

# New Tools for Electron Cryo-Microscopy in Phenix

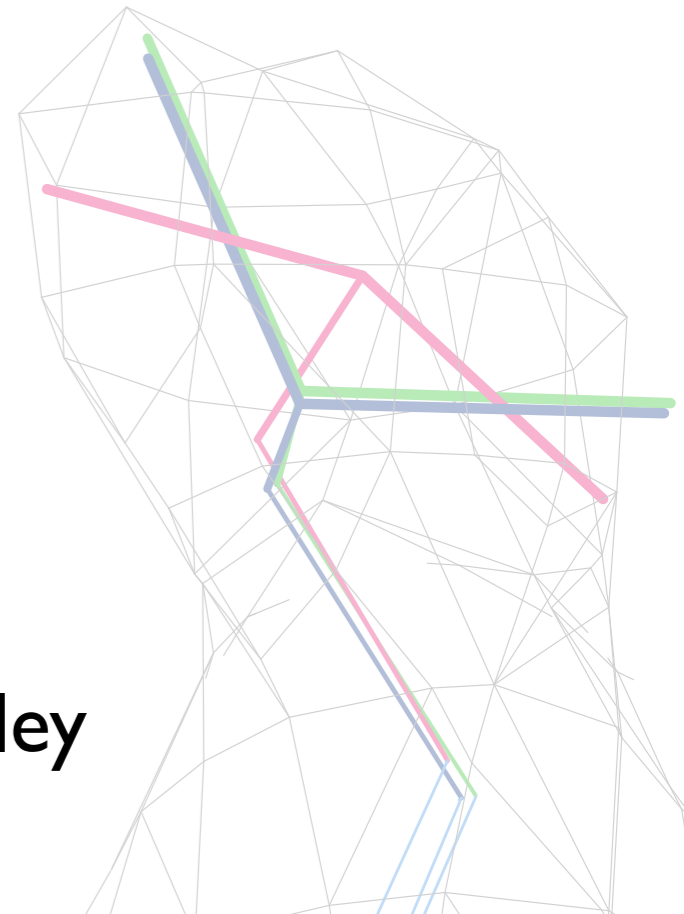
*Modern Concepts in Structural Biology Seminar Series*

*University of Vienna*

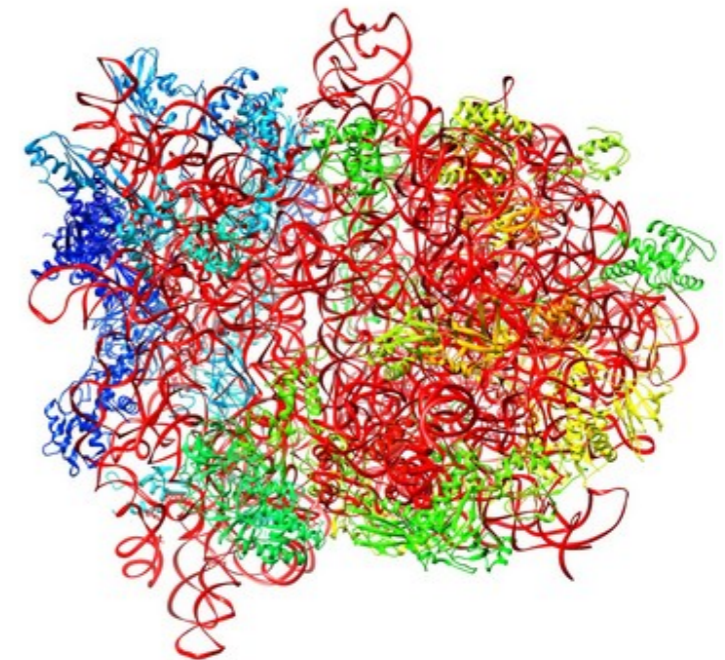
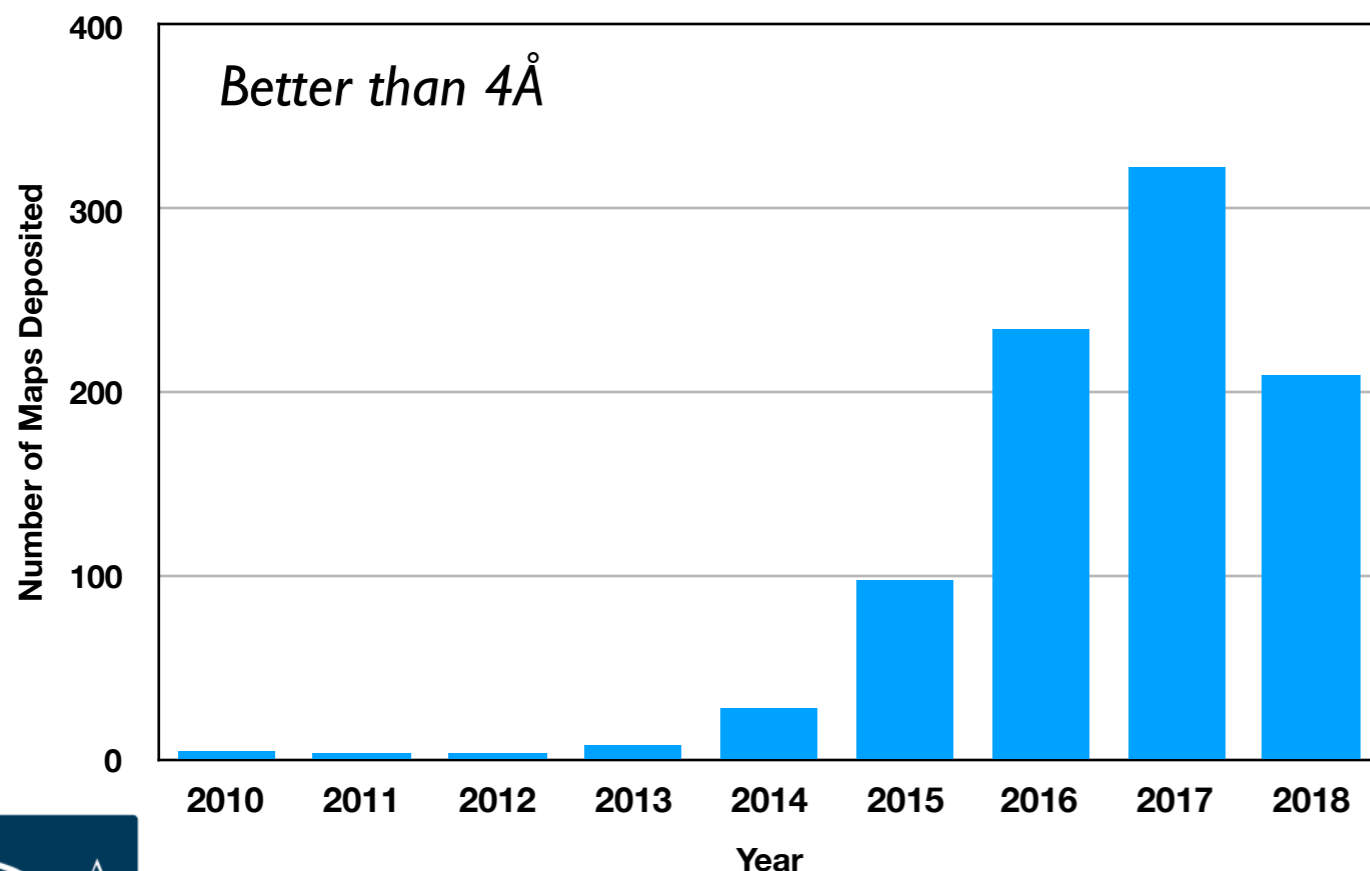
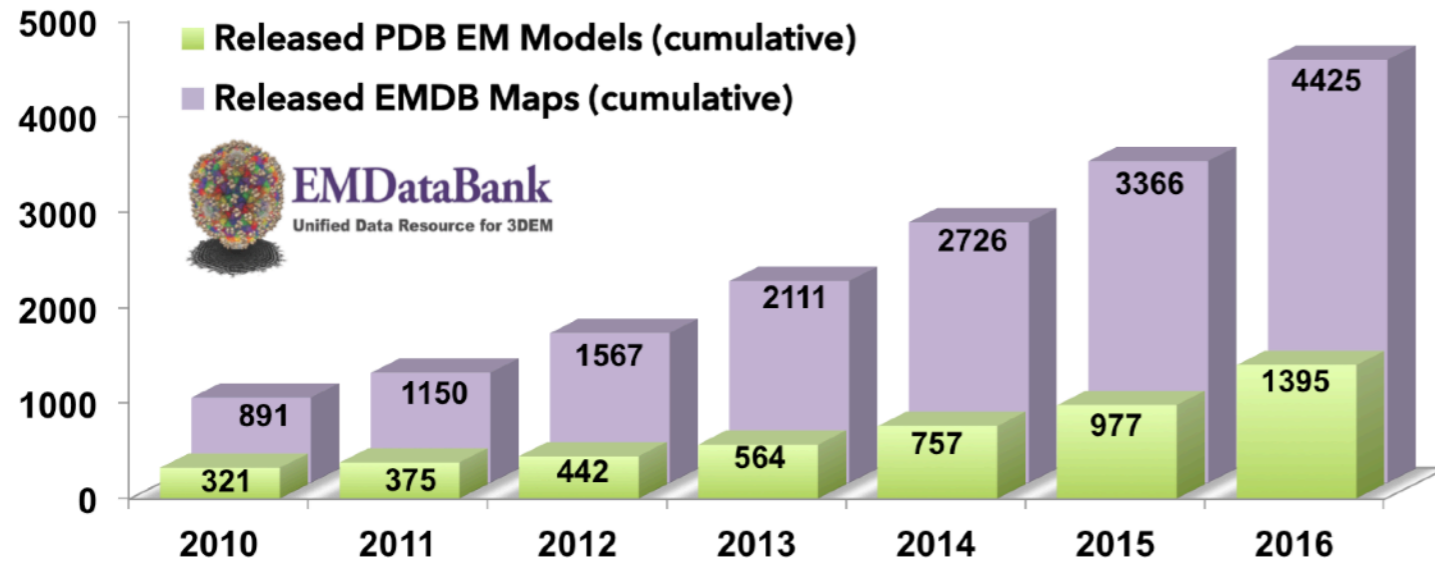
*May 2018*

**Paul Adams**

Lawrence Berkeley Laboratory and  
Department of Bioengineering UC Berkeley

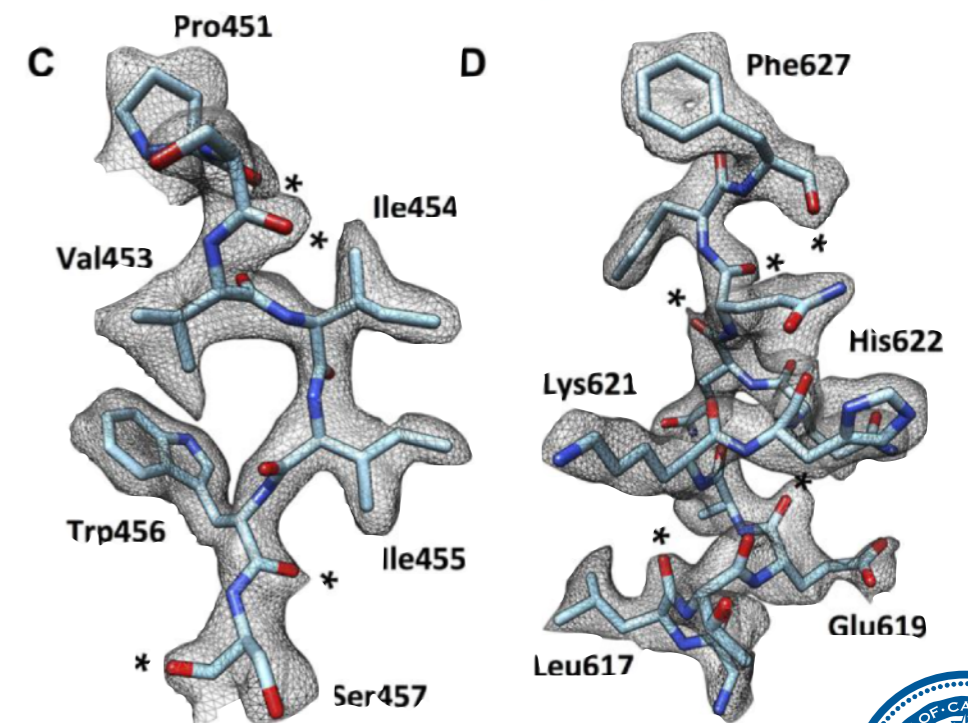


# Rapid Growth of Higher Resolution Cryo-EM



2.9Å Ribosome (5AFI)

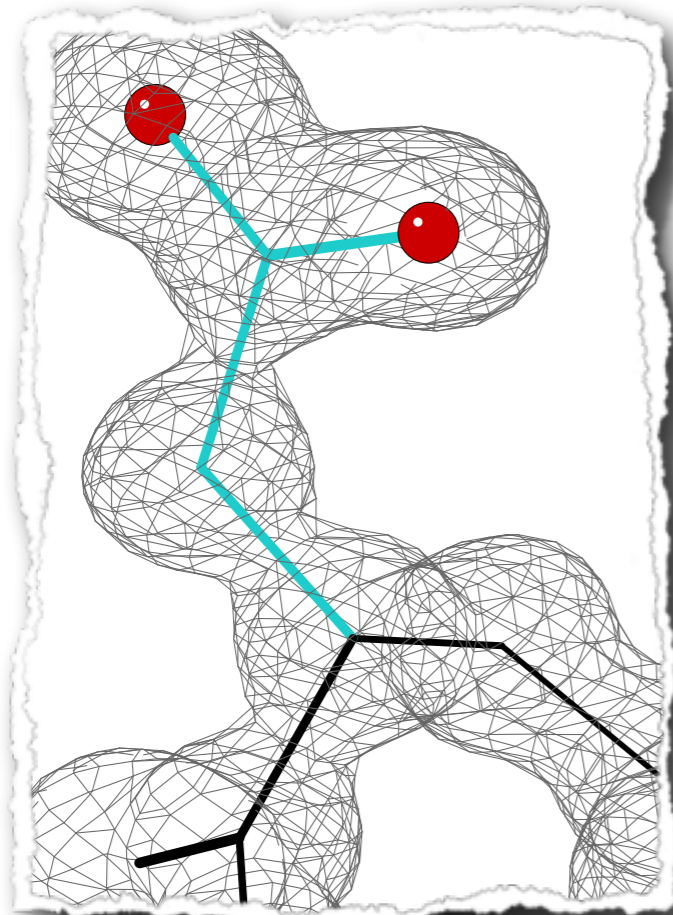
Fischer et al., *Nature* **520**: 567 (2015)



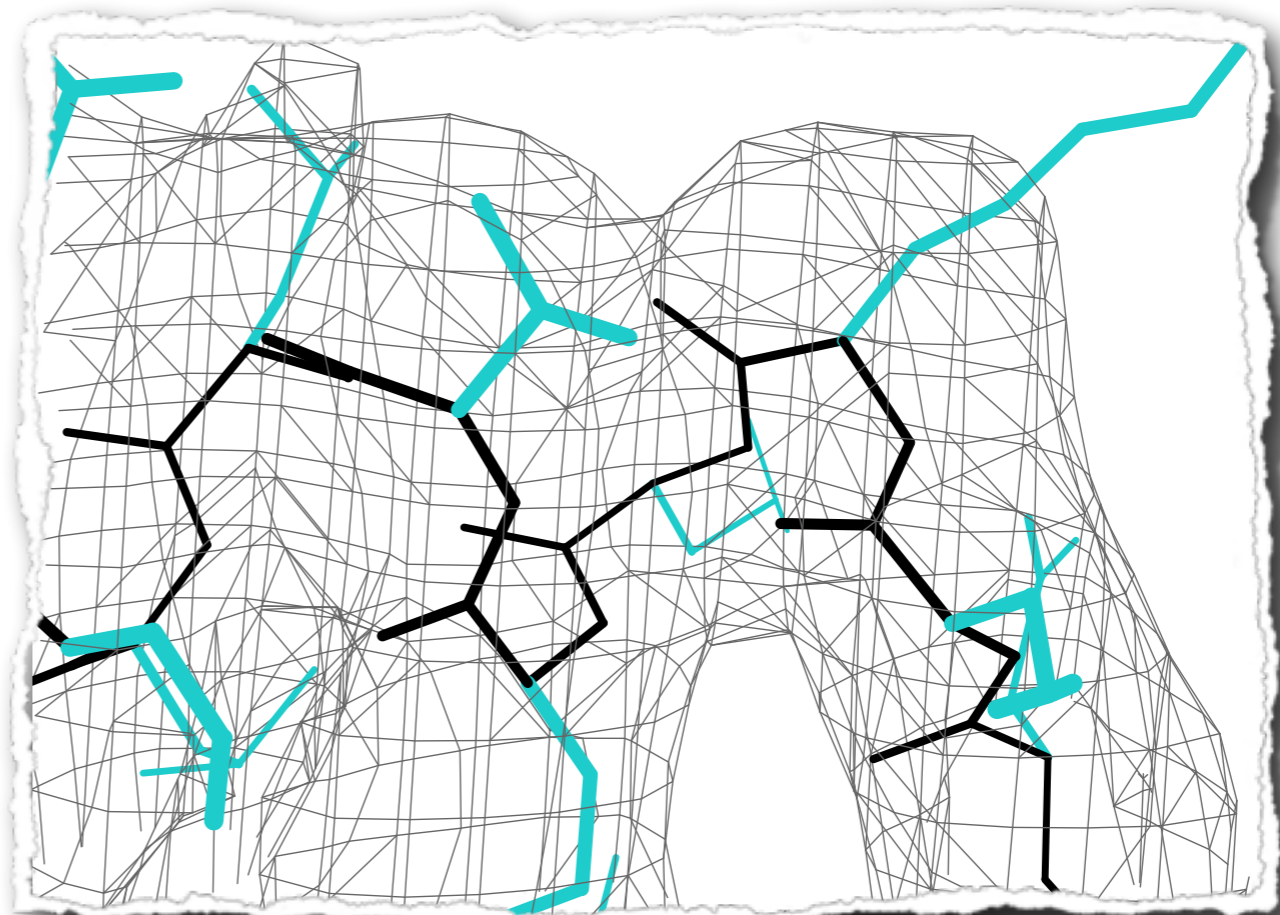
Bartesaghi et al., *Science*, 2015

# Low Resolution

PDB ID: 2gkg  
Resolution: 1.00Å



PDB ID: 3k7a  
Resolution: 3.80Å

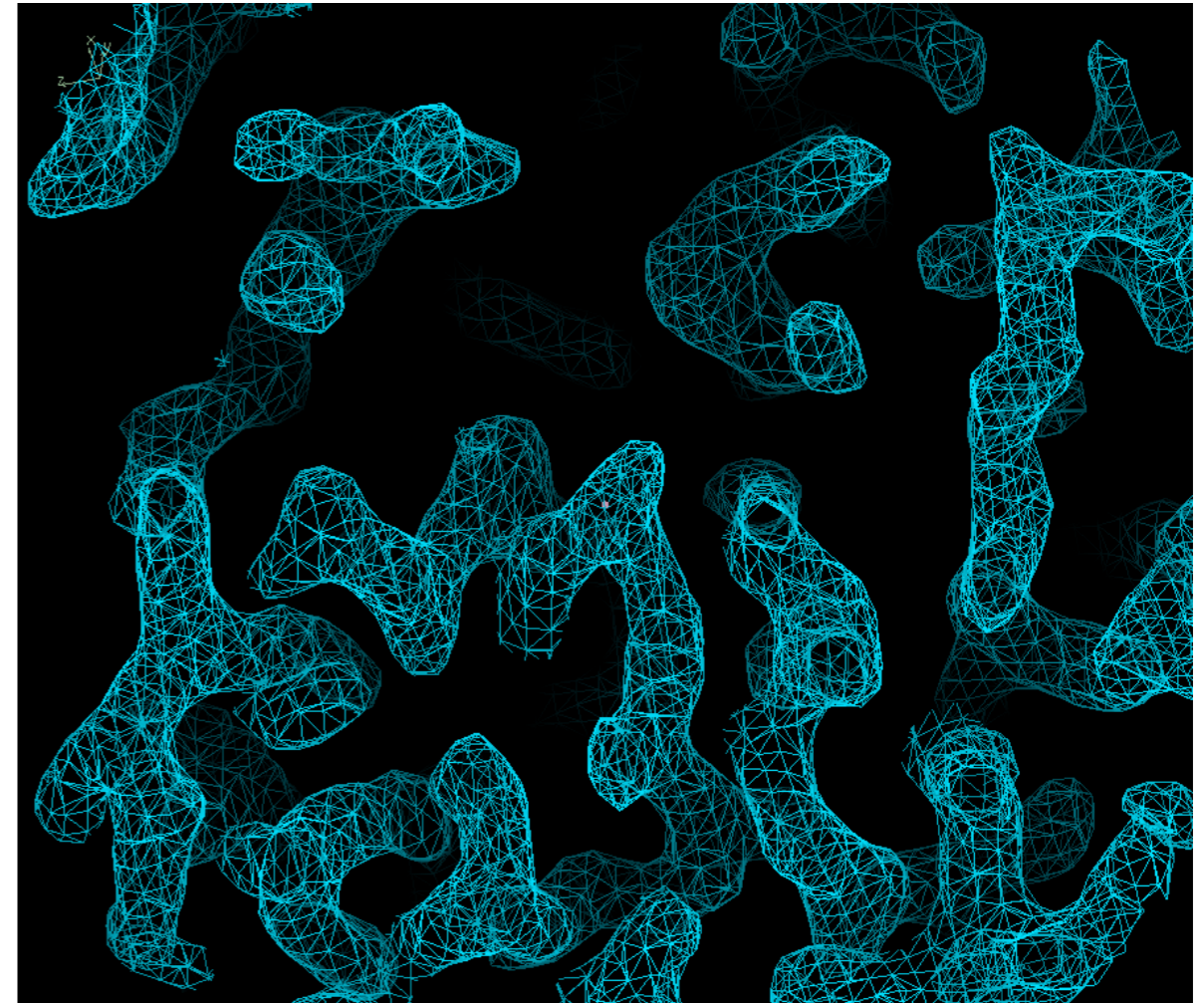
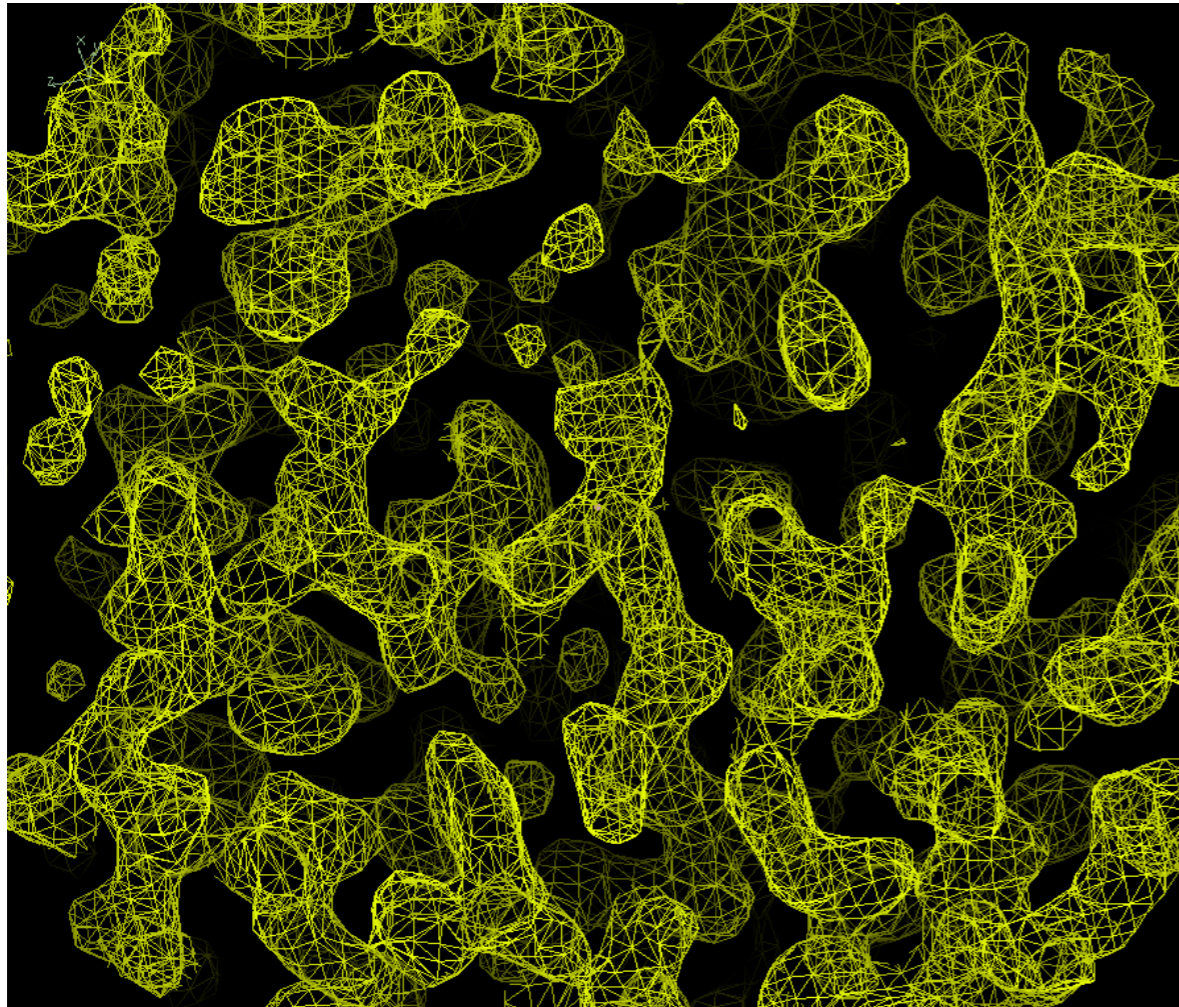


- Many challenges:
- How to interpret “featureless” maps (pattern matching, chemical constraints)
- How to optimize models with sparse data (prior information)

  
**Phenix**

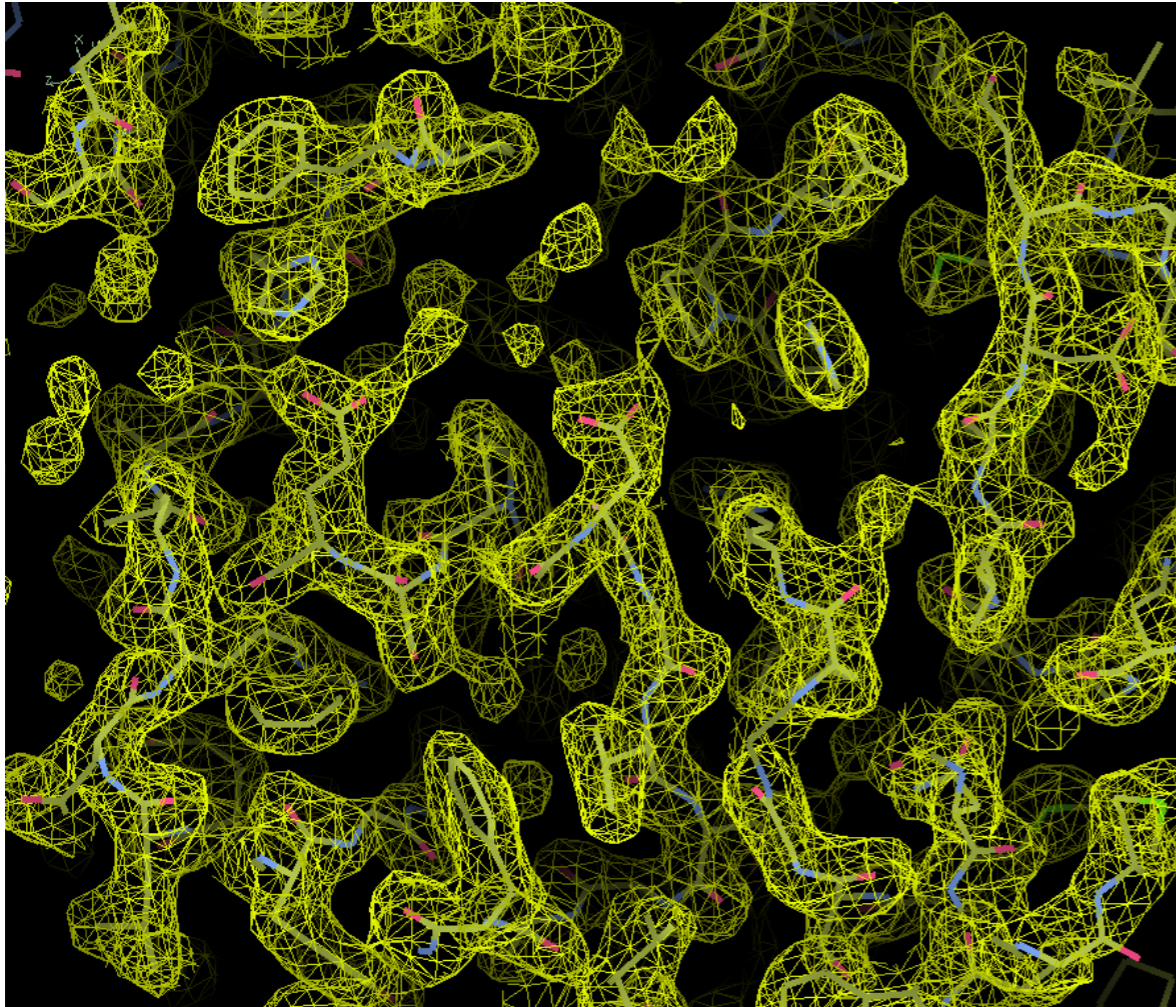
# Cryo-EM vs. Crystallographic Maps

Beta galactosidase at 2.2 Å

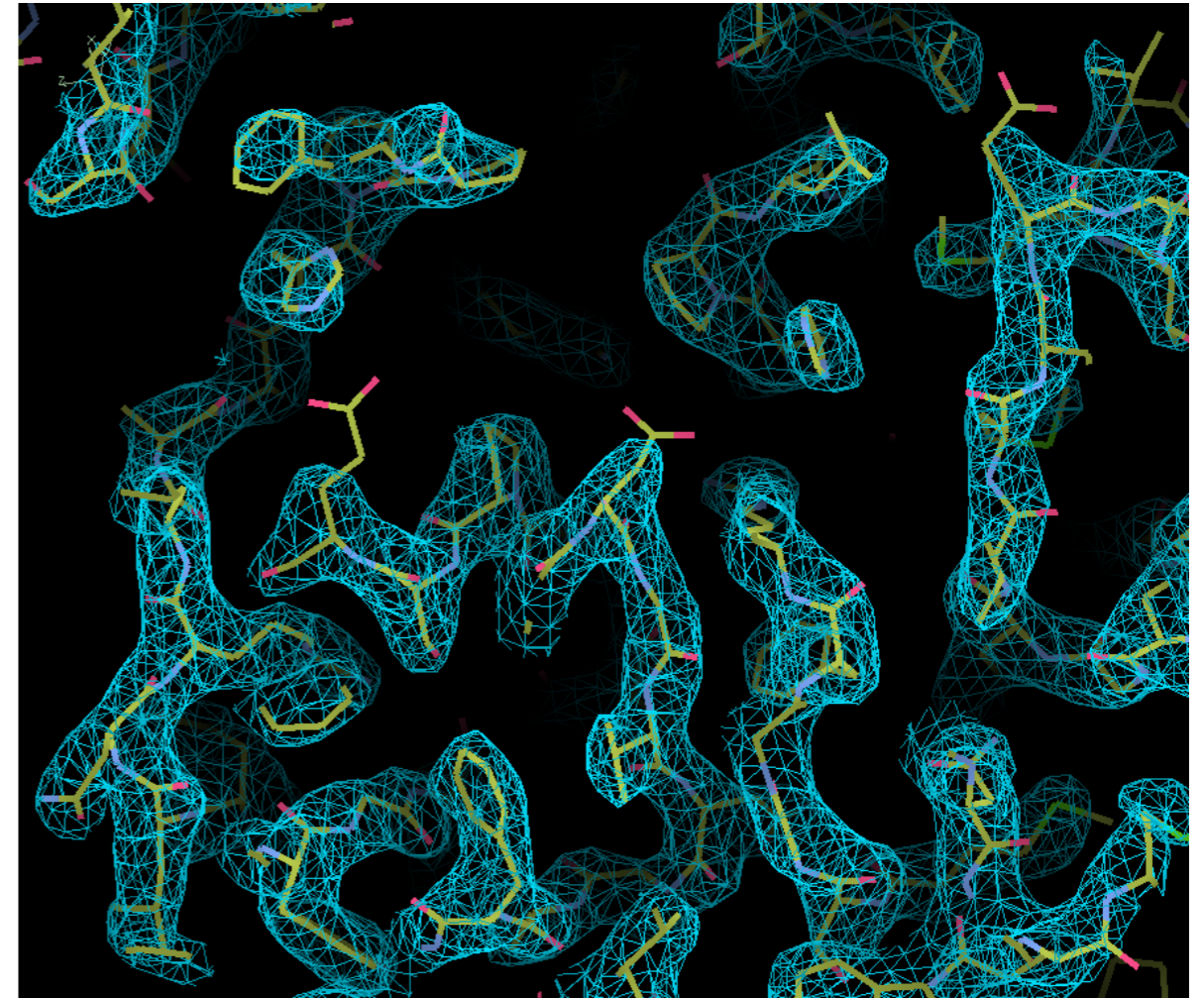


# Cryo-EM vs. Crystallographic Maps

Beta galactosidase at 2.2 Å



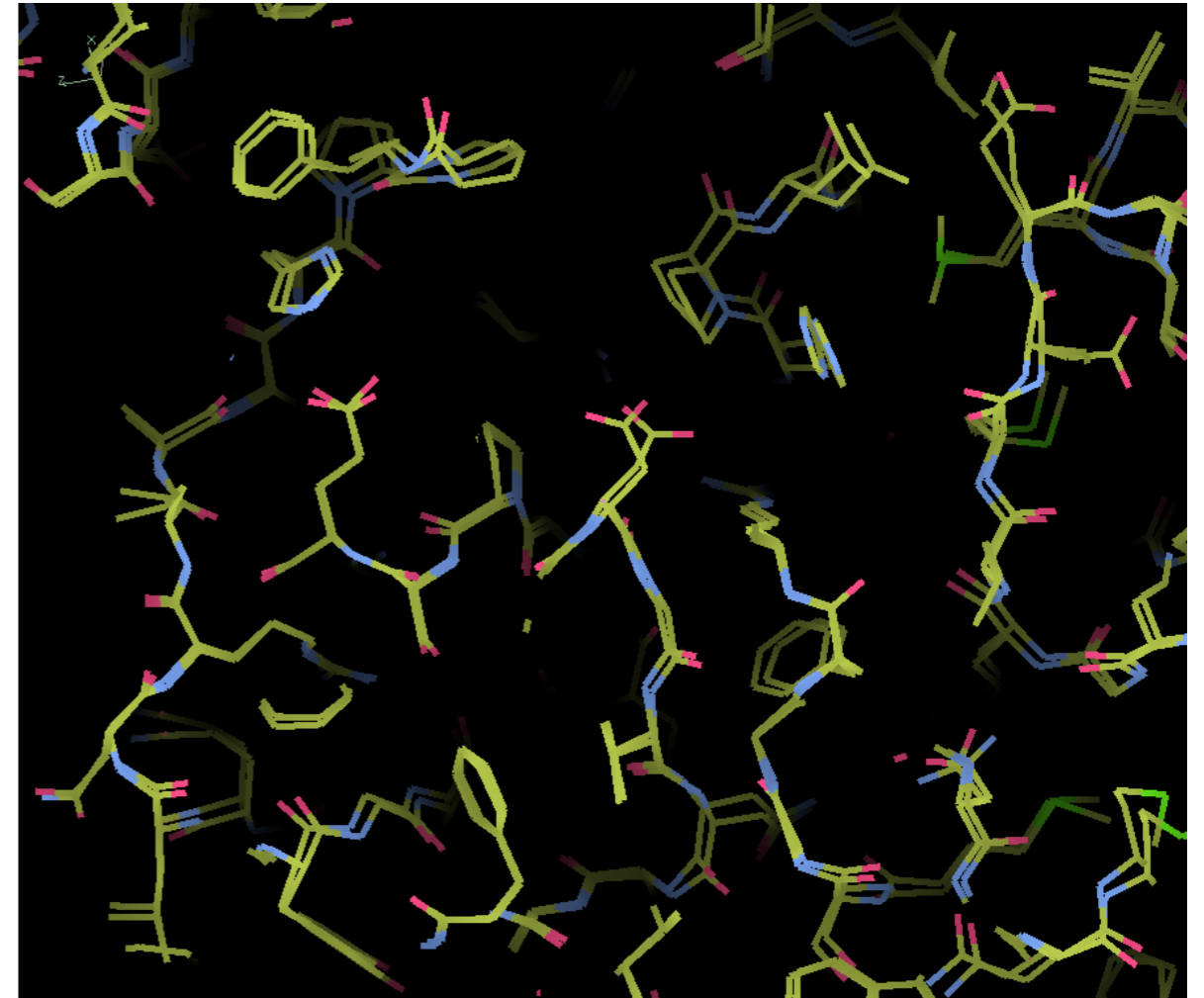
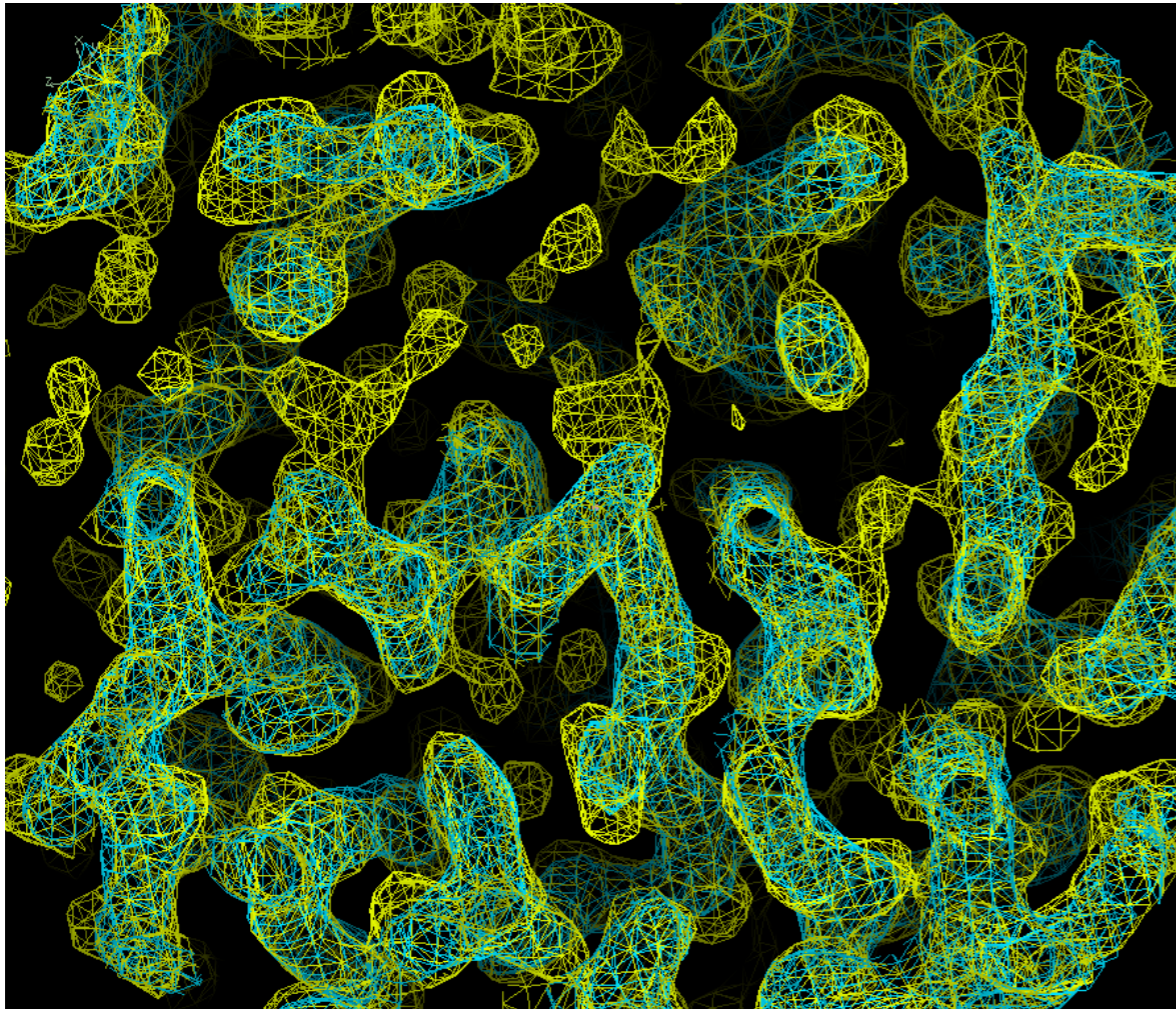
X-ray (PDB 3i3b)



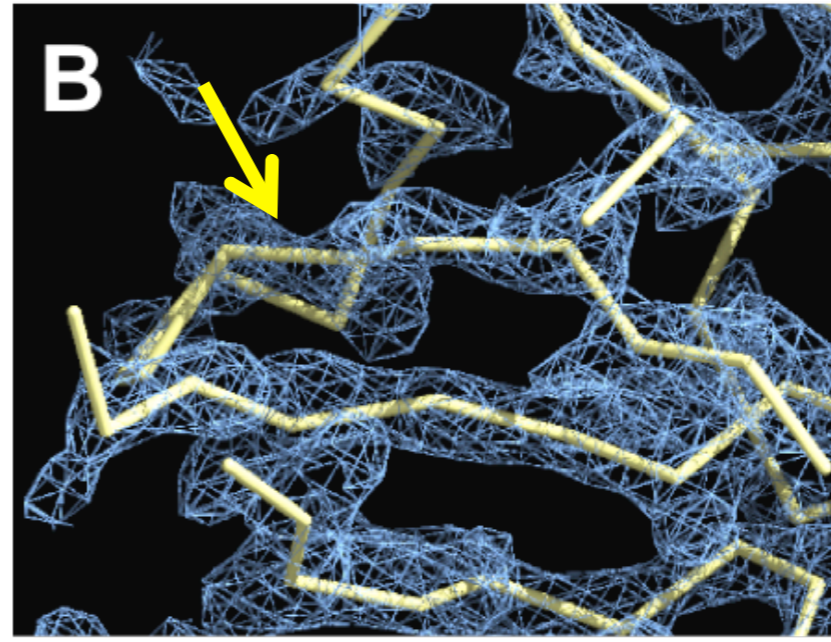
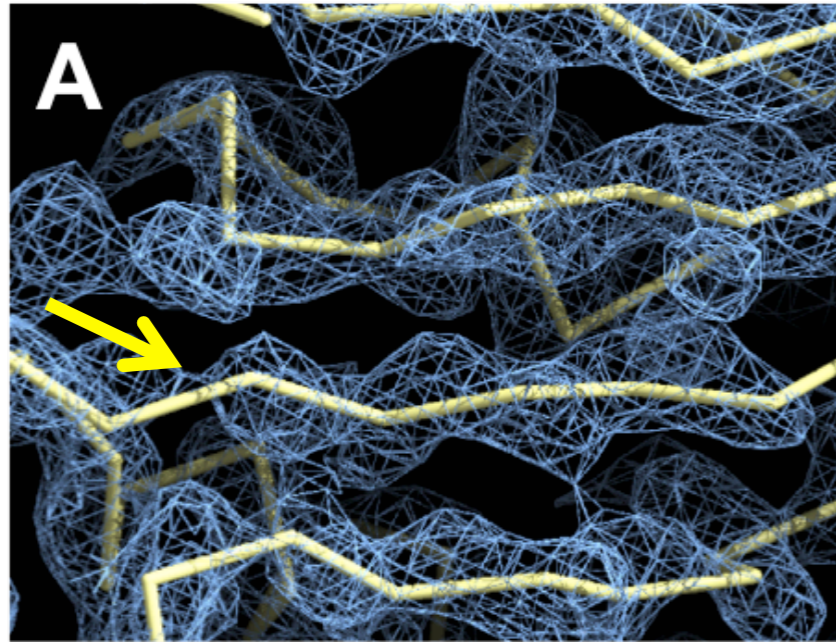
Cryo-EM (PDB 5a1a)

# Cryo-EM vs. Crystallographic Maps

- The maps are very similar



# Low Resolution Information in Cryo-EM Maps



**Original**

# Challenges

- Automated model building
  - What is the optimal sharpening of the map?
  - What is the region containing the molecule?
  - Low and variable resolution across maps
  - What is the magnification of the map? (can be 10% uncertainty)
- Model refinement
  - Variable resolution across maps
  - Large molecules
  - Poor initial models
- Validation
  - How to validate models against moderate resolution maps



# The Phenix Project

## Lawrence Berkeley Laboratory

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



## Los Alamos National Laboratory

Tom Terwilliger, Li-Wei Hung



Randy Read, Airlie McCoy,  
Tristan Croll, Rob Oeffner

## 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**



# Automated 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



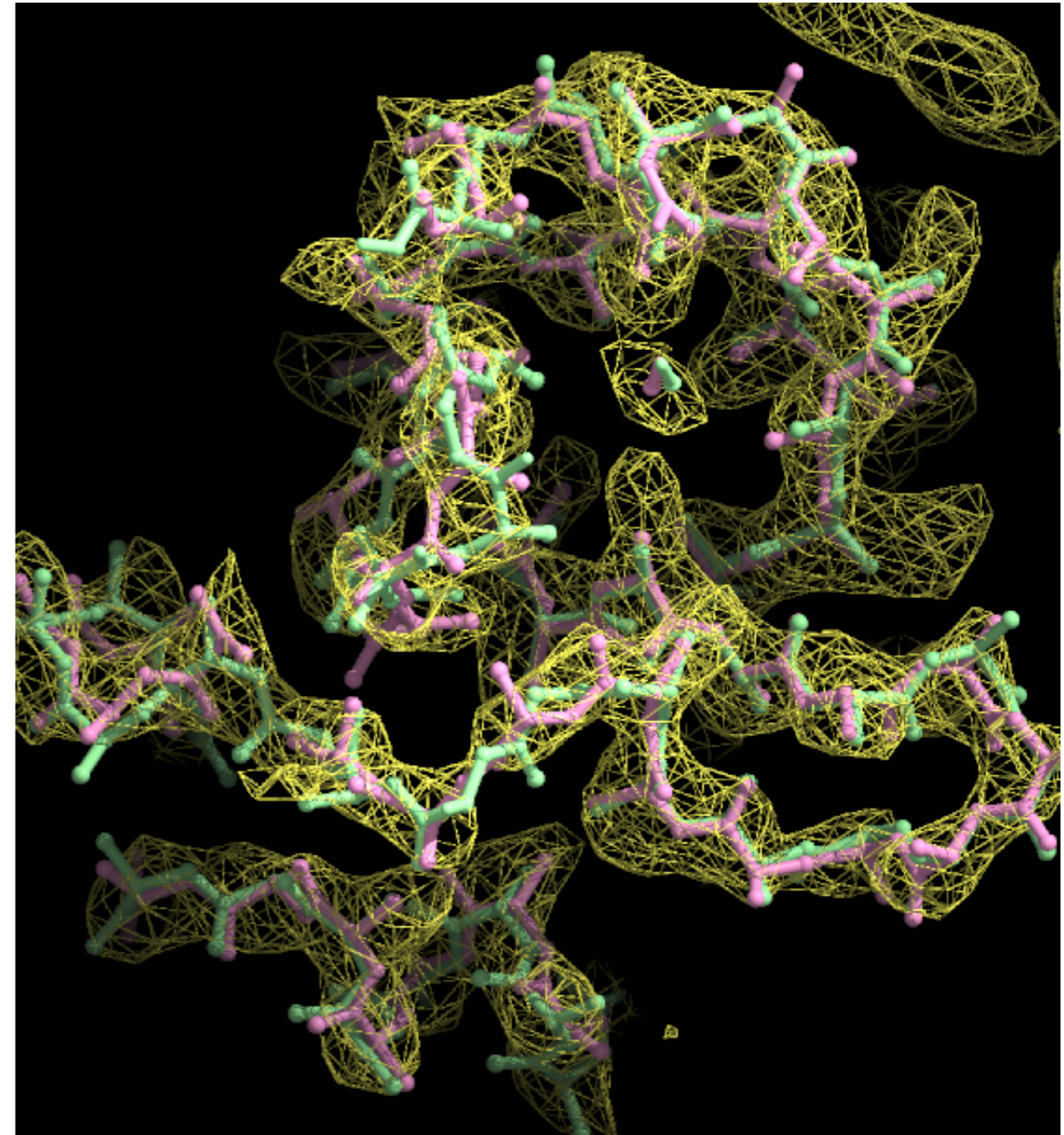
Idealize secondary structure and refine



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)

Autobuilt model (pink)  
Deposited model (green)

**Phenix**



# Automated Map Sharpening

Create series of maps with variable overall B-values

Analyze maps for detail and connectivity

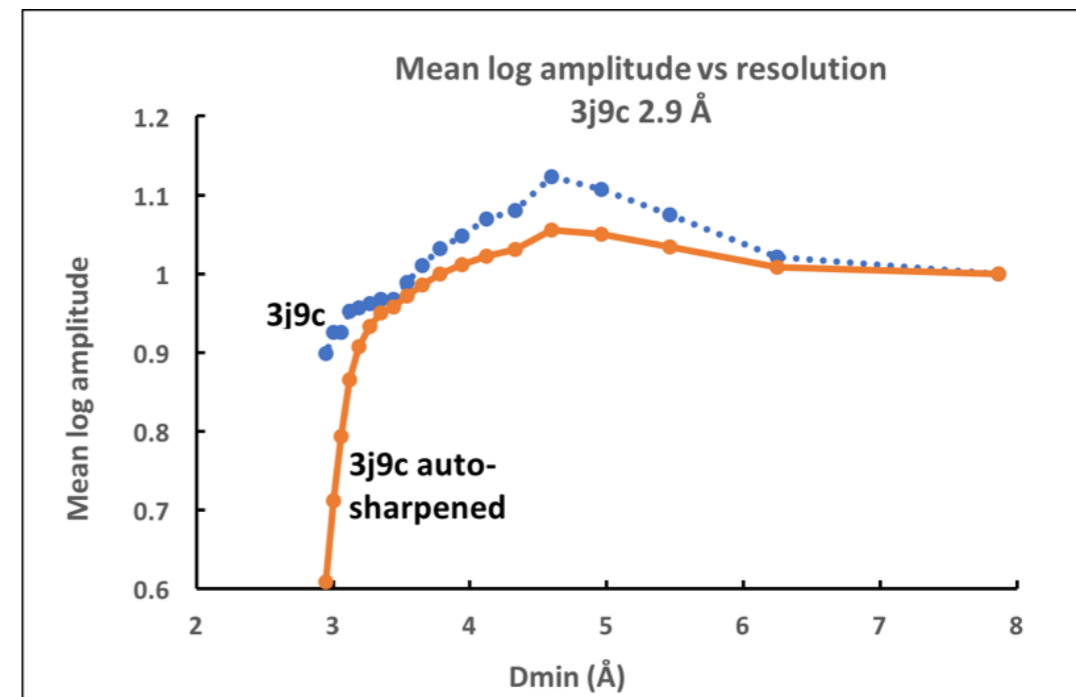
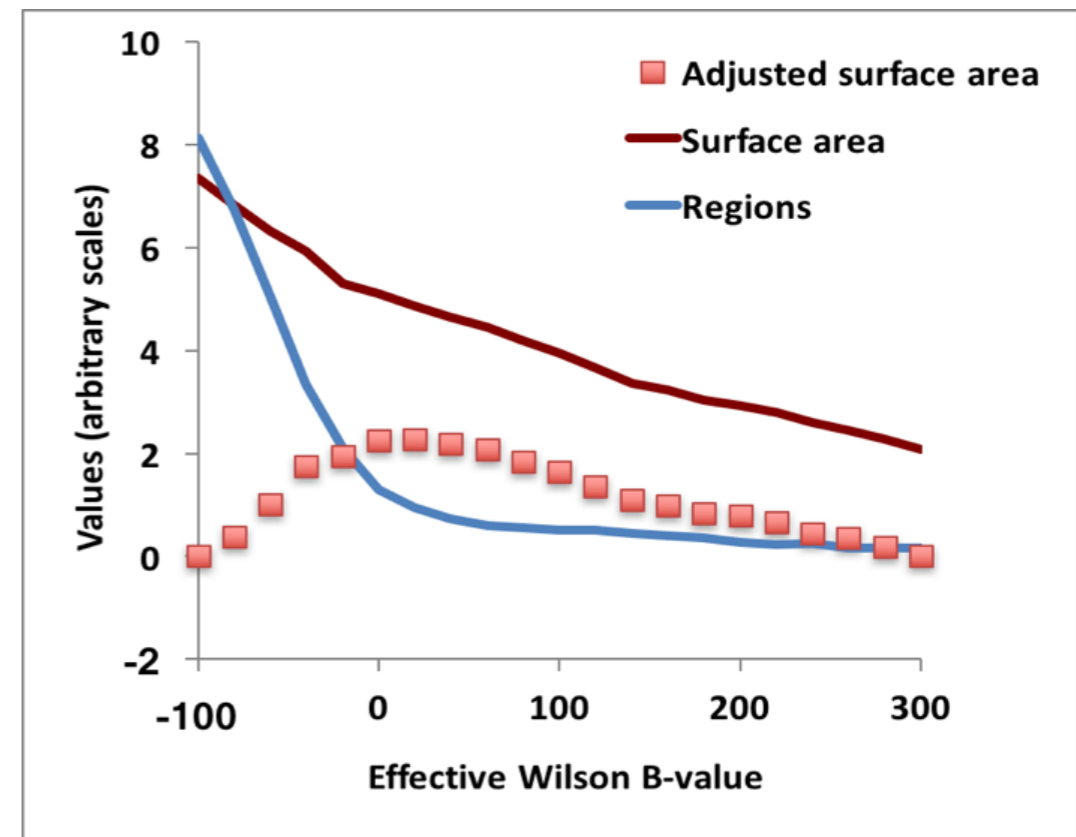
Set contour level enclosing 20% of molecular volume

Calculate surface area of contours

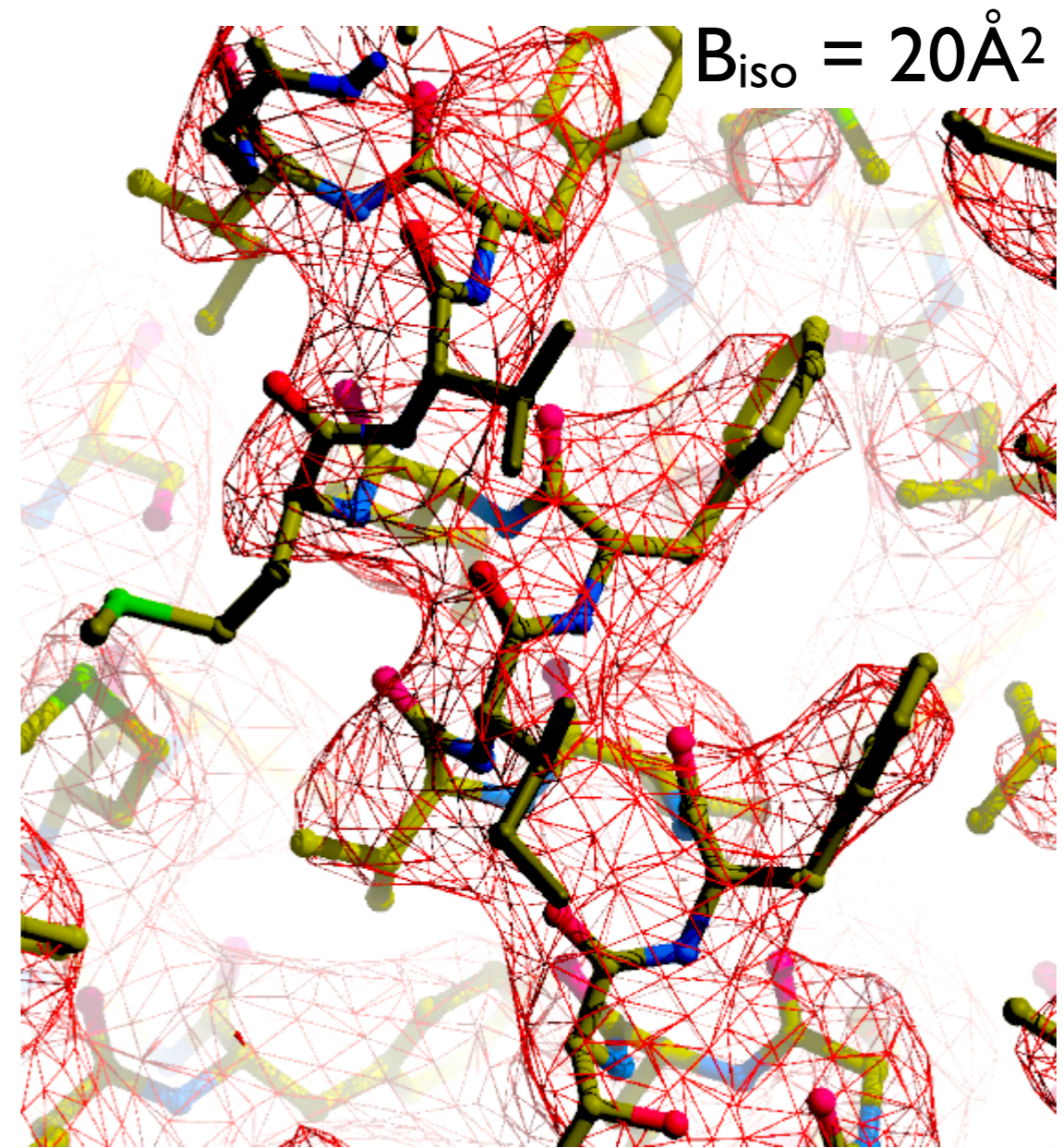
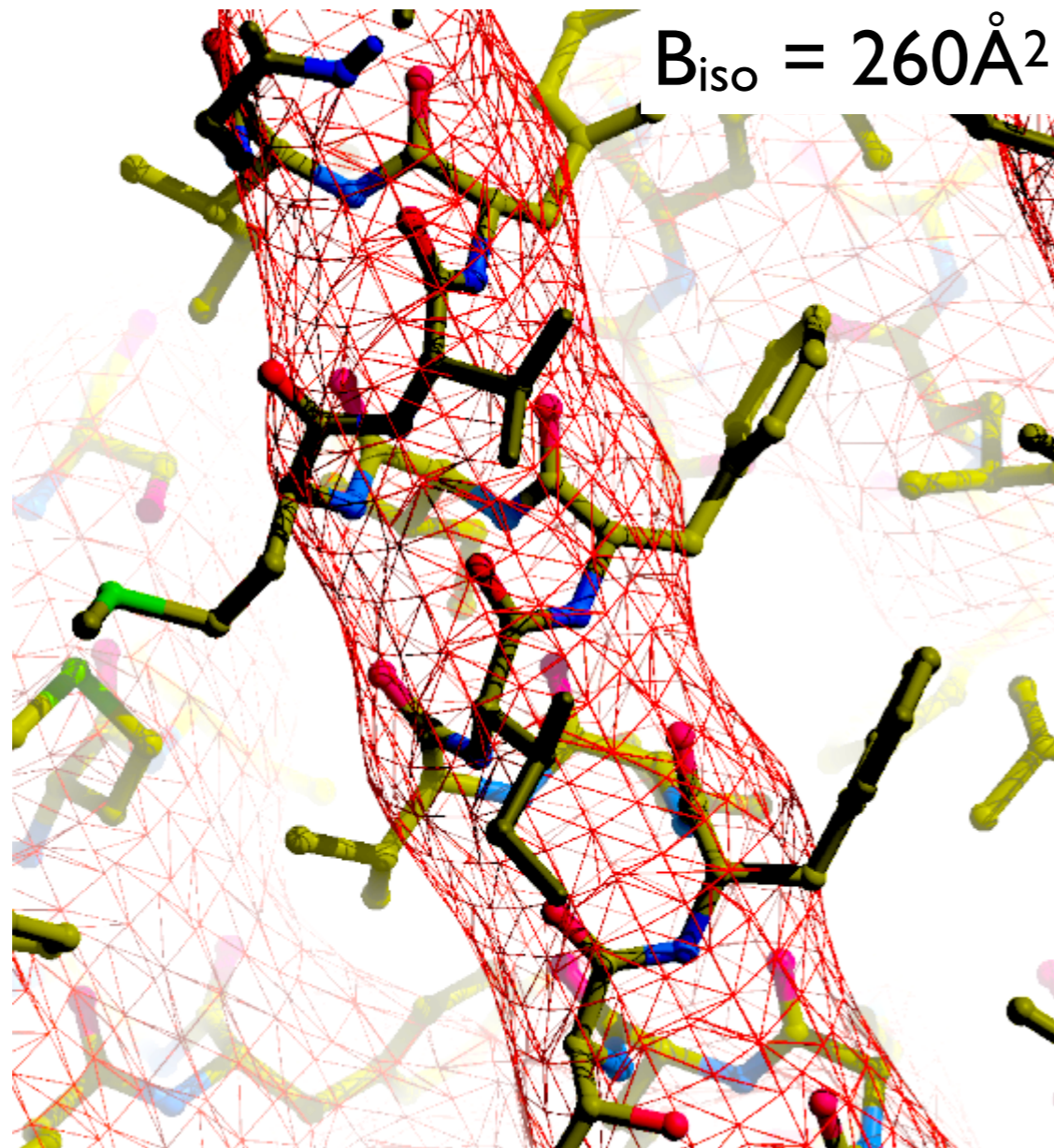
Count number of connected regions enclosed by contours

Choose map with maximum of adjusted surface area

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



# Automated Map Sharpening



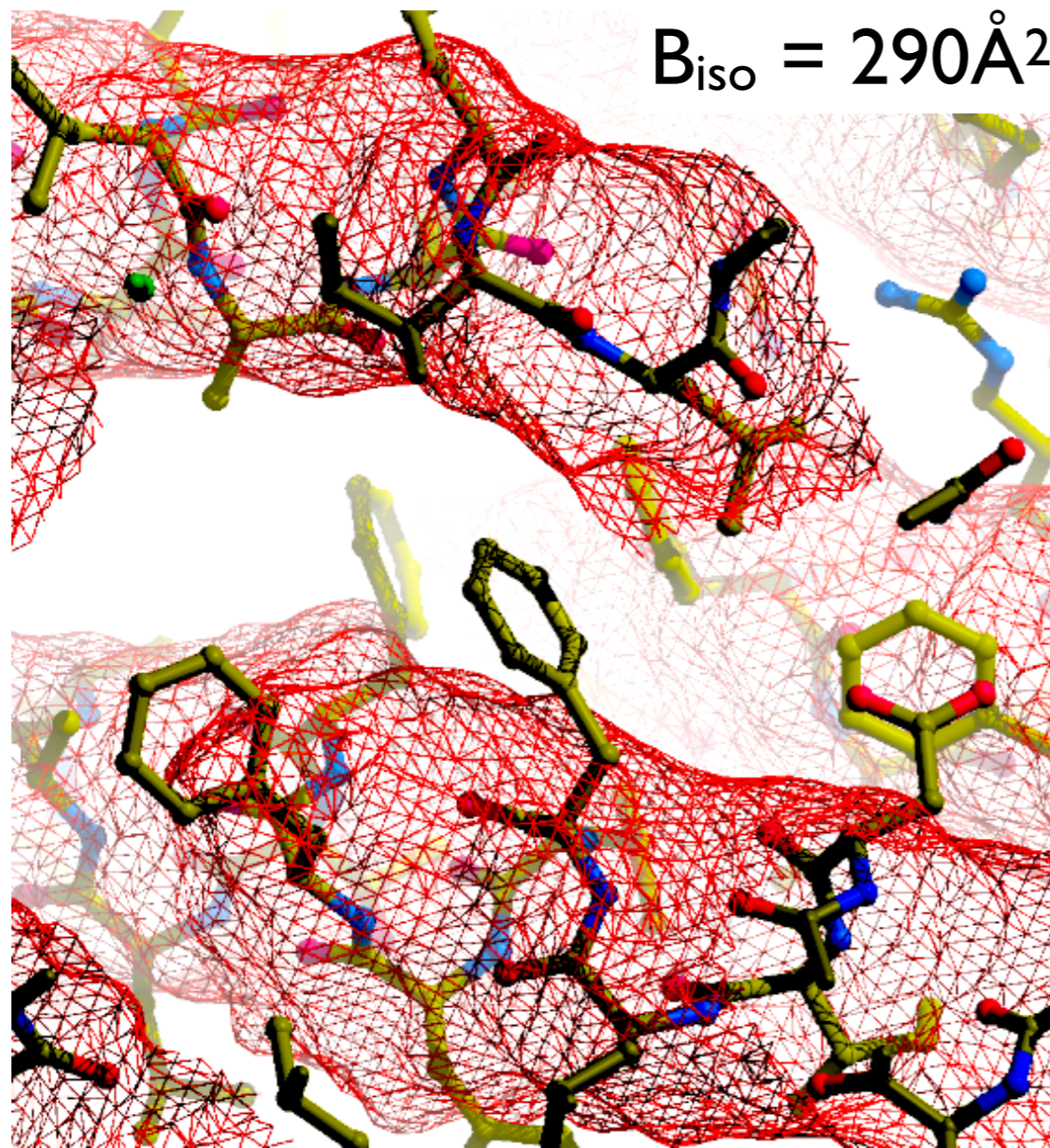
*Deposited Map*

*Autosharpened Map*

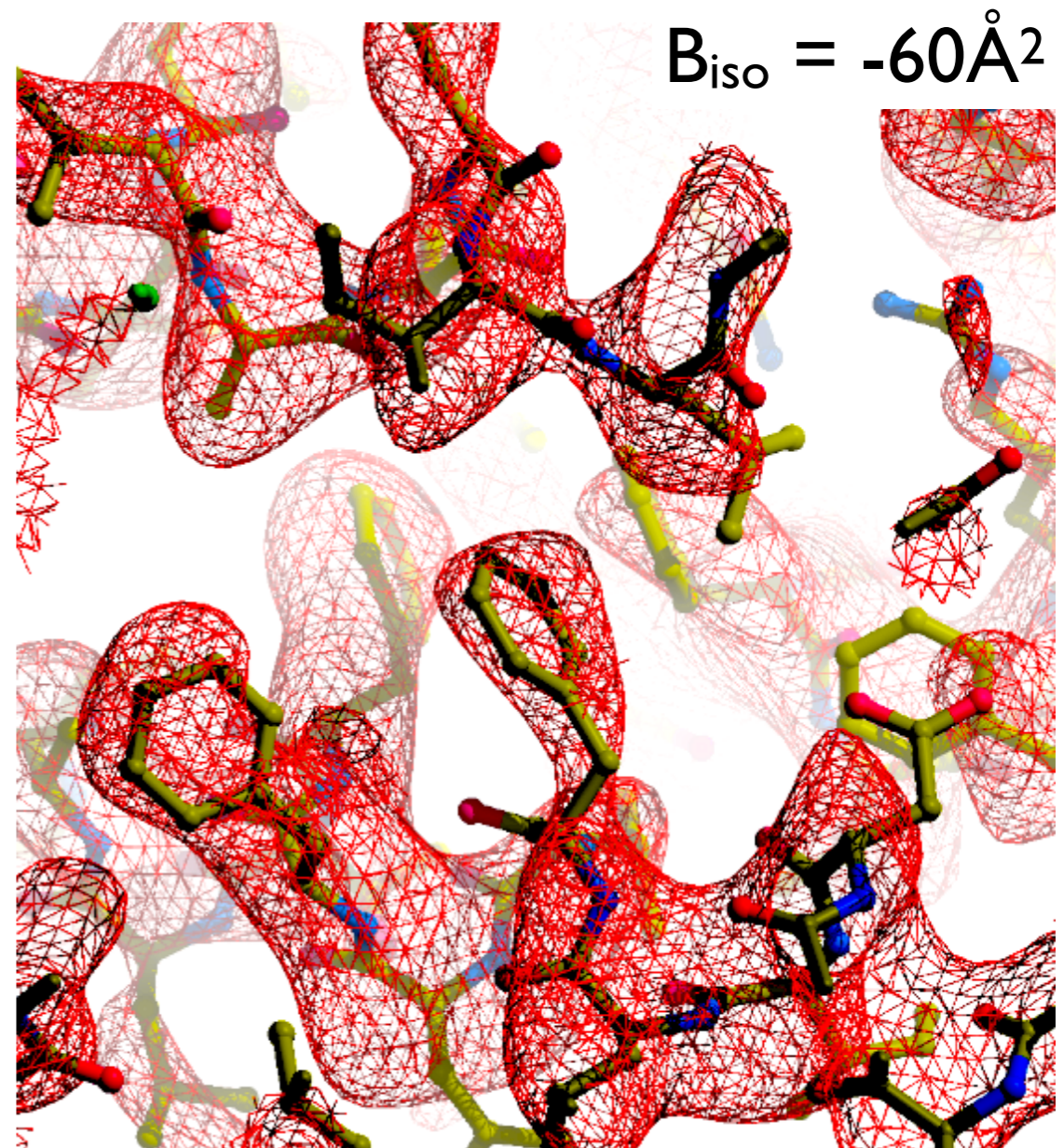
High-conductance Ca(2+)-activated K(+) channel (emd\_8414 and PDB entry 5tji; Hite et al., 2017)

**Phenix**

# Automated Map Sharpening



*Deposited Map*

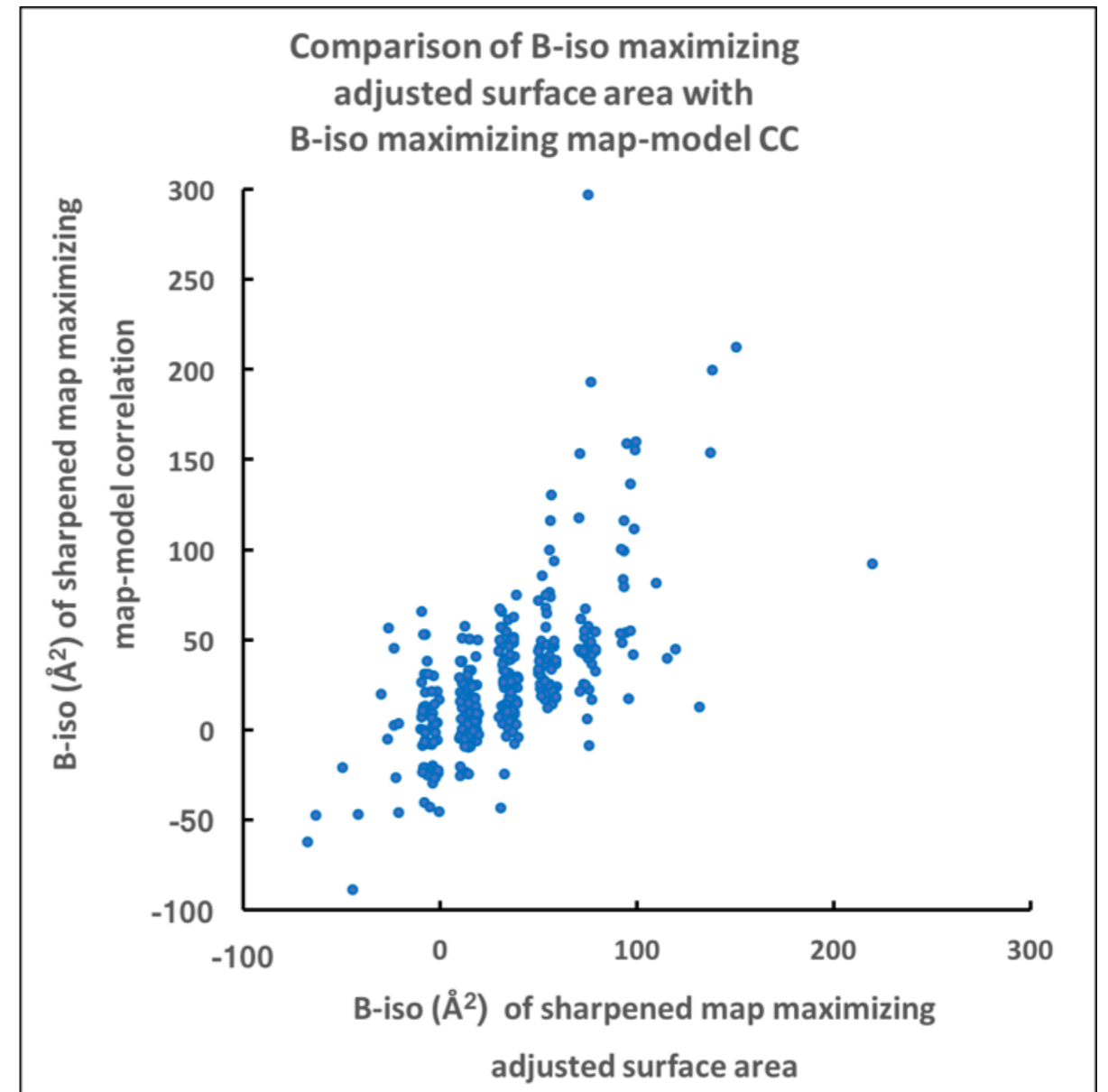
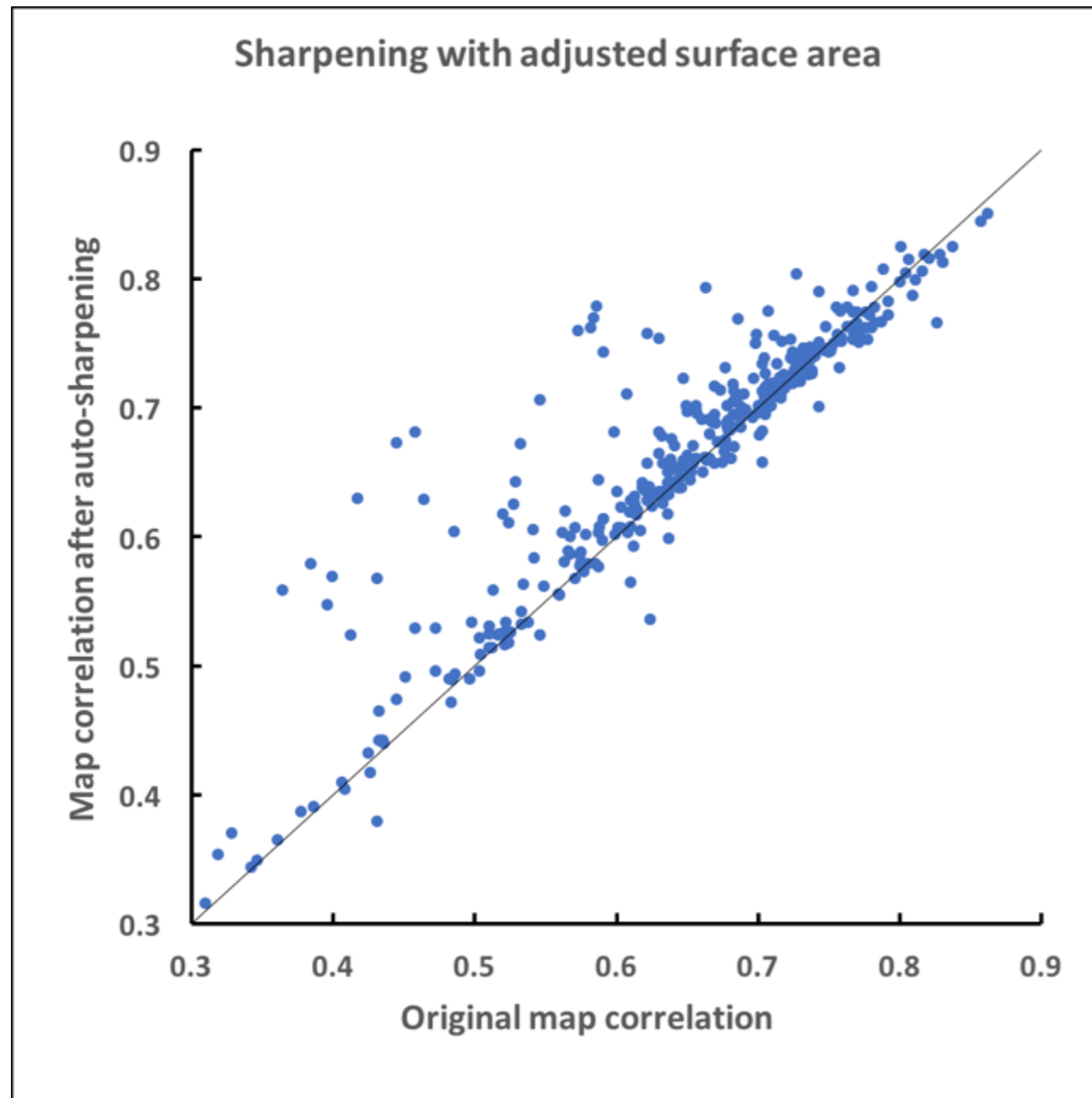


*Autosharpened Map*

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

**Phenix**

# Automated Map Sharpening



# Automated Segmentation

Determine optimal sharpening of the map

Cut out asymmetric unit of the map

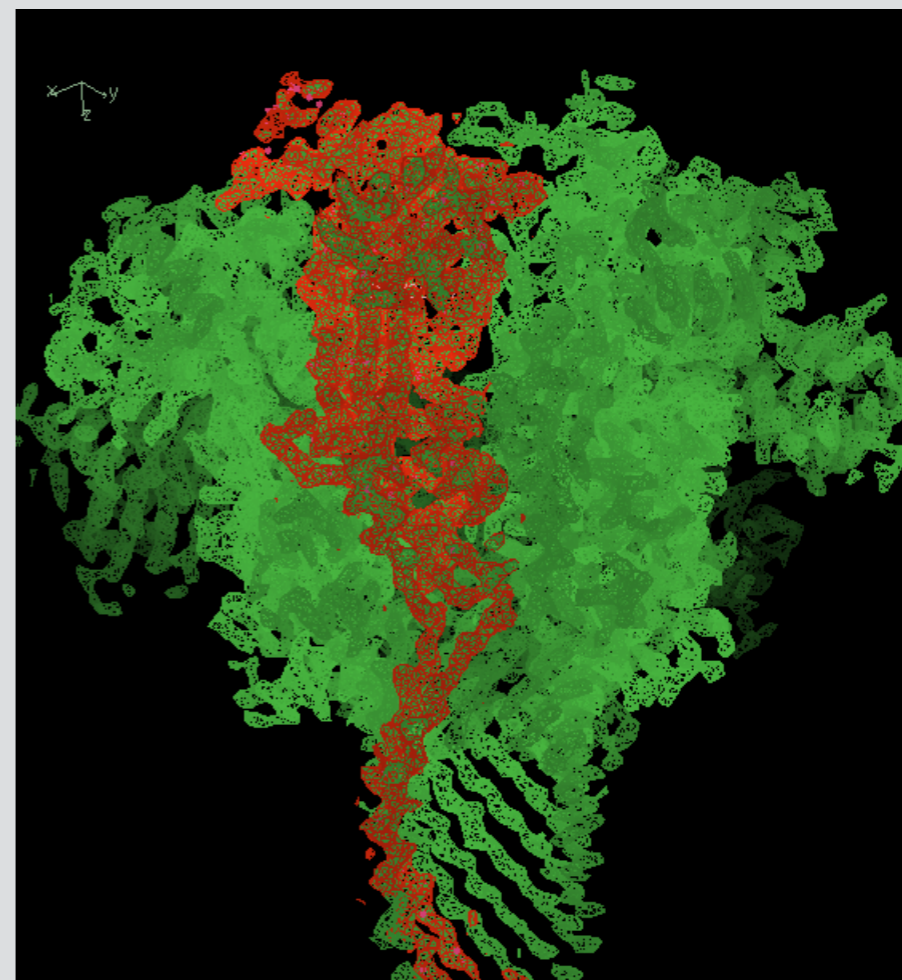
Trace chain and build model

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



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

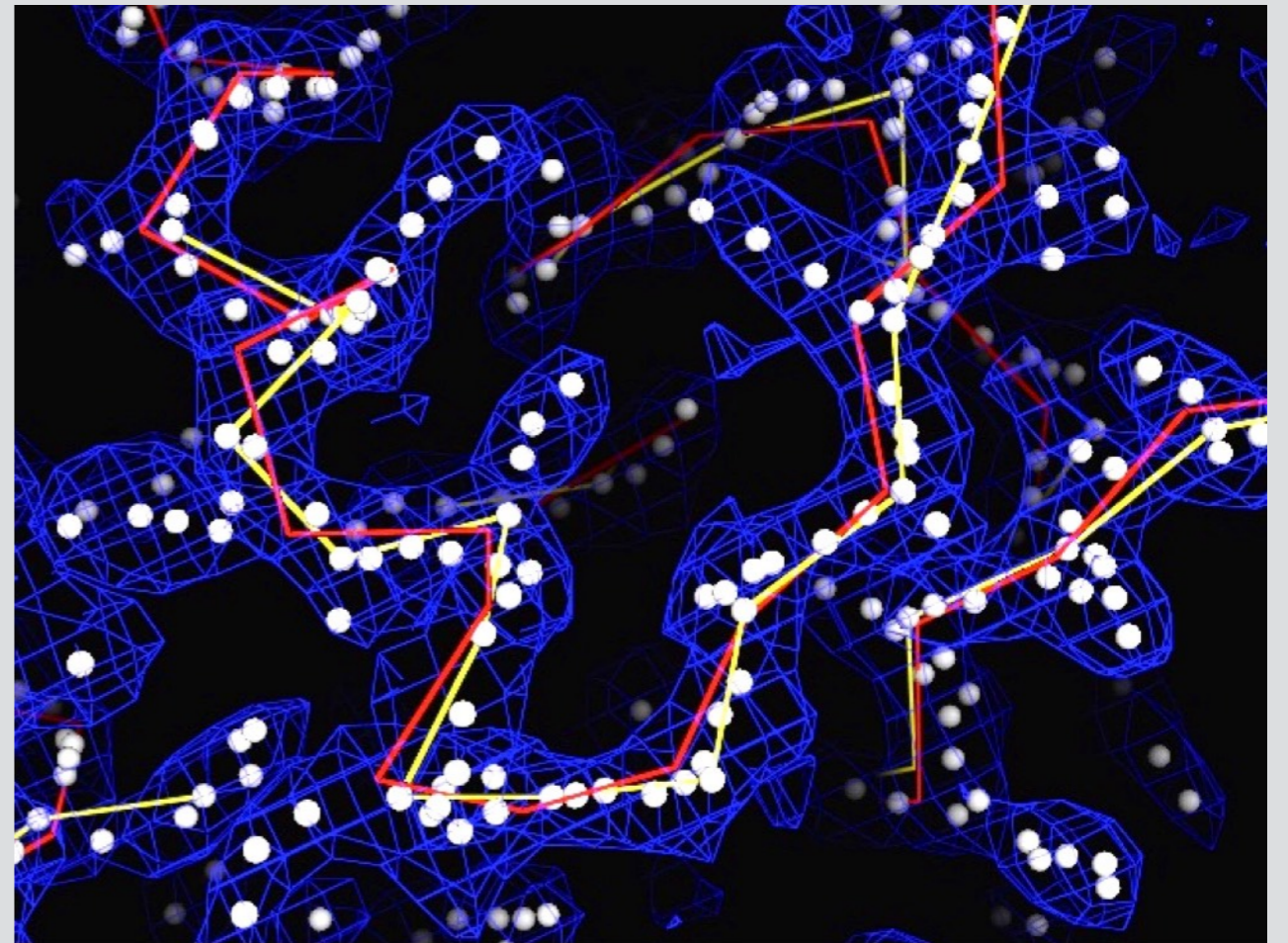
Trace chain and build model

Idealize secondary structure and refine

Assemble and refine (protein/RNA/DNA)

Apply molecular symmetry and re-refine

- Variable map sharpening
- 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

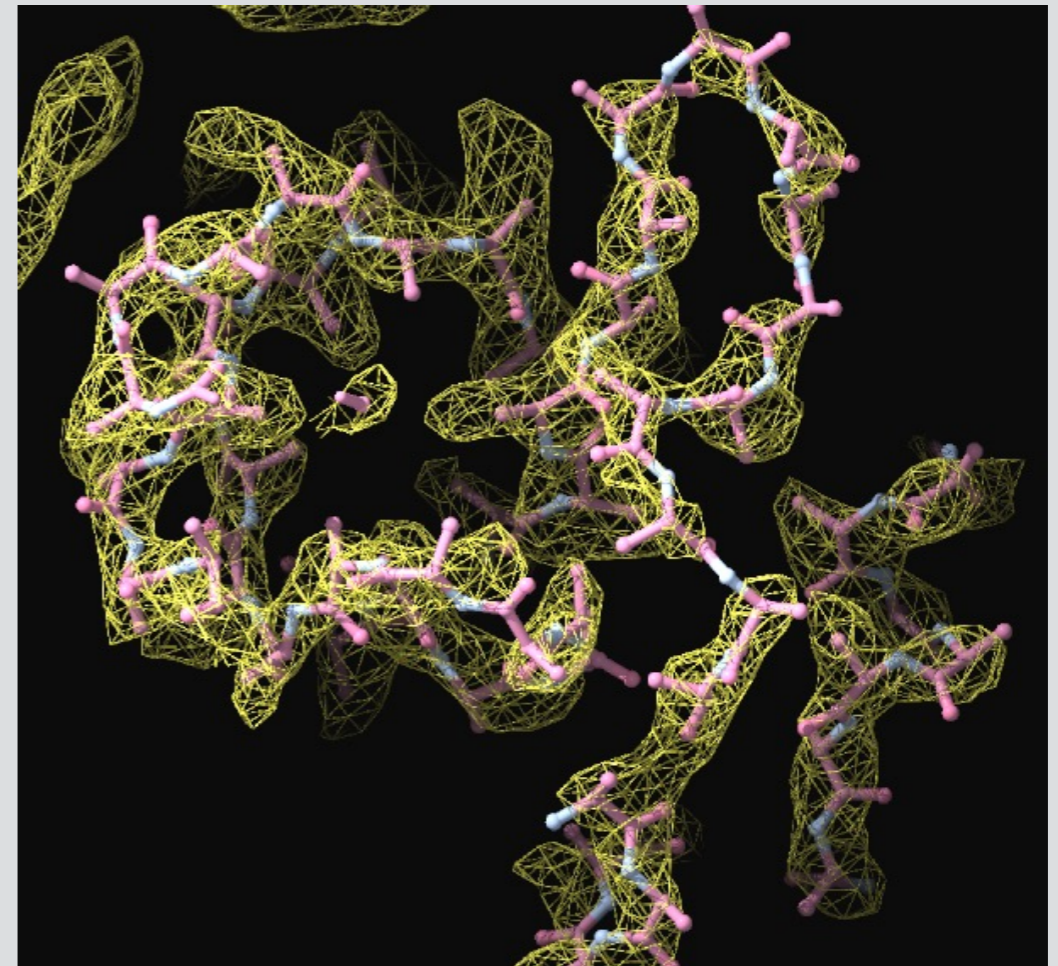
Trace chain and build model

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

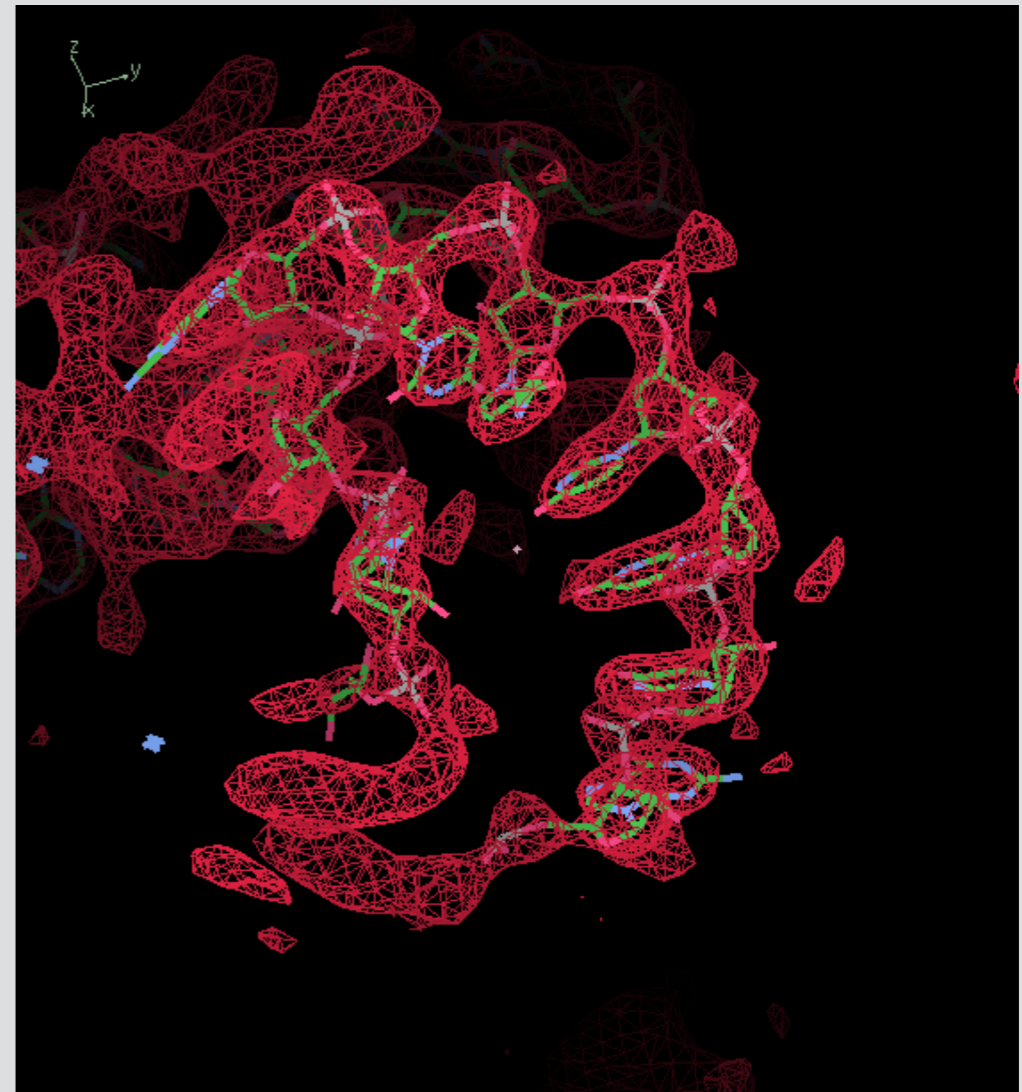
Trace chain and build model

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

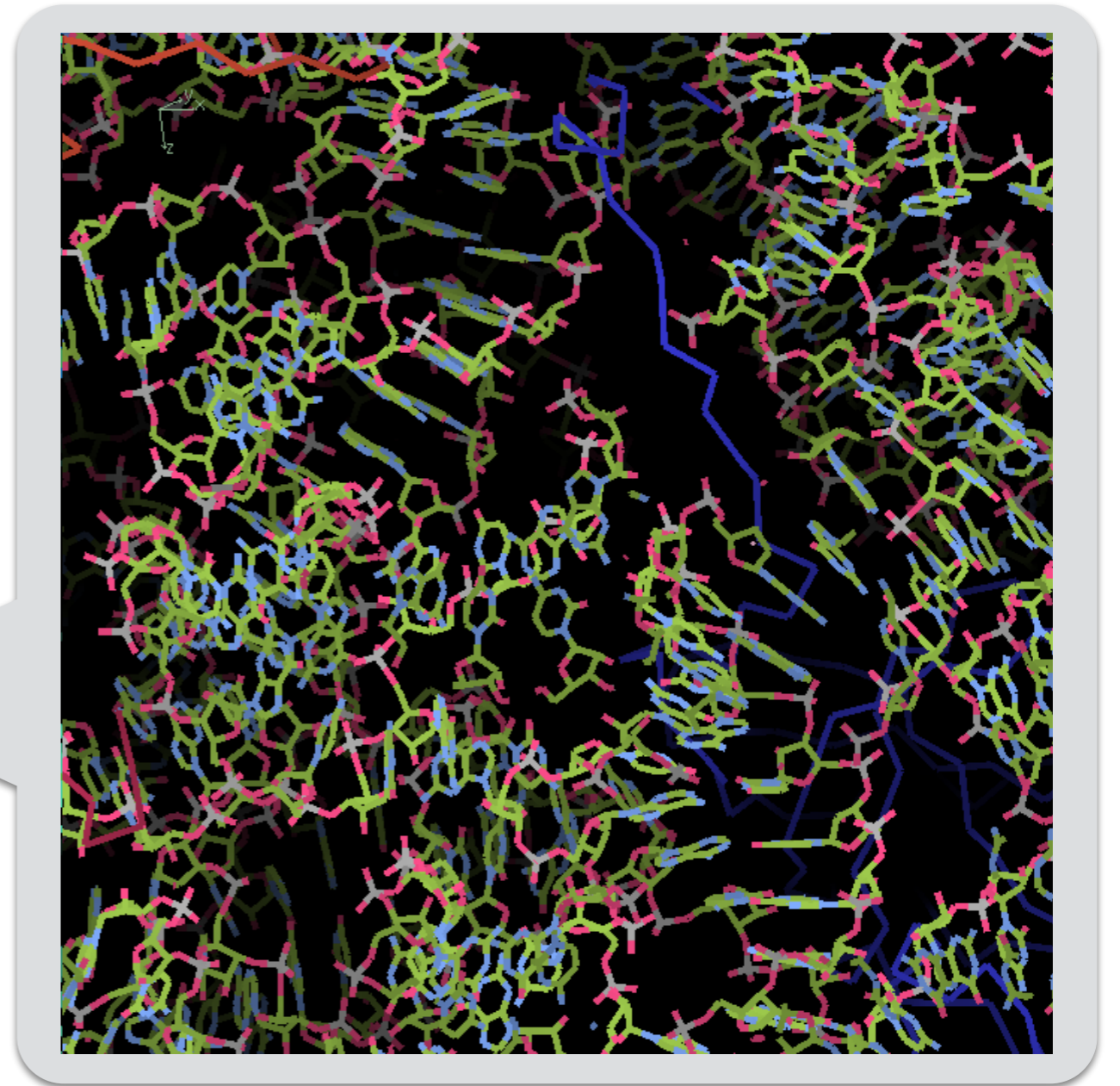


Assemble and refine (protein/RNA/DNA)



Apply molecular symmetry and re-refine

● `phenix.map_to_model`

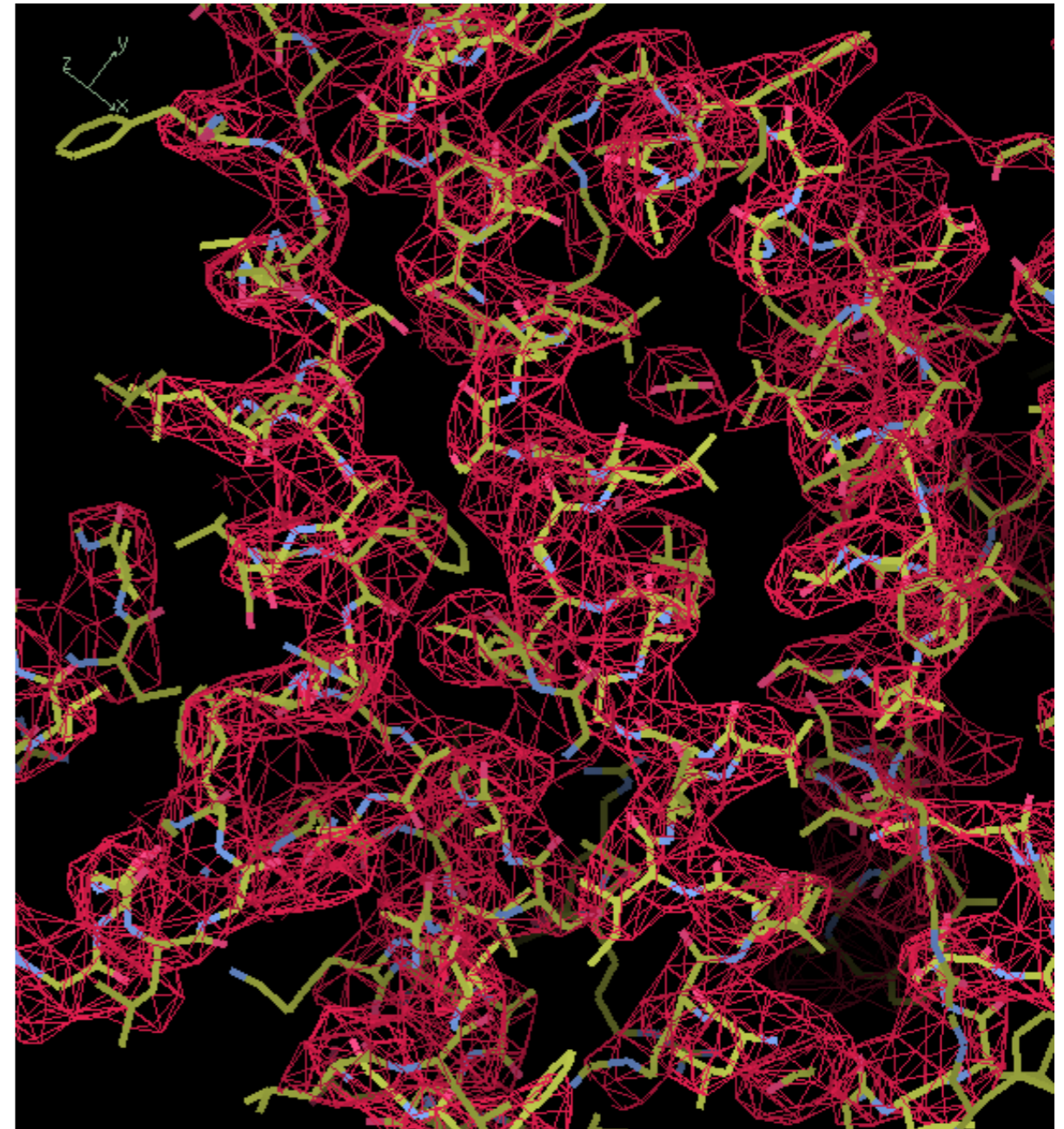


30S Ribosome (1j5e, 2.9 Å)

# Building at Low Resolution

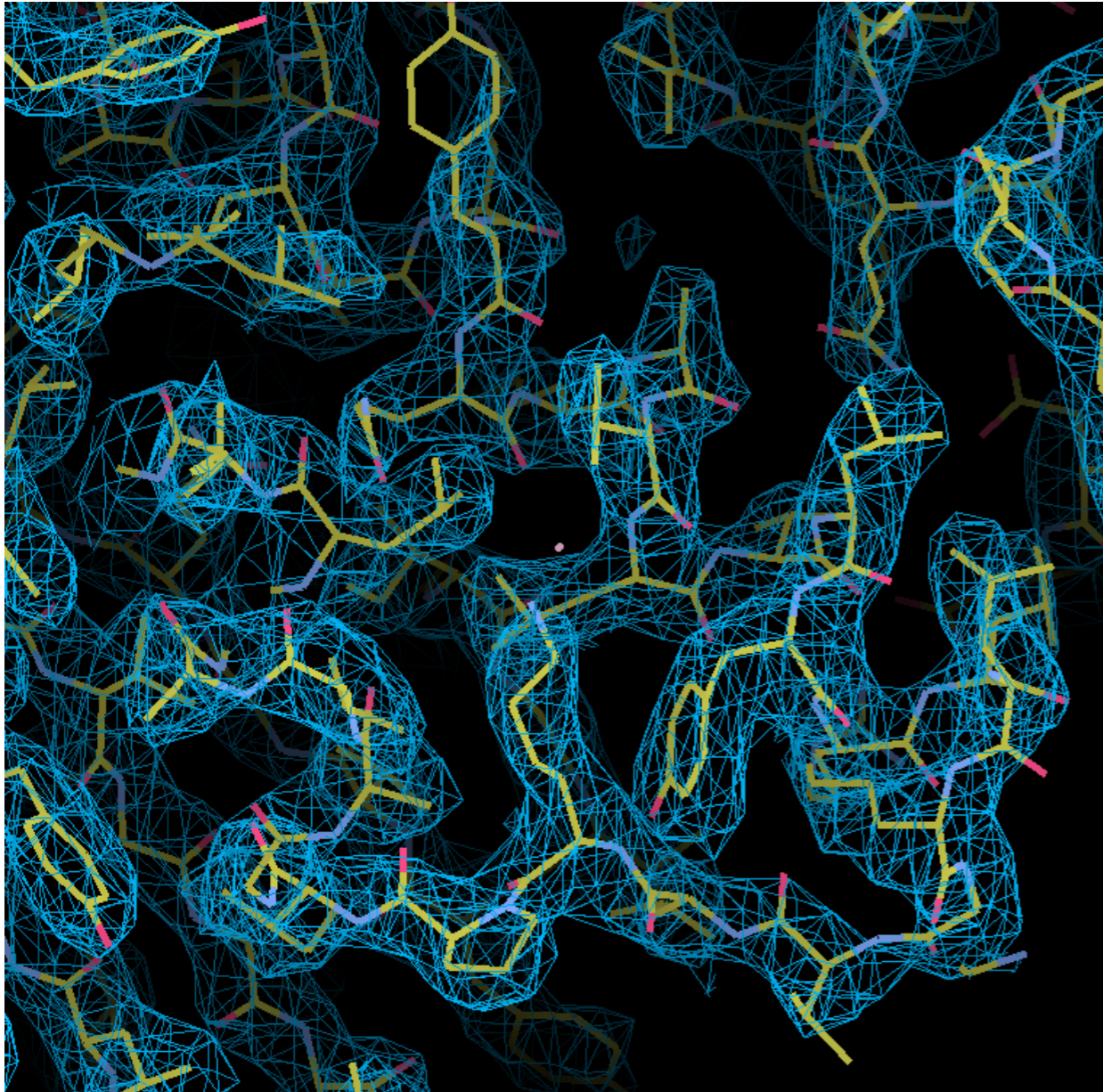


Gamma-secretase at 4.5 Å  
(autobuilt model; emd\_2677)

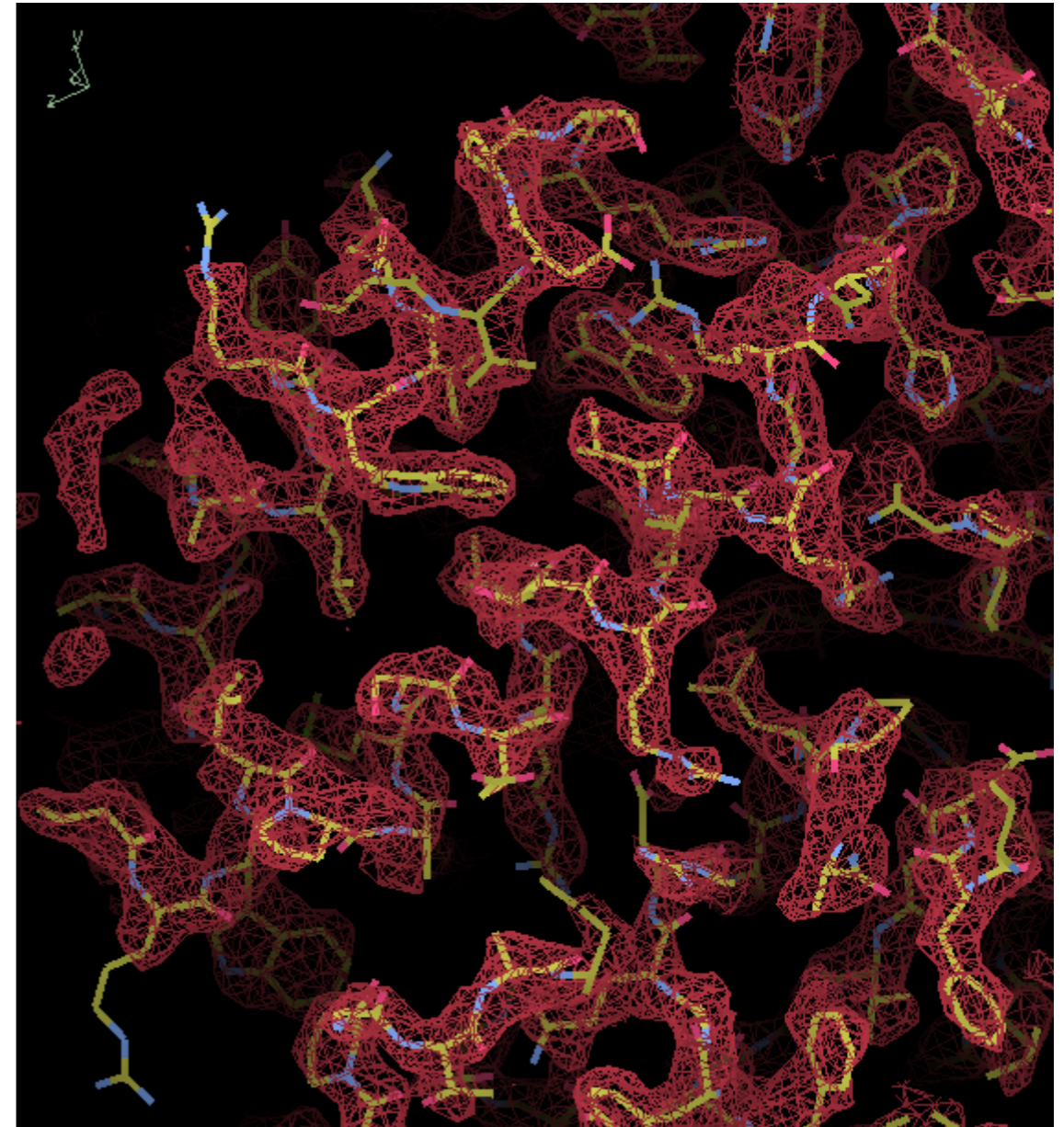


Gamma-secretase structure at 3.4 Å  
(autobuilt model; emd\_3061)

# Building at Medium/High Resolution

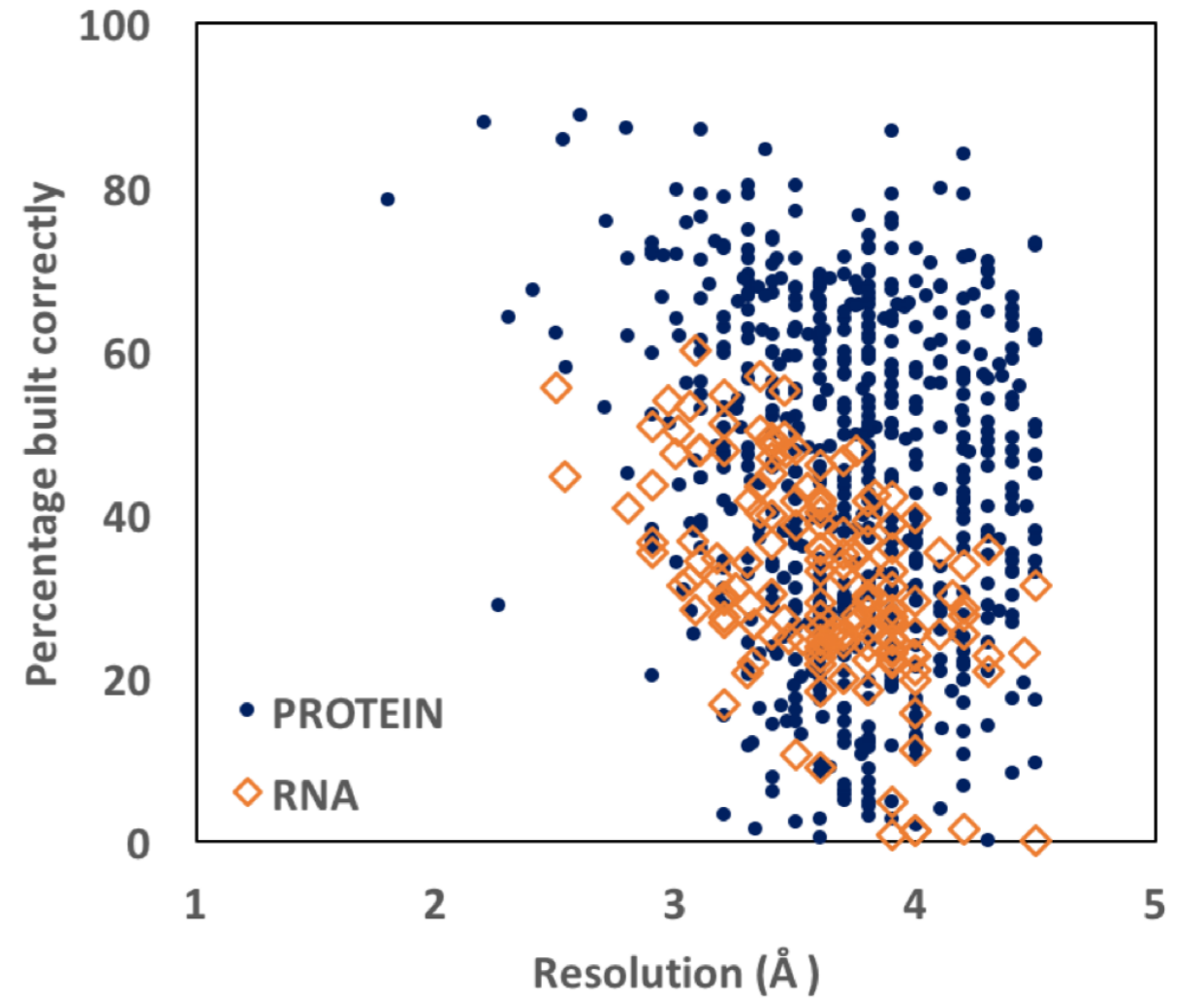
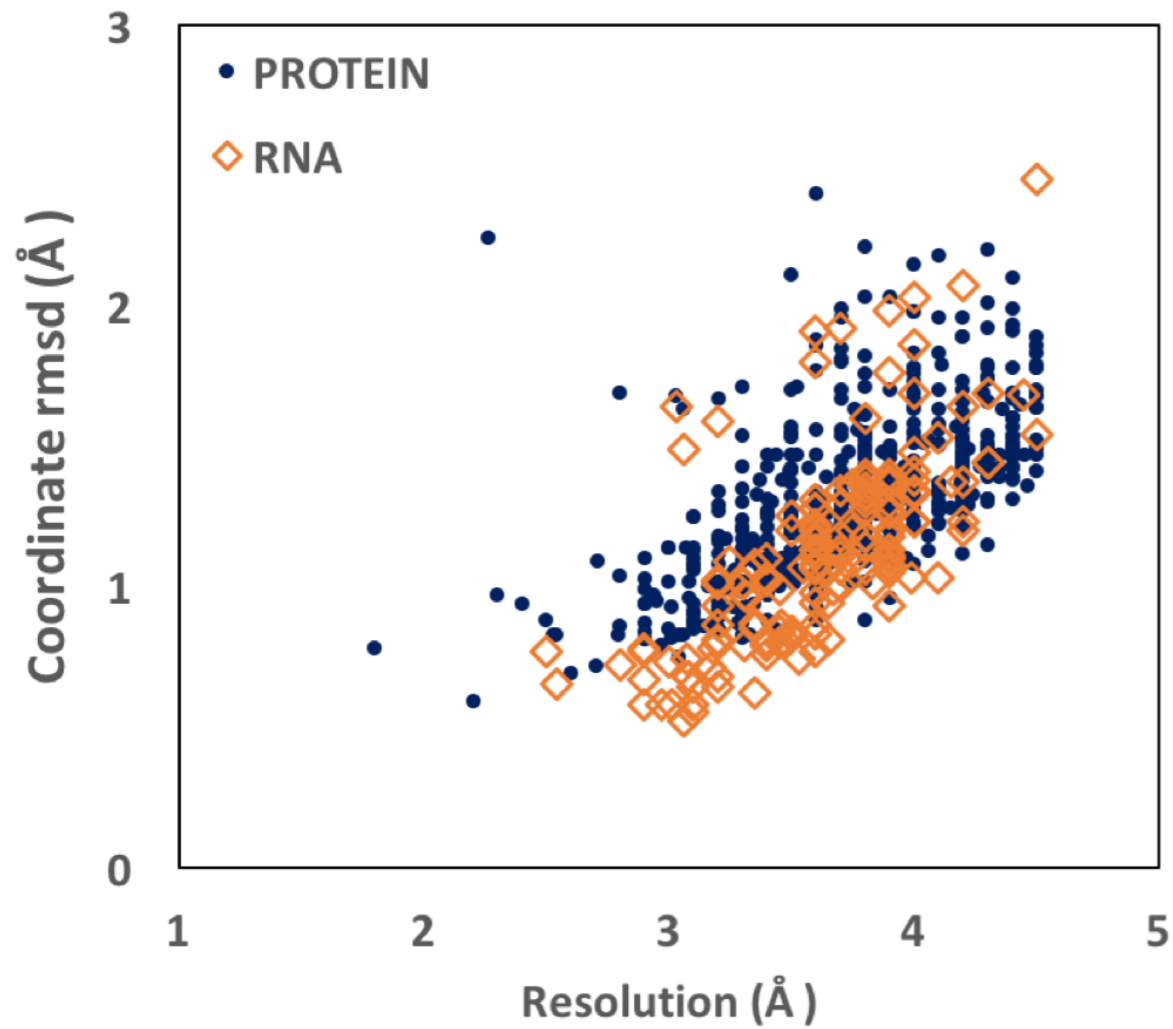


Proteasome at 2.8 Å  
(autobuilt model; emd\_6287)



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

# Autobuilding Performance



# Structure Optimization

**Pavel Afonine, Oleg Sobolev, Youval Dar, Nat  
Echols, Jeff Headd, Nigel Moriarty**

Lawrence Berkeley National Laboratory

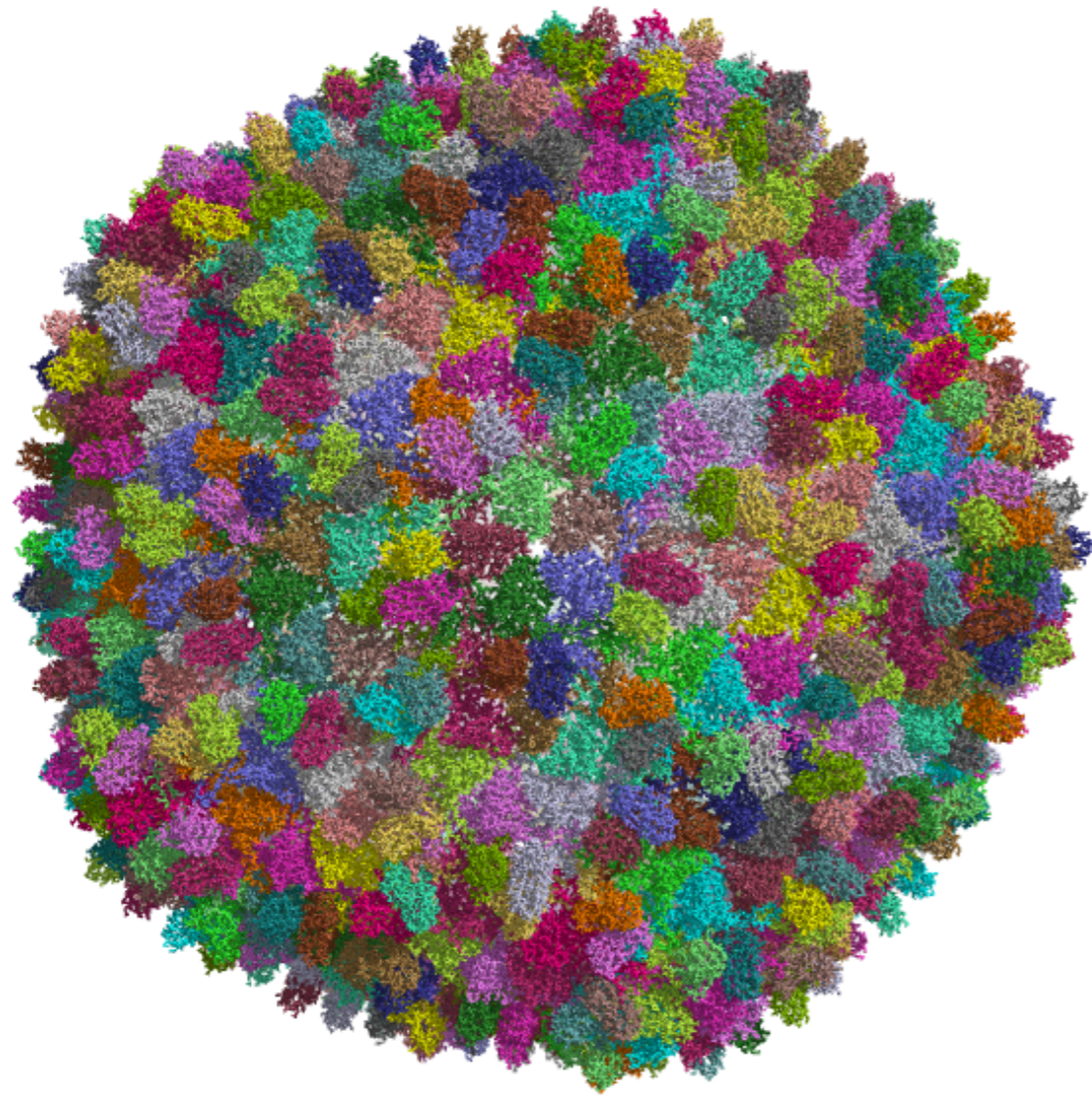
**Tom Terwilliger**

Los Alamos National Laboratory





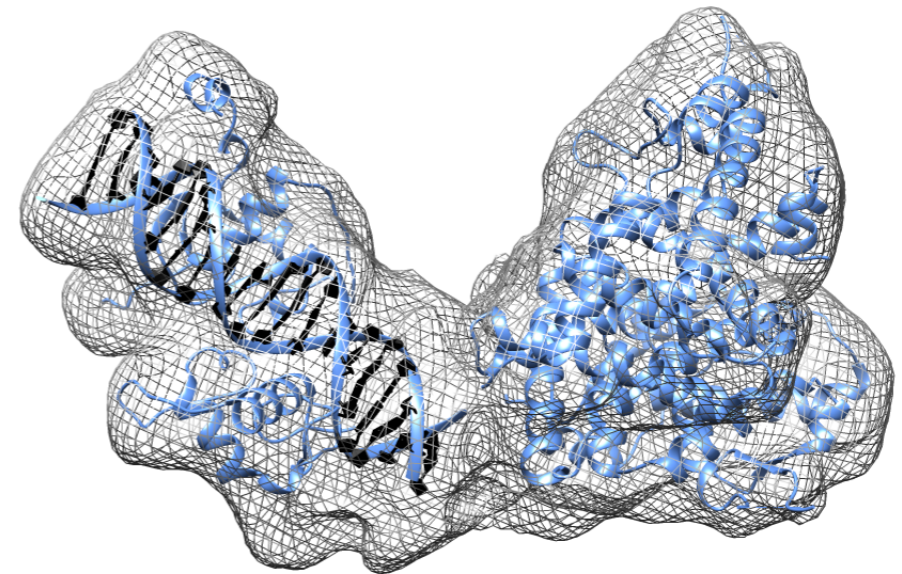
# Challenges



Resolution 4.5 Å

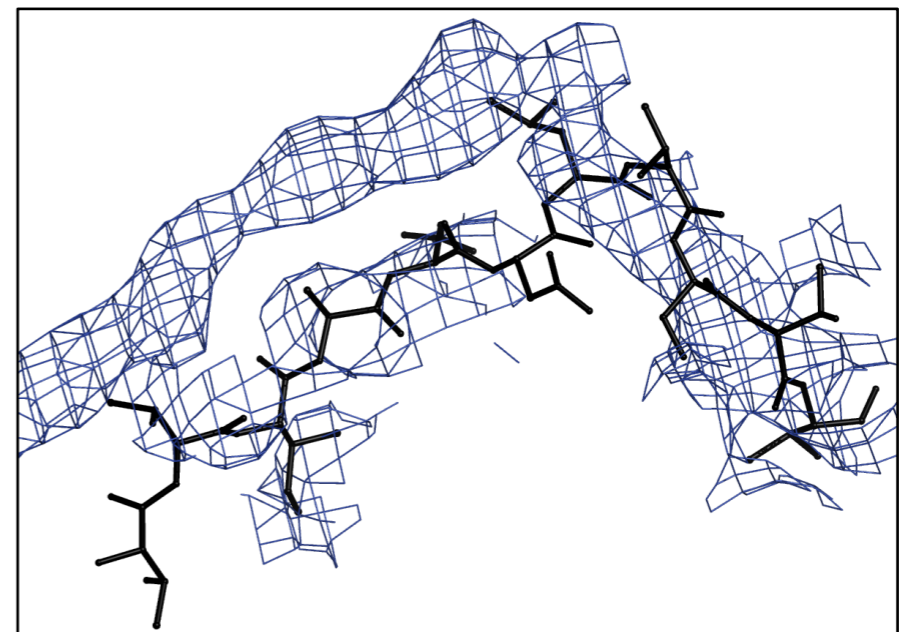
840 chains, 187,320 residues 1,443,960 atoms

**Size**



Resolution: 11.6 Å

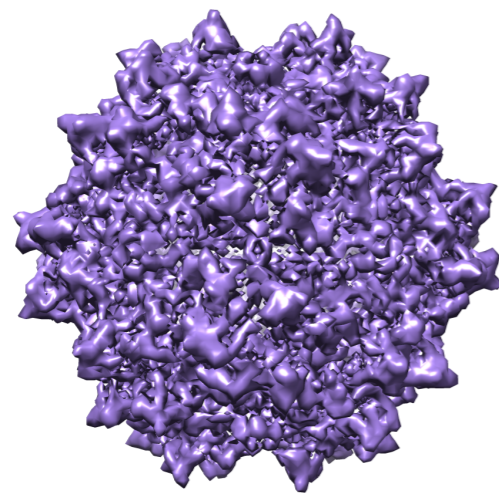
**Wide Resolution Range**



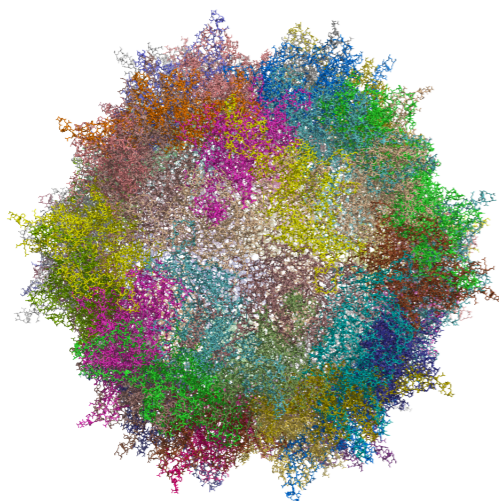
User data, resolution: 3.8 Å

**Poor Initial Fit**

# Direct Refinement Against the Map



Real space  
refinement



VOLUME FOUR



COMPUTATIONAL  
CRYSTALLOGRAPHY  
NEWSLETTER

JULY MMXIII

## ENSEMBLE REFINEMENT, CABLAM

### Table of Contents

• PHENIX News	28
• Crystallographic meetings	29
• Expert Advice	
• Fitting tips #6 - Potential misfitting by a switch of sidechain vs mainchain	30
• FAQ	32
• Short Communications	
• <i>CaBLAM</i> : Identification and scoring of disguised secondary structure at low resolution	33
• Structural Classification of Allergen IgE Epitopes by Hierarchical Clustering	36
• New Tool: <i>phenix.real_space_refine</i>	43
• <i>cctbx</i> tools for transparent parallel job execution in Python. III. Remote access	45
• <i>phenix.ensemble_refinement</i> : a test study of apo and holo BACE1	51

Editor  
Nigel W. Moriarty, [NWMoriarty@LBL.Gov](mailto:NWMoriarty@LBL.Gov)

### PHENIX News

#### New programs

**FEM: Feature Enhanced Maps (Pavel V. Afonine)**  
Interpretation of a crystallographic map is a means of obtaining an atomic representation of a crystal structure or the map itself may serve as the crystal model. There are number of factors that affect quality of crystallographic maps that in turn affect difficulty (or even feasibility) of their interpretation and quality of resulting model of crystal structure, and include:

- finite resolution of measured reflections;
- incompleteness of data (missing reflections within the resolution range of the measured data);
- experimental errors in measured reflections;
- errors in atomic model parameters.

These factors a) result in artificial peaks in the map that may be confused with the signal and therefore erroneously interpreted in terms of atomic model, b) introduce noise that may obscure the signal and c) may distort the signal in various ways.

Another fundamentally different contributor to the difficulty of map interpretation is that not all the signal has the same strength. For example, a strong signal arising from a heavy atom derivative may easily obscure a very weak signal (that may be at or below the noise level) arising from a partially occupied very mobile ligand or residue side chain alternative conformation or even hydrogen atoms.

The Computational Crystallography Newsletter (CCN) is a regularly distributed electronically via email and the PHENIX website, [www.phenix-online.org/newsletter](http://www.phenix-online.org/newsletter). Feature articles, meeting announcements and reports, information on research or other items of interest to computational crystallographers or crystallographic software users can be submitted to the editor at any time for consideration. Submission of text by email or word-processing files using the CCN templates is requested. The CCN is not a formal publication and the authors retain full copyright on their contributions. The articles reproduced here may be freely downloaded for personal use, but to reference, copy or quote from it, such permission must be sought directly from the authors and agreed with them personally.

Computational Crystallography Newsletter (2013). Volume 4, Part 2. 28

# Real Space Refinement

- Has a long history in both X-ray crystallography and cryo-EM
  - Early crystallographic refinement programs (Diamond)
  - Alternative to reciprocal space refinement, then applied to EM maps (Chapman)
  - Regularly used in model building (O, Coot)
- New structure fitting approaches make use of real space refinement
  - Molecular dynamics flexible fitting (MDFF)
  - Deformable elastic network fitting (DireX)
  - Rosetta model building and model refinement

# Goal for Cryo-EM Model Refinement

- Stable refinement against any density map (Cryo-EM or X-ray)
- End result should be an improvement in the model
- Large radius of convergence
- Final models with good fit to density and physically reasonable geometry (Ramachandran distribution, rotamers, packing)
- Fast: no more than one second per residue

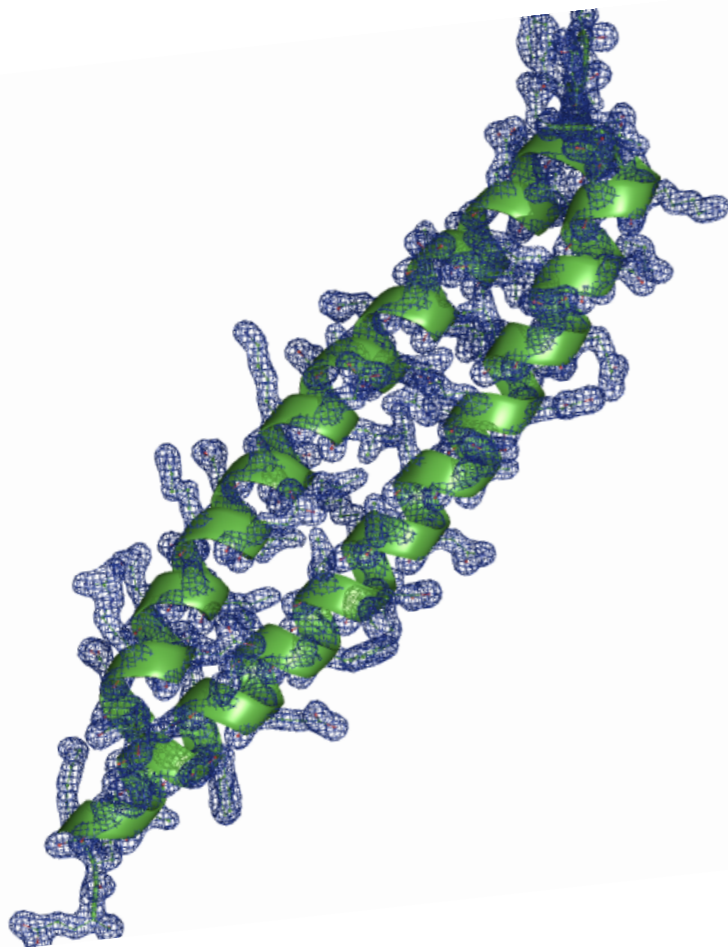
$$E = E_{\text{chemistry}} + \omega \sum (\rho_o - \rho_c)^2$$

  
**Phenix**

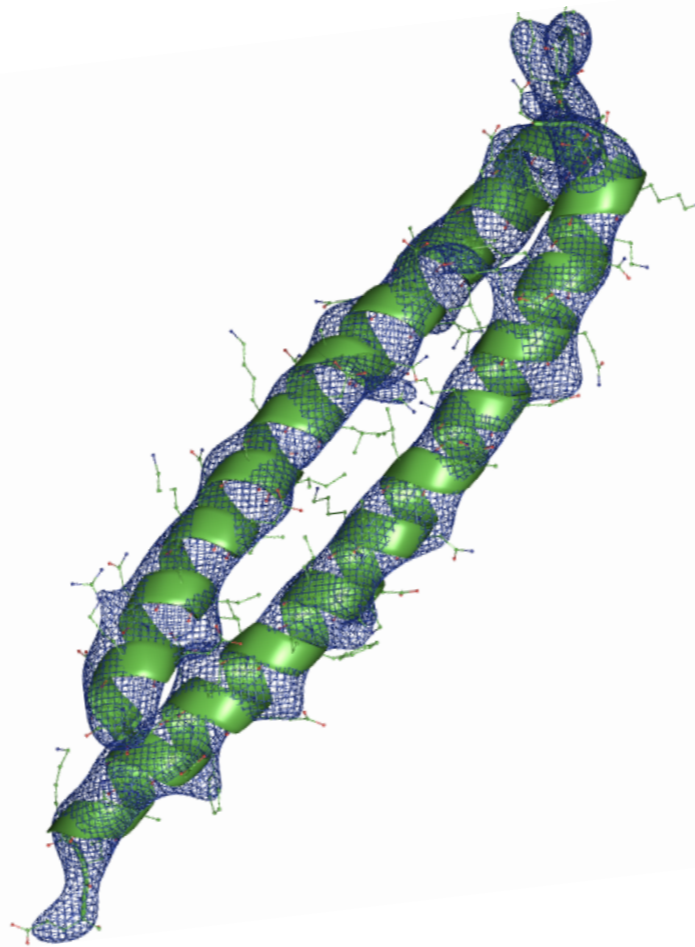
# Lower Resolution Requires Additional Information

**High Resolution**

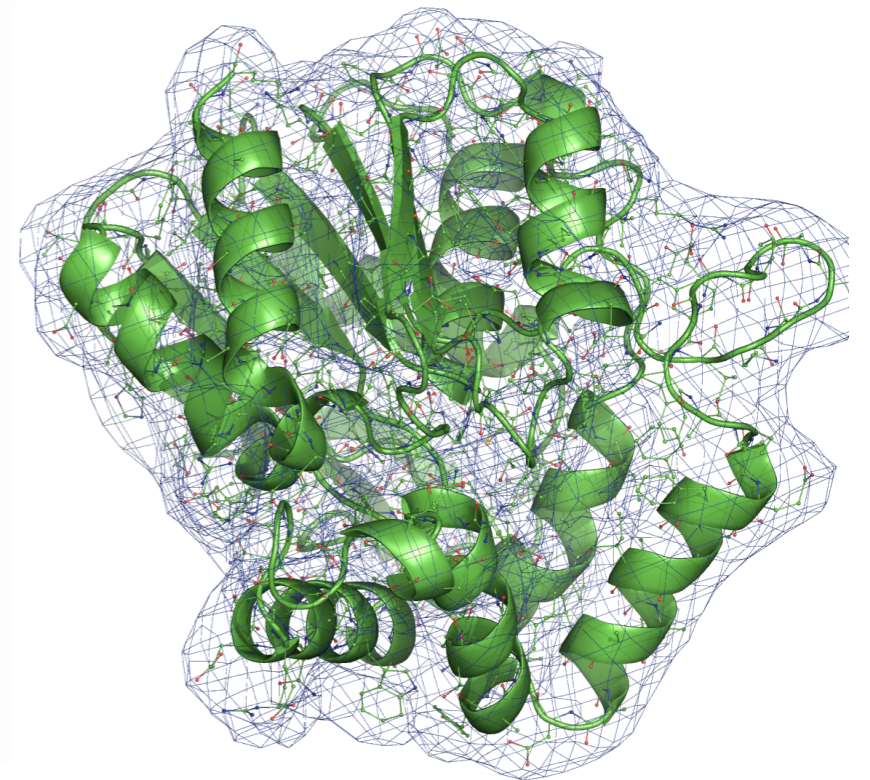
**Low Resolution**



*Side chains*

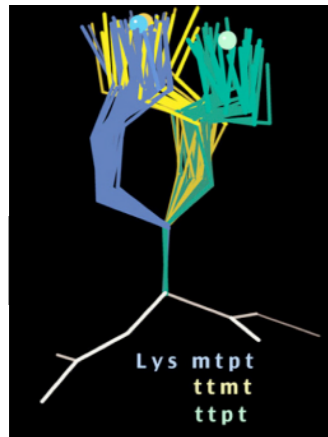


*Secondary Structure*

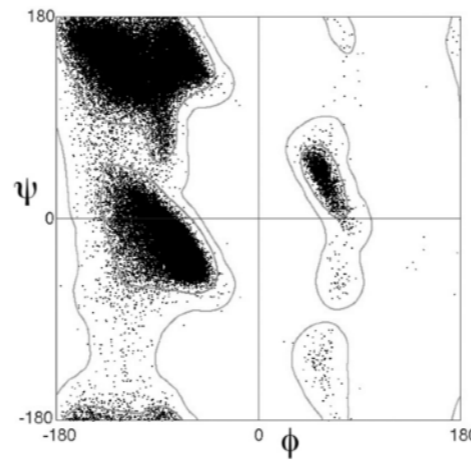
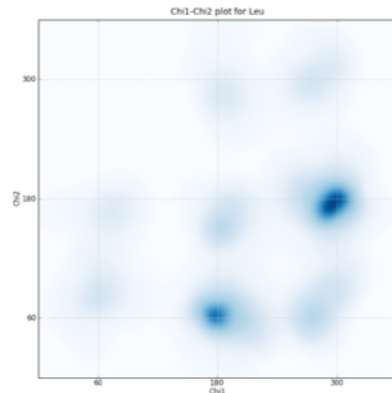


*Molecule*

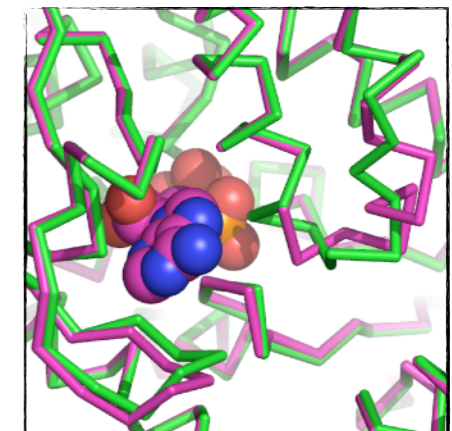
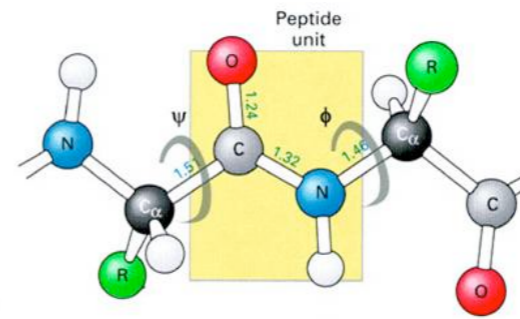
# Sources of Prior Information



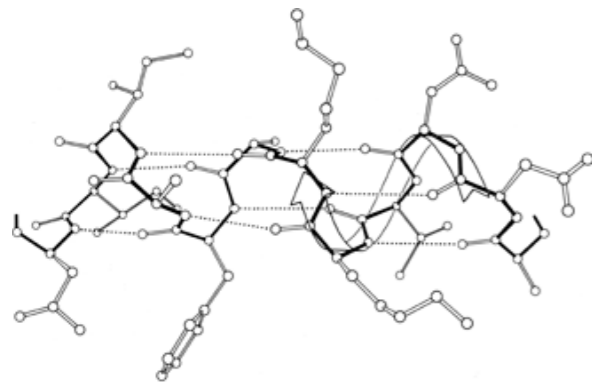
Sidechain distributions



Mainchain distributions

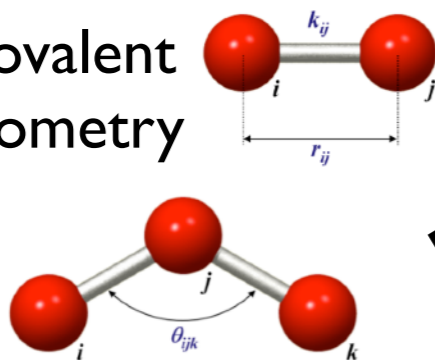


Related structures

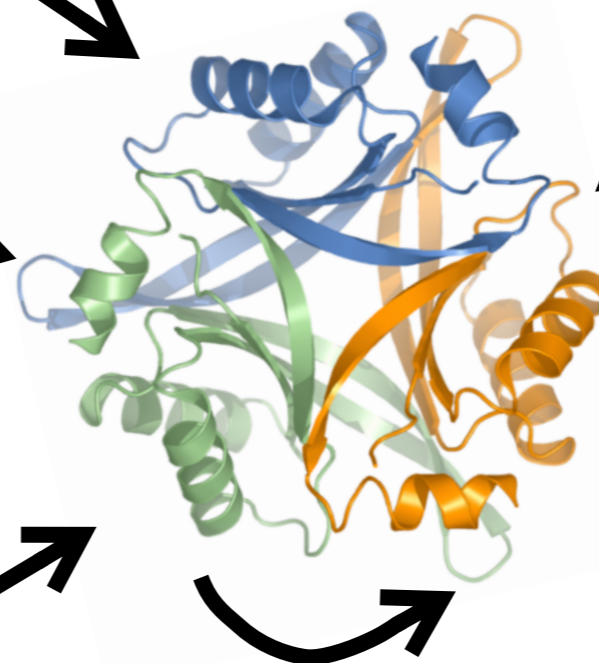


Secondary structure

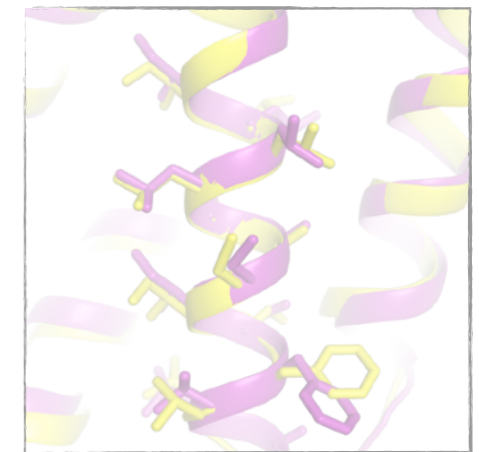
Covalent geometry



Images from PumMa web site (<http://www.pumma.nl>)



Internal symmetry



Physically Realistic Potentials

DiMaio et al., *Nature Methods* 2013, **10**:1102-1104

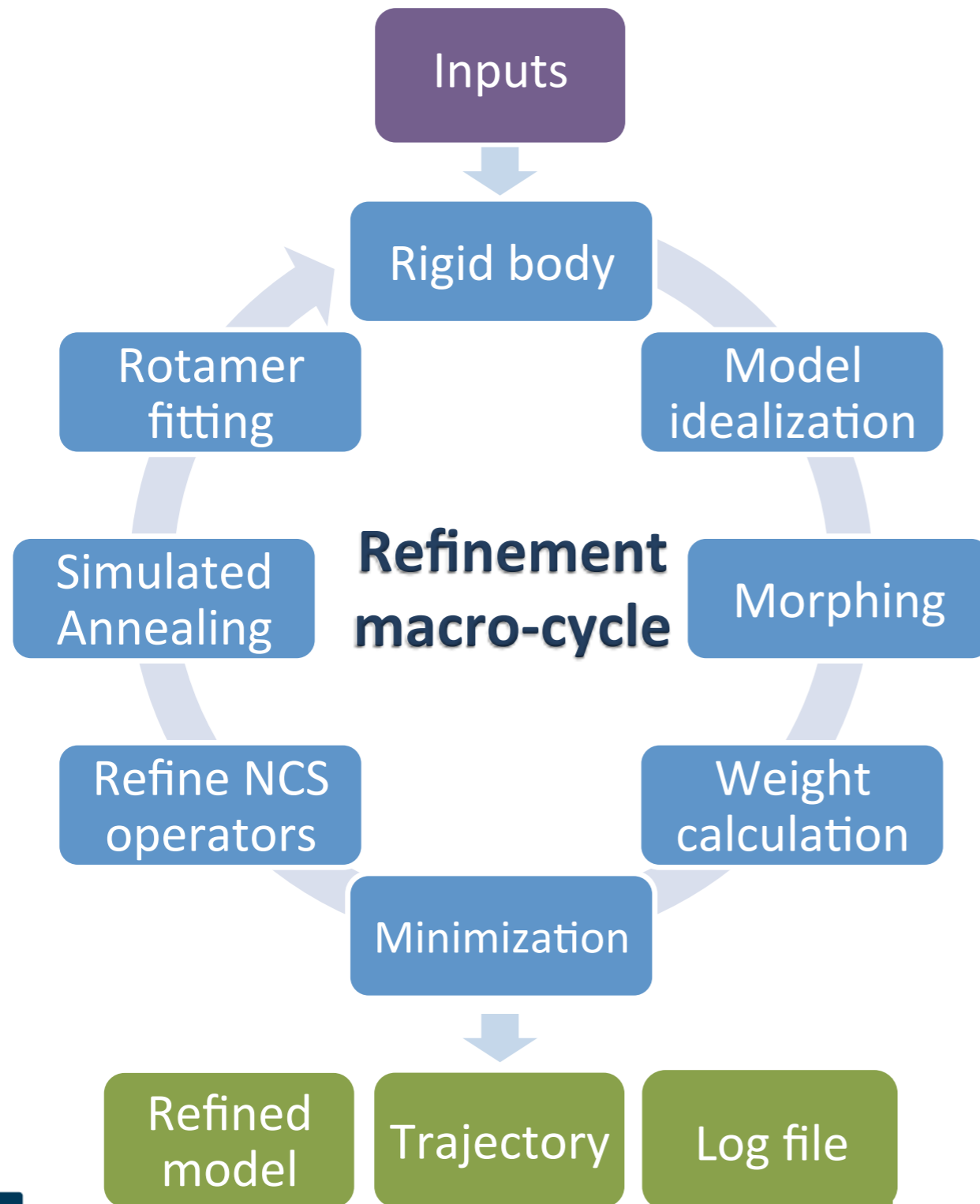
Headd et al., *Acta Cryst.* 2014, **D70**:1346-1356

Headd et al., *Acta Cryst.* 2012,

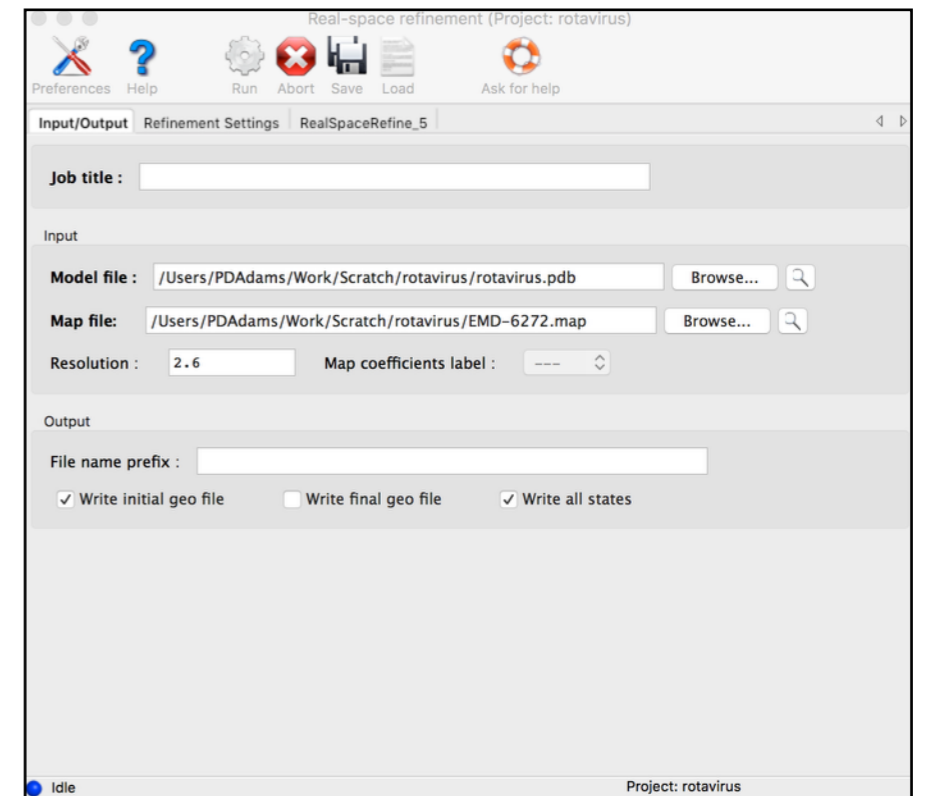
**D68**:381-390

**Phenix**

# Real Space Refinement Procedure



● `phenix.real_space_refine`

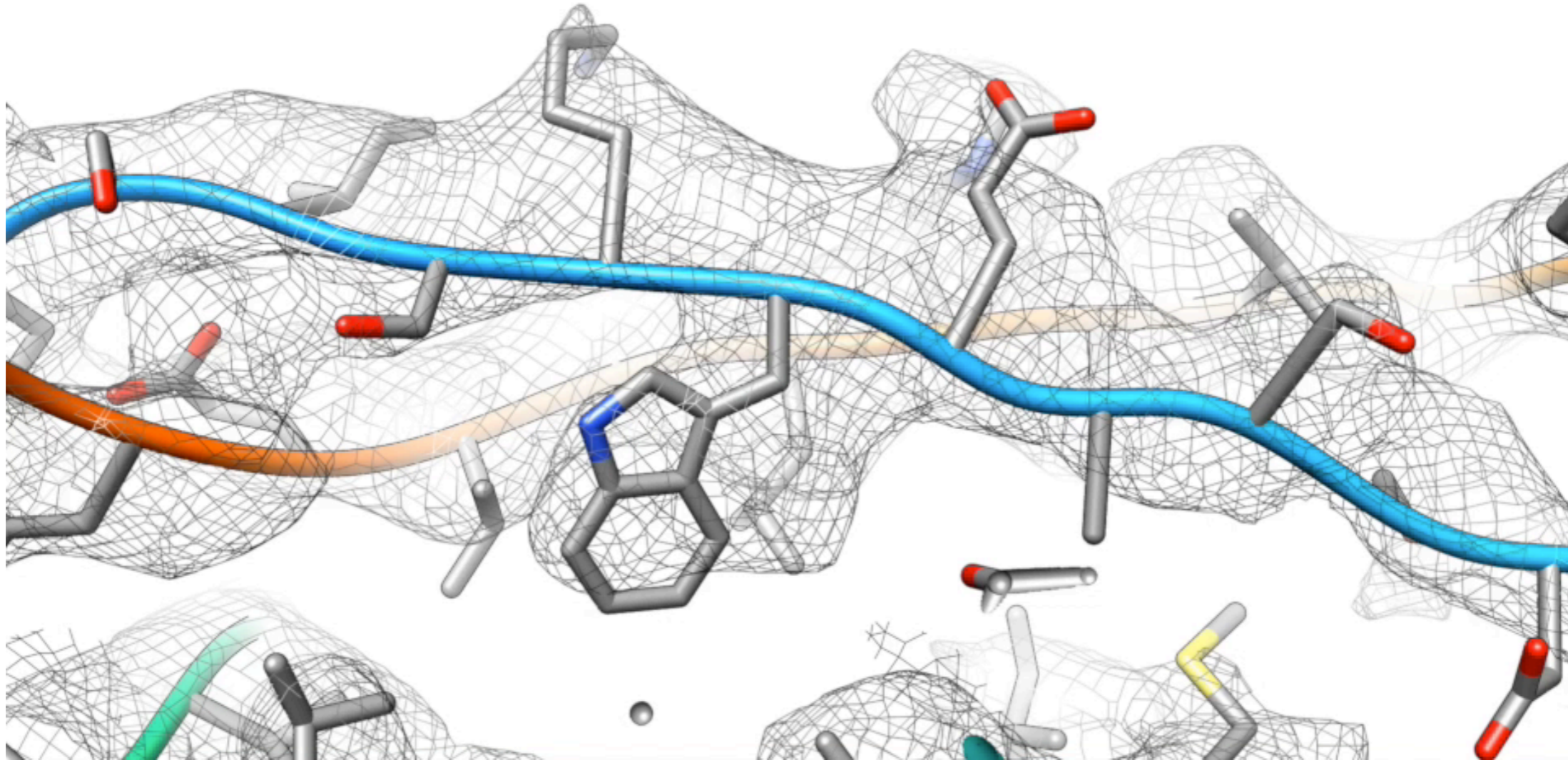


*Pavel Afonine, Youval Dar, Oleg Sobolev, Billy Poon (LBNL), Tom Terwilliger (LANL)*

Wang Z, et al. An atomic model of brome mosaic virus using direct electron detection and real-space optimization. *Nature Commun.* 2014, **5**:4808

Hryc CF et al. Accurate model annotation of a near-atomic resolution cryo-EM map. *Proc Natl Acad Sci U S A* 2017.

# Virus Structure Refinement



Collaboration with Wah Chiu, Zhao Wang, Corey Hryc, Matthew Baker (Baylor College of Medicine)

Pavel Afonine (LBNL), Corey Hryc (BCM)

Wang et al., *Nature Commun.* 2014, **5**:4808

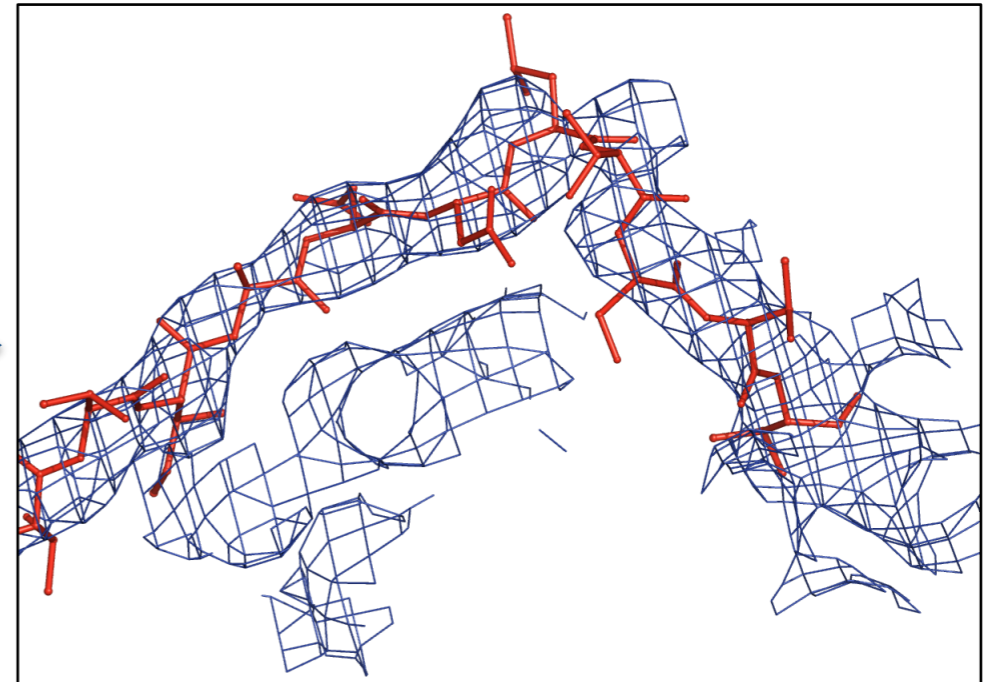
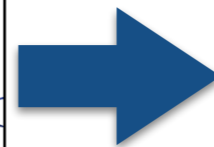
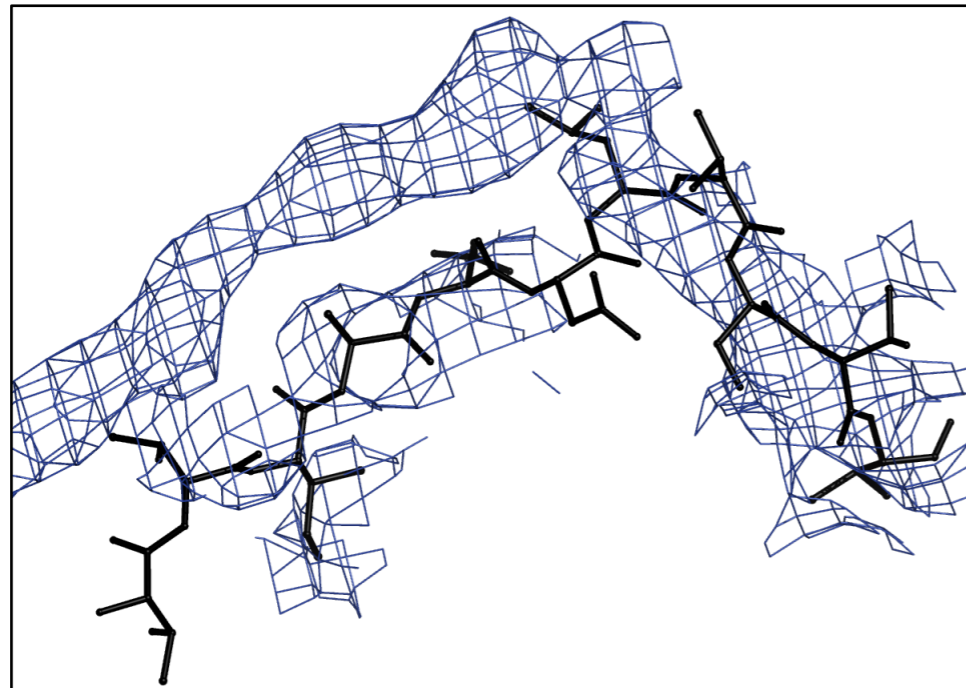
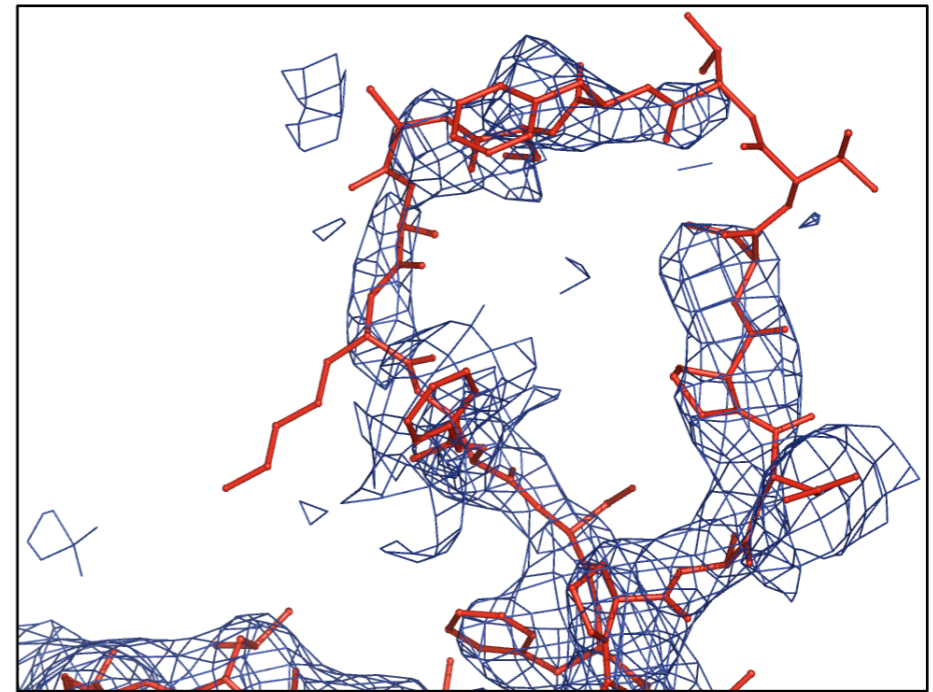
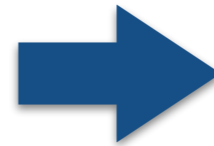
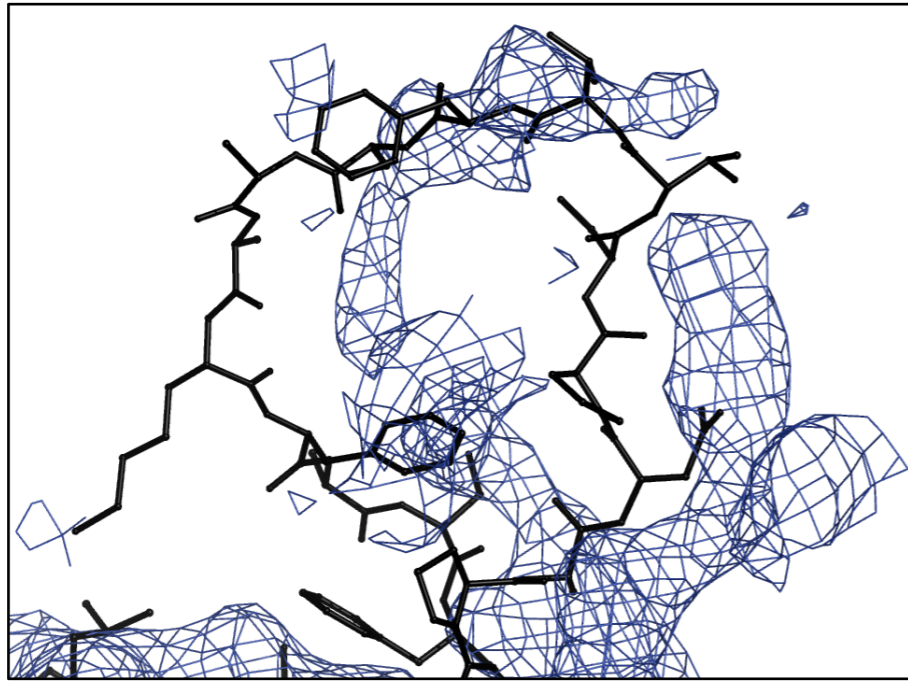
  
**Phenix**





# Real Space Refinement Improves Fit to Data

- Models are moved to better fit the Cryo-EM map



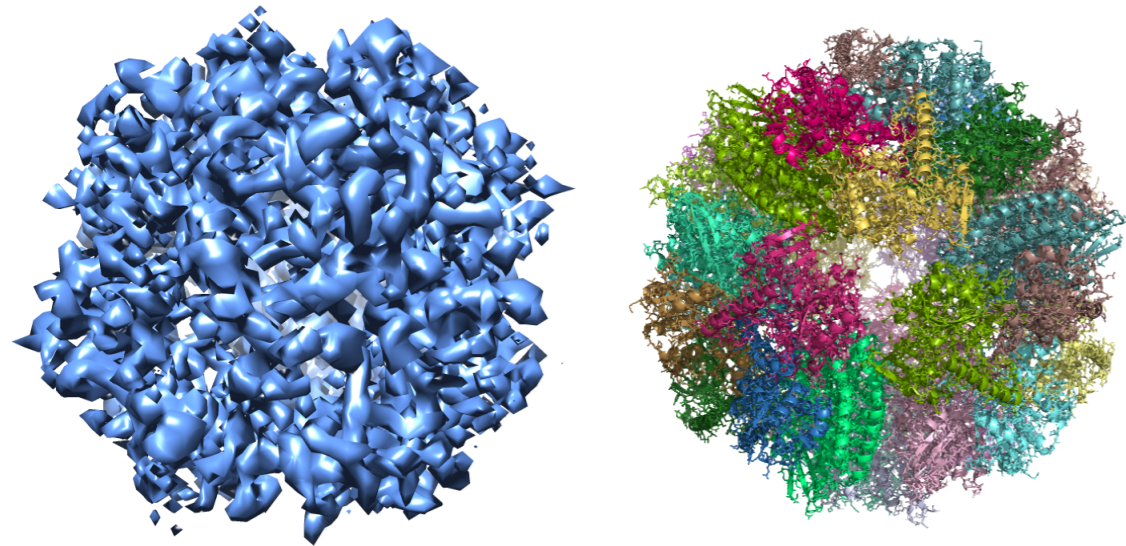
# While Also Improving Stereochemistry

- Standard metrics (MolProbity) are all improved

	Starting Model	After RS Refinement
Clashscore	47	18
Ramachandran Outliers (%)	6	2
Ramachandran Allowed (%)	16	8
Ramachandran Favoured (%)	78	90
Rotamer Outliers (%)	20	0
C <sub>β</sub> Deviations (%)	2	0

# Typical Results at Higher Resolution

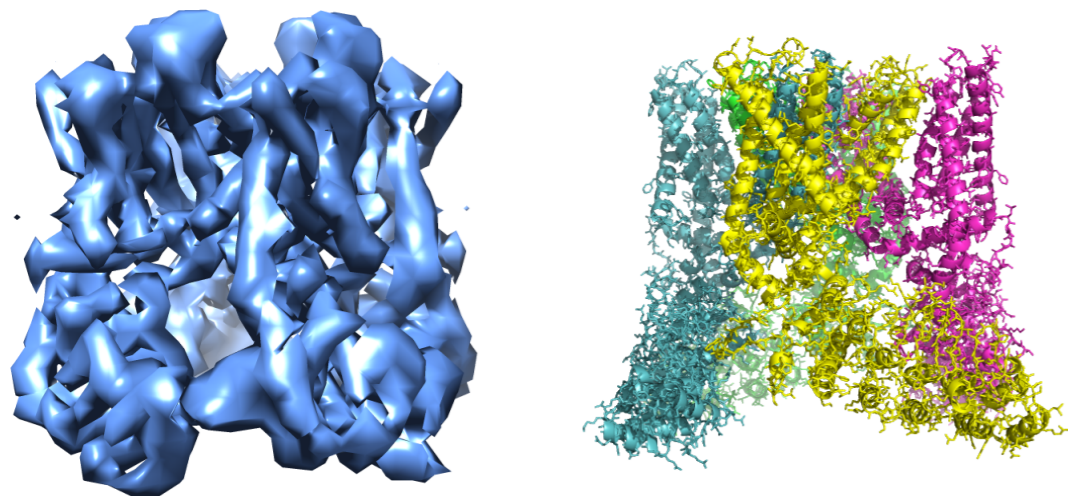
**Resolution: 3.36 Å**



**Residues/atoms: 10,716/82,404**  
**Refinement: 173 min**

METRIC	Original	<i>Phenix</i>
Map CC	0.645	0.783
RMSD (bonds/angles)	0.02/2.05	0.01/1.21
Clashscore	117.1	18.79
Rama. outl., %	0.11	0.11
Rotamer outl., %	35.51	0
C-beta deviations	24	0

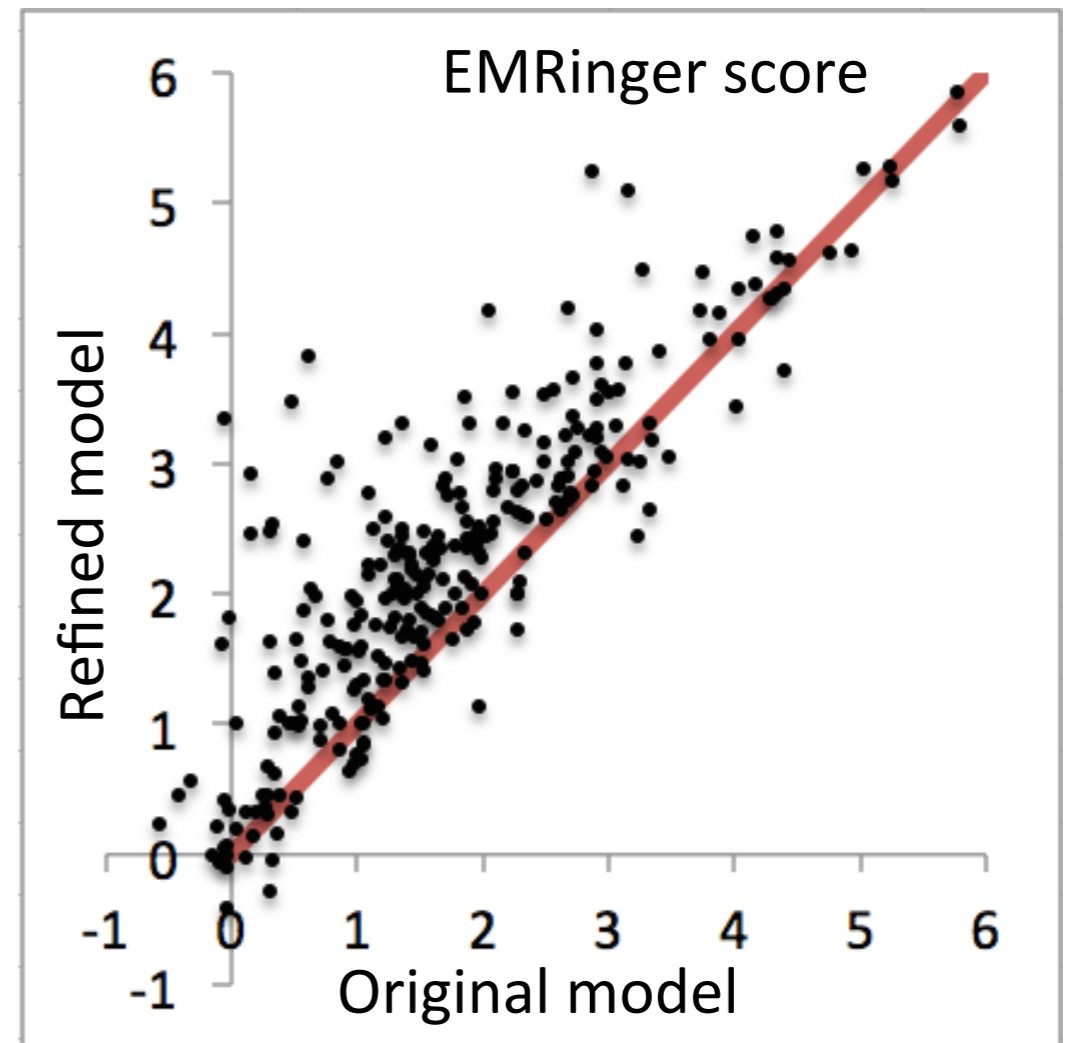
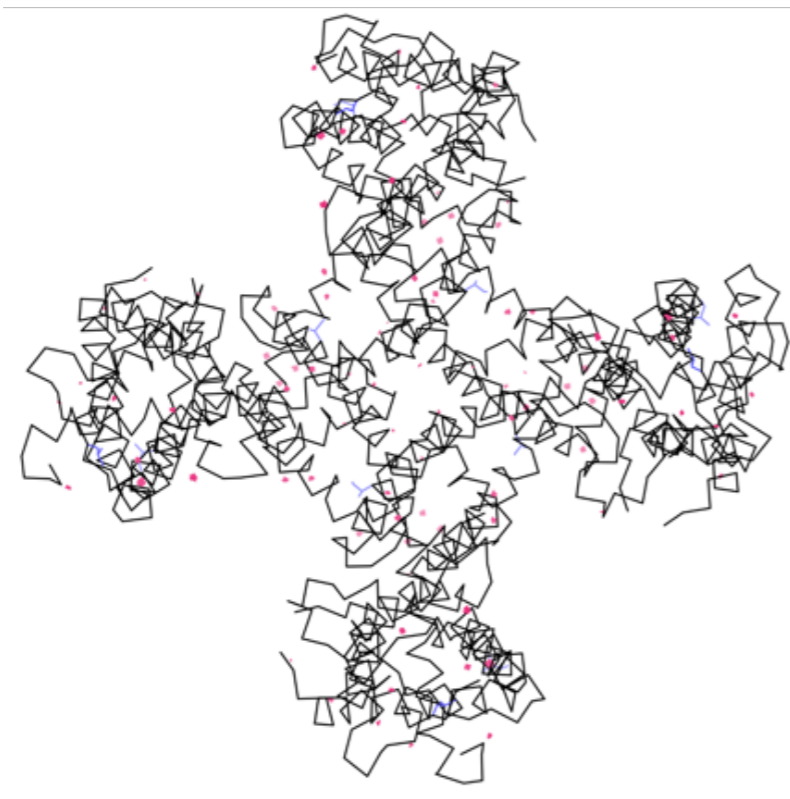
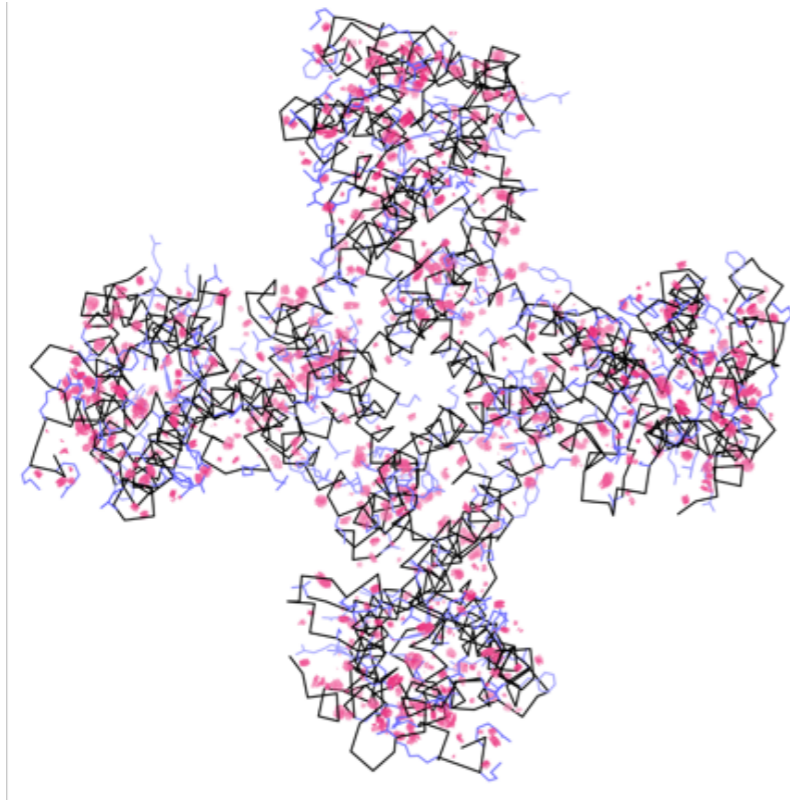
**Resolution: 3.8 Å**



**Residues/atoms: 2,324/17,424**  
**Refinement: 20 min**

METRIC	Original	<i>Phenix</i>
Map CC	0.650	0.714
RMSD (bonds/angles)	0.01/1.34	0.01/1.31
Clashscore	100.9	32.84
Rama. outl., %	0.52	0
Rotamer outl., %	27.99	0
C-beta deviations	0	0

# Improved Models



# Validation

**Benjamin Barad, Yifan Cheng, Jaime Fraser**

University of California San Francisco

**Ray Yu-Ruei Wang, Frank DiMaio**

University of Washington

**Nat Echols**

Lawrence Berkeley National Laboratory

**UCSF**

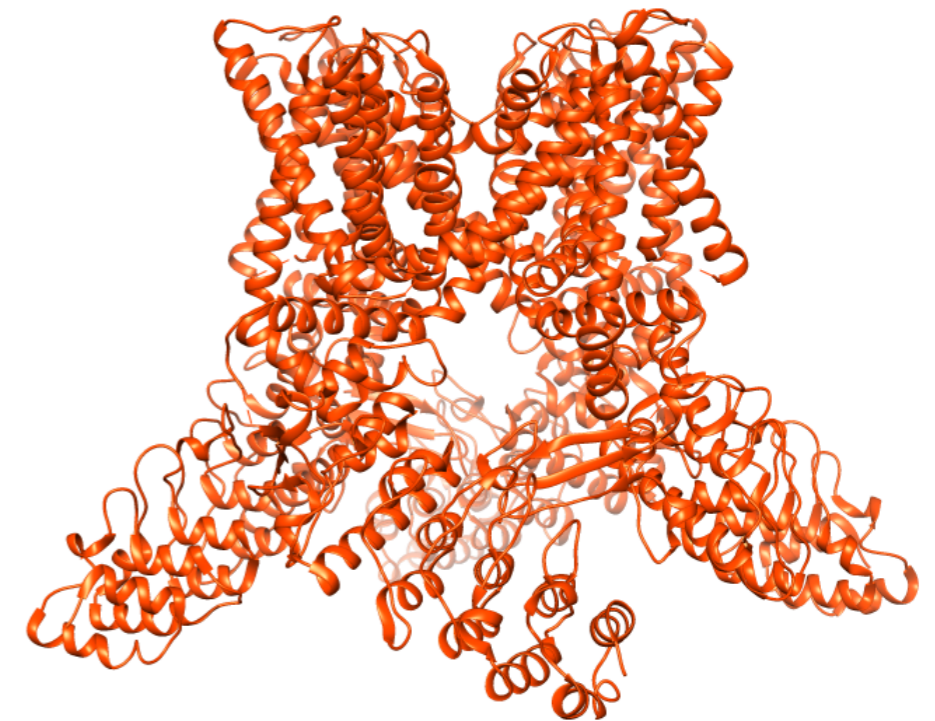
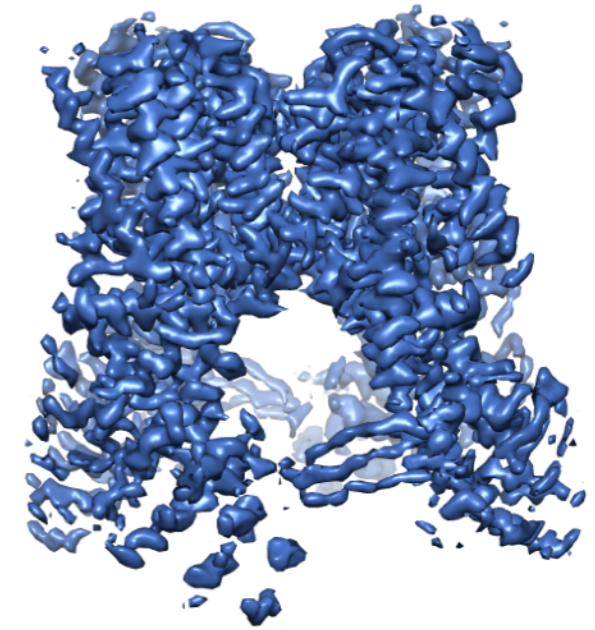
University of California  
San Francisco

**Phenix**



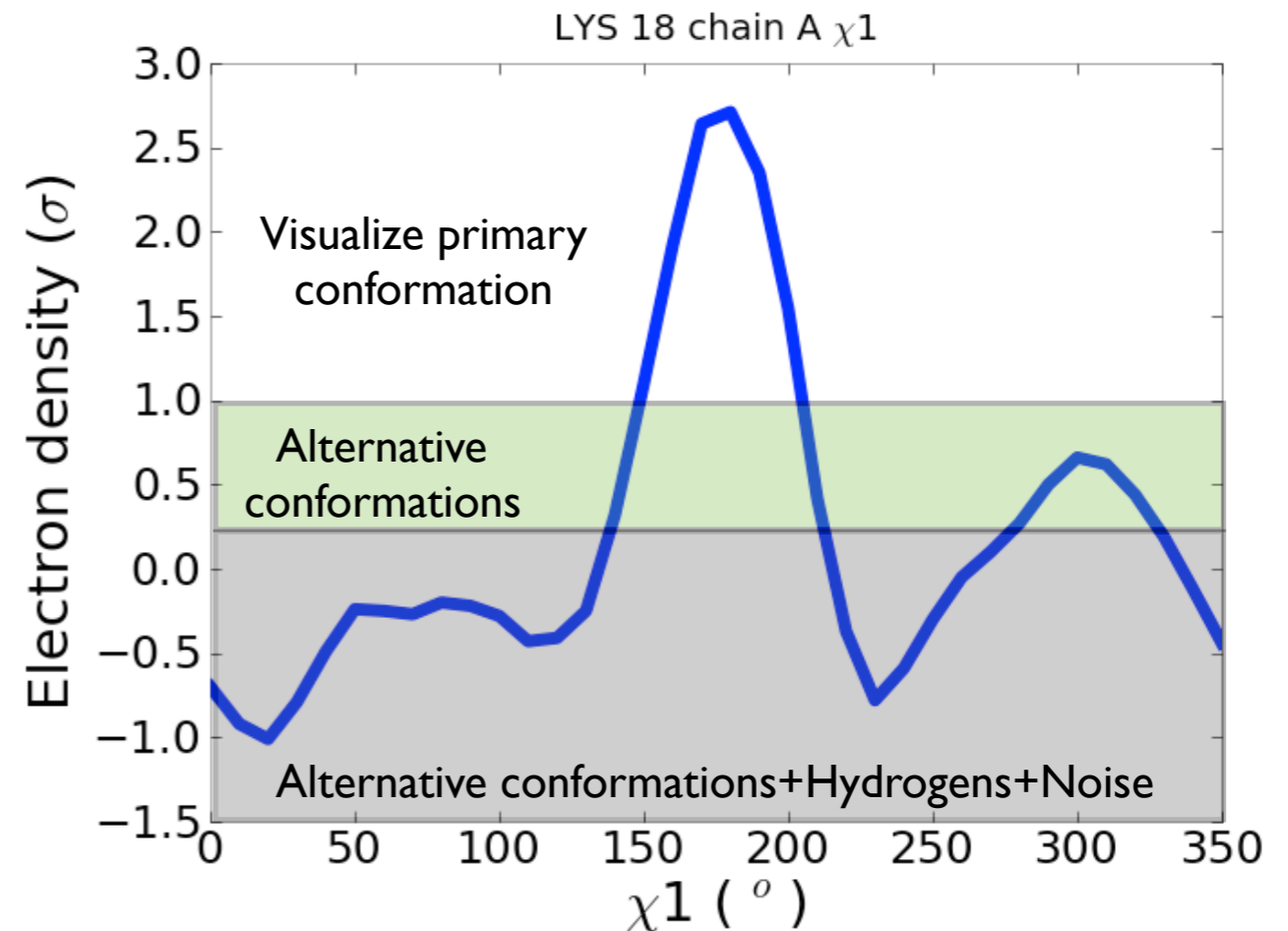
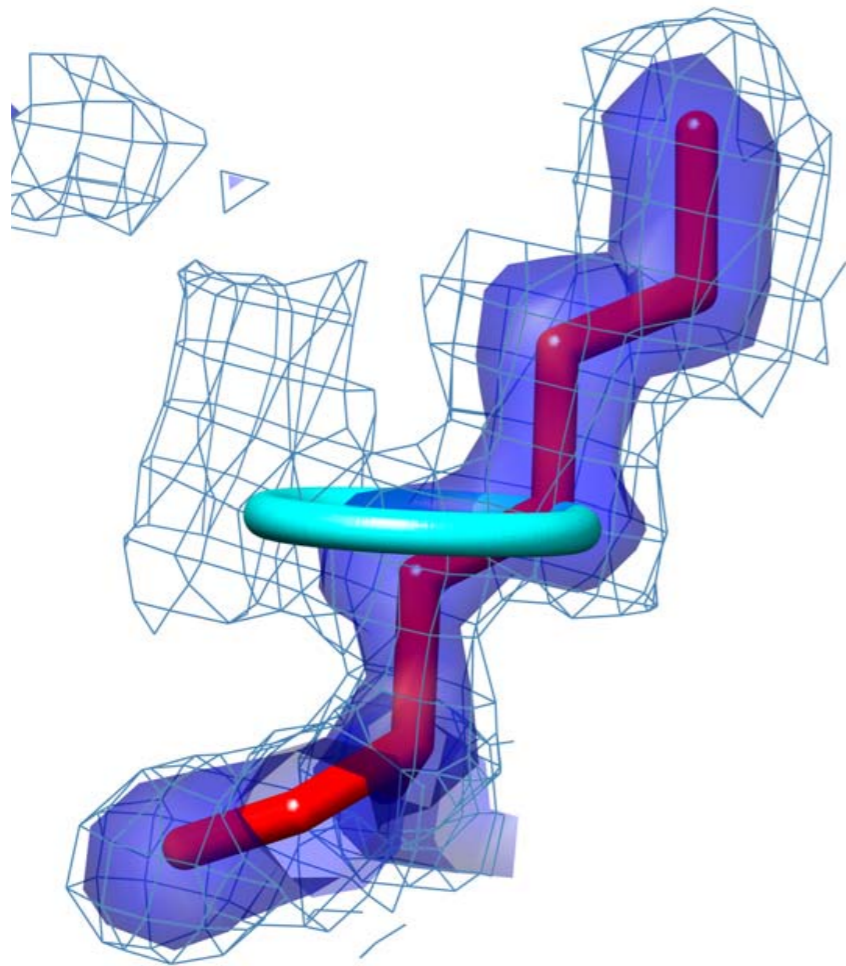
# Validation and Cryo-EM

- Do the map make sense?
  - Gold Standard FSC of half maps
- Does the model make sense?
  - MolProbity
- Does the model fit the map?
  - Overall and local density correlation
  - What about the detailed local fit?



# Look at the Density Around Sidechains

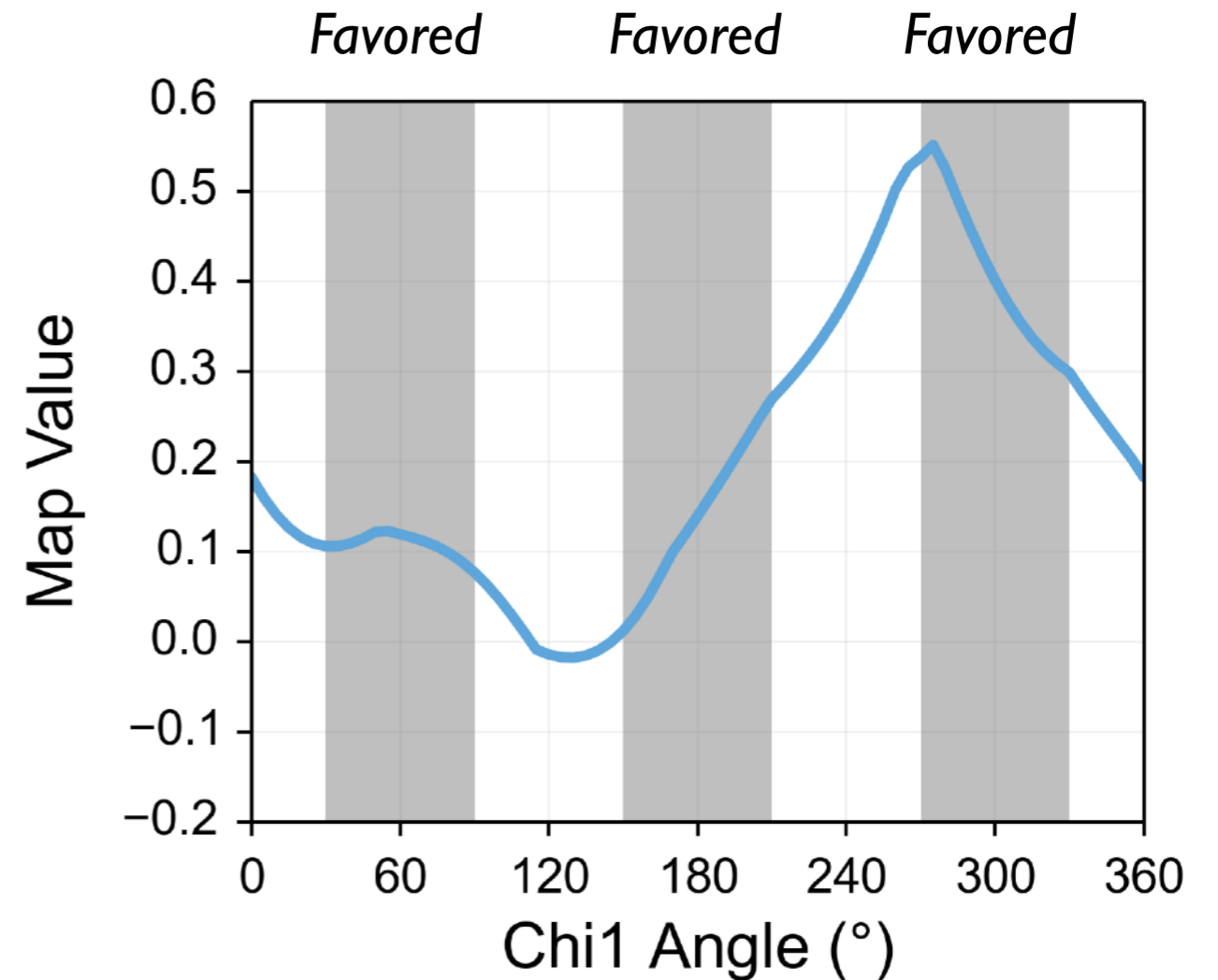
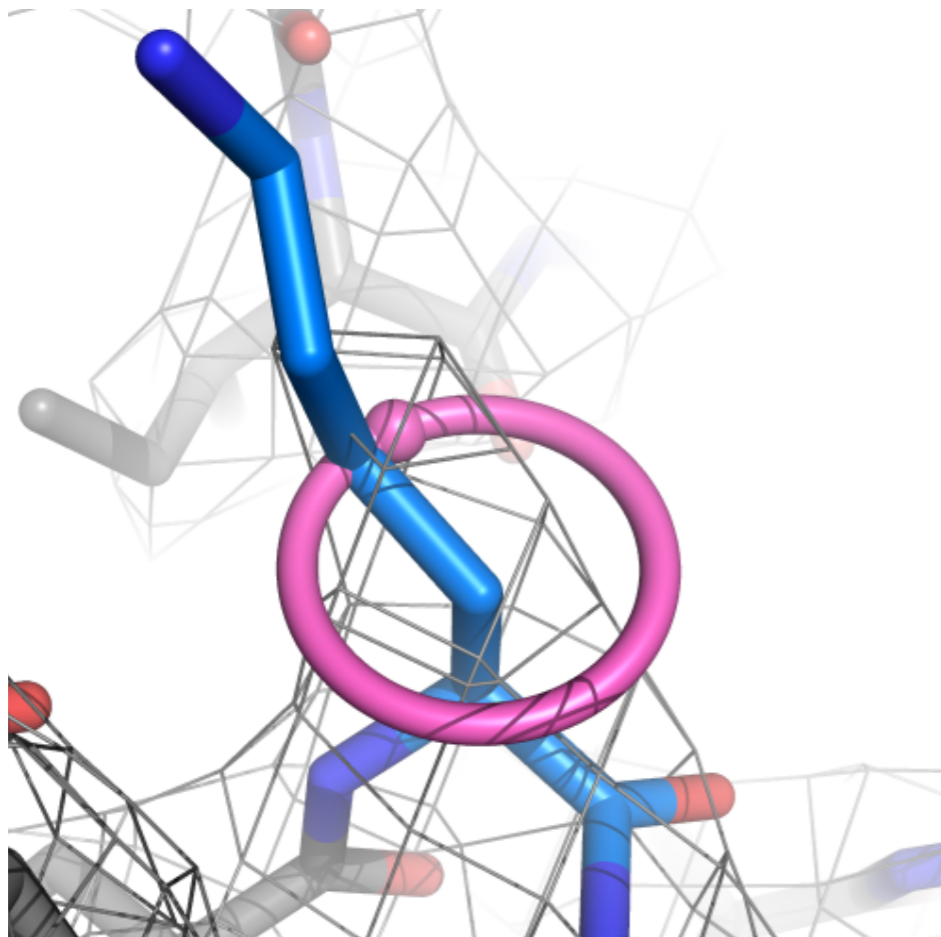
*Ringer*



Lang PT, et al. Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein Science*. 2010.

# Look at the Density Around Sidechains

*EMRinger*



Barad BA, et al. EMRinger: Side-chain-directed model and map validation for 3D Electron Cryomicroscopy. *Nature Methods*. 2015

**UCSF**

University of California  
San Francisco

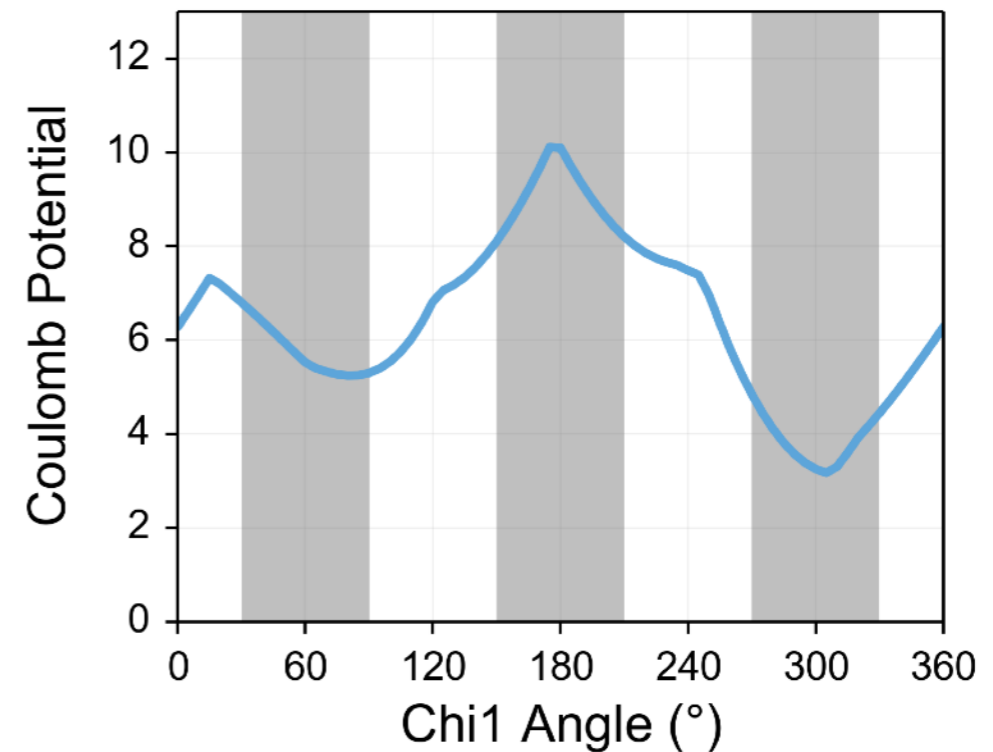
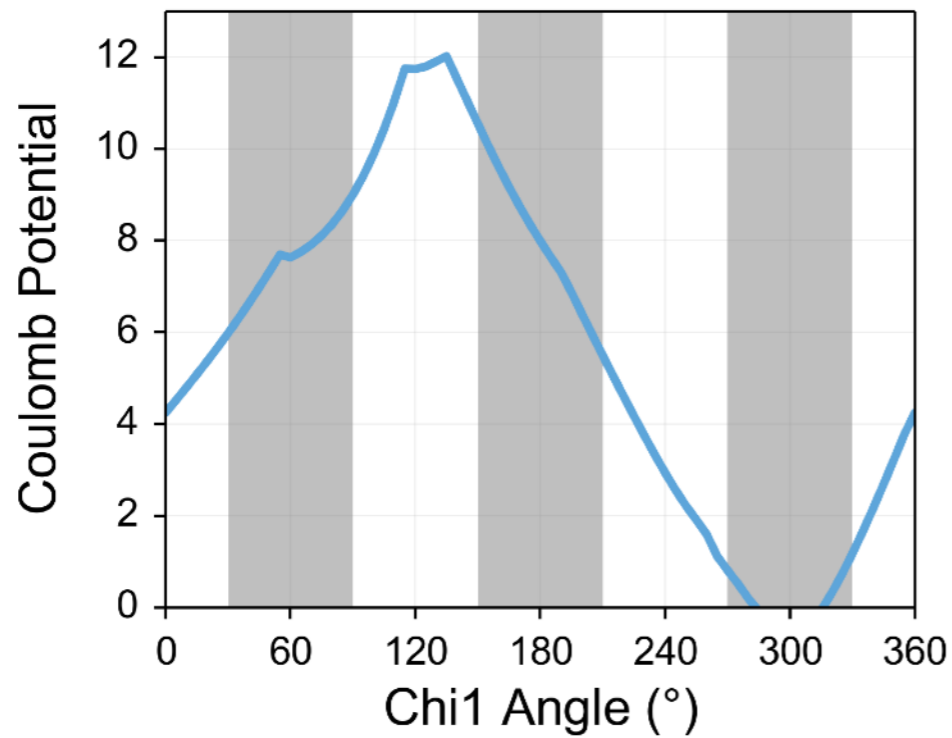
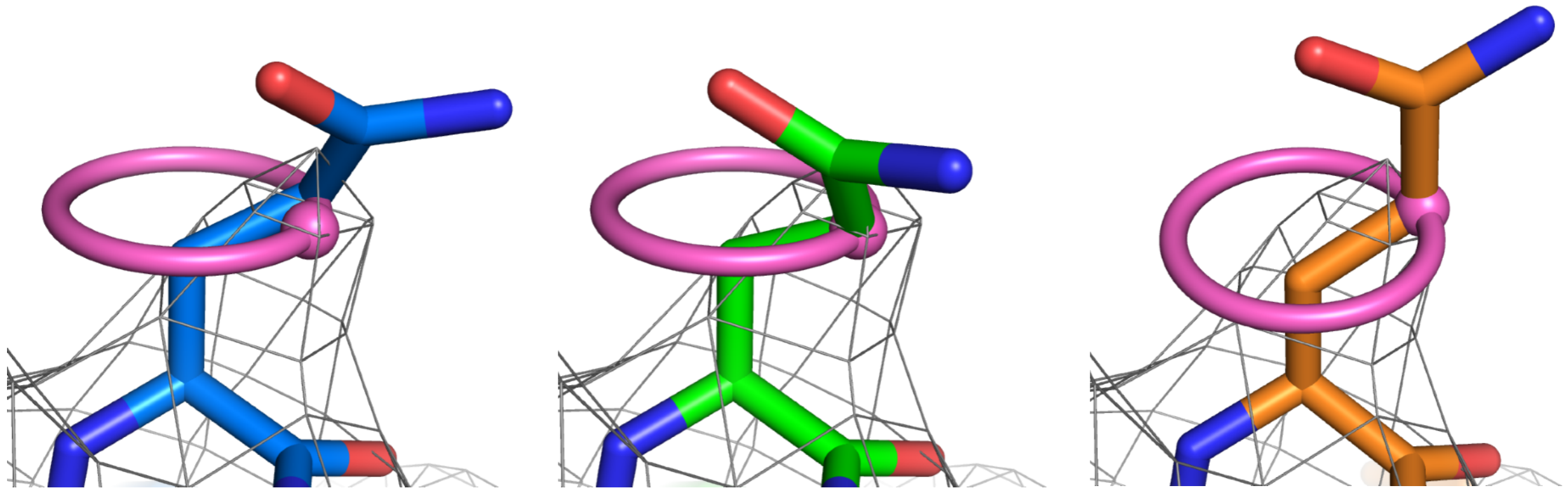
**Phenix**

Ben Barad, Jaime Fraser, UCSF



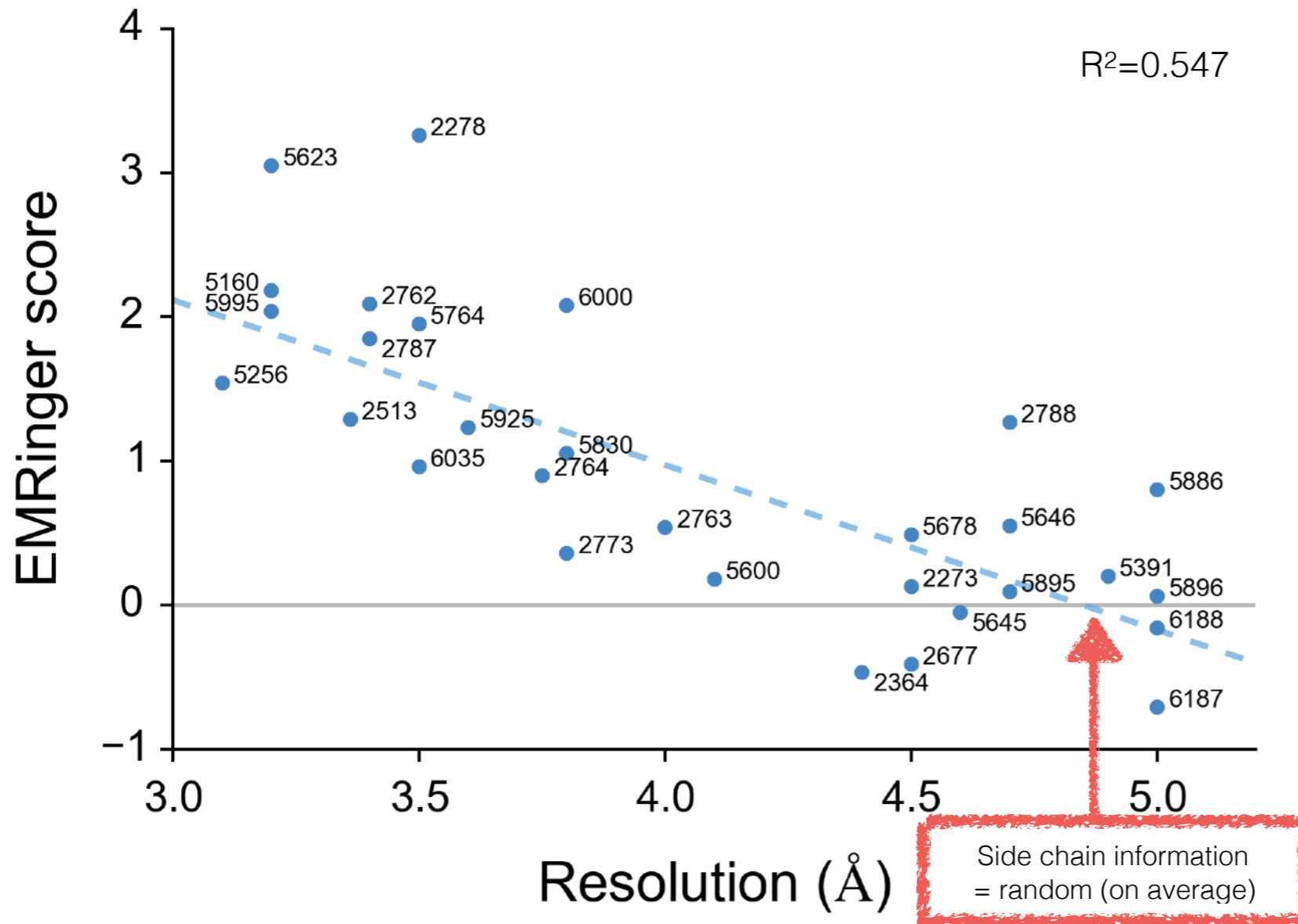


# EMRinger reports on backbone placement



# EMRinger Score to Validate Model vs Data

- Quantify how well the model backbone puts side chains in places where there are density peaks consistent with rotameric conformations



<http://emringer.com>

**Phenix**

- Available in GUI and command line
- `phenix.emringer`  
`model.pdb`  
`map.ccp4`

UCSF

University of California  
San Francisco

**Phenix**

Ben Barad, Jaime Fraser, UCSF



# Summary

- Automated model building is possible, but can be improved
  - Secondary structure prediction, evolution etc.
  - Introduction of structure-modeling tools (e.g. Rosetta) in model-building
- The application of prior or complementary information improves refinement of Cryo-EM structures in real space
- Local analysis of side chain positions w.r.t. density provides a metric for model quality
- Many challenges remain:
  - Local variation in resolution leads to uncertainties in interpretation
  - Efficiently accounting for atomic displacements in models
  - Additional validation metrics for the model w.r.t. the data are needed
  - Reliably accounting for uncertainty in magnification

# Acknowledgments

## ● Lawrence 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 National Laboratory

- Tom Terwilliger, Li-Wei Hung

## ● Cambridge University

- Randy Read, Airlie McCoy, Laurent Storoni, Gabor Bunkoczi, Robert Oeffner

## ● Duke University

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

## ● EMRinger

- Ben Barad, Yifan Cheng, Jaime Fraser (UCSF)
- Ray Wang, Frank DiMaio (UW)

## ● Rosetta

- Frank DiMaio, David Baker (U Washington)

## ● Oak Ridge National Laboratory

- Marat Mustyakimov, Paul Langan

## ● Cryo-EM Structure Refinement

- Corey Hryc, Zhao Wang, Matthew Baker, Wah Chiu (Baylor College of Medicine)
- Adam Frost (UC San Francisco), Gino Cingolani (Jefferson University, PA), Valerie Pye (Cancer Research UK), Darcie Miller (St. Jude Research Hospital), Anastasia Aksyuk (NIH), Mavis Agbandje-McKenna (University of Florida), Stuart Howes (Eva Nogales Lab), Ali Andres Malay (RIKEN, Japan)

## ● Other Collaborators

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

## ● Funding:

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

# Thank You

