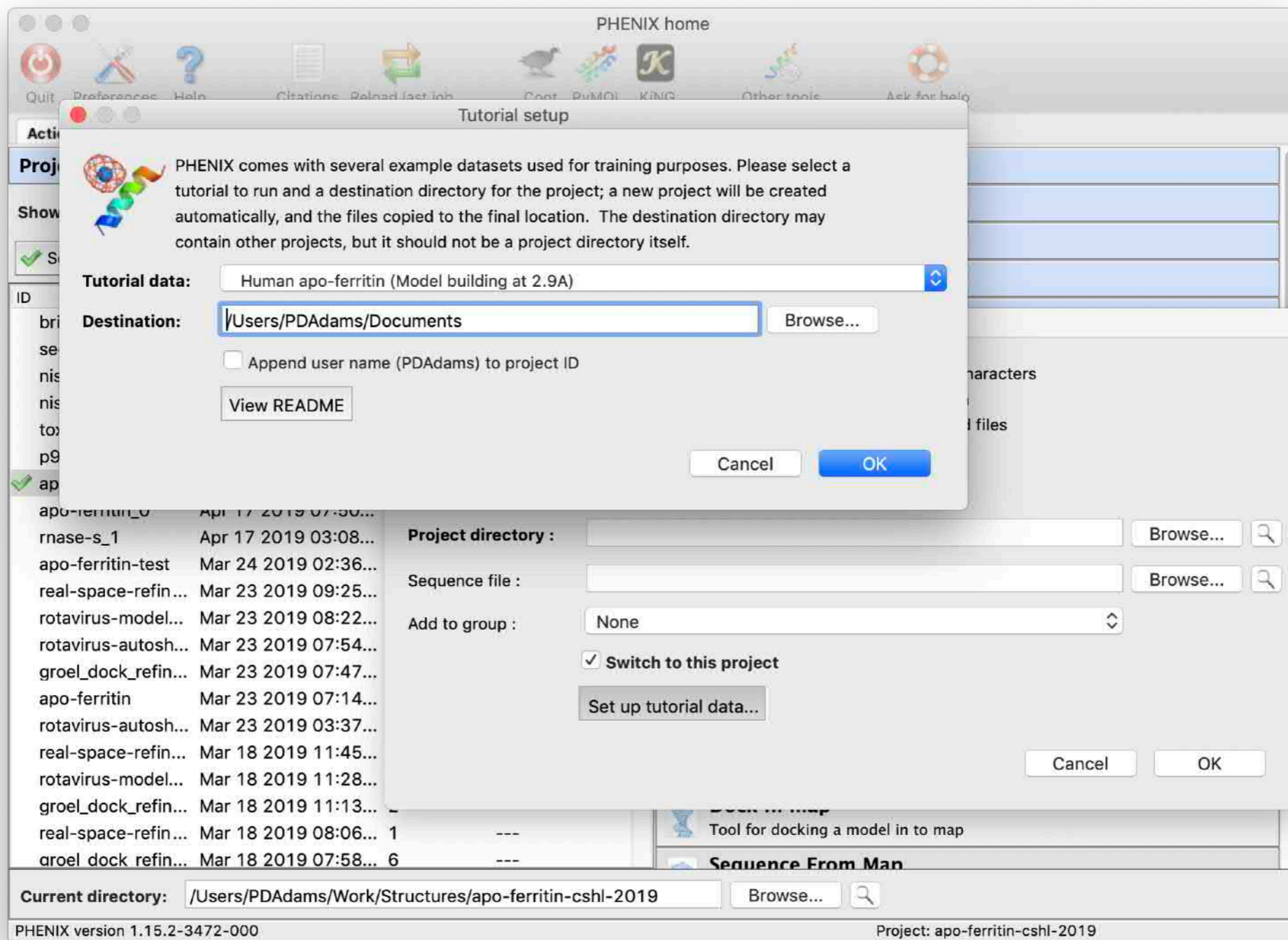
- Use graphical user interface

The screenshot displays the PHENIX home window. At the top, there is a menu bar with icons for Quit, Preferences, Help, Citations, Reload last job, Coot, PyMOL, KING, Other tools, and Ask for help. Below the menu bar, there are tabs for 'Actions' and 'Job history'. The main area is divided into two panels. The left panel, titled 'Projects', shows a table of project entries with columns for ID, Last modified, # of jobs, and R-free. The right panel, titled 'Favorites', lists various analysis tools such as Data analysis, Experimental phasing, Molecular replacement, Model building, Refinement, Cryo-EM, Mtriage, Map to Model, Real-space refinement, Comprehensive validation (cryo-EM), EMRinger, Autosharpen Map, Dock in map, and Sequence From Map. At the bottom, there is a 'Current directory' field showing the path /Users/PDAdams/Work/Structures/apo-ferritin-cshl-2019 and a 'Project: apo-ferritin-cshl-2019' label.

ID	Last modified	# of jobs	R-free
brink	May 17 2019 01:0...	5	0.3868
sec17-sad_0	May 02 2019 06:2...	3	0.3684
nisha2	Apr 29 2019 11:54...	57	0.5146
nisha	Apr 28 2019 07:21...	67	0.4645
toxd-mr_0	Apr 26 2019 10:53...	6	0.4918
p9-sad_0	Apr 25 2019 10:26...	1	0.2898
✓ apo-ferritin-csh...	Apr 25 2019 09:43...	26	---
apo-ferritin_0	Apr 17 2019 07:50...	9	---
rnase-s_1	Apr 17 2019 03:08...	1	0.2644
apo-ferritin-test	Mar 24 2019 02:36...	7	---
real-space-refin...	Mar 23 2019 09:25...	2	---
rotavirus-model...	Mar 23 2019 08:22...	1	---
rotavirus-autosh...	Mar 23 2019 07:54...	1	---
groel_dock_refin...	Mar 23 2019 07:47...	3	---
apo-ferritin	Mar 23 2019 07:14...	29	---
rotavirus-autosh...	Mar 23 2019 03:37...	1	---
real-space-refin...	Mar 18 2019 11:45...	1	---
rotavirus-model...	Mar 18 2019 11:28...	1	---
groel_dock_refin...	Mar 18 2019 11:13...	2	---
real-space-refin...	Mar 18 2019 08:06...	1	---
groel dock refin...	Mar 18 2019 07:58...	6	---

# 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



# Automated Model Docking

**Tom Terwilliger**

Los Alamos National Laboratory

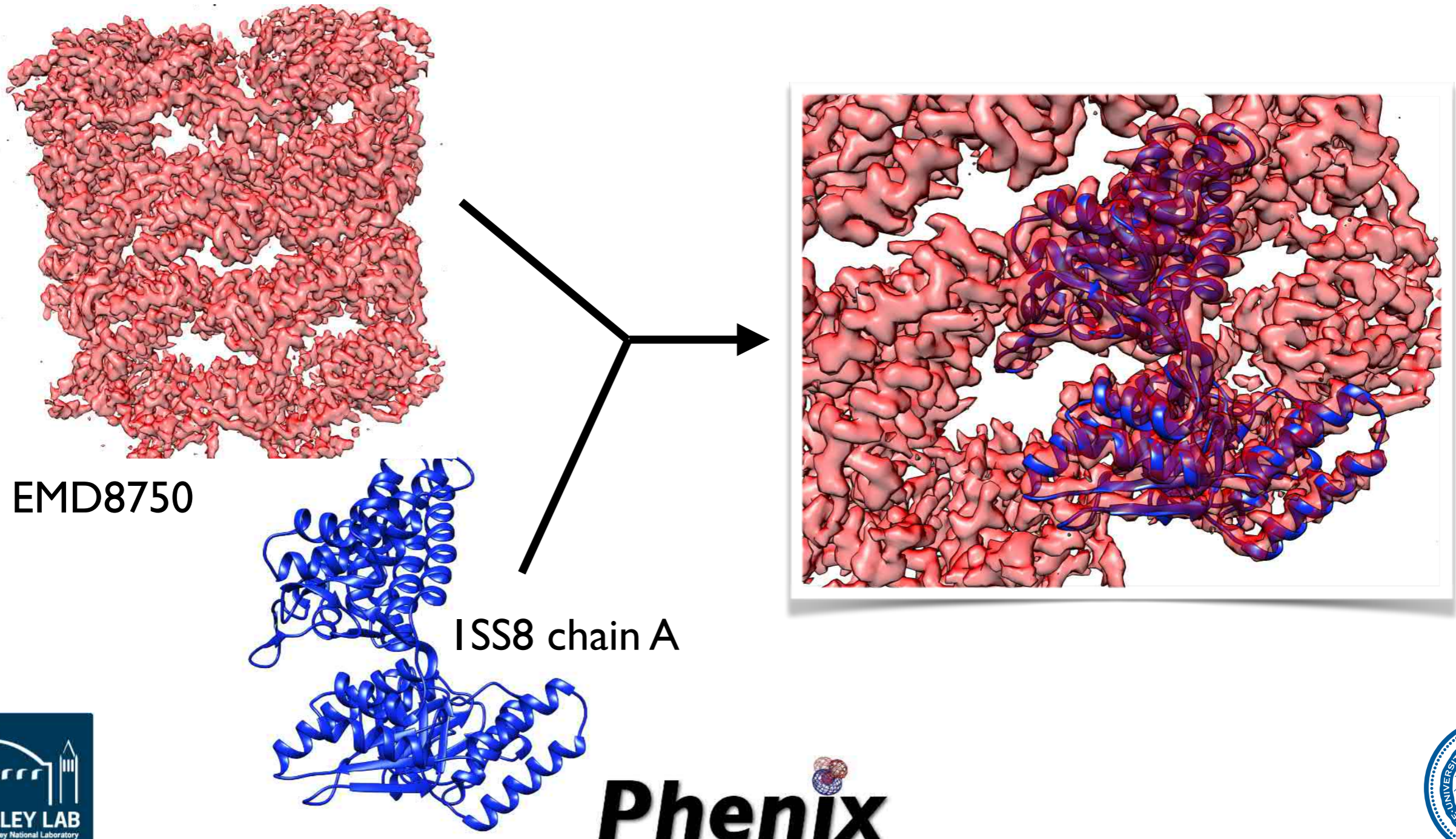
**Pavel Afonine, Oleg Sobolev**

Lawrence Berkeley National Laboratory



# Automated Model Docking

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





# Automated Model Sharpening, Segmentation and Model Building

**Tom Terwilliger**

Los Alamos National Laboratory

**Pavel Afonine, Oleg Sobolev**

Lawrence Berkeley National Laboratory



# Automated Model Building Procedure

Determine optimal sharpening of the map



Cut out asymmetric unit of the map



Trace chain and build model



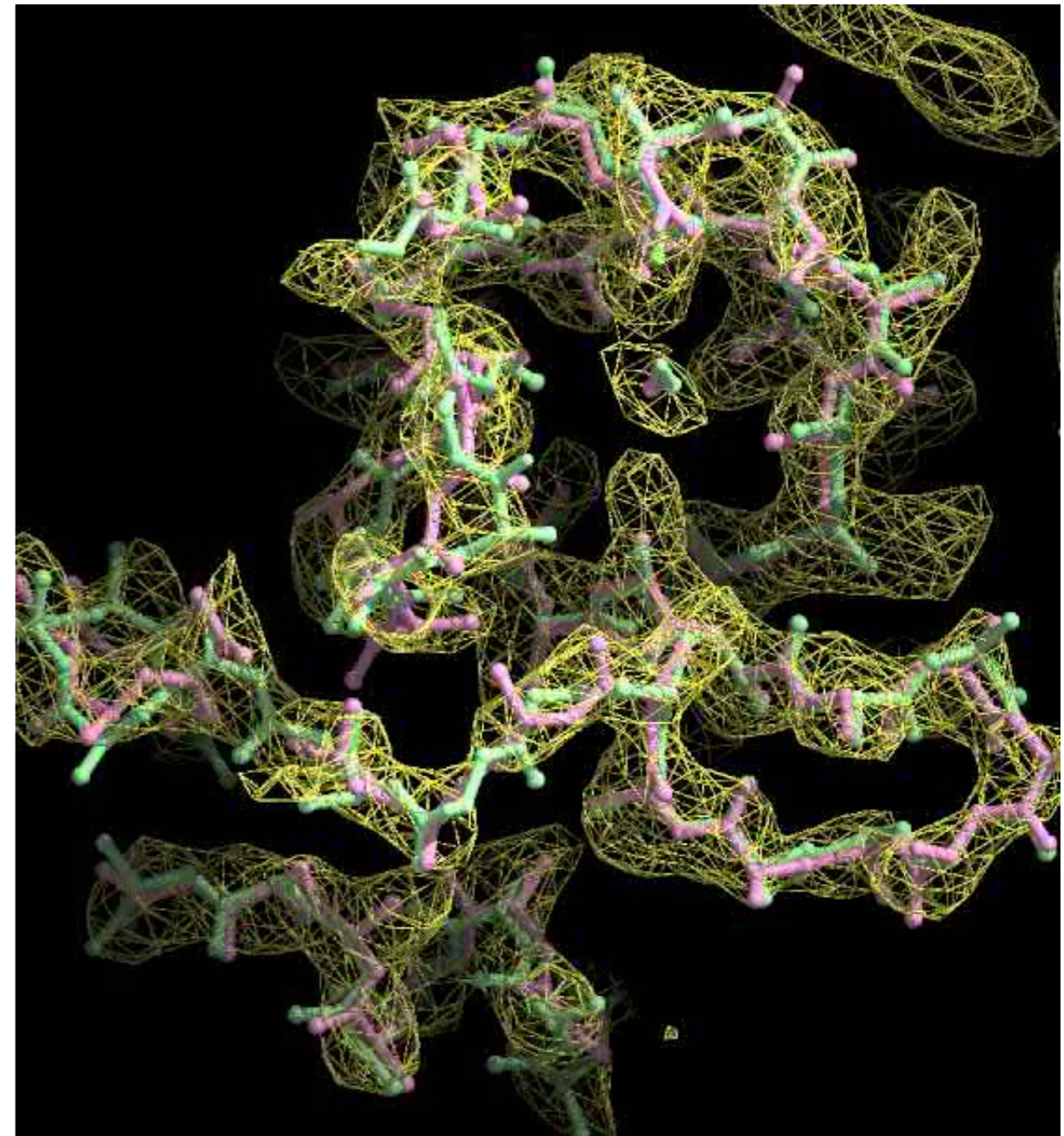
Idealize secondary structure and refine



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)

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

  
**Phenix**

Autobuilt model (pink)  
Deposited model (green)





# Automated Map Sharpening

Create series of maps with variable overall B-values

Analyze maps for detail and connectivity

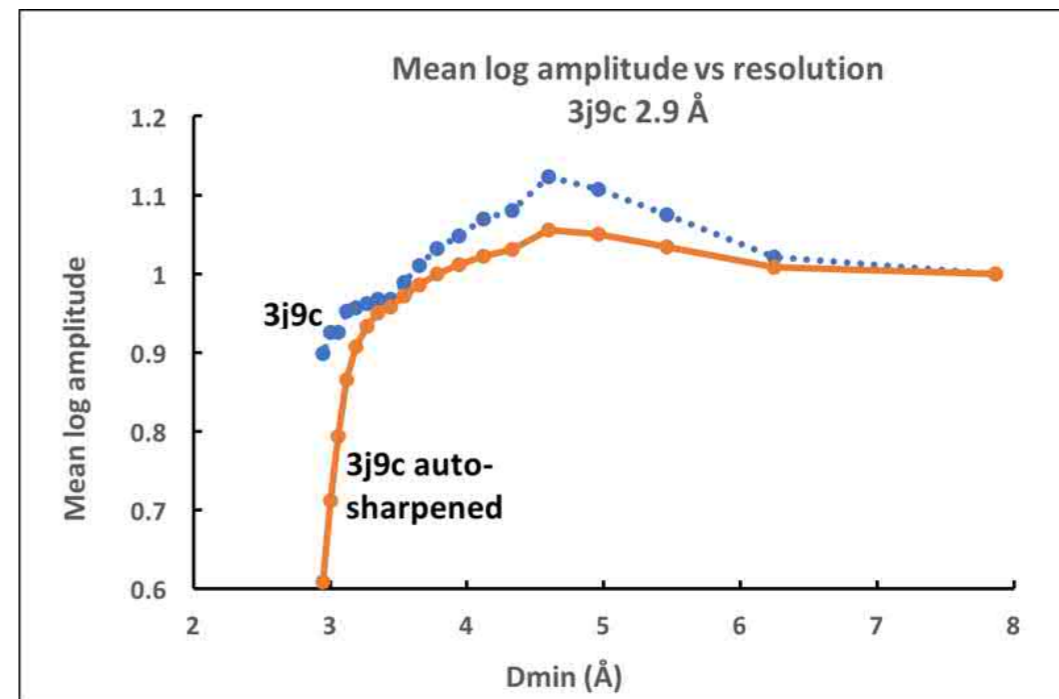
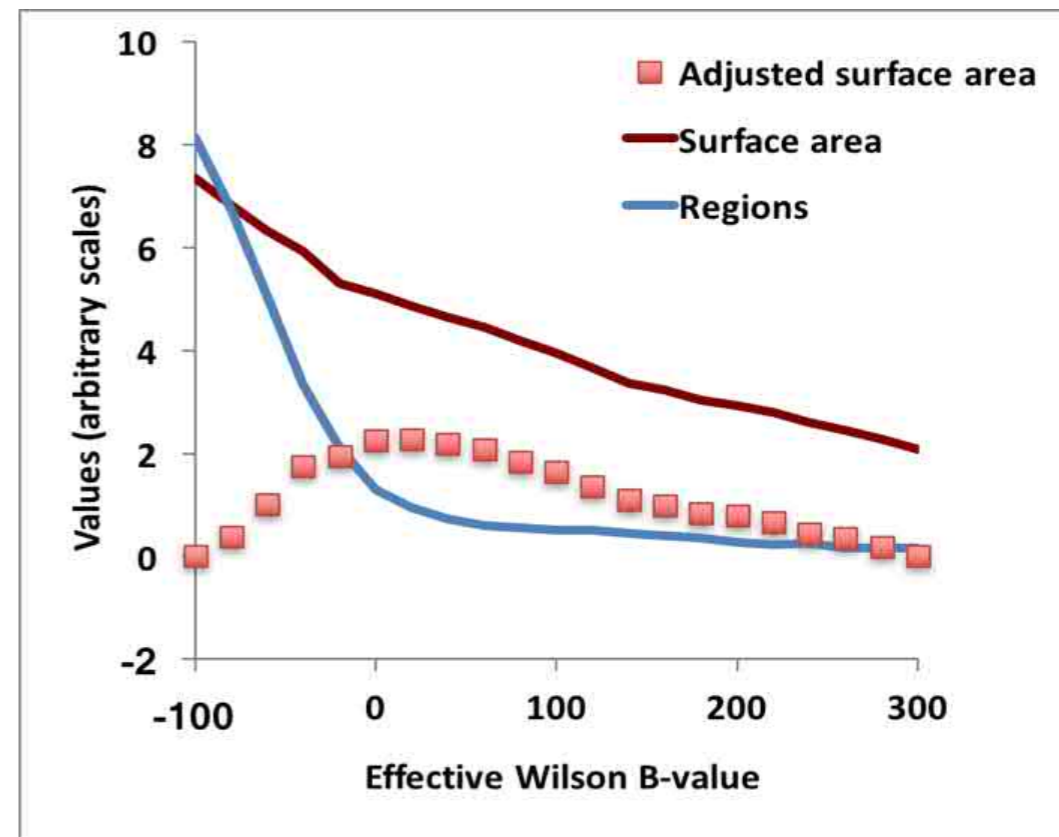
Set contour level enclosing 20% of molecular volume

Calculate surface area of contours

Count number of distinct regions enclosed by contours

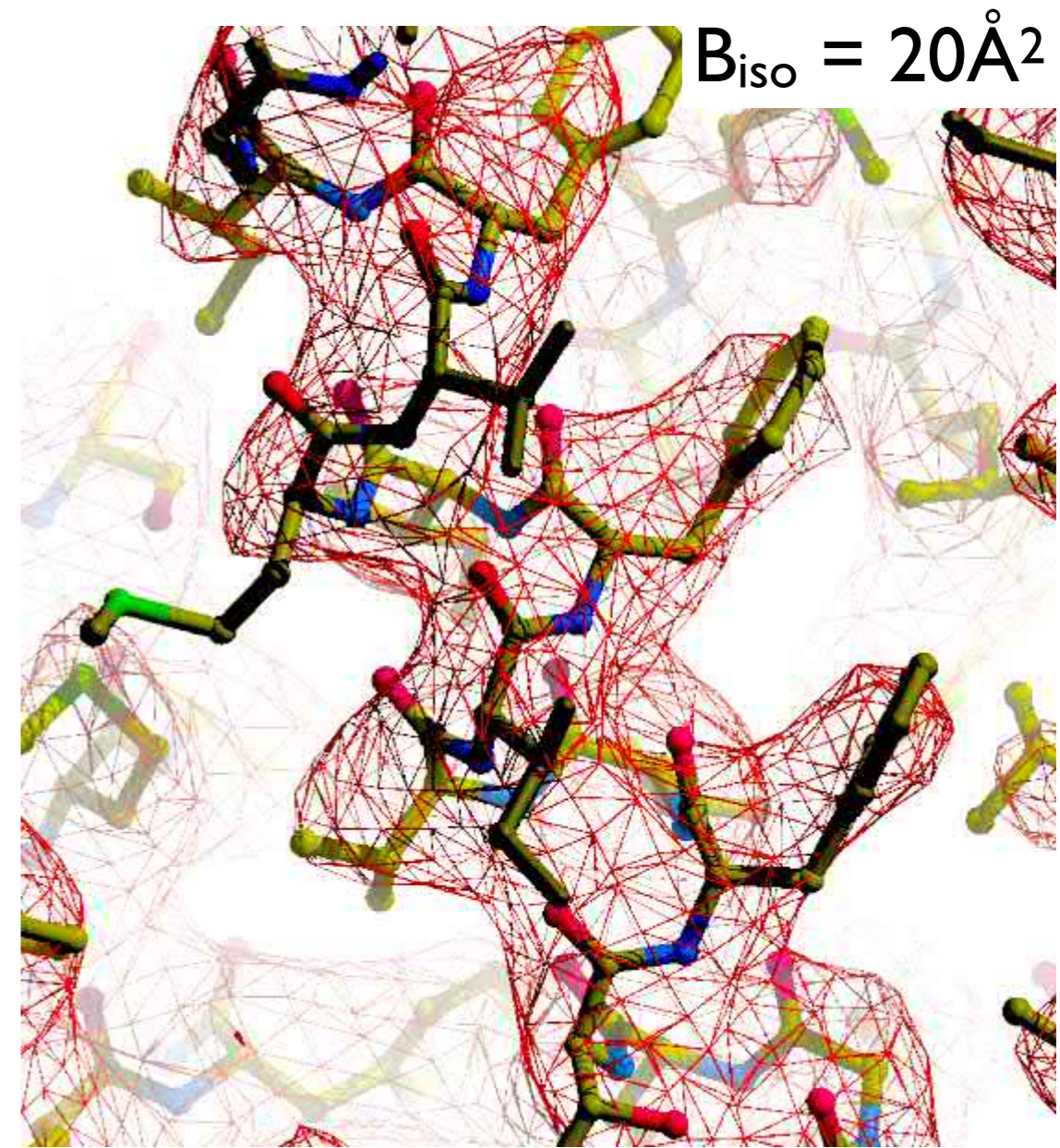
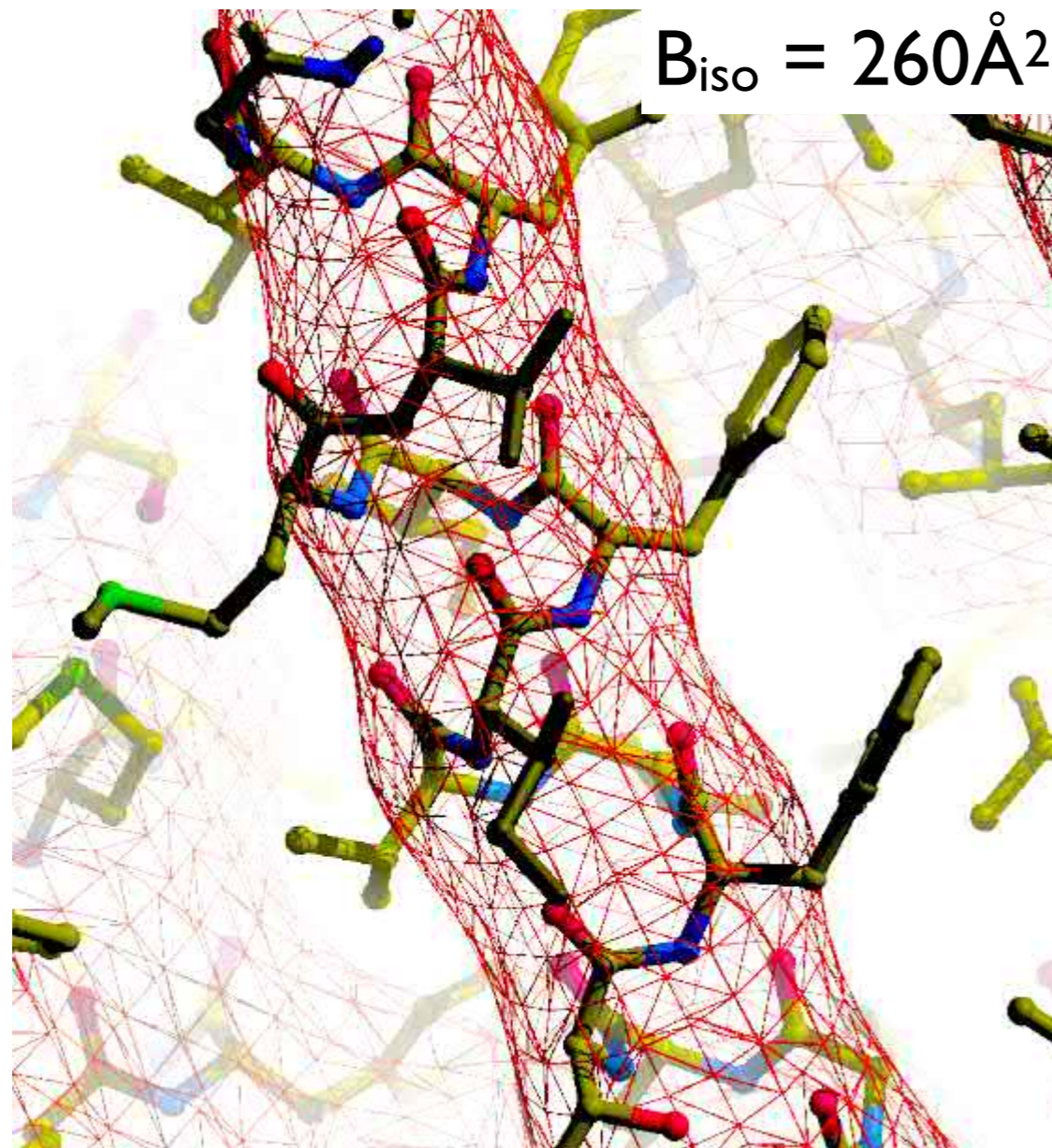
Choose map with maximum of adjusted surface area

adjusted area = surface area – weight \*  
number of regions





# Automated Map Sharpening



*Deposited Map*

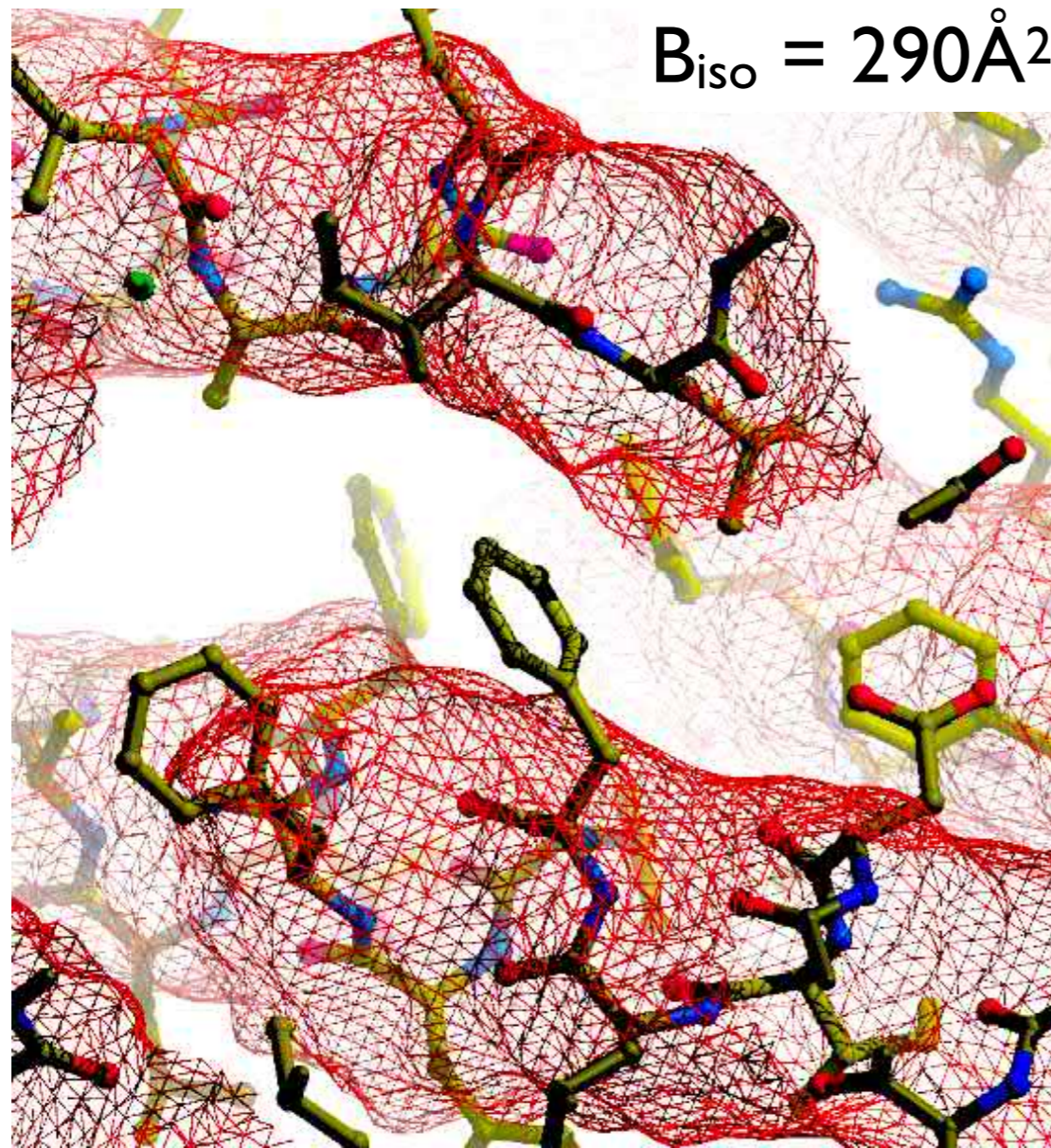
*Autosharpened Map*

High-conductance Ca(2+)-activated K(+) channel (emd\_8414 and PDB entry 5tji; Hite et al., 2017)

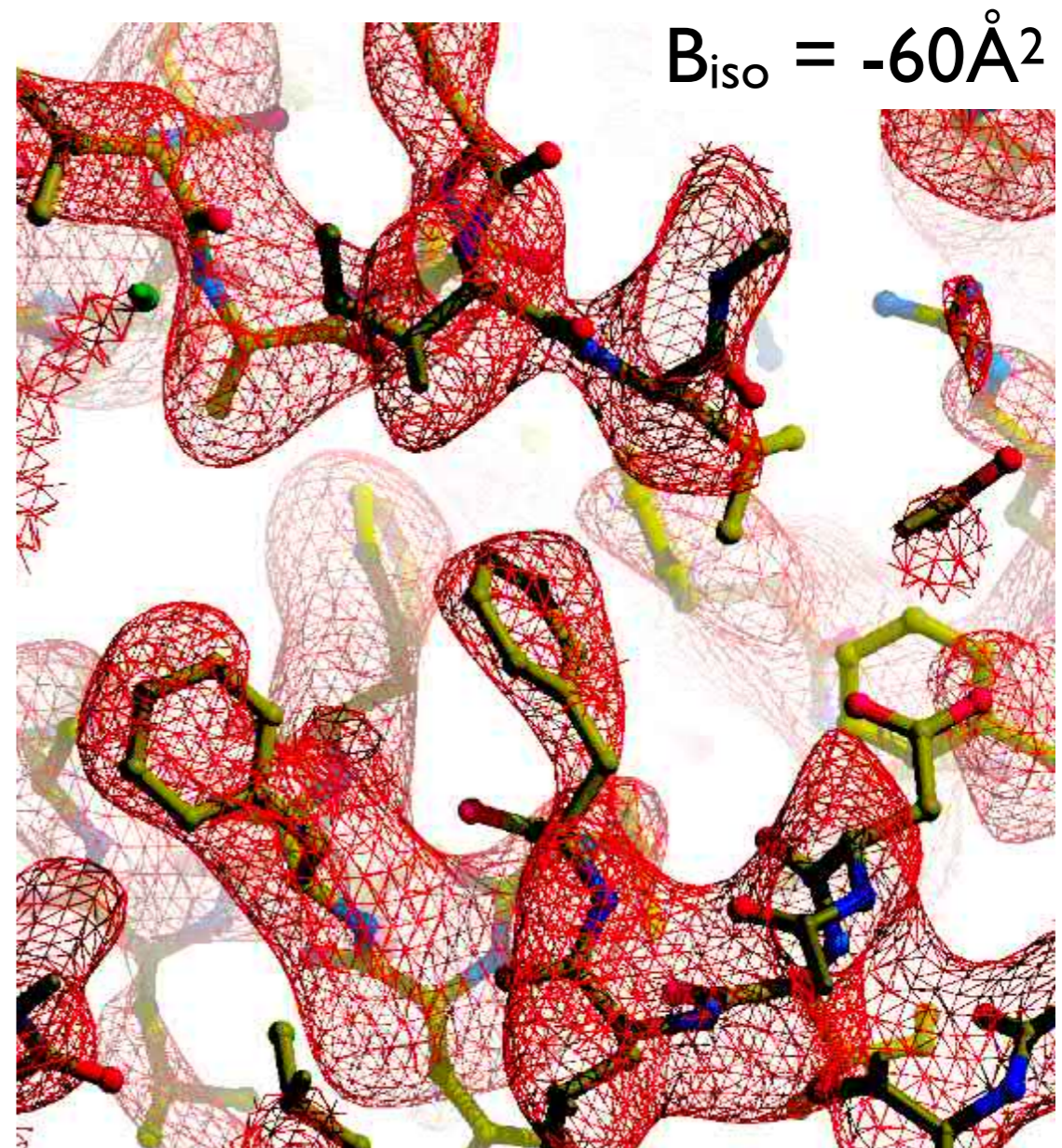
**Phenix**



# Automated Map Sharpening



*Deposited Map*

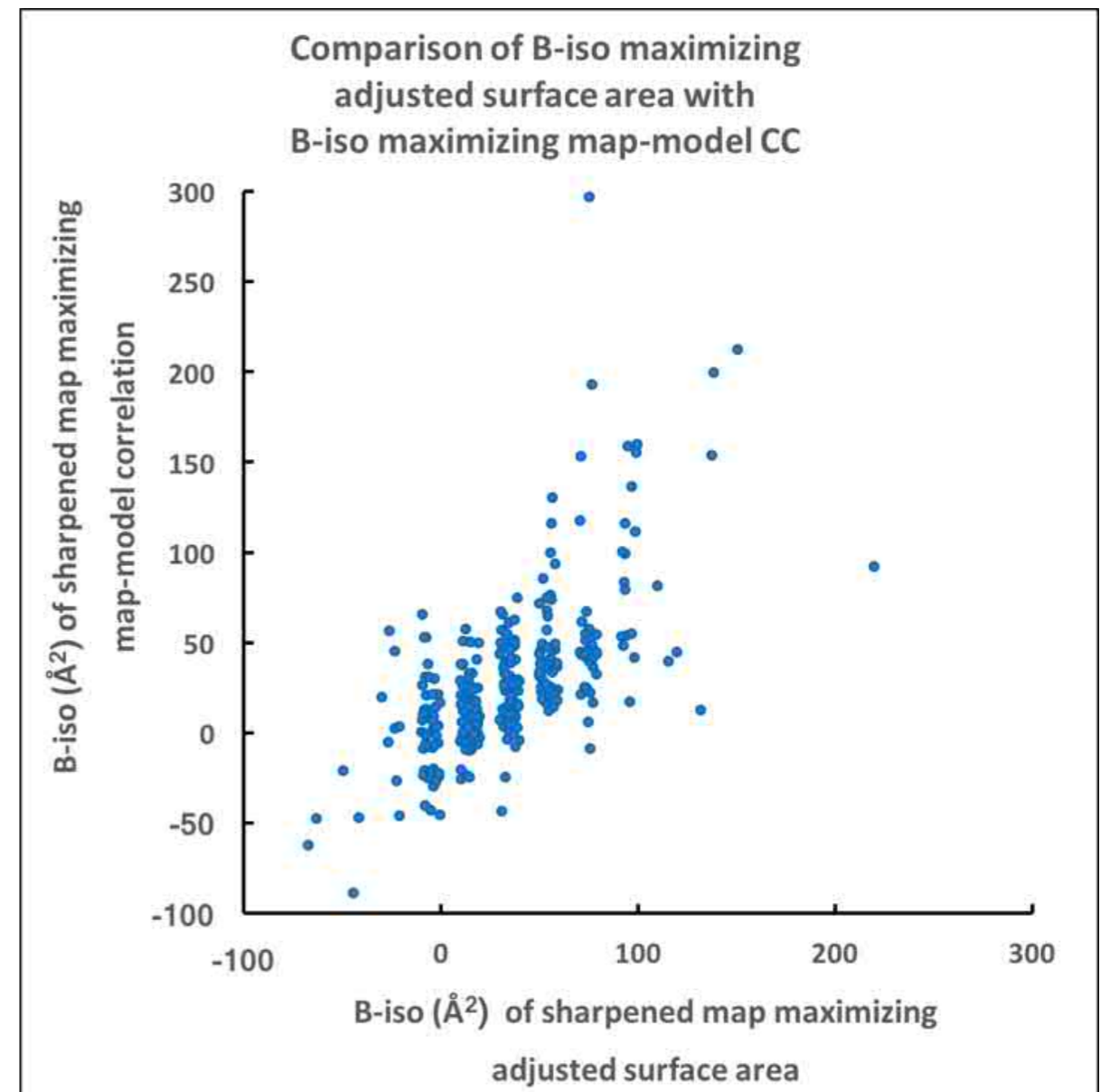
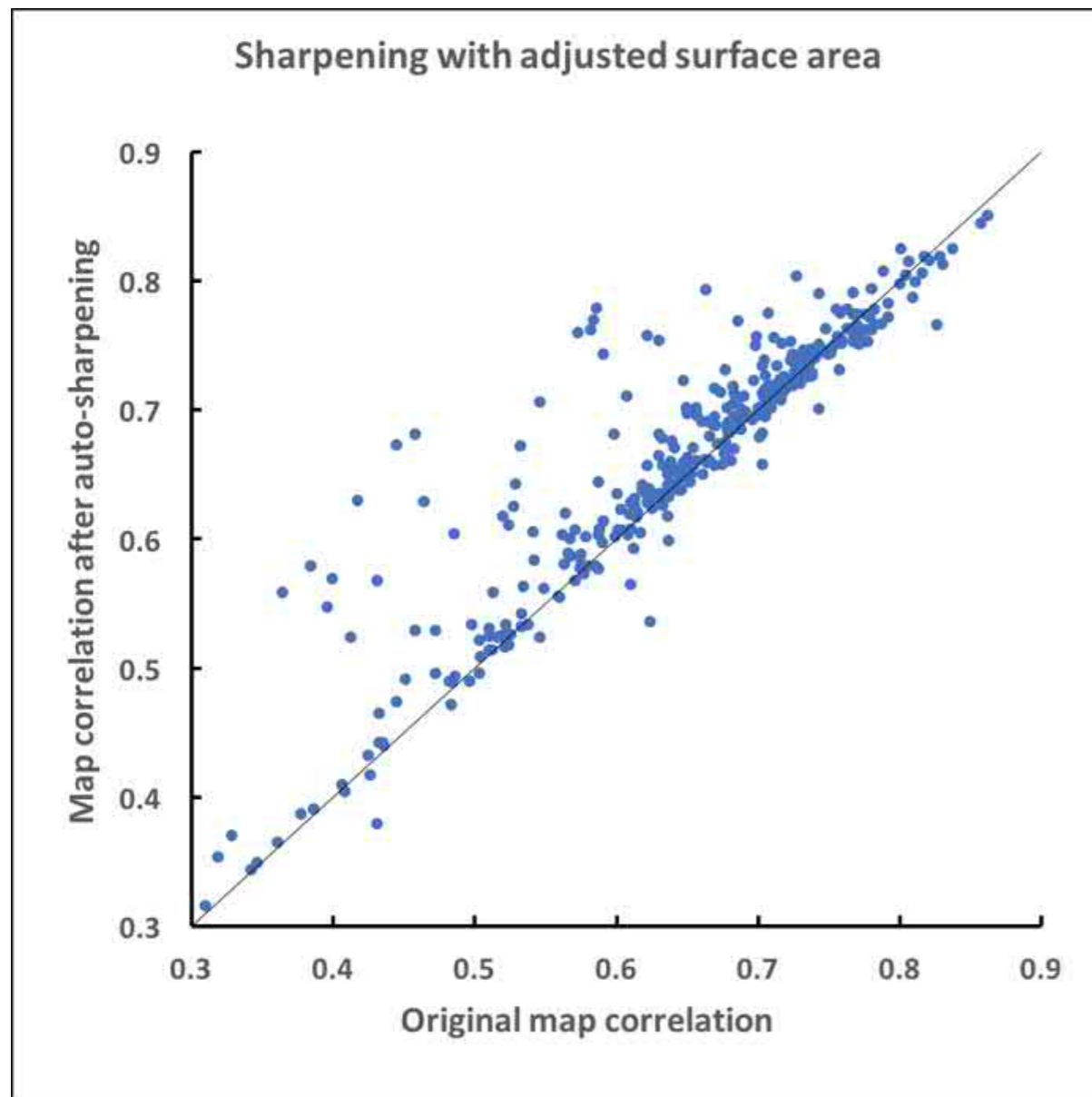


*Autosharpened Map*

Cystic fibrosis transmembrane conductance regulator  
(emd\_8461 and PDB entry 5uar; Zhang and Chen, 2016)



# 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

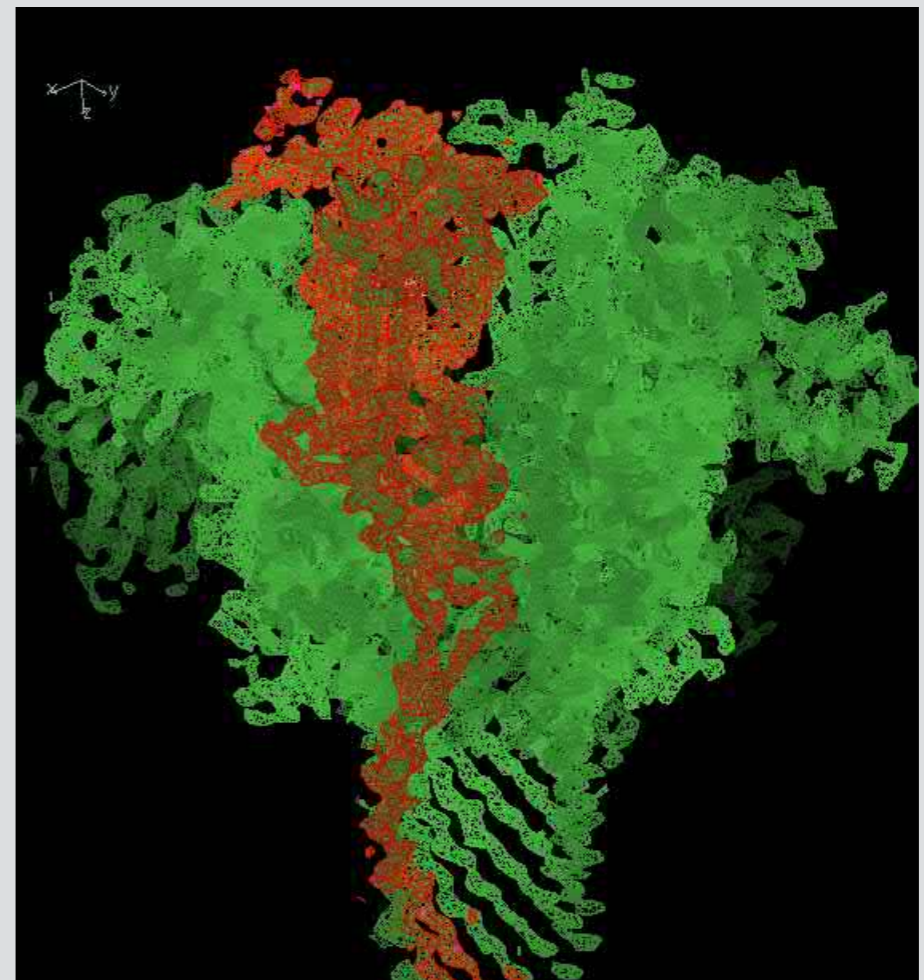
Idealize secondary structure and refine

Assemble and refine (protein/RNA/DNA)

Apply molecular symmetry and re-refine

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

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



emd\_6224 (anthrax toxin protective antigen pore at 2.9 Å; Jiang et al. 2015)

# 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

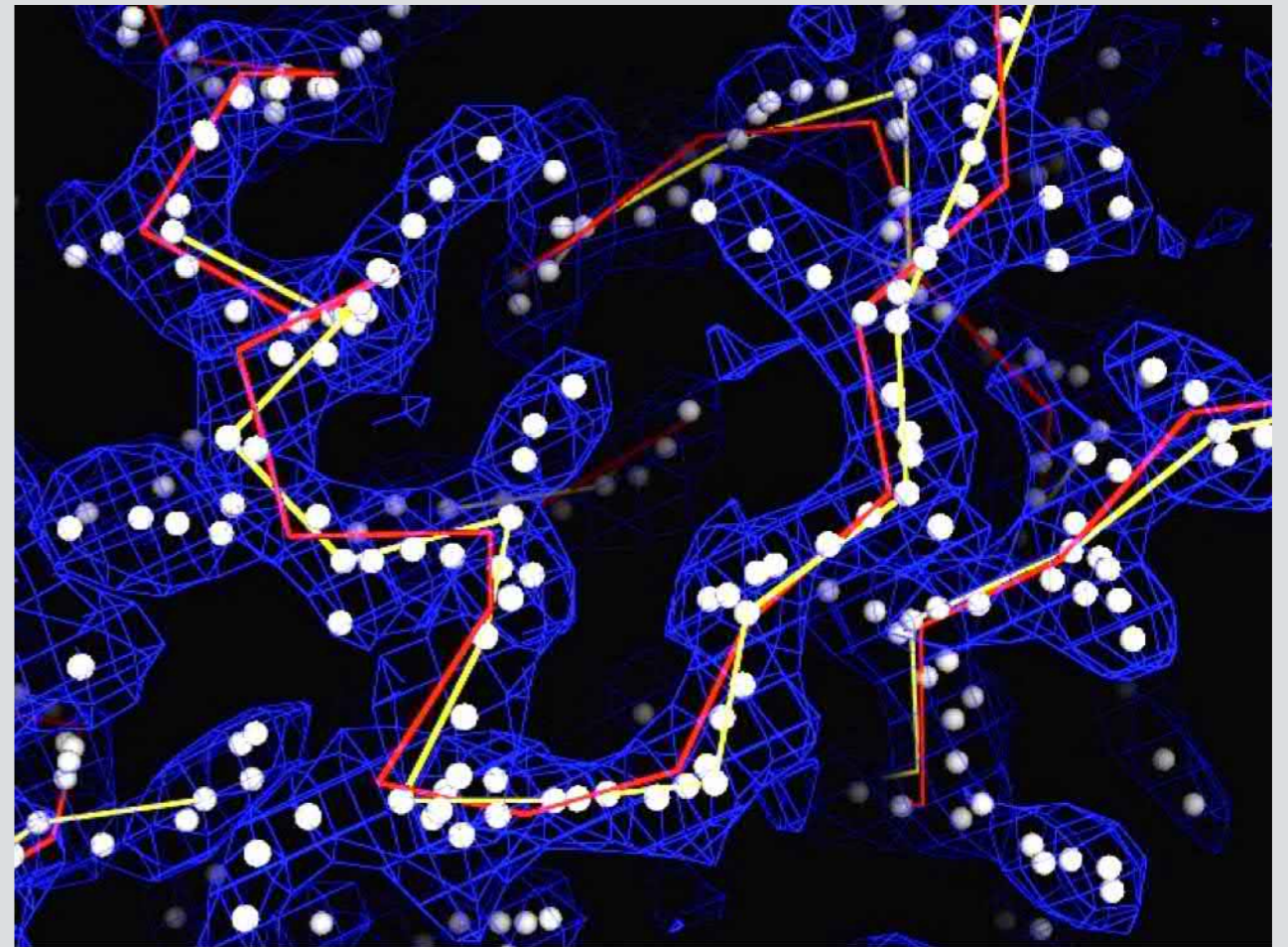


Assemble and refine (protein/RNA/DNA)



Apply molecular symmetry and re-refine

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





# Idealization and Refinement

Determine optimal sharpening of the map

Cut out asymmetric unit of the map

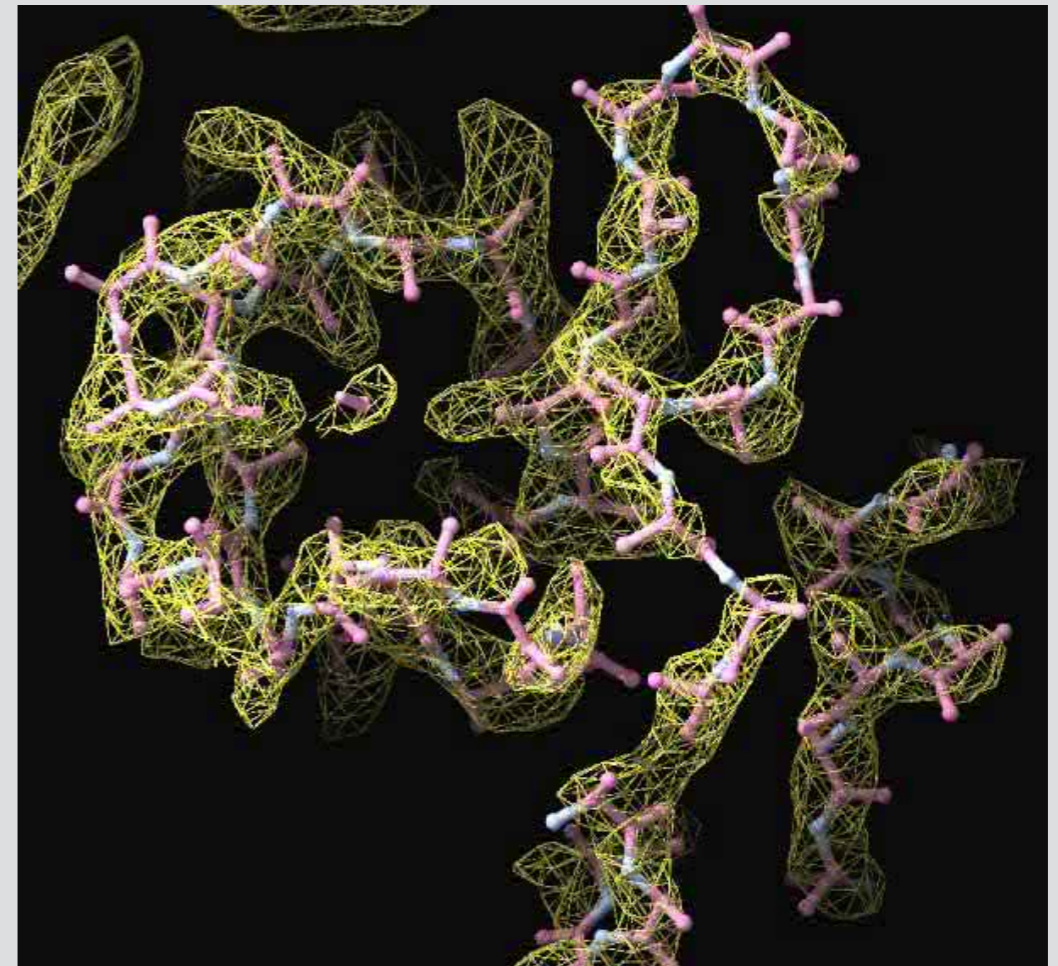
Trace chain and build model

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 ( $\beta$ -sheets,  $\alpha$ -helices) and use them as restraints in real-space refinement



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

# Assembly and Polymer Recognition

Determine optimal sharpening of the map

Cut out asymmetric unit of the map

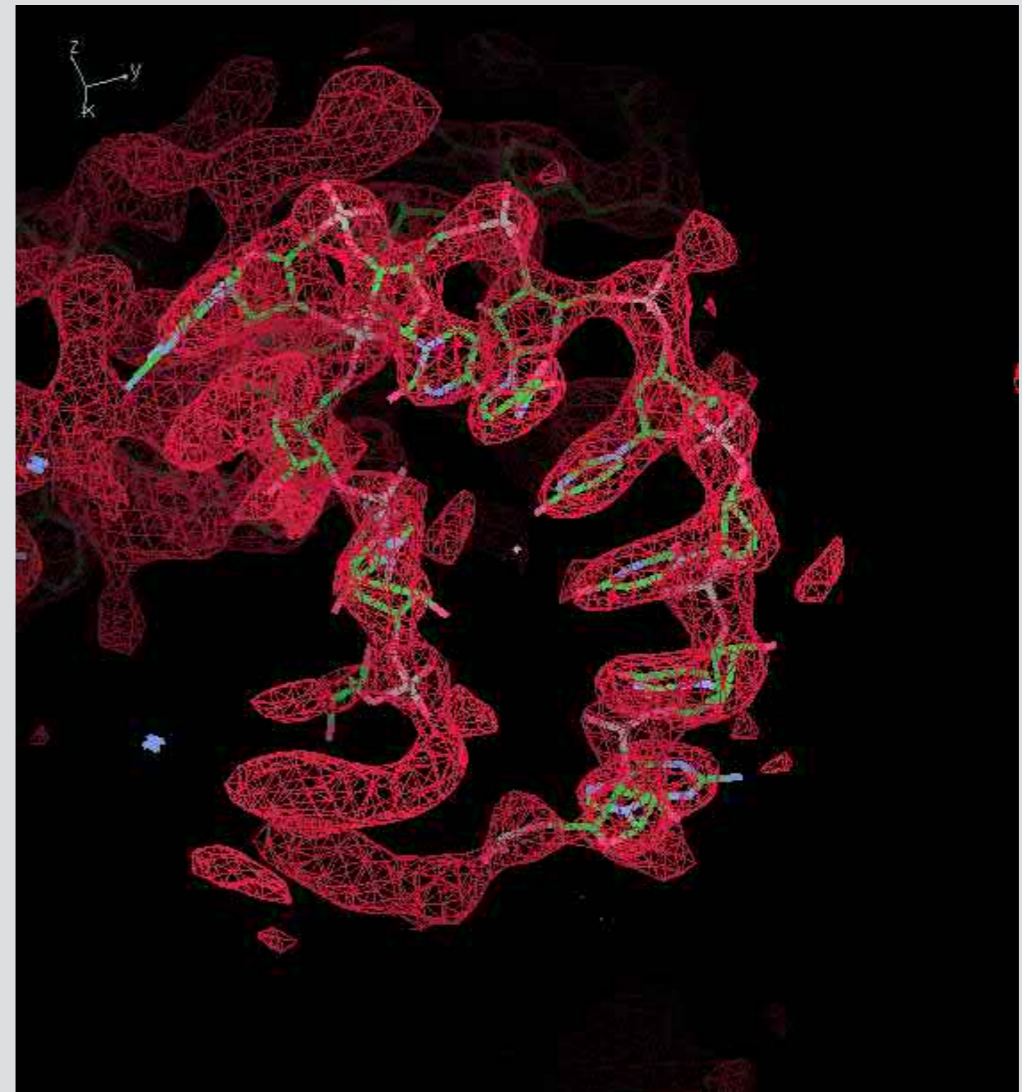
Trace chain and build model

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



# 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

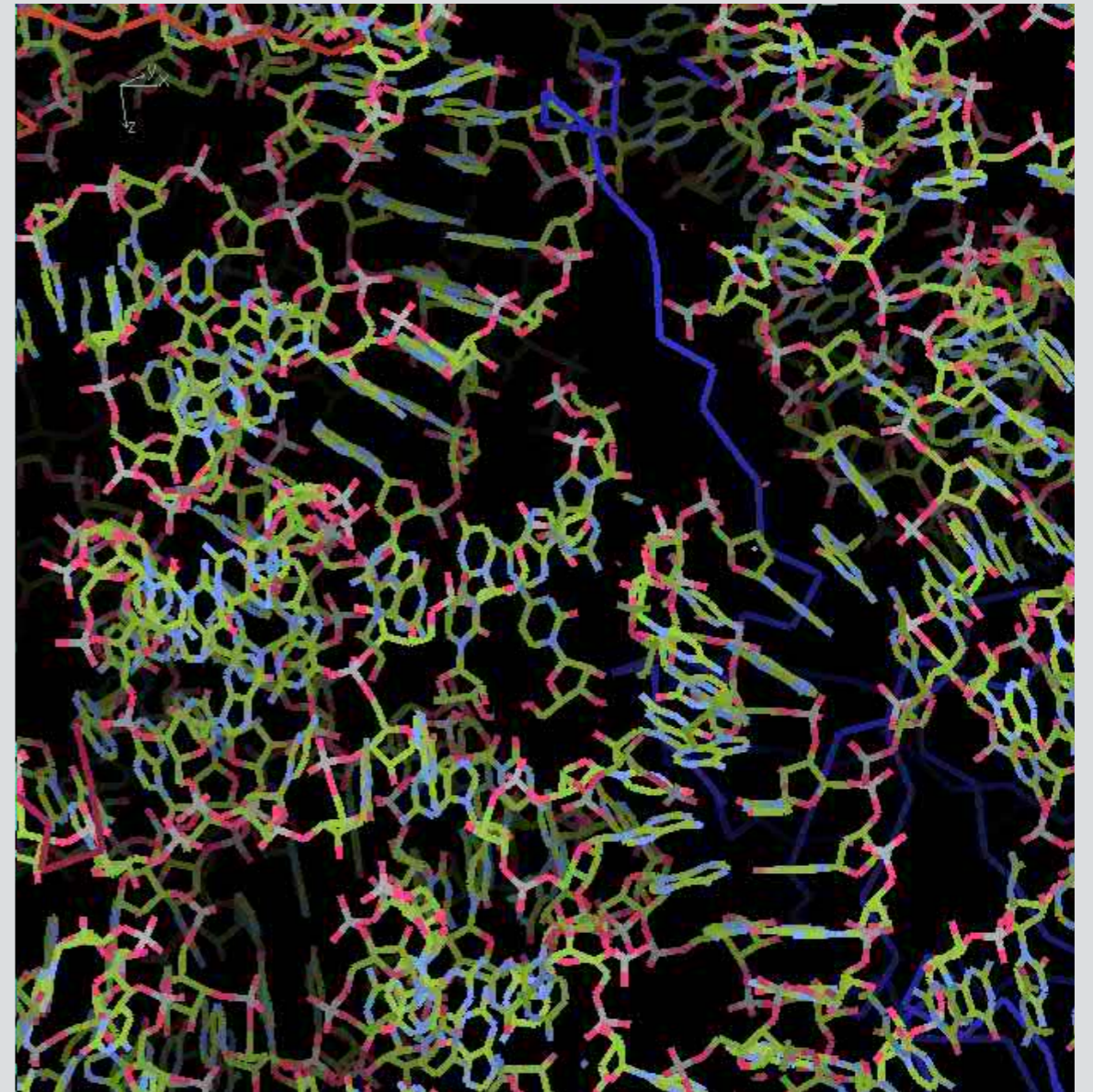


Assemble and refine (protein/RNA/DNA)



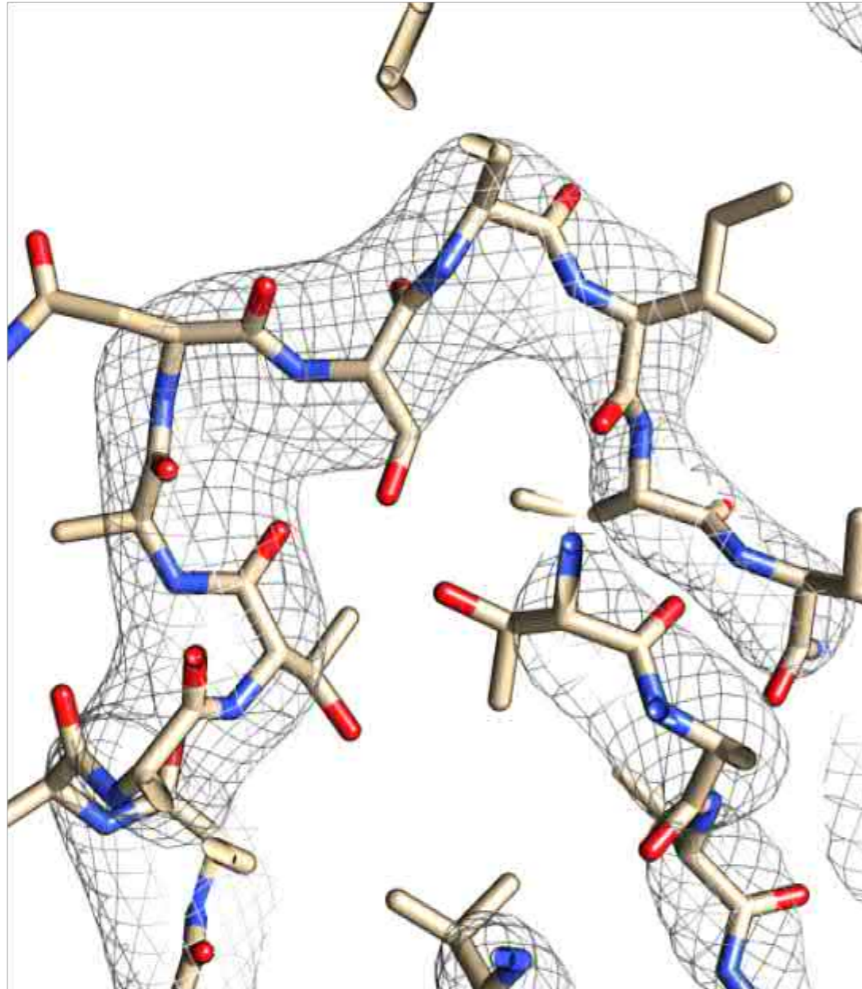
Apply molecular symmetry and re-refine

- `phenix.map_to_model`

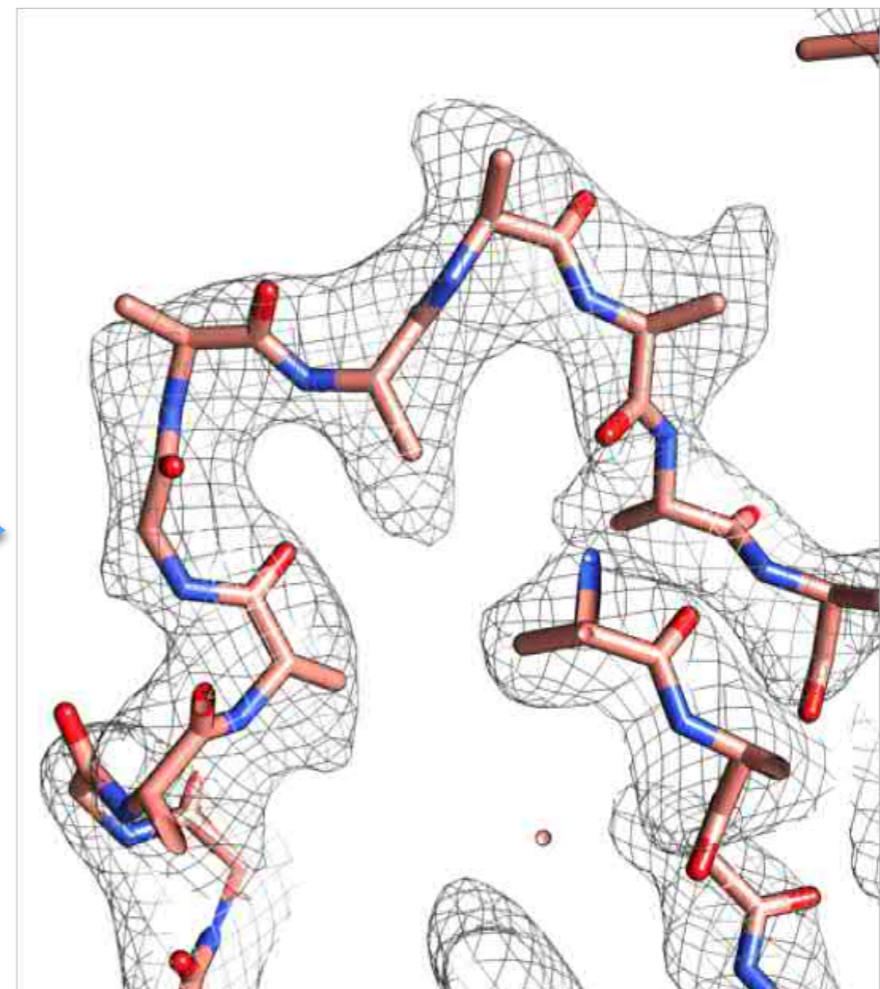
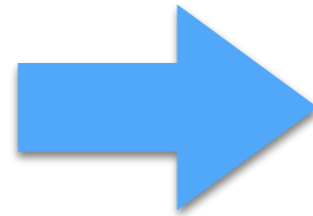


30S Ribosome (1j5e, 2.9 Å)

# Automated Building - Sharpening



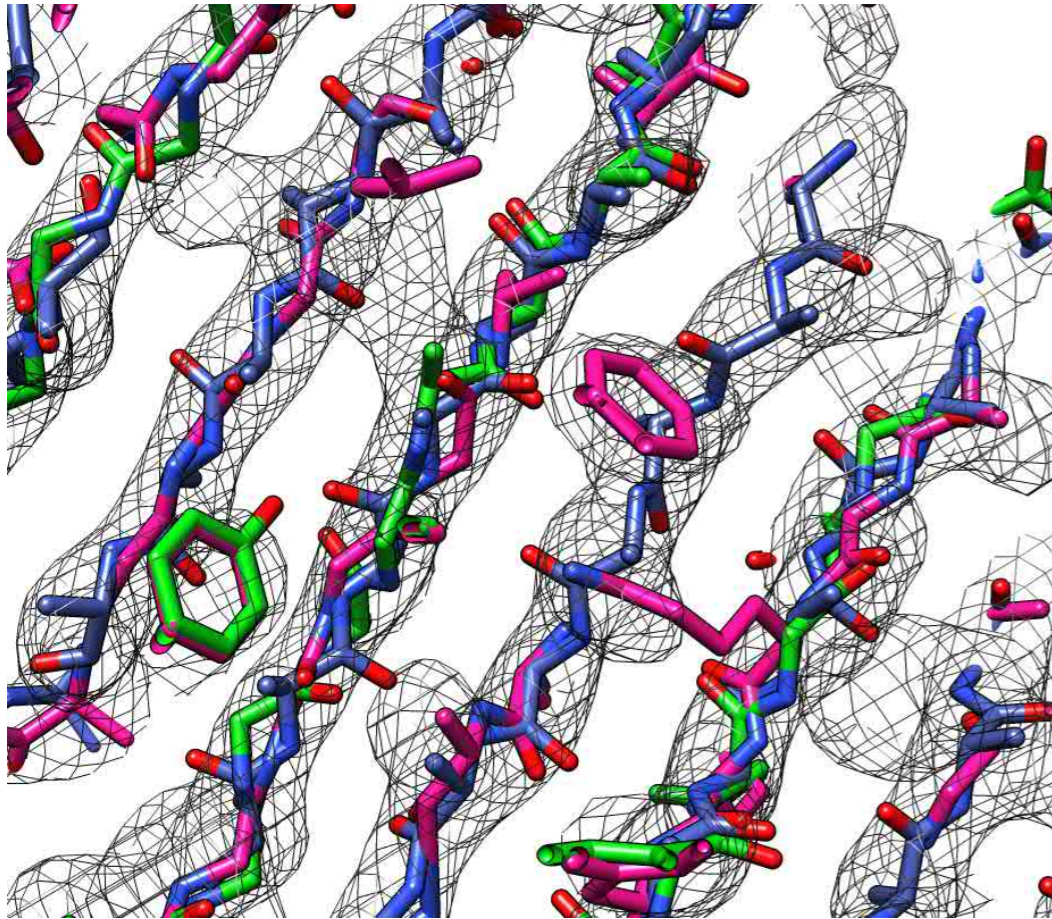
Original



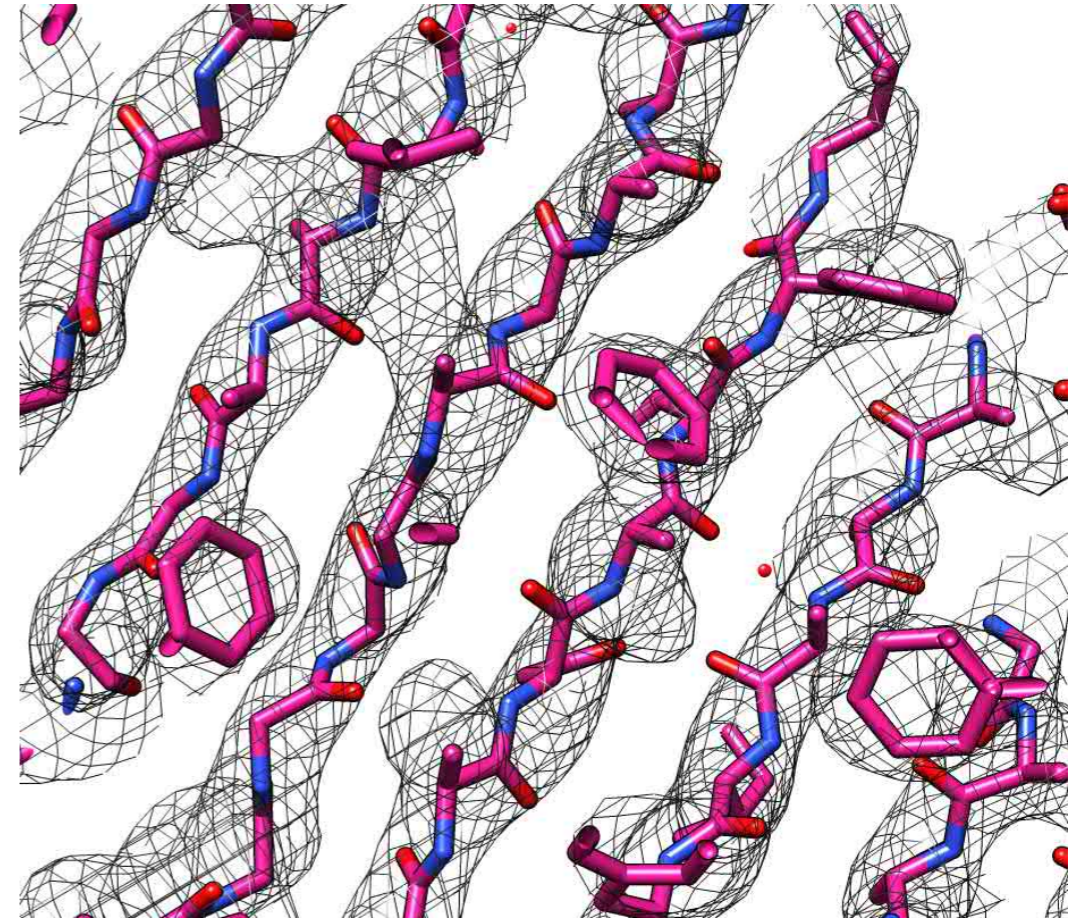
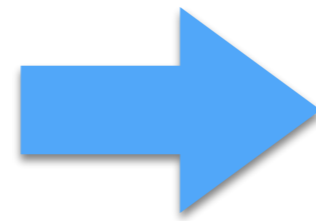
Automatically Sharpened



# Automated Building - Combining Multiple Models



Three Independently Built Models



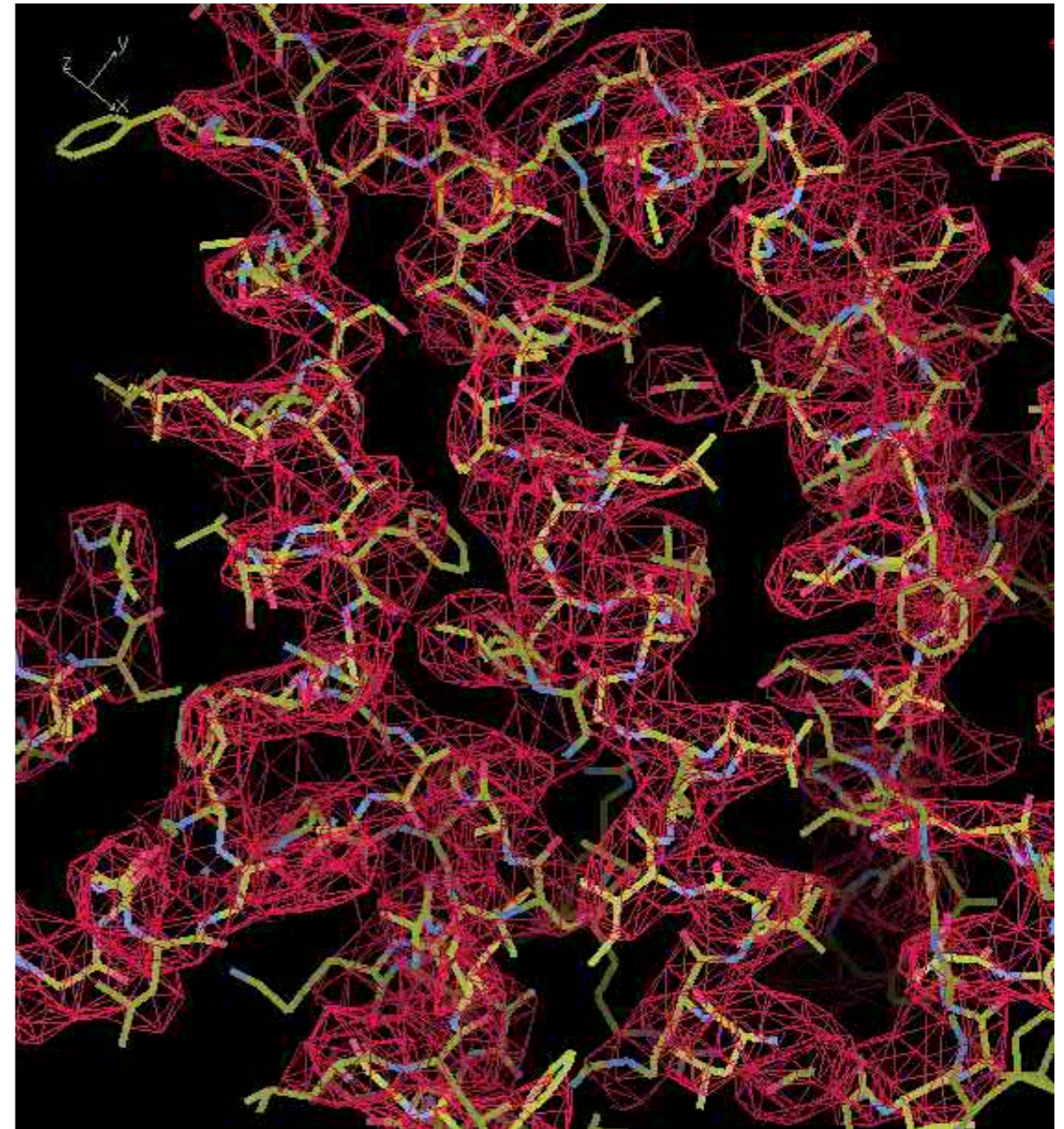
Composite Model



# Building at Low Resolution



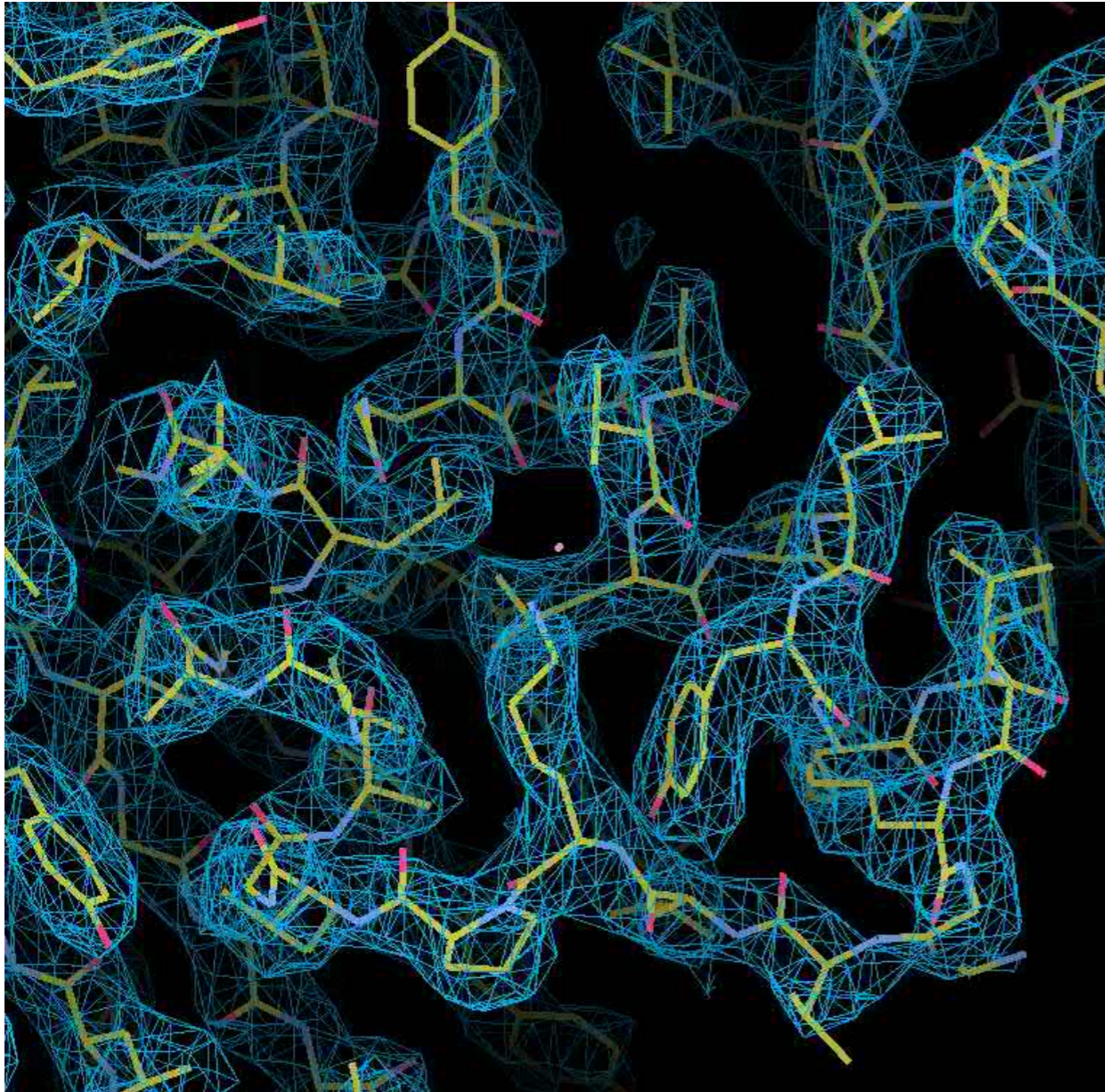
Gamma-secretase at 4.5 Å  
(autobuilt model; emd\_2677)



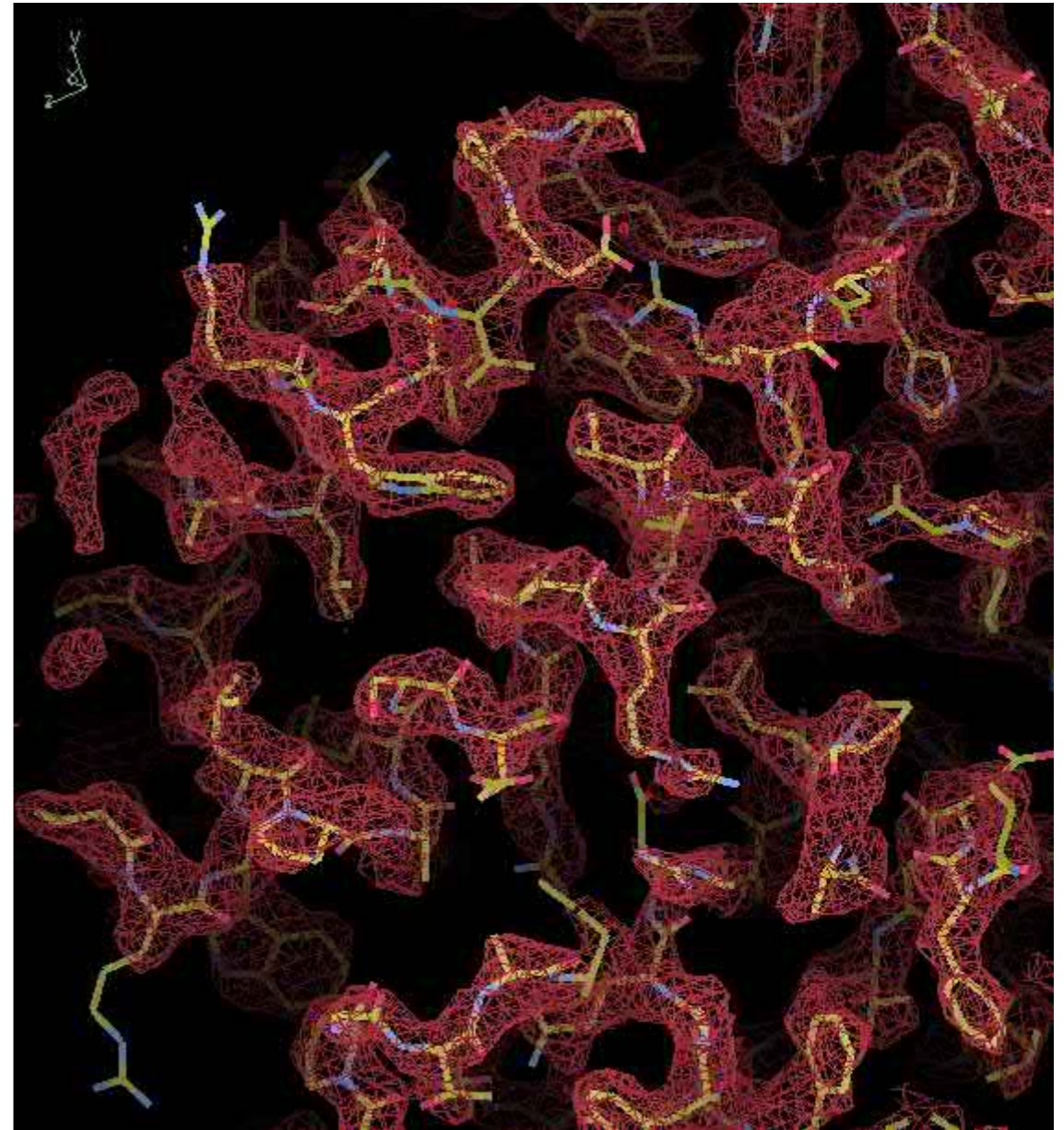
Gamma-secretase structure at 3.4 Å  
(autobuilt model; emd\_3061)



# Building at Medium/High Resolution



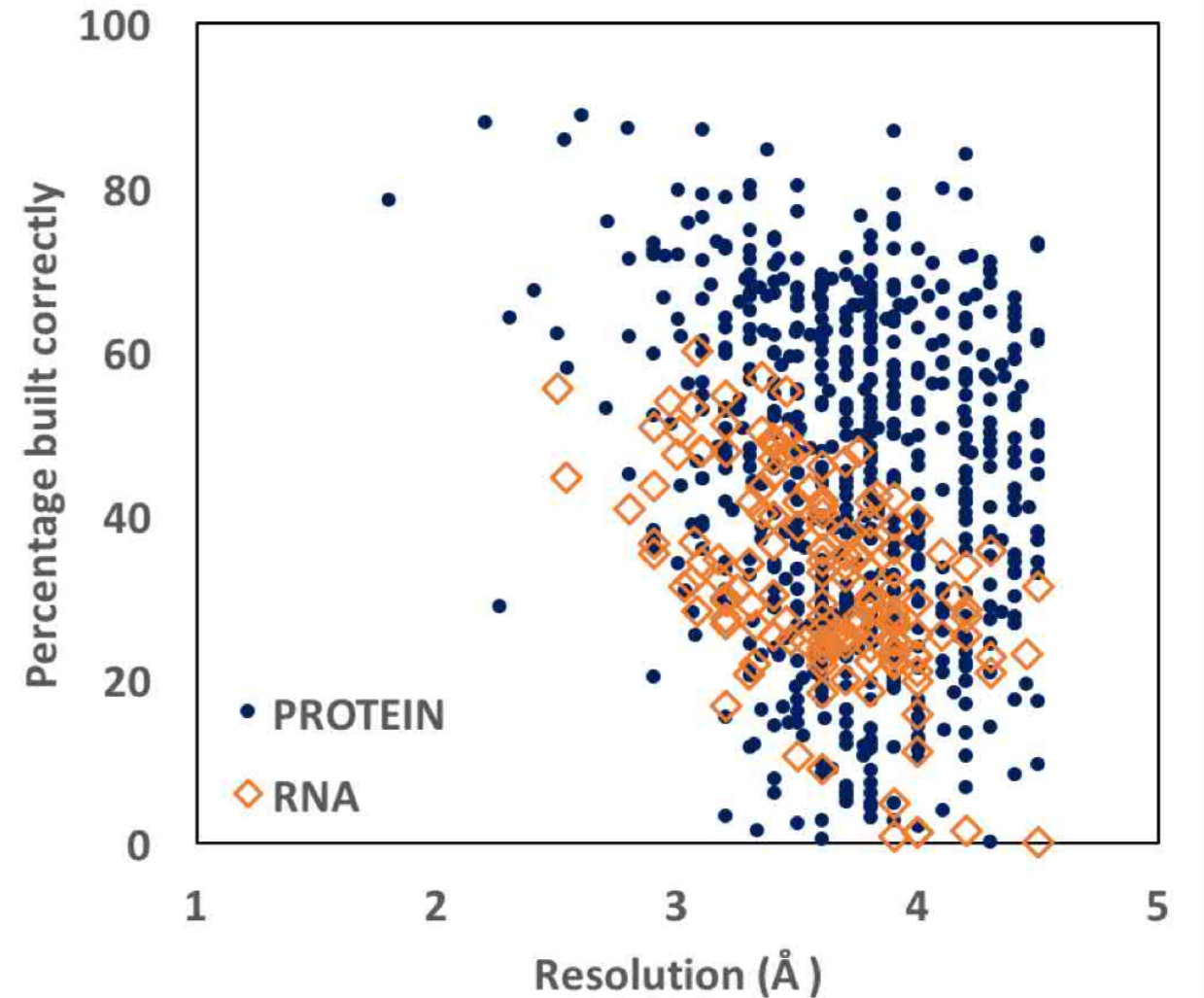
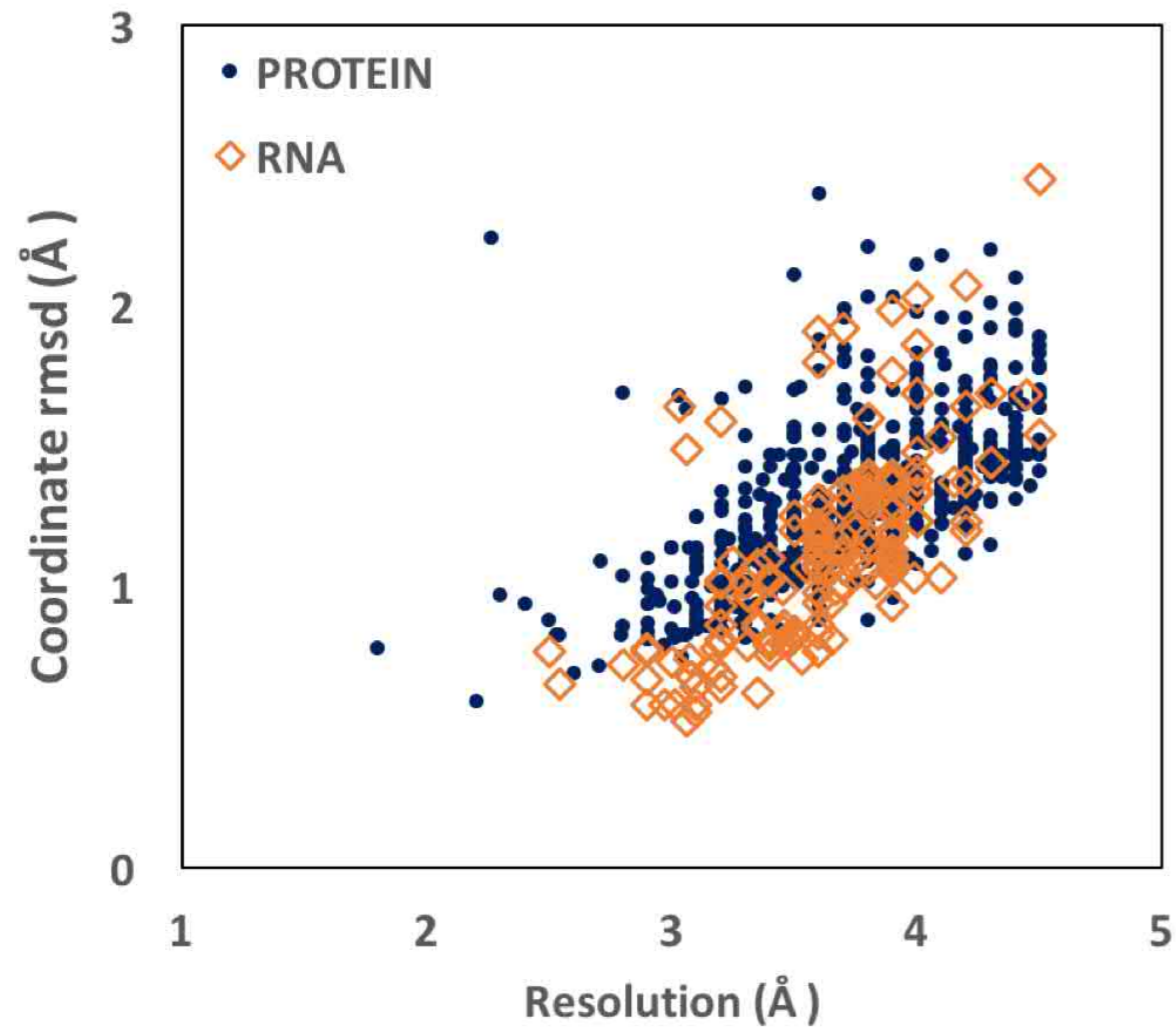
Proteasome at 2.8 Å  
(autobuilt model; emd\_6287)



Beta-galactosidase at 2.2 Å  
(autobuilt model; emd\_2984)



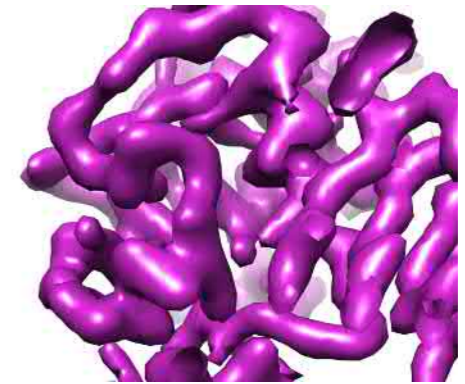
# Autobuilding Performance



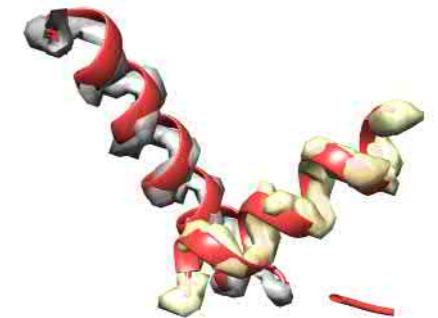


# Model Building Version 2

Trace chain the way a person does

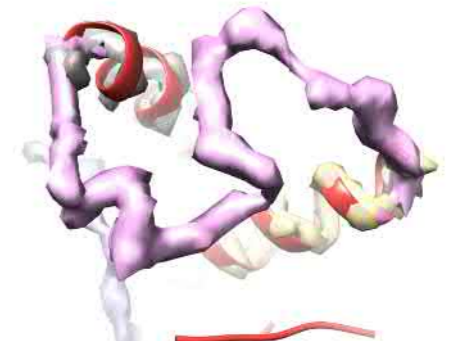


Find secondary structure

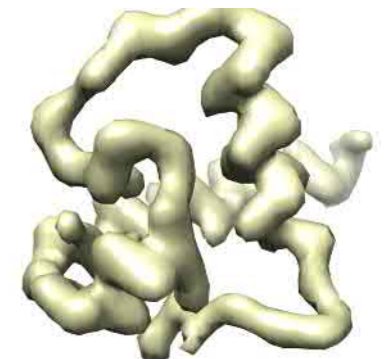


Find clear regions of density

Adjust contour level until a region just connects to another

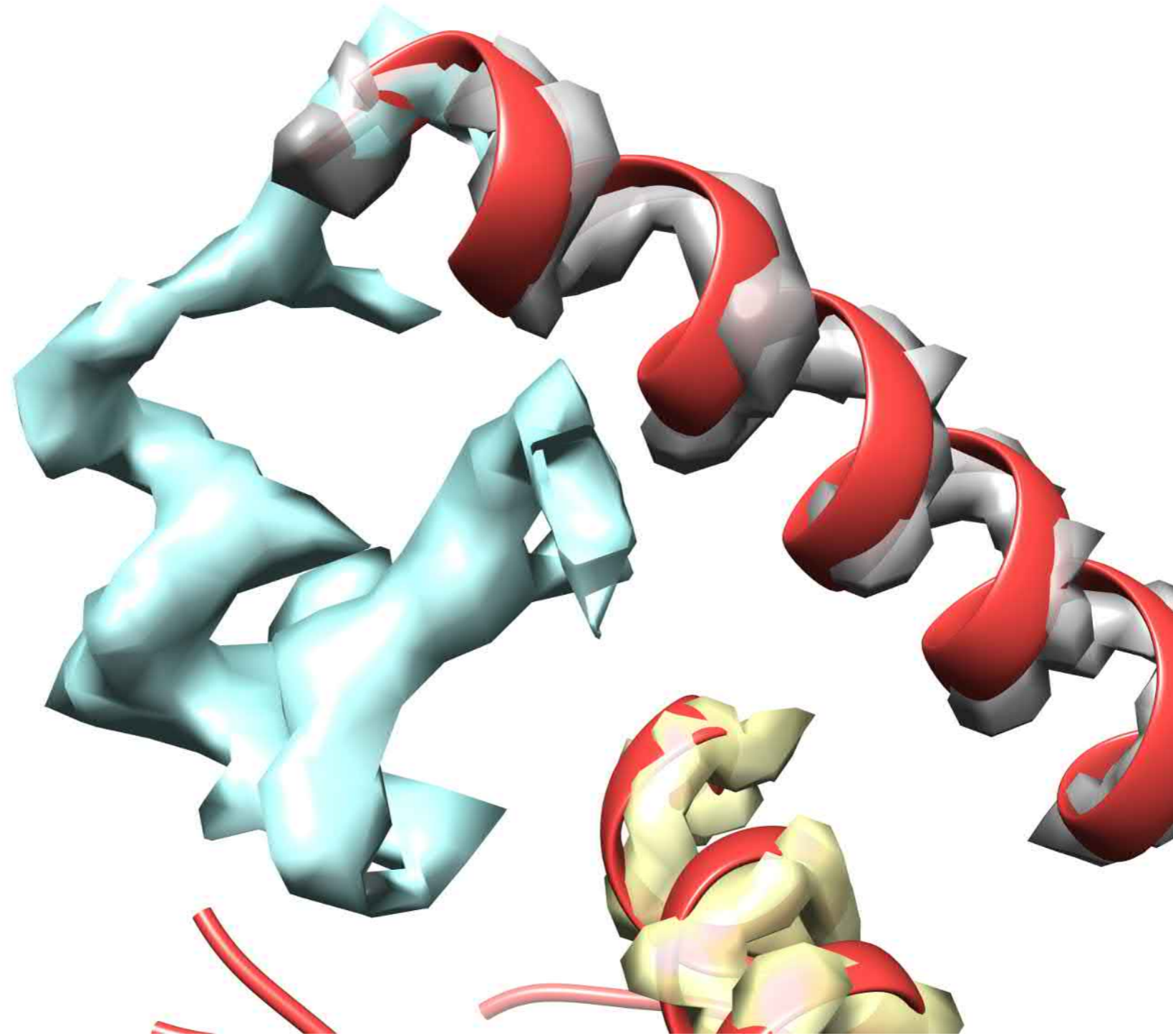


Iterate to build up a connected chain



**Phenix**

# Model Building Version 2



  
**Phenix**



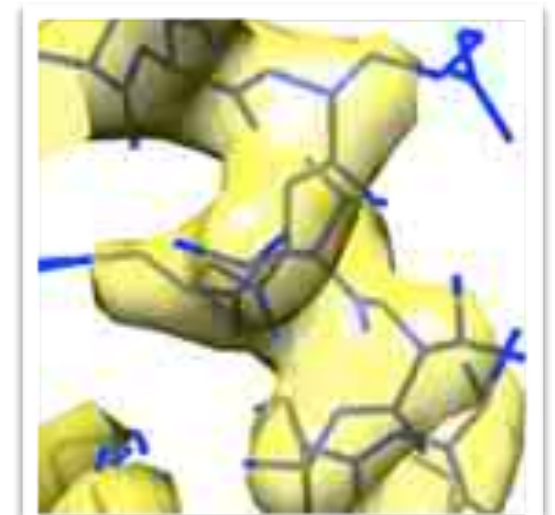
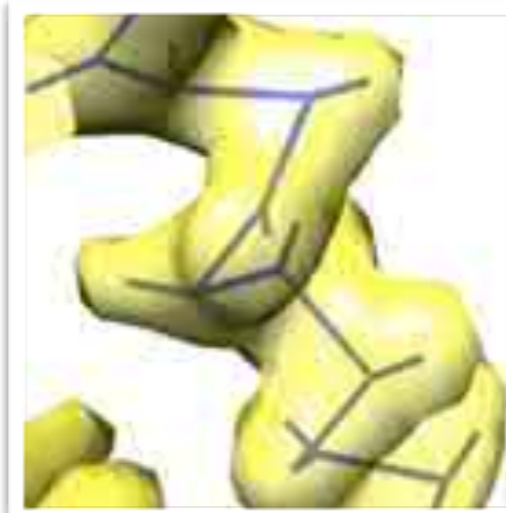
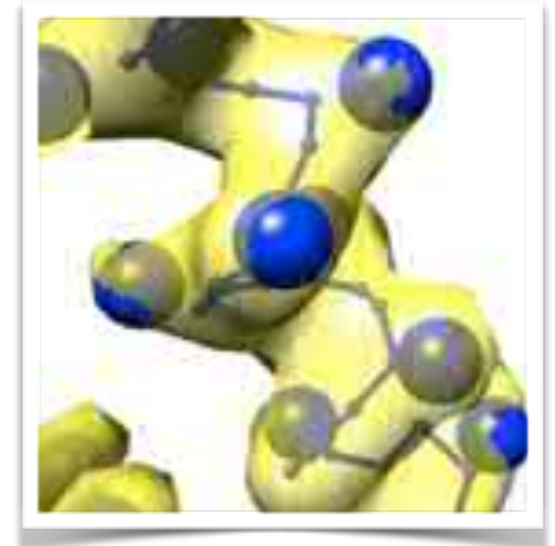
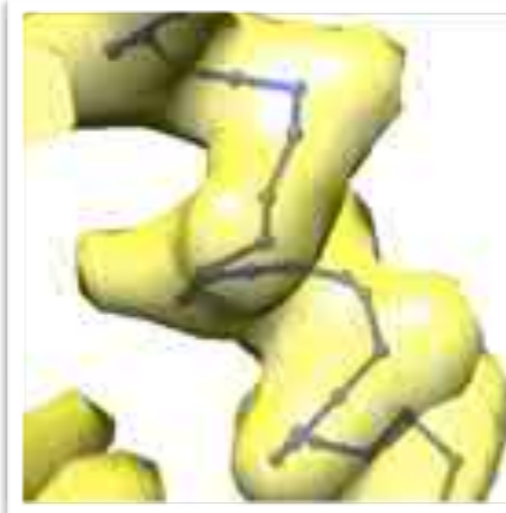
# Finding $C_{\alpha}$ and $C_{\beta}$ positions

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



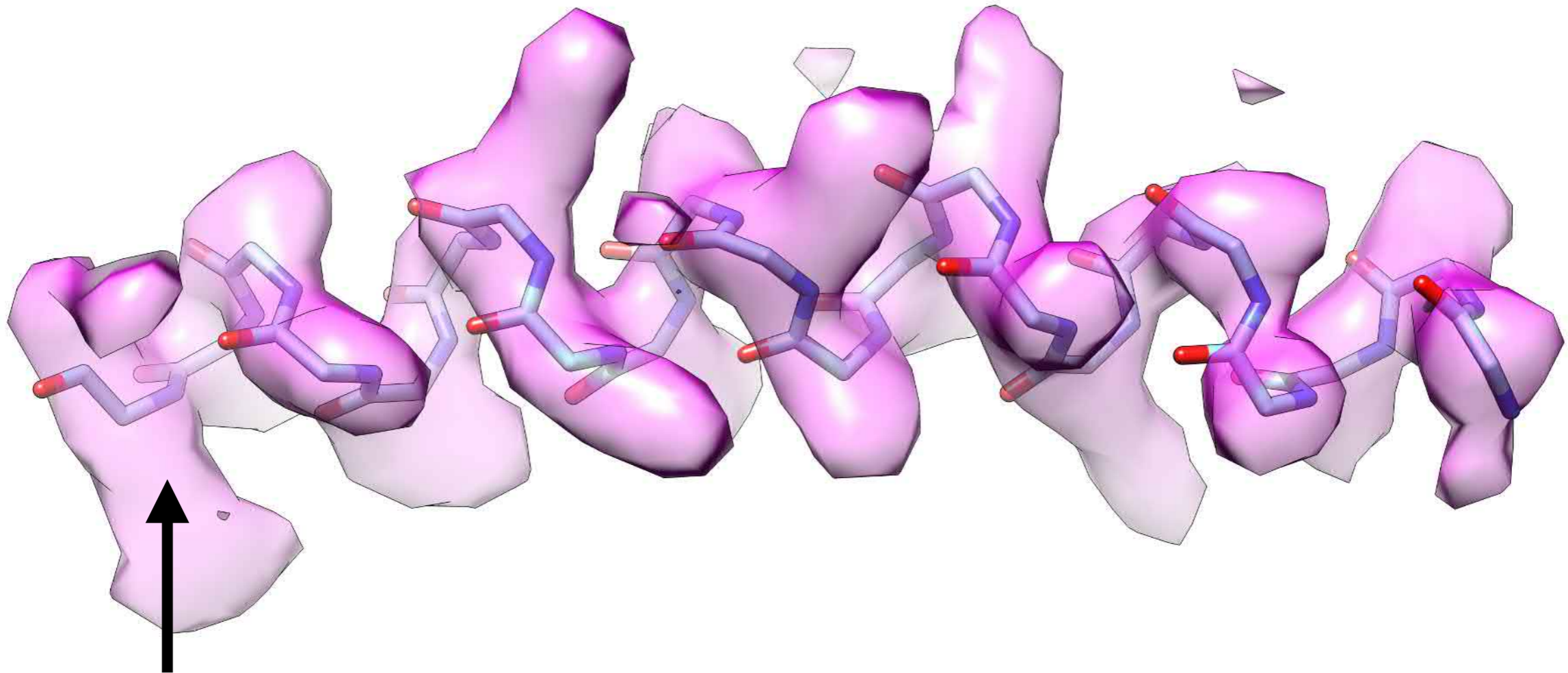
\*Rotkiewicz & Skolnick (2008).  
*J. Comp. Chem.* 29, 1460.

  
**Phenix**





# Sequence Assignment



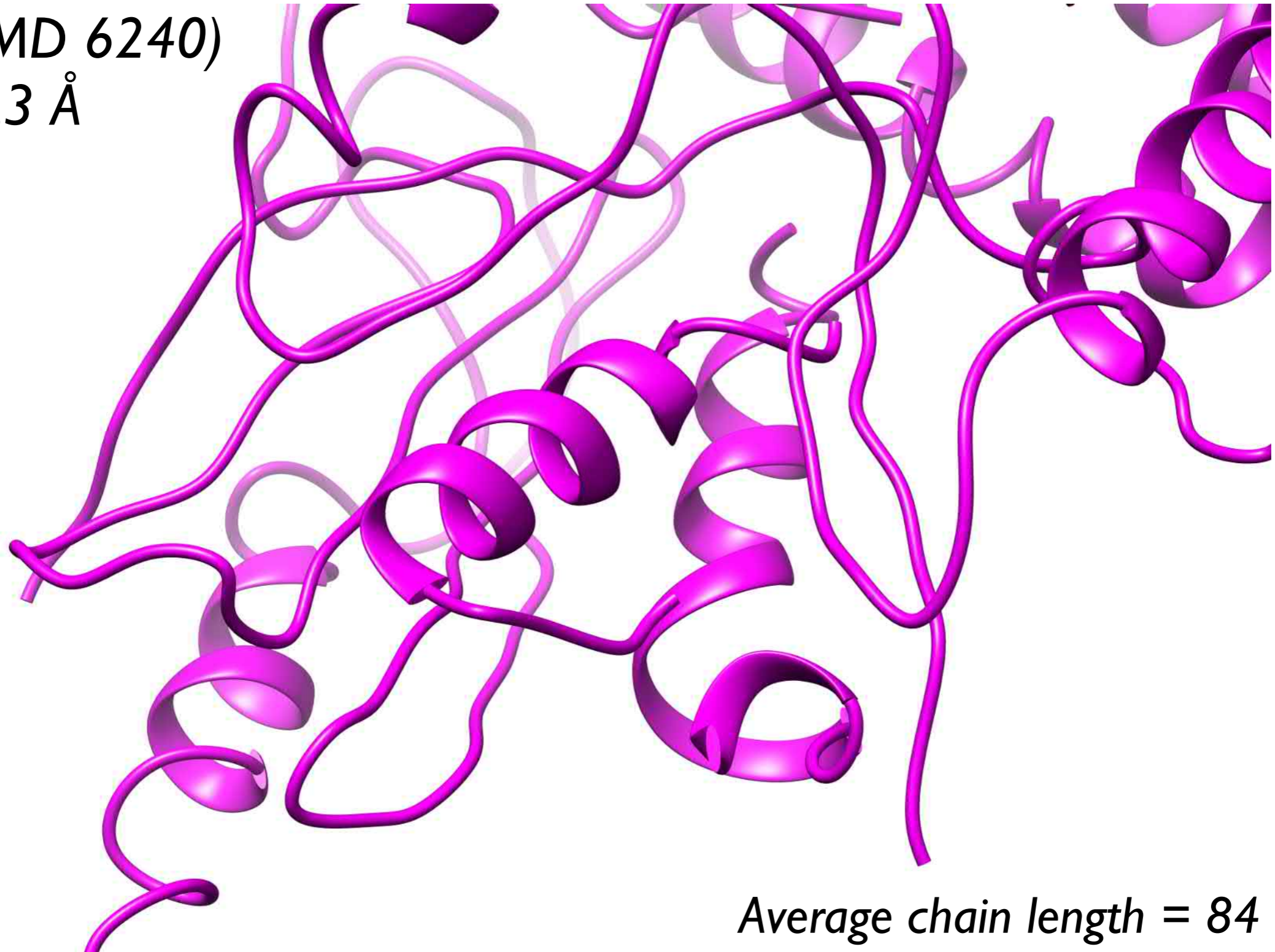
Residue	G	A	S	V	I	L	M	C	F	Y	K	R	W	H	E	D	Q	N	P	T
CC	0.30	0.50	0.53	0.47	0.58	0.62	0.68	0.59	0.83	0.77	0.71	0.69	0.70	0.82	0.65	0.64	0.60	0.60	0.35	0.47
Prob	3	0	0	0	0	0	1	0	40	23	5	5	4	9	2	2	1	0	2	0

- Determine probability of side chain at each  $C_{\alpha}$
- Align sequence to maximize total probability for the chain



# Improved Connectivity

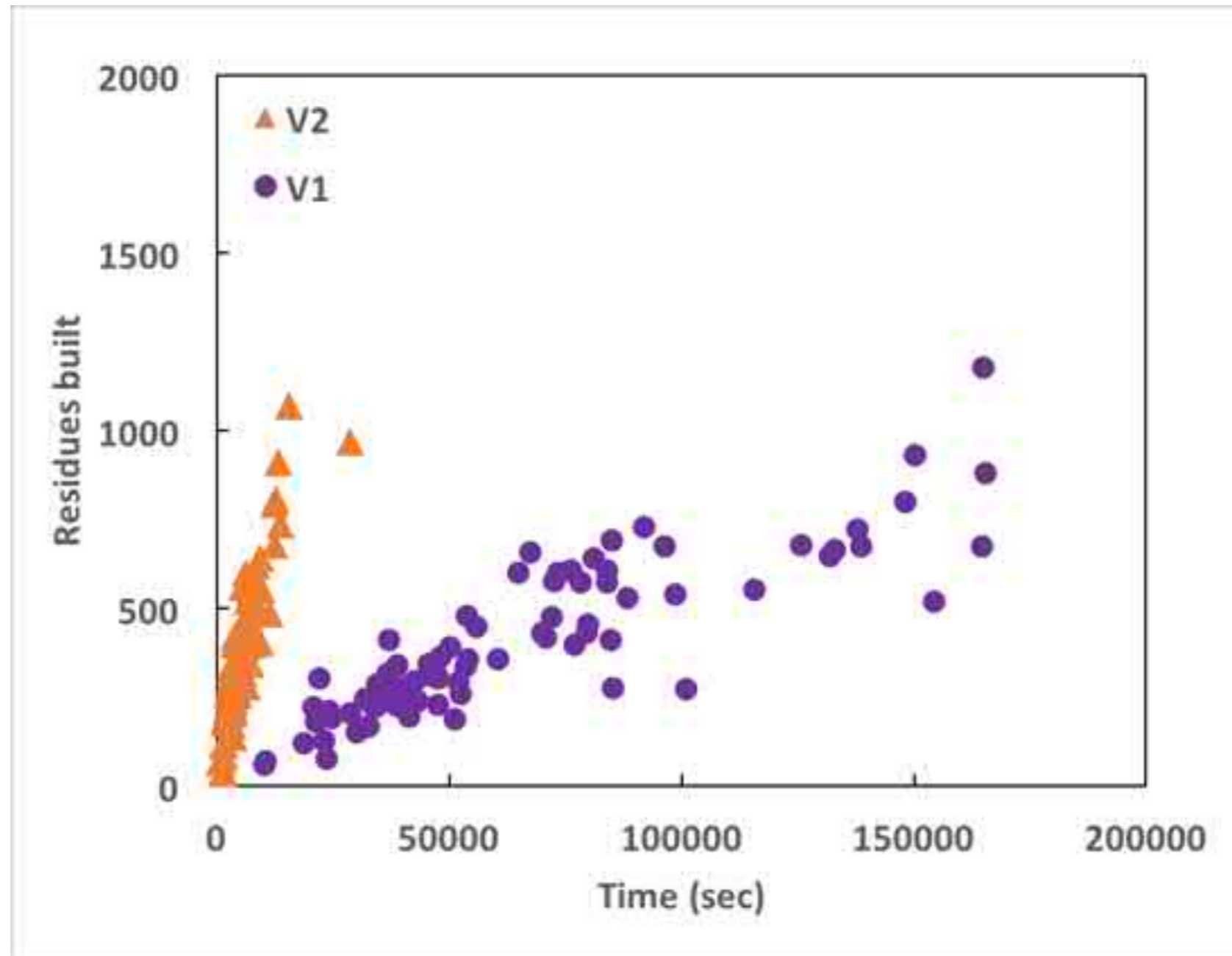
3j9e (EMD 6240)  
3.3 Å



Average chain length = 84

  
**Phenix**

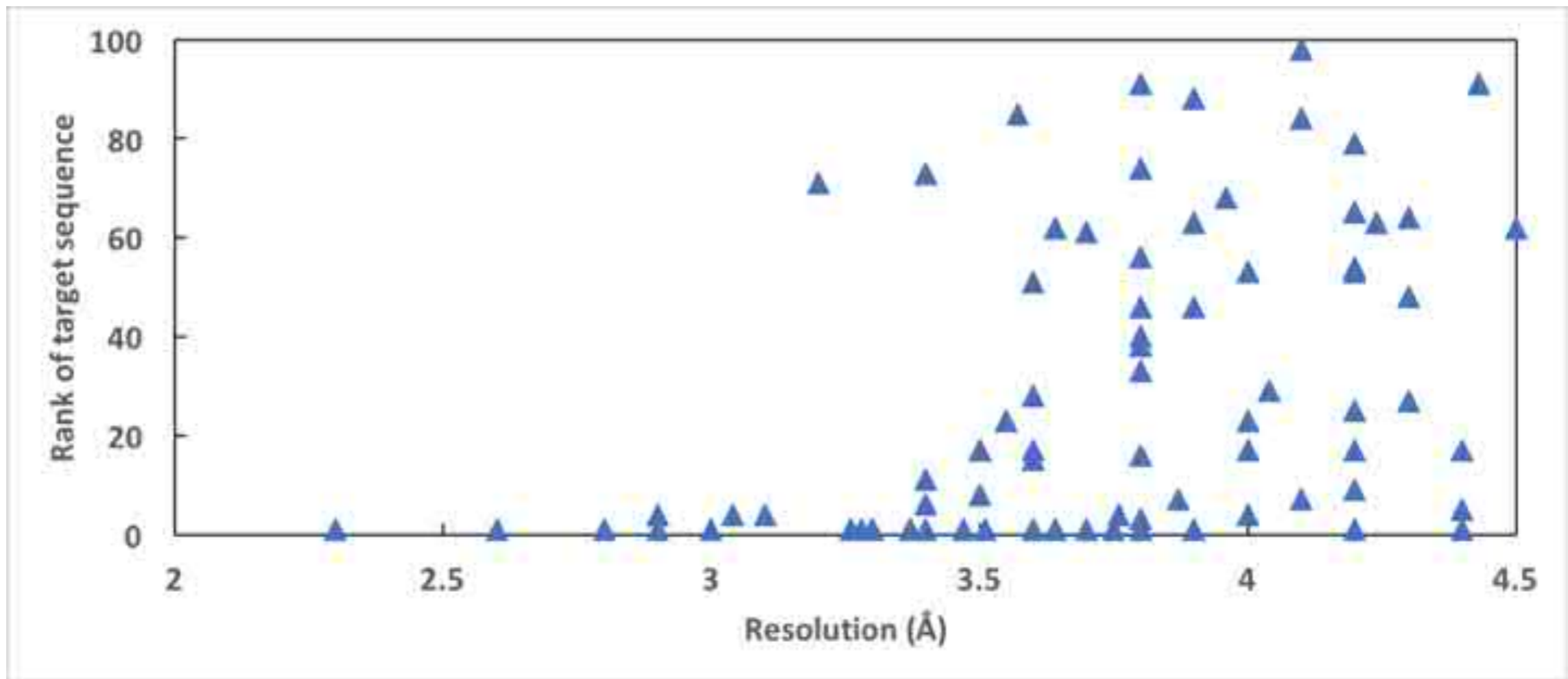
# Improved Performance





# 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*

**Phenix**

# 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



# Acknowledgements

## **Berkeley Laboratory**

Pavel Afonine, Youval Dar, Nat Echols, Jeff Headd, Richard Gildea, Ralf Grosse-Kunstleve, Dorothee Liebschner, Nigel Moriarty, Nader Morshed, Billy Poon, Ian Rees, Nicholas Sauter, Oleg Sobolev, Peter Zwart

## **Los Alamos Laboratory/New Mexico Consortium**

Tom Terwilliger, Li-Wei Hung

## **Baylor College of Medicine**

Matt Baker

## **Cambridge University**

Randy Read, Airlie McCoy, Gabor Bunckozi, Tristan Croll, Rob Oeffner, Kaushik Hatti, Massimo Sammito, Duncan Stockwell, Laurent Storoni

## **Duke University**

Jane Richardson & David Richardson, Ian Davis, Vincent Chen, Jeff Headd, Chris Williams, Bryan Arendall, Bradley Hintze, Laura Murray

## **UC San Francisco**

Ben Barad, Yifan Cheng, Jaime Fraser

## **University of Washington**

Frank DiMaio, Ray Wang, David Baker

## **Oak Ridge National Laboratory**

Marat Mustyakimov, Paul Langan

## **Other Collaborators**

Corey Hryc, Zhao Wang, Wah Chiu  
Pawel Janowski, David Case  
Dale Tronrud, Donnie Berholz, Andy Karplus  
Alexandre Urzhumtsev & Vladimir Lunin  
Garib Murshudov & Alexi Vagin  
Paul Emsley, Bernhard Lohkamp, Kevin Cowtan  
David Abrahams  
PHENIX Testers & Users

## **Funding**

- NIH/NIGMS: P01GM063210, P50GM062412, P01GM064692, R01GM071939
- PHENIX Industrial Consortium
- Lawrence Berkeley Laboratory

