New *Phenix* Tools for Cryo-EM

Pavel Afonine

***Phenix* software developer**

LBNL, Berkeley, California, USA

UTMB, Galveston, Texas, May 9th, 2019

Phenix

Lawrence Berkeley Laboratory

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

Los Alamos National Laboratory

Tom Terwilliger, Li-Wei Hung

Cambridge University

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

Duke University

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

www.phenix-online.org

Macromolecular Neutron Consortium

Paul Langan
Brendan Sullivan
Vickie Lynch
(*Marat Mustyakimov*)

Paul Adams
Pavel Afonine

https://mnc.ornl.gov/

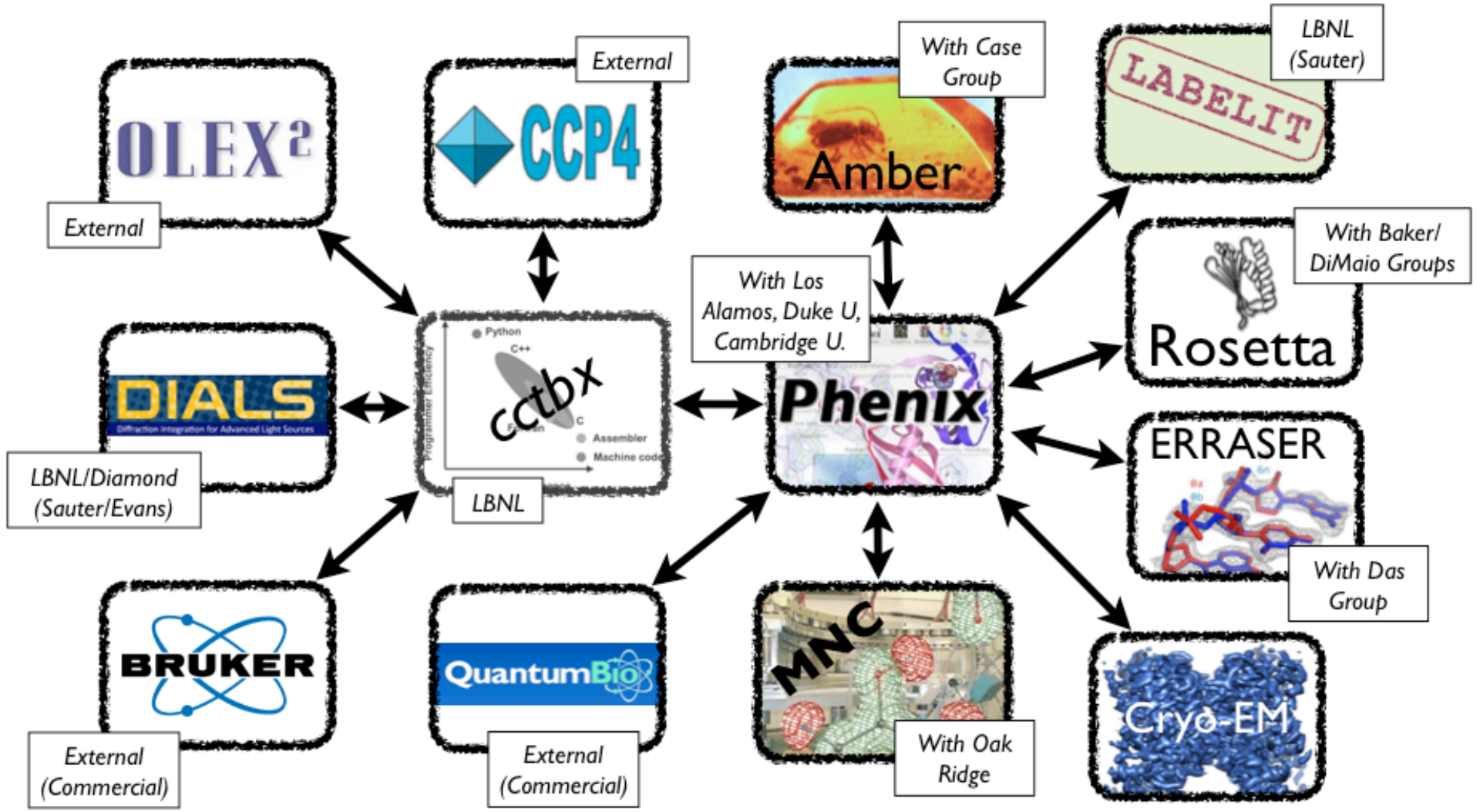
Quantum Refinement

Shanghai University

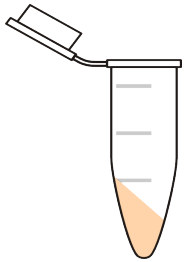
Mark Waller
Pavel Afonine
Nigel Moriarty
Min Zheng
& associates

www.qrefine.com

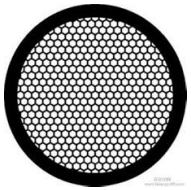
Phenix - a Structural Biology Hub



Cryo-EM tools in *Phenix*



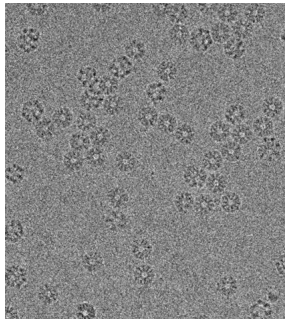
Sample



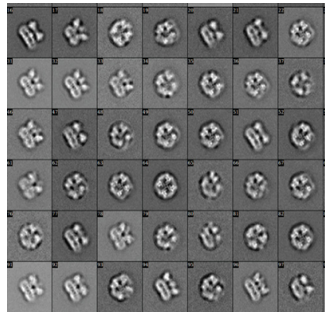
Grid

Freeze

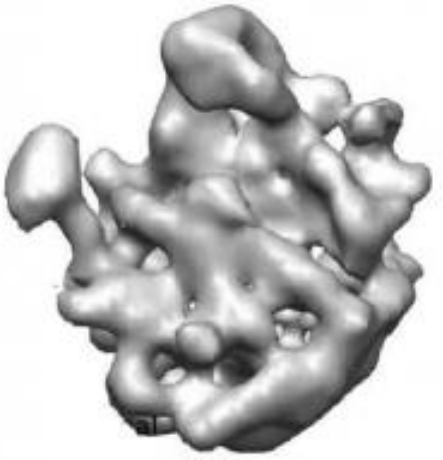
Electron microscope



2D projections

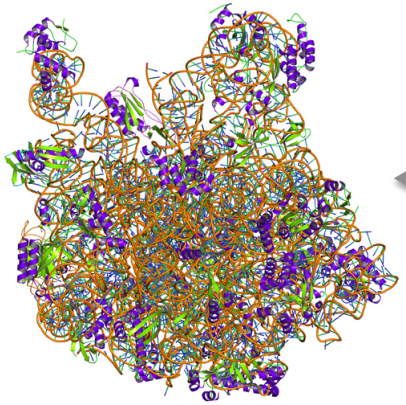


Pick particles, align and average



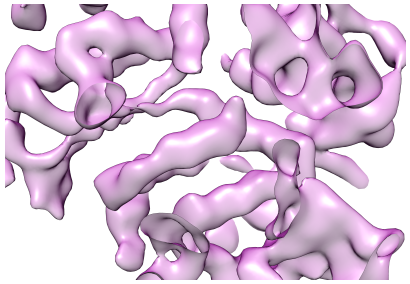
3D map

Model building, refinement, validation

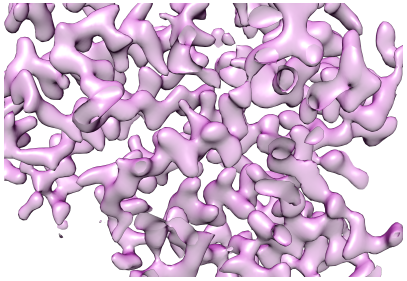


Atomic model

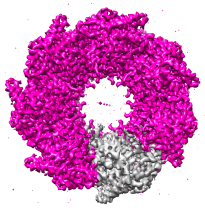
Cryo-EM tools in *Phenix*



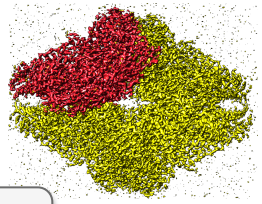
Starting map



Map improvement



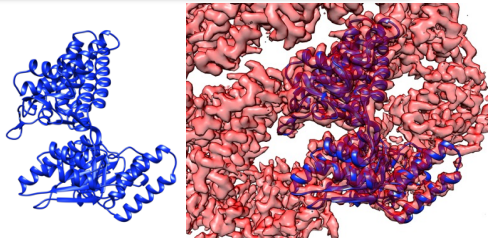
Map symmetry



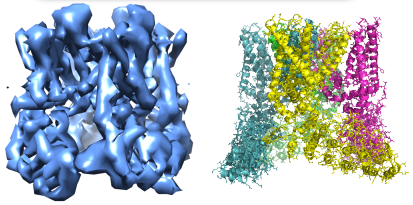
Map manipulation

Extract unique part

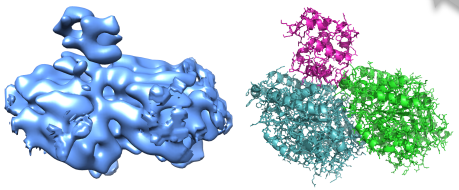
Docking, model building



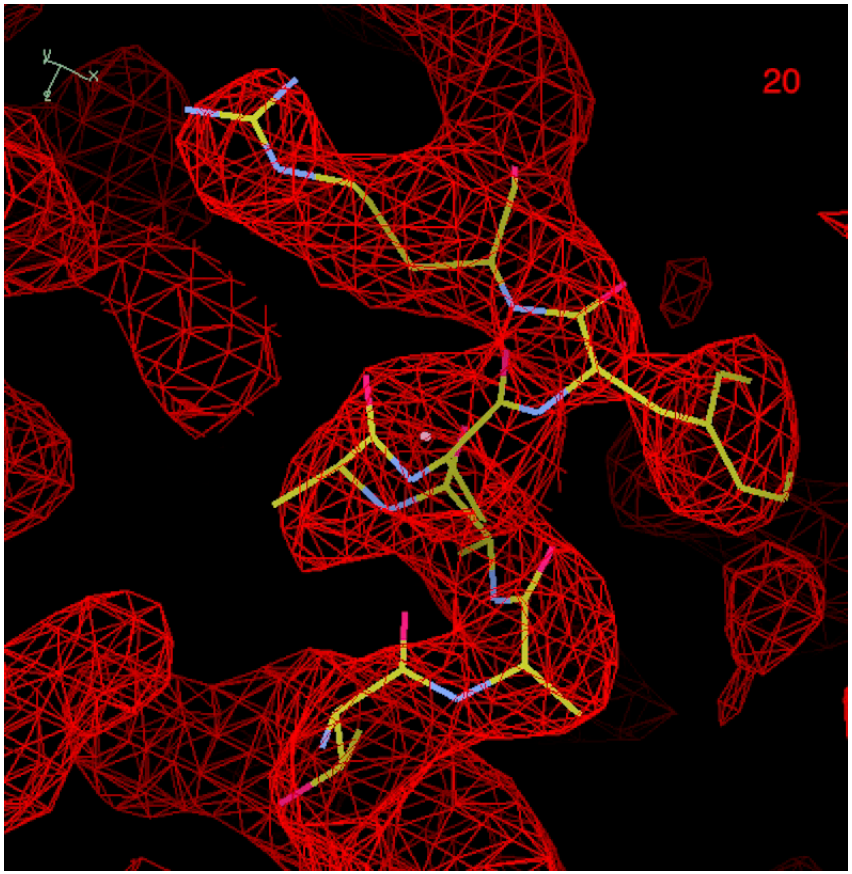
Refinement



Validation



Automatic map sharpening: *phenix.auto_sharpen*



Maximize detail in the map

... and connectivity of map

Adjusted surface area

Optimally sharpened map

Fully automatic:

- No manual trial-and-error
- No parameters to adjust
- Only inputs: map and resolution

Acta Cryst D STRUCTURAL BIOLOGY
ISSN 2059-7983

Automated map sharpening by maximization of detail and connectivity

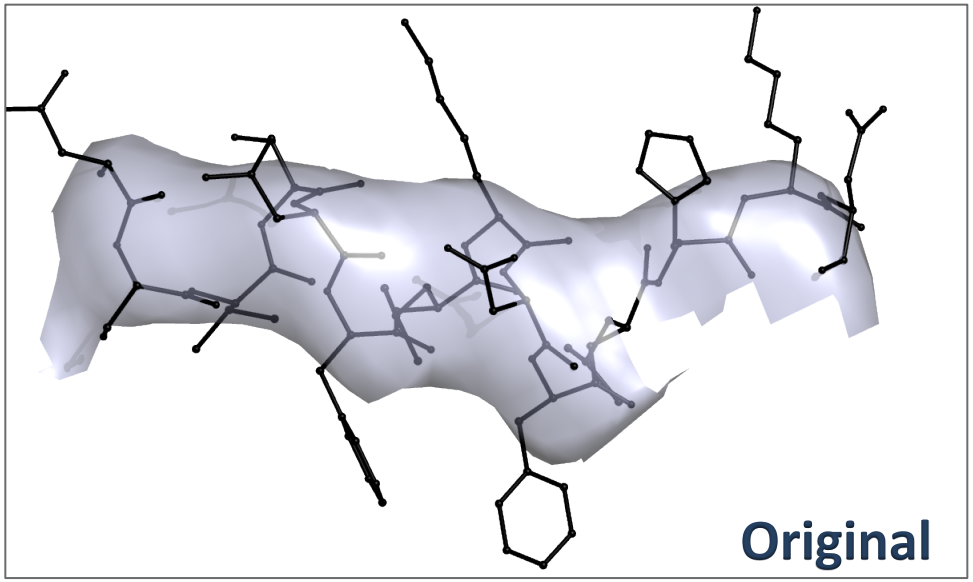
Thomas C. Terwilliger,[a,b]* Oleg V. Sobolev,[c] Pavel V. Afonine[c,d] and Paul D. Adams[d,e]

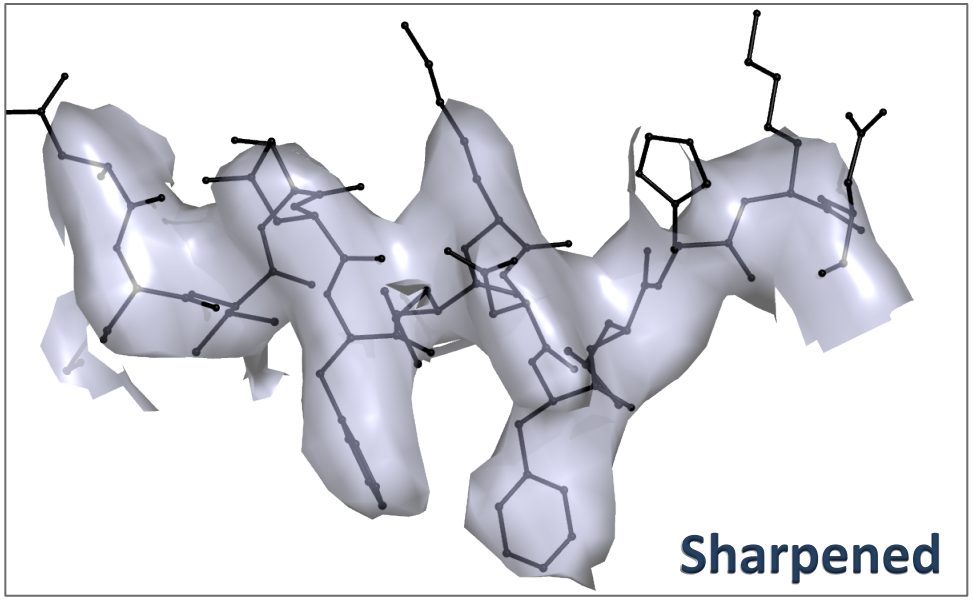
Automatic map sharpening: *phenix.auto_sharpen*

PDB / EMDB:

5tji / 8414

Resolution:
3.8 Å



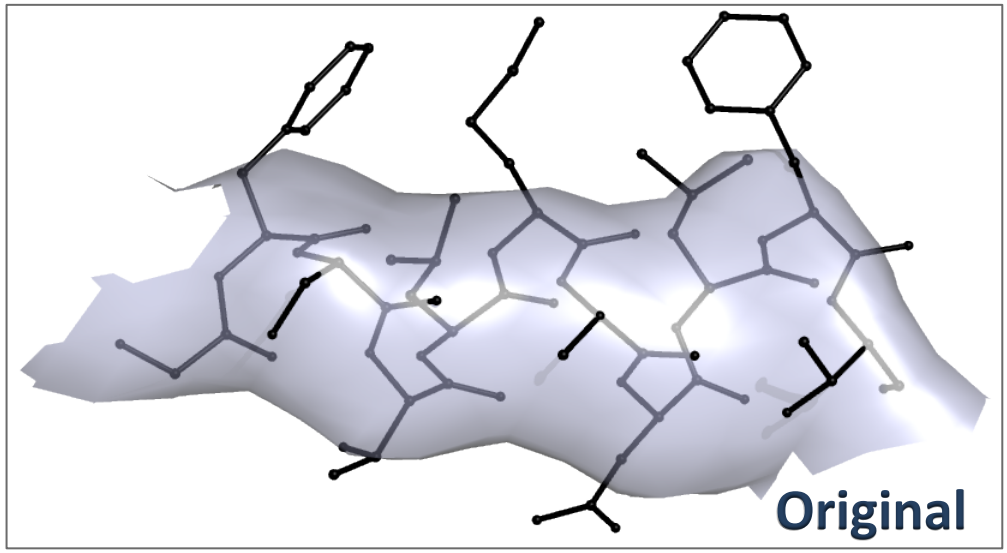


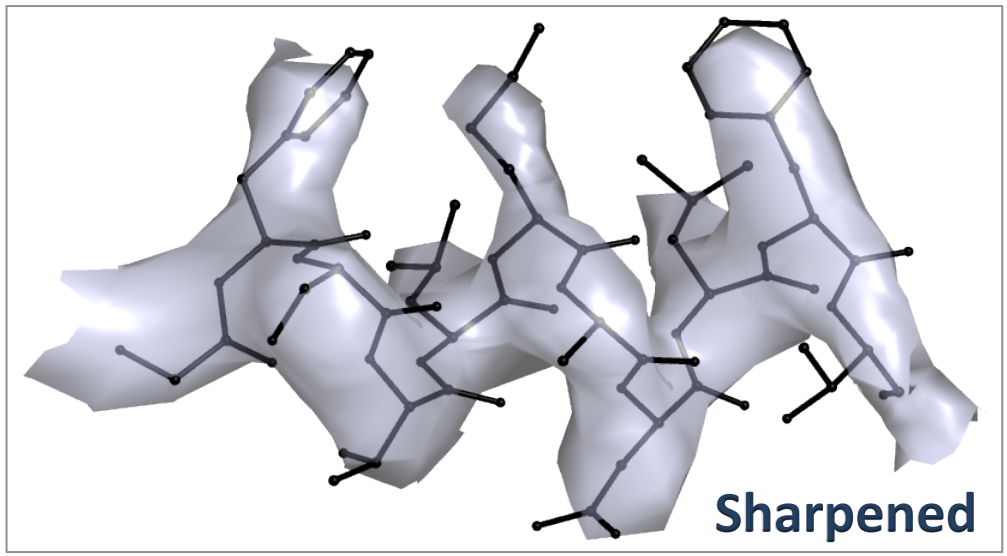
Automatic map sharpening: *phenix.auto_sharpen*

PDB / EMDB:

5tji / 8414

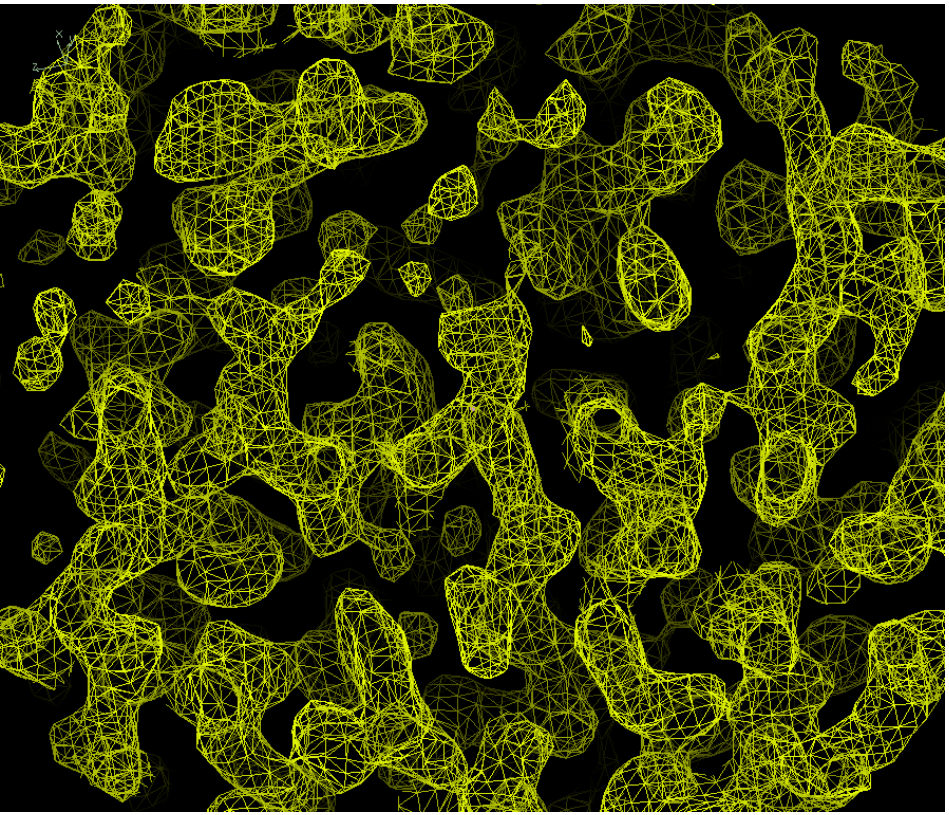
Resolution: 3.8 Å

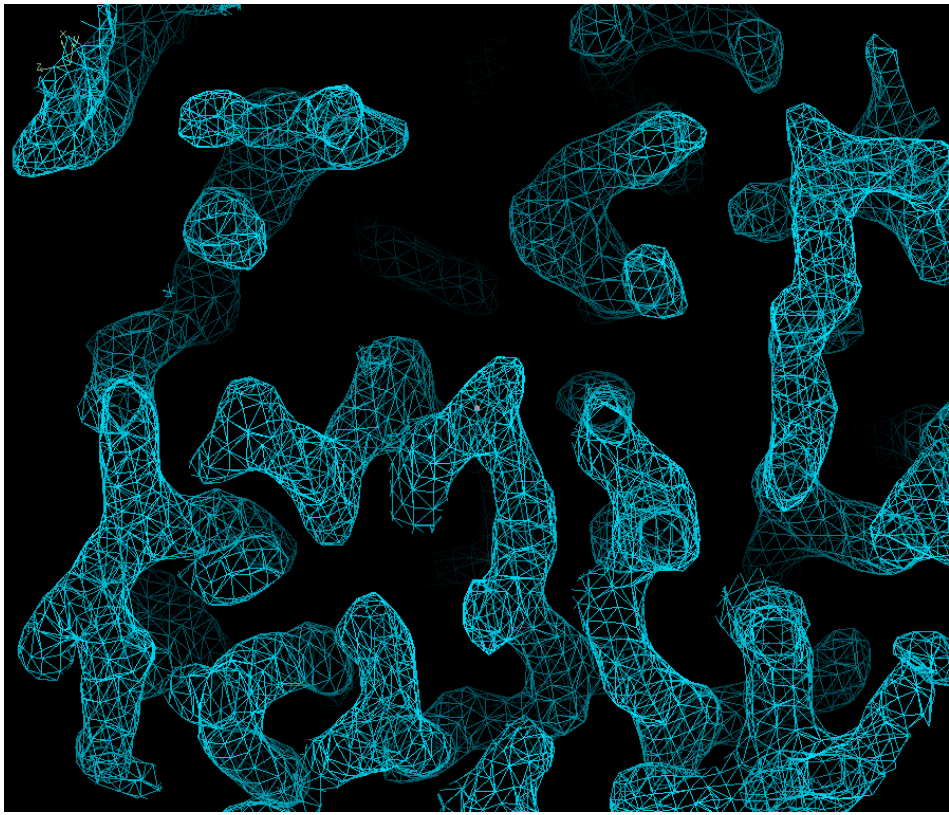




X-ray vs cryo-EM

Beta galactosidase at 2.2 Å

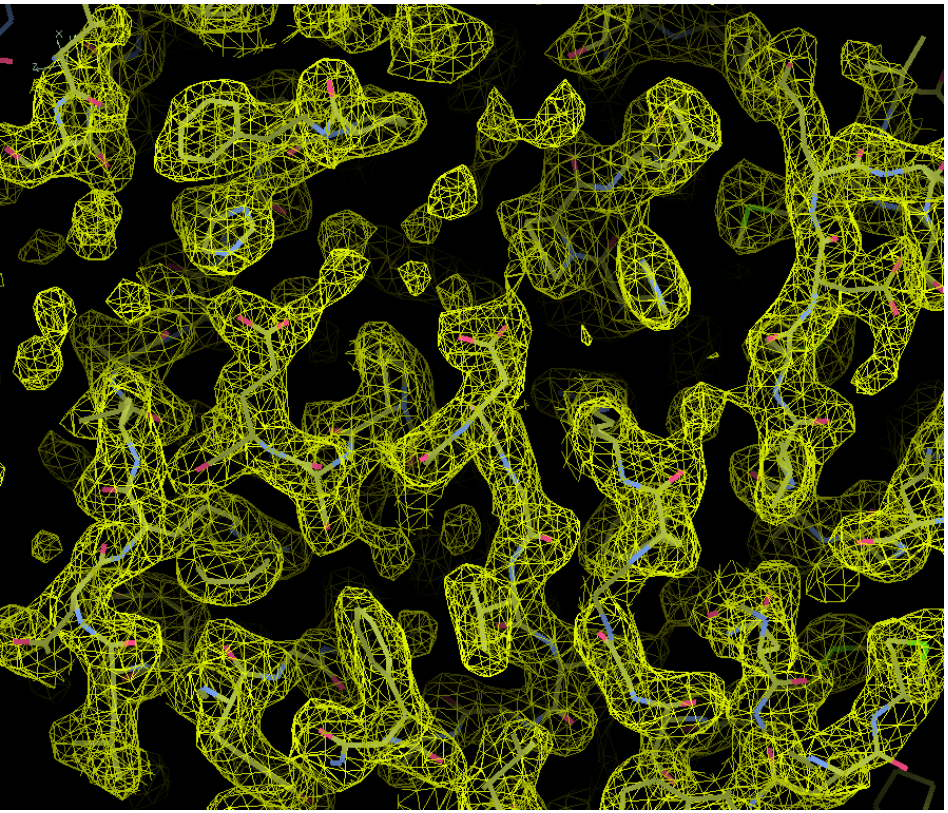




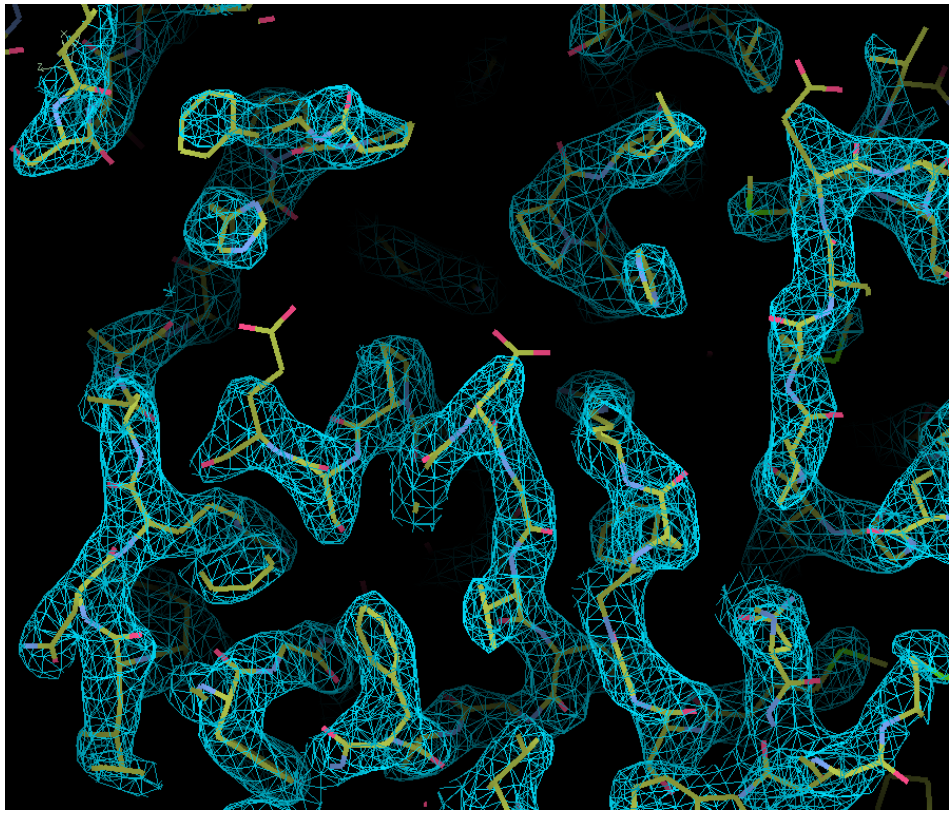
(which is the cryo-EM map?)

X-ray vs cryo-EM

Beta galactosidase at 2.2 Å



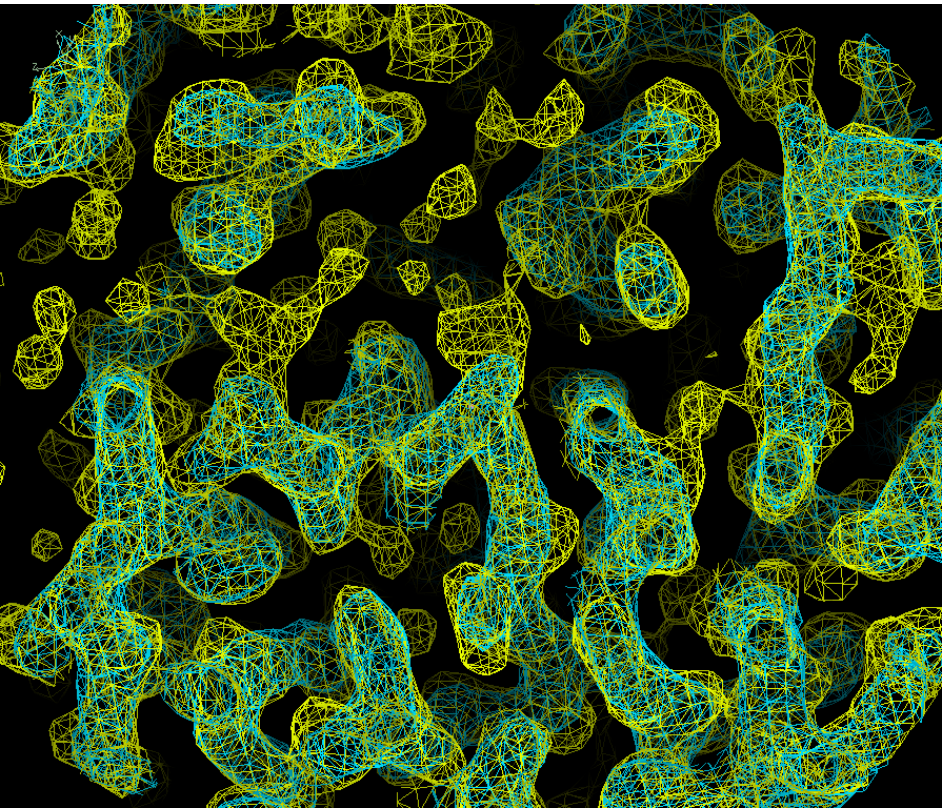
X-ray (PDB 3i3b)

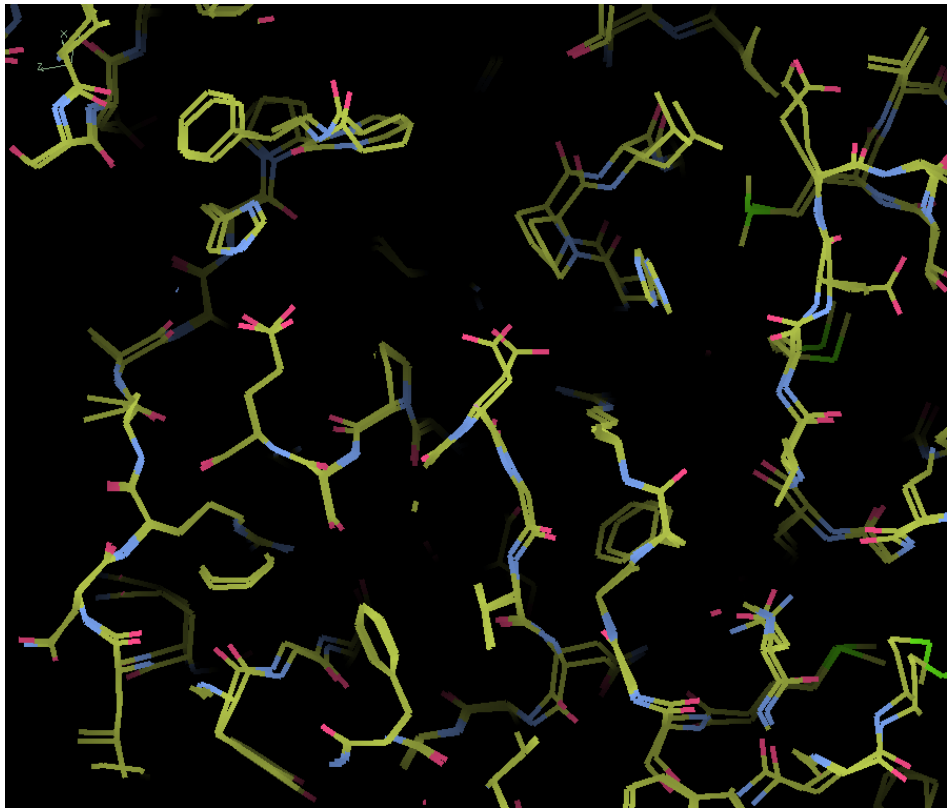


Cryo-EM (PDB 5a1a)

X-ray vs cryo-EM

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



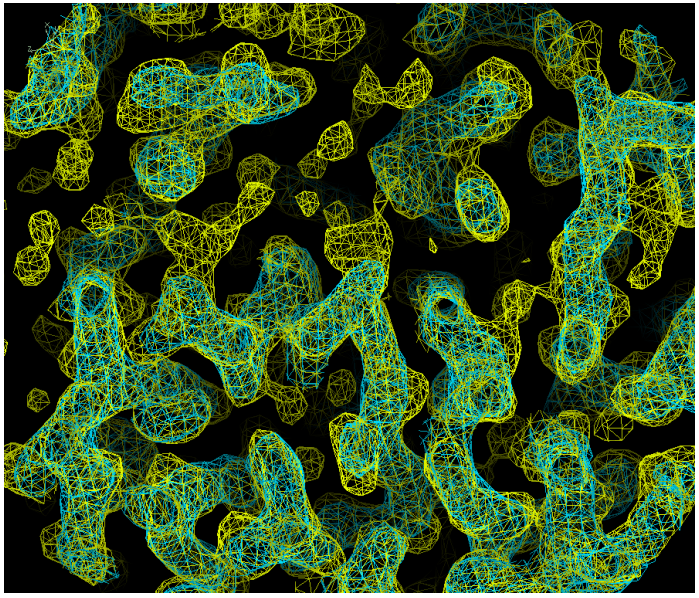


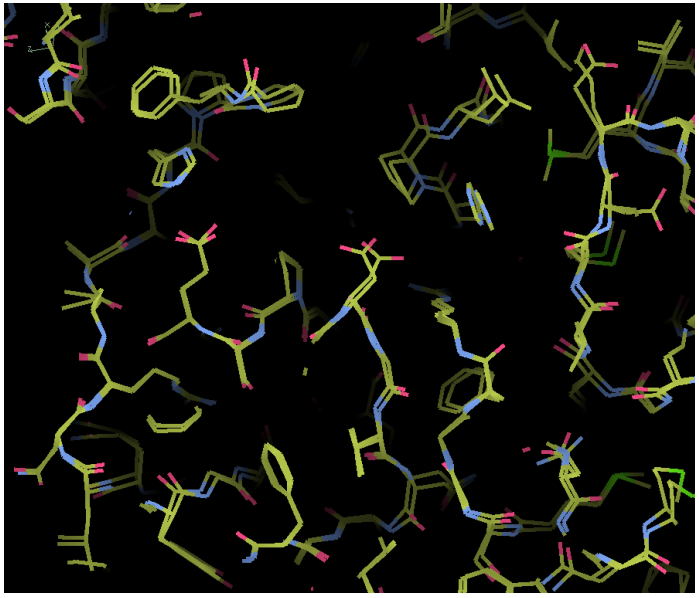
X-ray vs cryo-EM

…but have different limitations

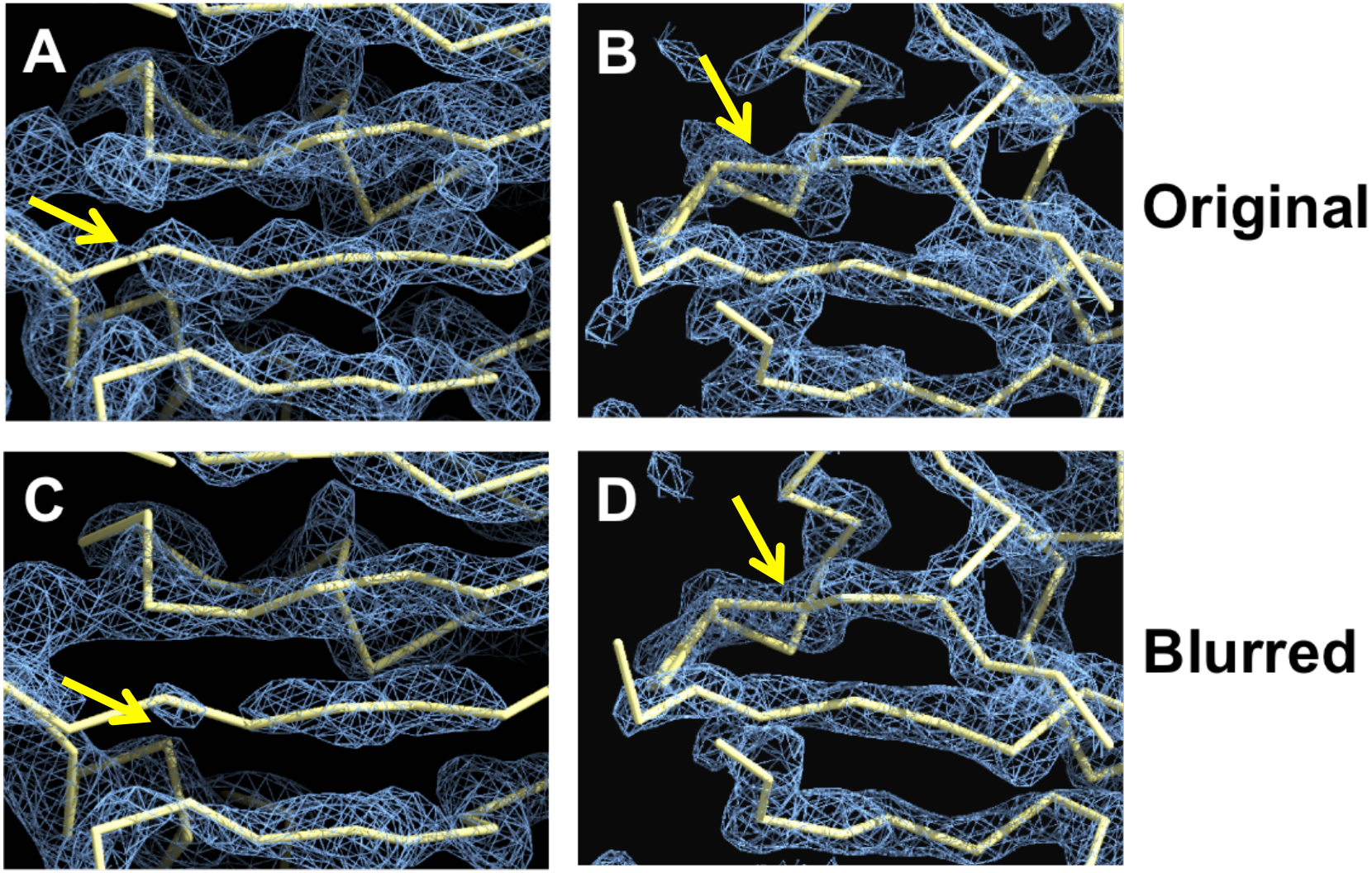
Cryo-EM maps cannot (yet) be improved by density modification—what you start with is what you get

X-ray maps have less accurate low-resolution information

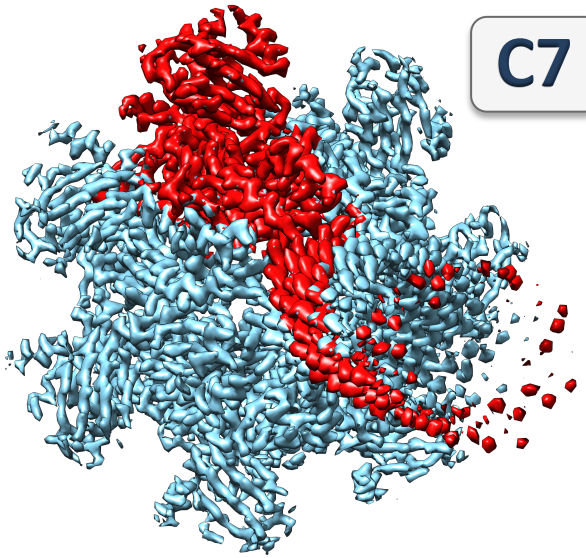


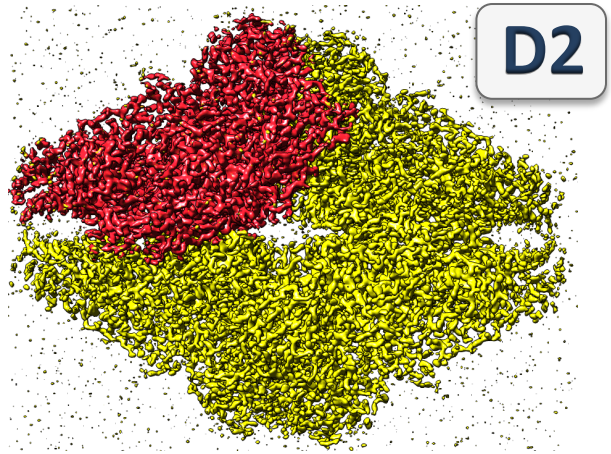


X-ray vs cryo-EM



Finding map symmetry: *phenix.symmetry_from_map*





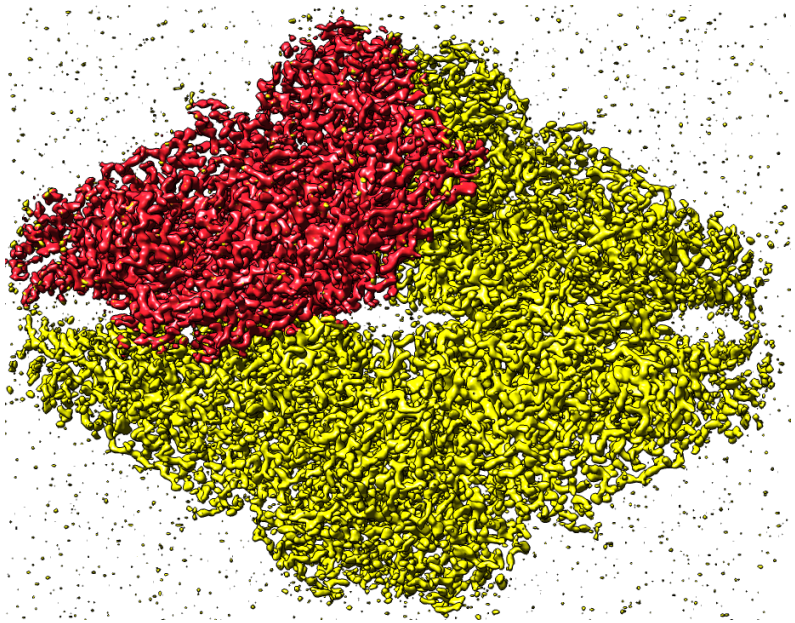
Procedure for finding symmetry:

- Test point group symmetries (e.g., C7, D2, I, O, T)
- Helical symmetry
- Principal rotation axes along z, x, y
- Score based on map correlation for symmetry-related points and number of operators

http://phenix-online.org/newsletter/

Tools for interpreting cryo-EM maps using models from the PDB

Extracting unique part of map: *phenix.map_box*



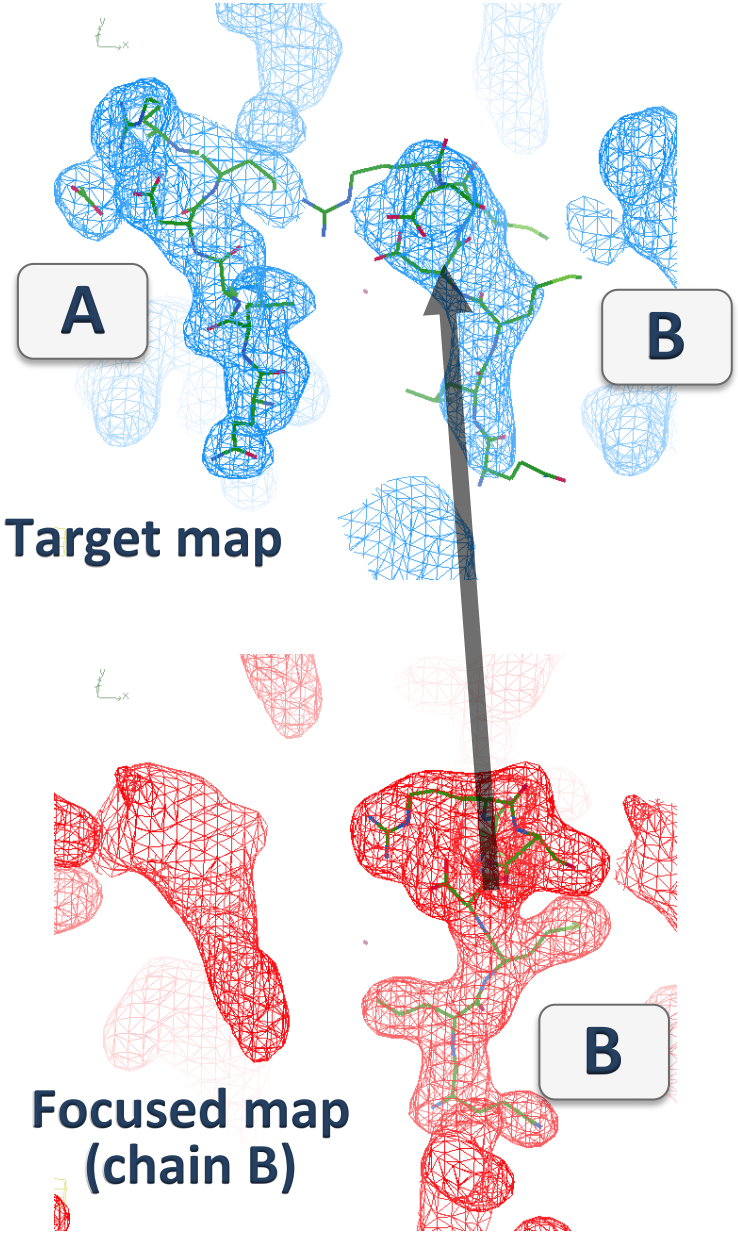
Procedure:

- Use symmetry of map
- Contour map at level that yields regions about 50 residues in size
- Group symmetry-related regions
- Choose one member of each group
- Optimize compactness and connectivity of unique part of map

http://phenix-online.org/newsletter/

Tools for interpreting cryo-EM maps using models from the PDB

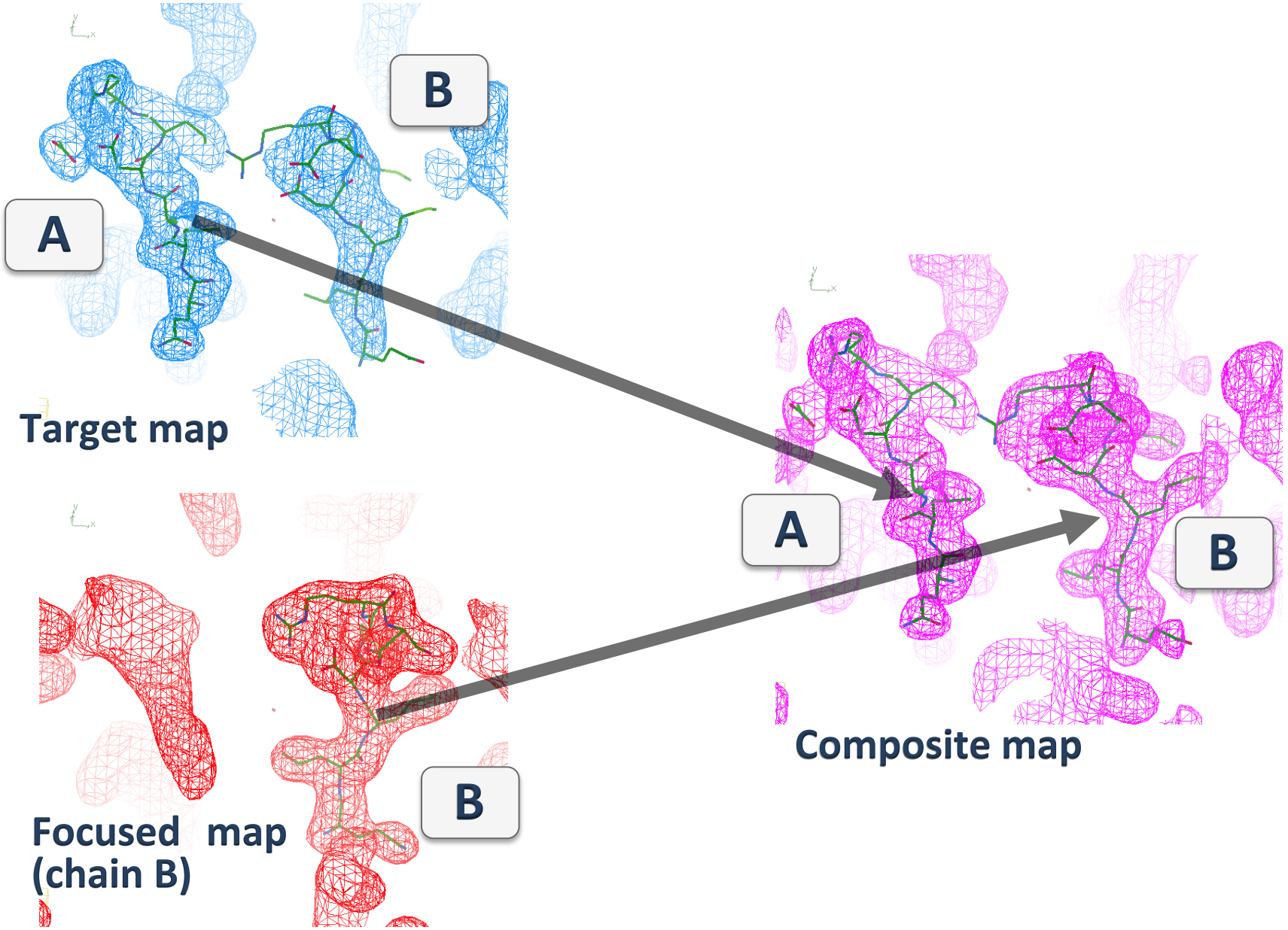
Combining maps with *phenix.combine_focused_maps*



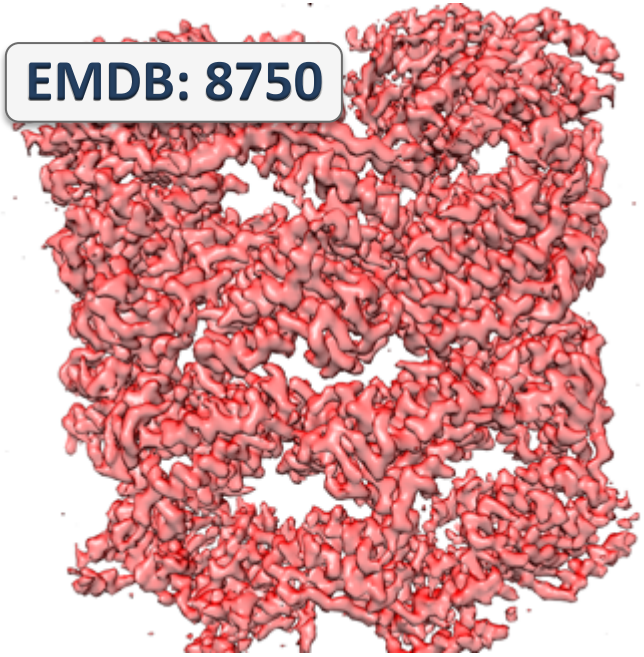
Procedure for combining maps:

- Superpose density
- Rotation/translation from refined models
- Average target and focused map density
- Weight using map-model correlations

Combining maps with *phenix.combine_focused_maps*



Docking models with *phenix.dock_in_map*



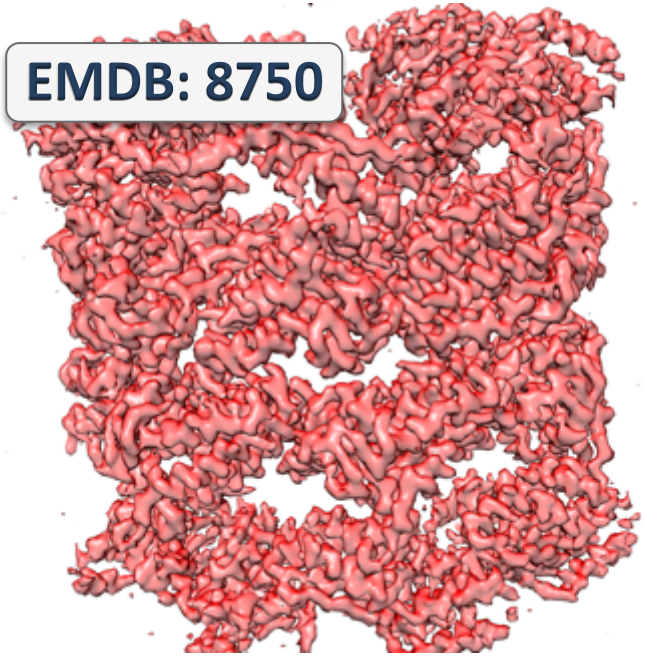
Search procedure:

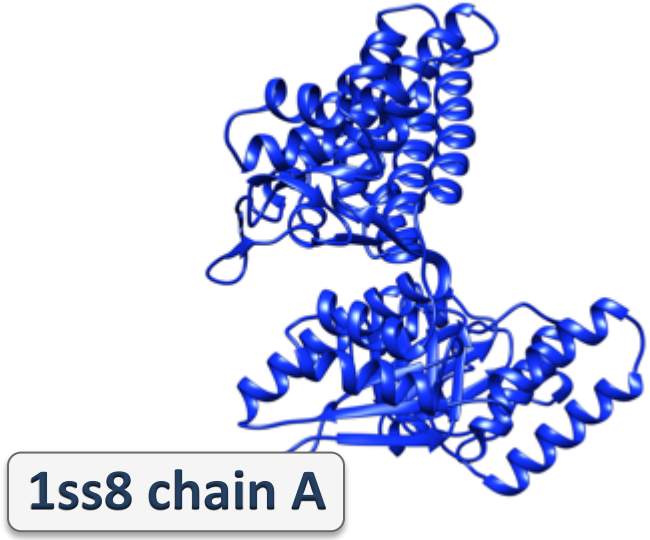
- **Pure translation:**
  - **low-res**
  - **high-res**
- **Rotation/translation:**
  - **low-res**
  - **high-res**

Score based on rigid-body refinement map-model correlation

http://phenix-online.org/newsletter/
Tools for interpreting cryo-EM maps using models from the PDB

Docking models with *phenix.dock_in_map*

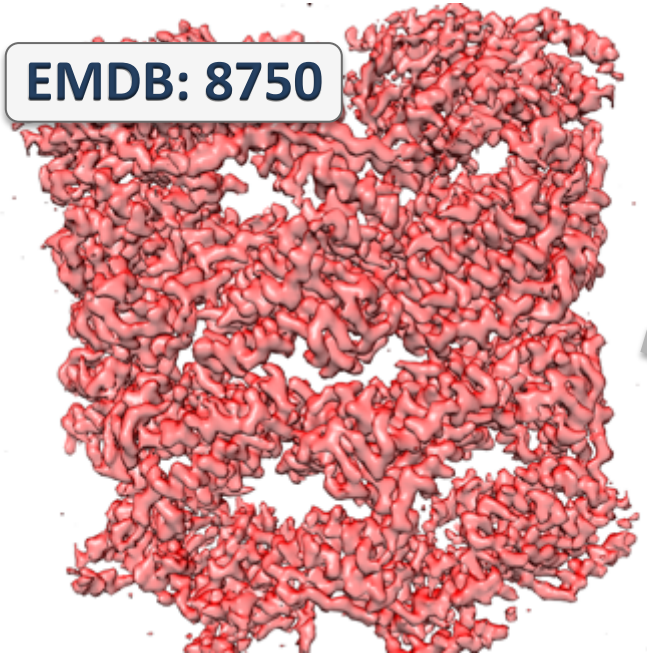


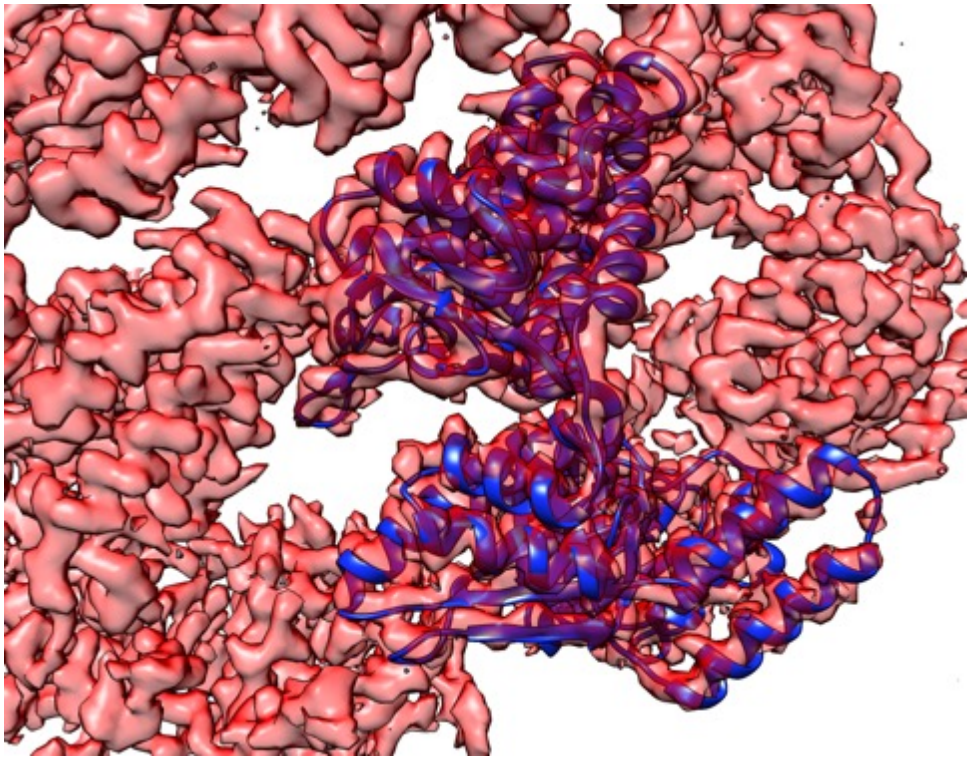


Features:

- Multiple chains
- Density search
- Symmetry
- Multiprocessing

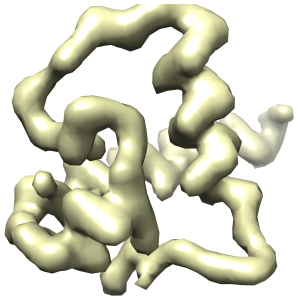
Docking models with *phenix.dock_in_map*



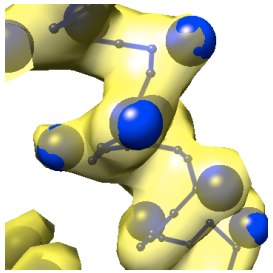


Automated model building: *phenix.map_to_model*

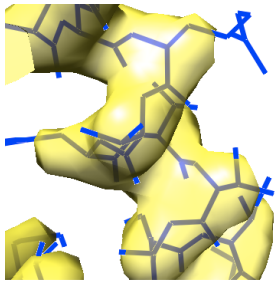
Isolate density for a chain



Identify Cα and Cβ positions from side-chain density



Construct and refine all-atom model



nature methods

BRIEF COMMUNICATION

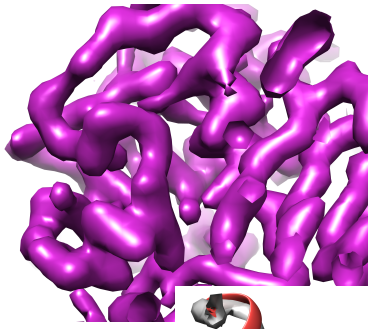
https://doi.org/10.1038/s41592-018-0173-1

A fully automatic method yielding initial models from high-resolution cryo-electron microscopy maps

Thomas C. Terwilliger[1,2*], Paul D. Adams[3,4], Pavel V. Afonine[3,5] and Oleg V. Sobolev[3]

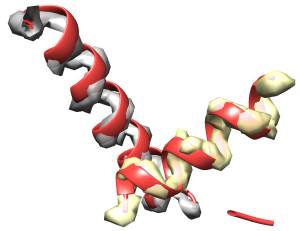
Automated model building: *phenix.map_to_model*

Trace chain the way a person does

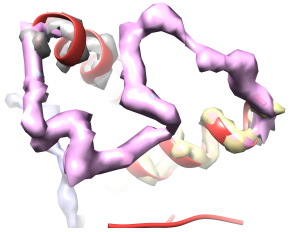


Find secondary structure

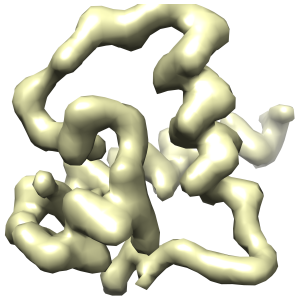
Find clear regions of density



Adjust contour level until a region just connects to one other



Iterate to build up chain



Automated model building: *phenix.map_to_model*

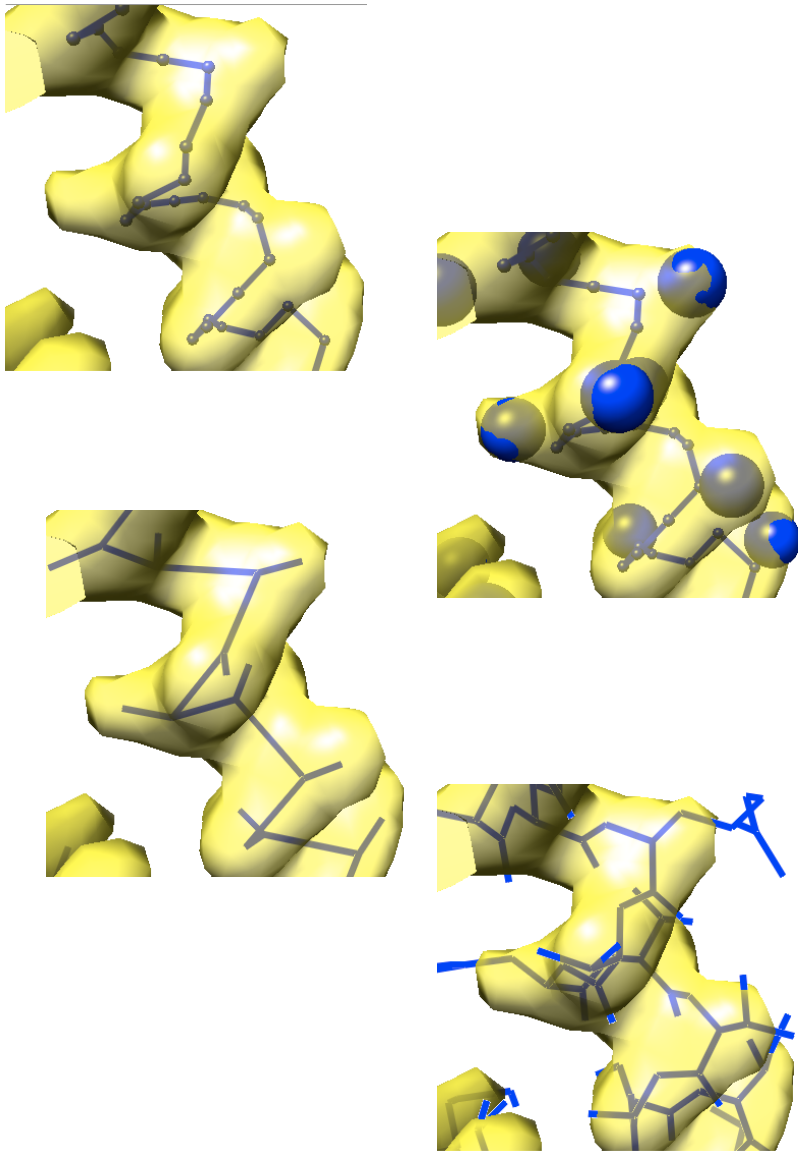
Finding Cα and Cβ positions

Trace chain path through high density

Find Cβ positions from side-chain density

Choose Cα positions 3.8 Å apart and next to Cβ positions

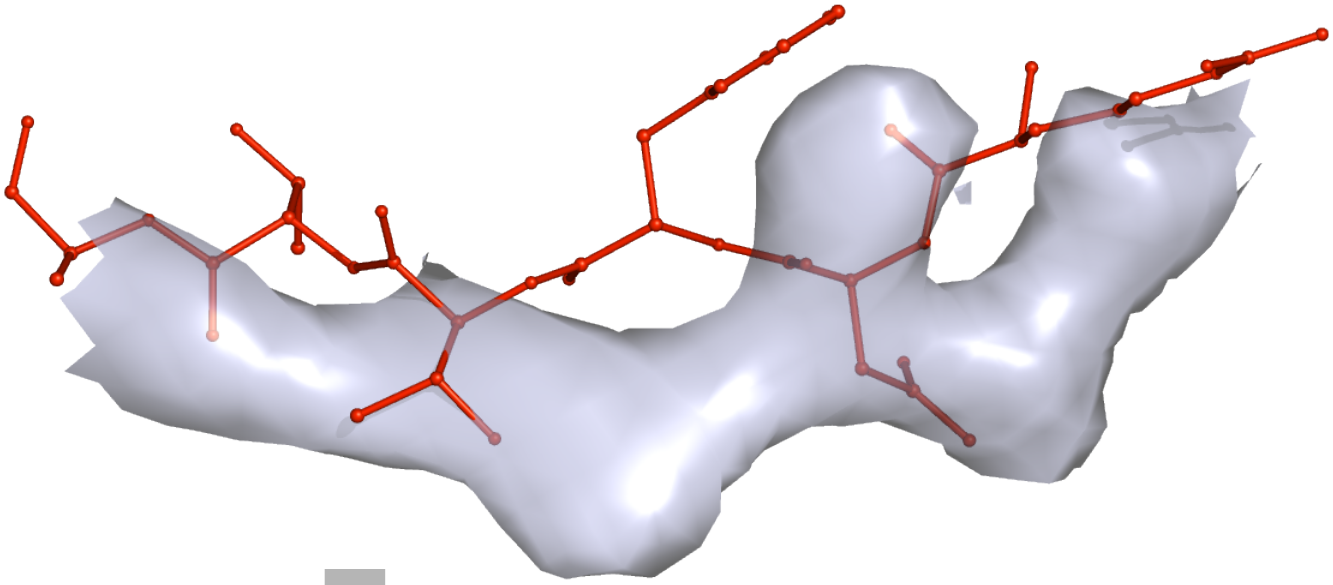
Construct all-atom model with Pulchra* and refine



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

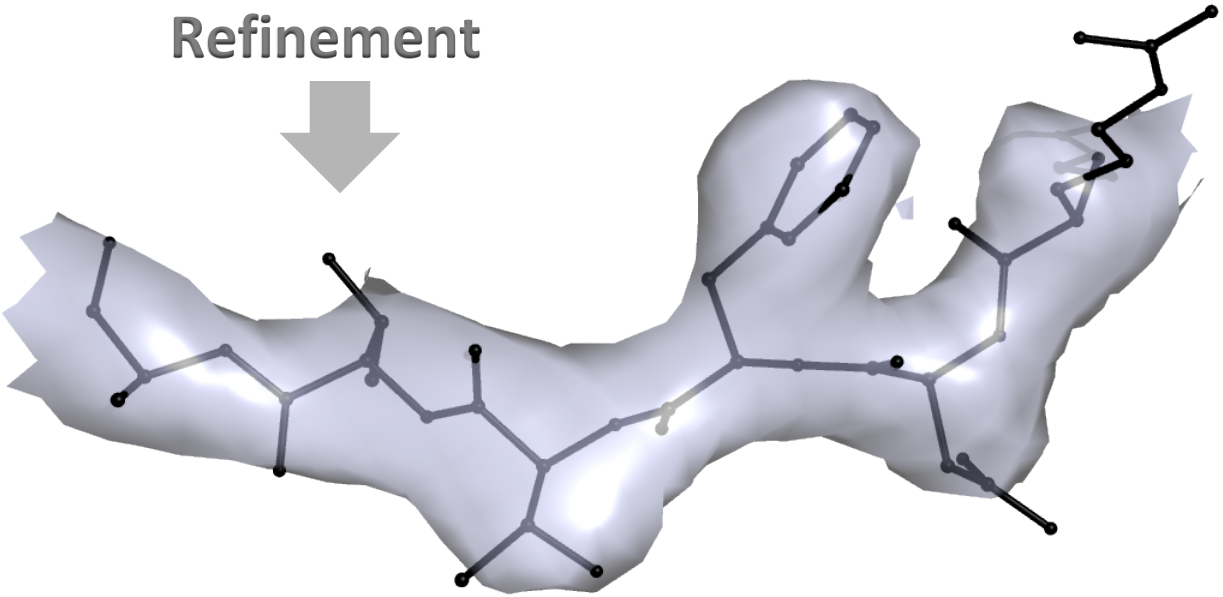
Structure refinement at a glance

Initial (poor) model



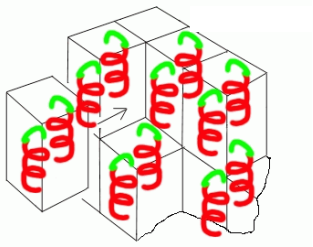
Refinement

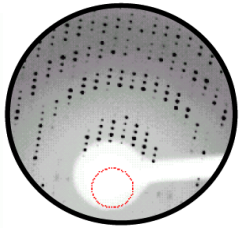
Improved (refined) model



Automated model refinement: *phenix.real_space_refine*

Crystallography





Initial model

Experimental data

A priori knowledge

Score

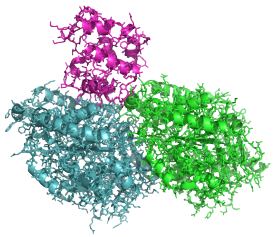
Modify model parameters

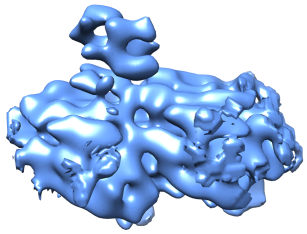
Improved model

phenix.refine

Available since 2005

Cryo-EM





Initial model

Experimental data

A priori knowledge

Score

Modify model parameters

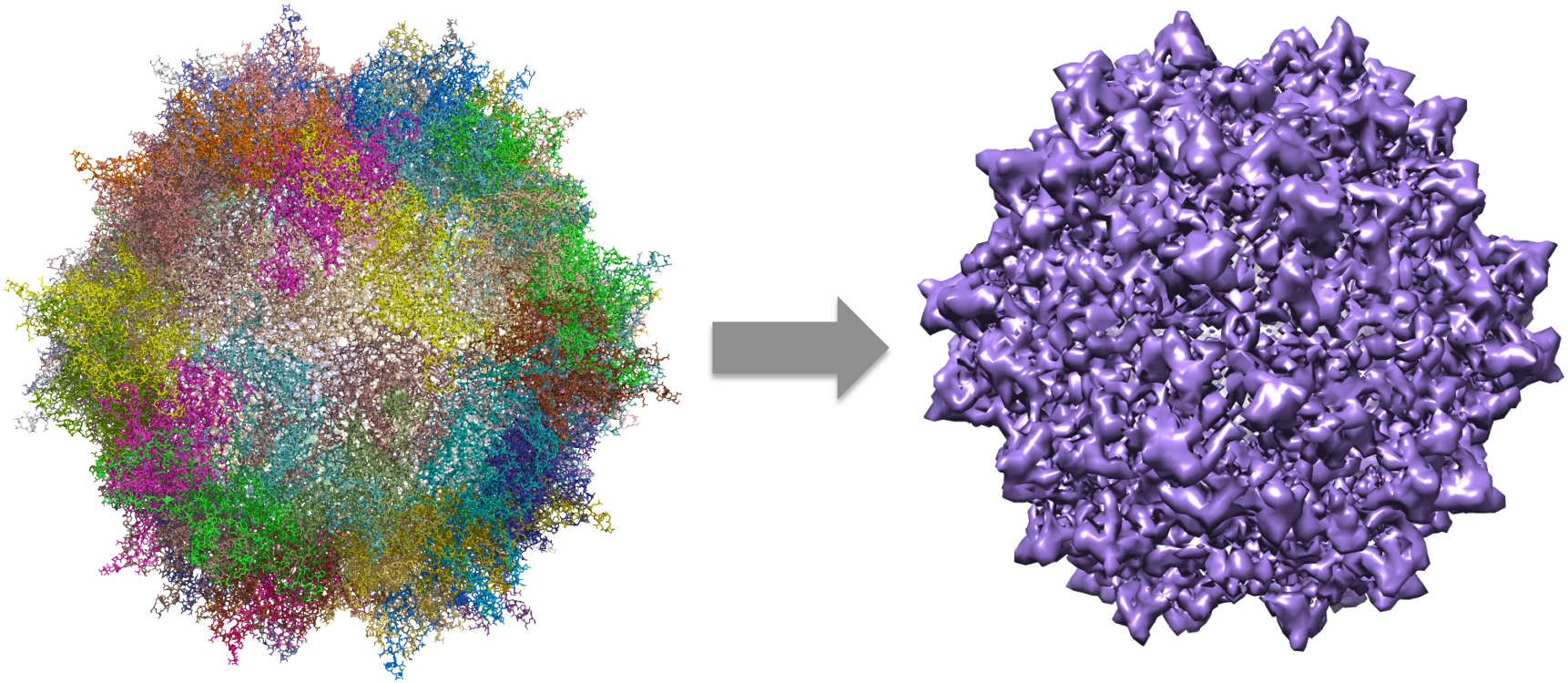
Improved model

phenix.real_space_refine

Available since 2013

Automated model refinement: *phenix.real_space_refine*

- Direct refinement against the map
  - No Fourier space involved



Acta Cryst D STRUCTURAL BIOLOGY
ISSN 2059-7983

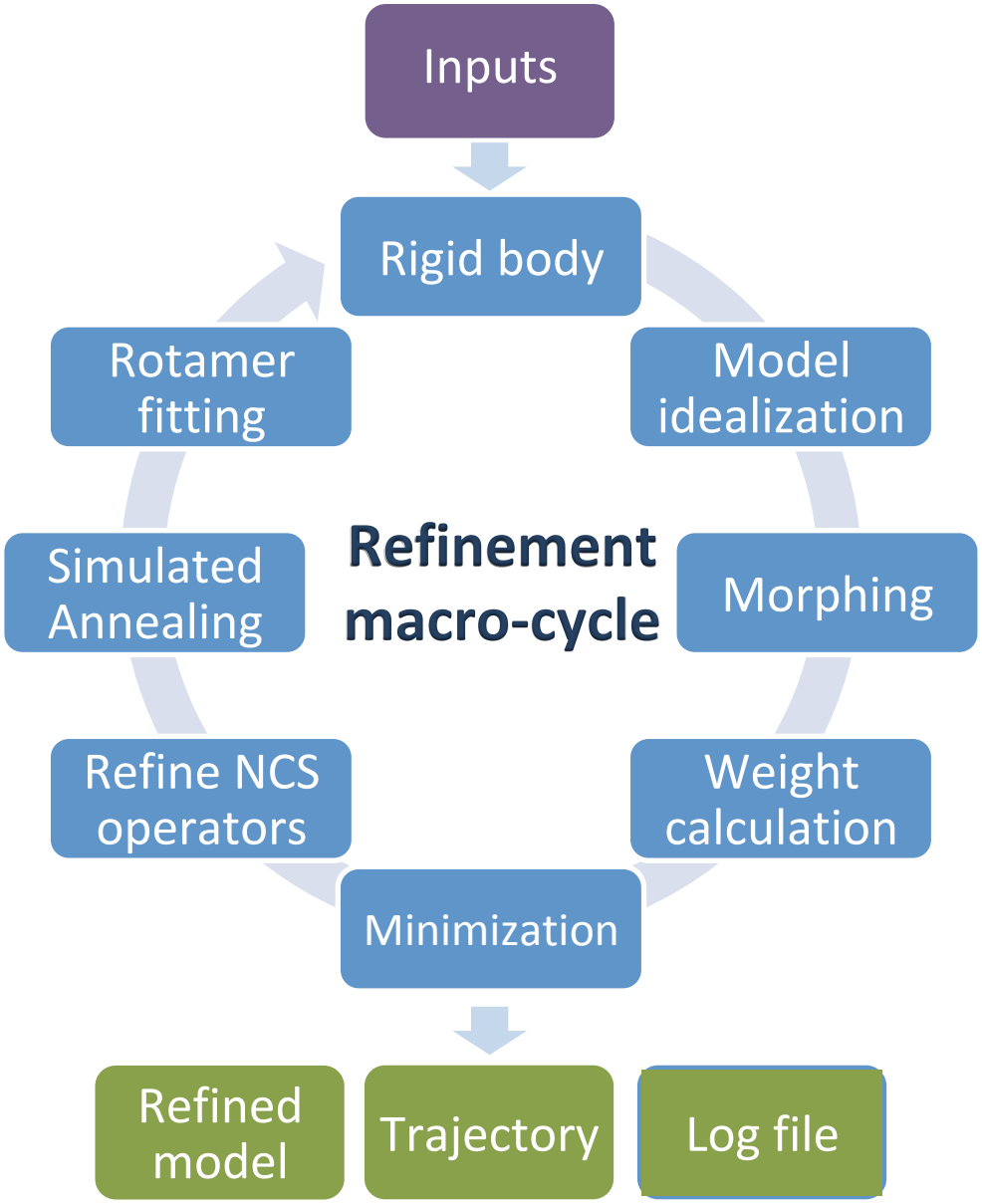
Real-space refinement in *PHENIX* for cryo-EM and crystallography

Pavel V. Afonine,[a,b]* Billy K. Poon,[a] Randy J. Read,[c] Oleg V. Sobolev,[a] Thomas C. Terwilliger,[d,e] Alexandre Urzhumtsev[f,g] and Paul D. Adams[a,h]

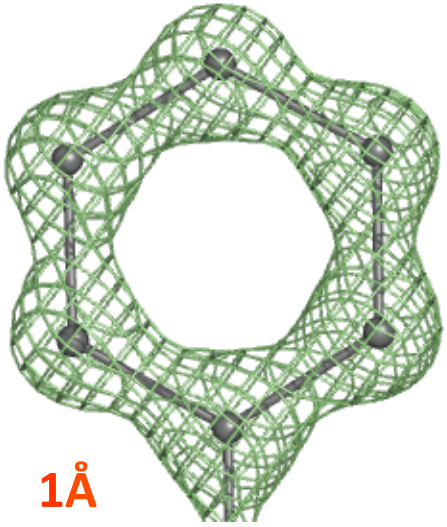
Automated model refinement: *phenix.real_space_refine*

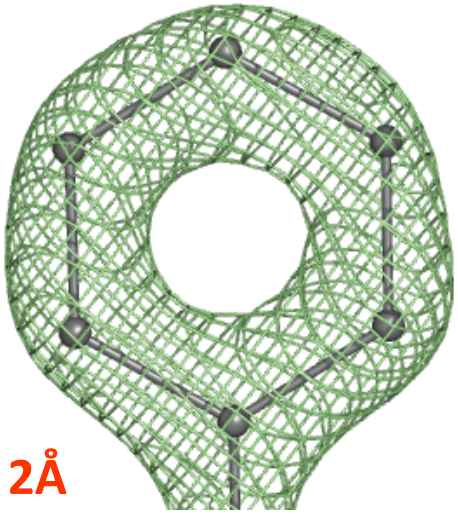
- Best model-map fit. Any map: X-ray, neutron, EM. Any resolution
- Refined models: no poor validation metrics
- Fast (minutes – a few hours, not days or many hours)
  - Make use of multiple CPUs: as many as available
- Large convergence radius
- Easy to use: map and model in, refined model out
- Accessible: no special hardware requirements

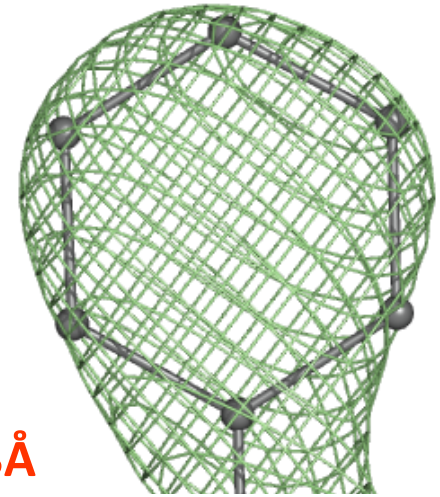
Automated model refinement: *phenix.real_space_refine*



Automated model refinement: *phenix.real_space_refine*







- Lower the resolution, less detailed the map
  - Need extra information to keep correct geometry during refinement

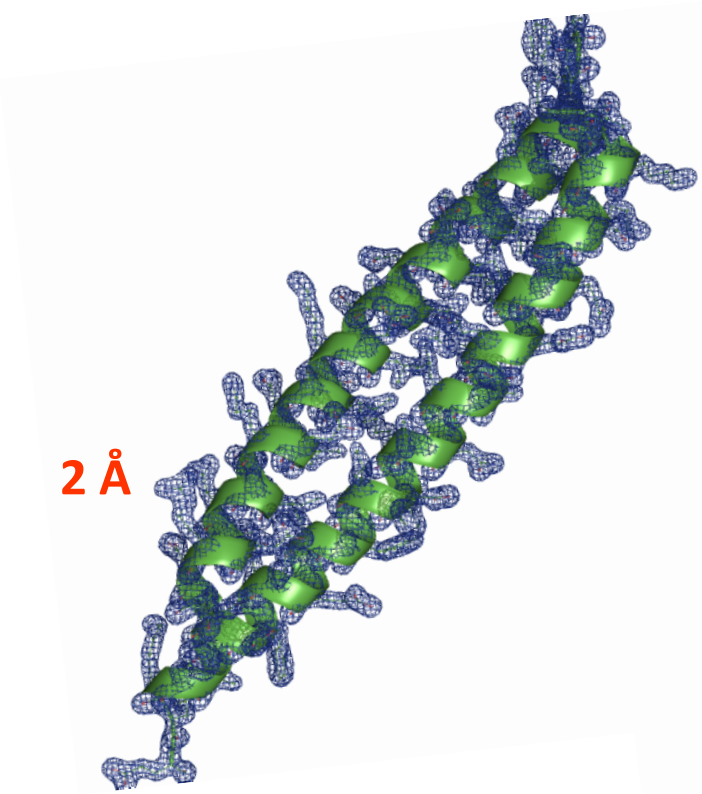
$$T = T_{\mathrm{DATA}} + wT_{\mathrm{RESTRAINTS}}$$

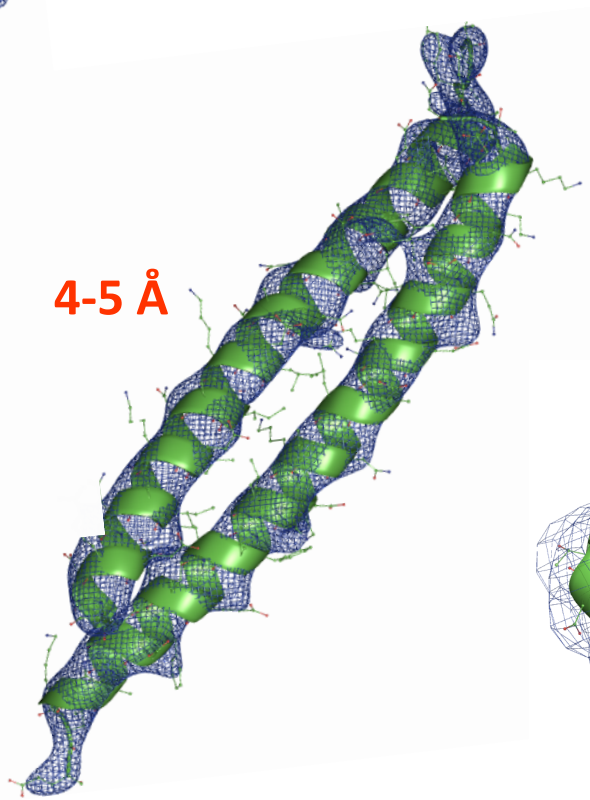
$$T_{RESTRAINTS} = T_{BOND} + T_{ANGLE} + T_{DIHEDRAL} + T_{PLANARITY} + T_{NONBONDED} + T_{CHIRALITY}$$

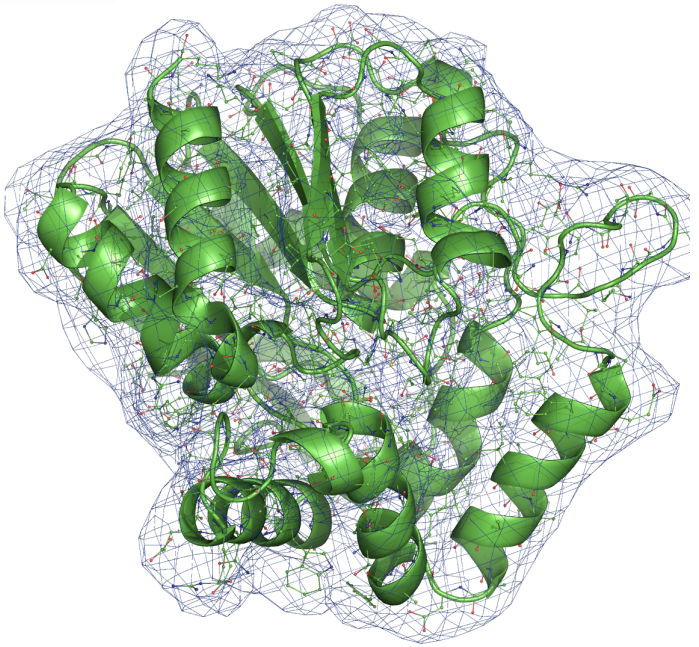
$$T_{BOND} = \Sigma_{\text{all bonded pairs}} w(d_{ideal} - d_{model})^2$$

Automated model refinement: *phenix.real_space_refine*

- Low resolution map is not sufficient to maintain secondary

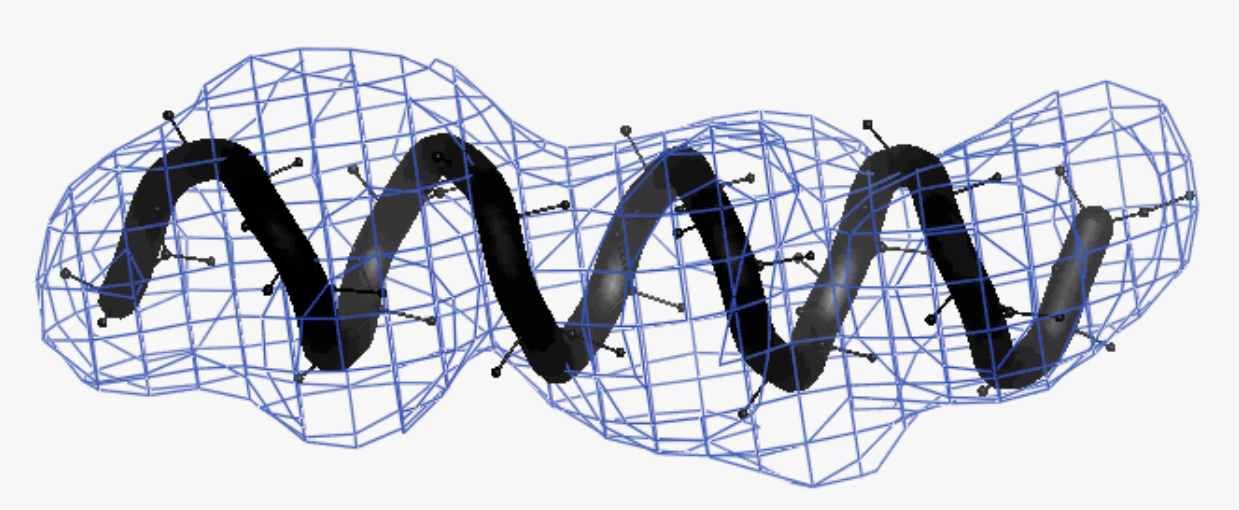






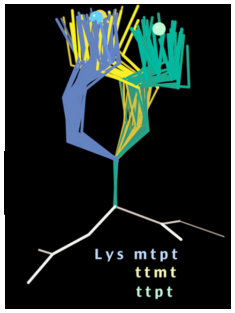
Automated model refinement: *phenix.real_space_refine*

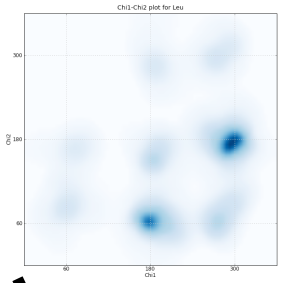
- Example: refinement of a perfect α-helix into low-res map
  - Using standard restraints on covalent geometry is insufficient
    - Model geometry deteriorates as result of refinement



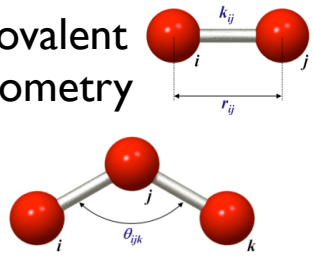
Automated model refinement: *phenix.real_space_refine*

Sidechain distributions

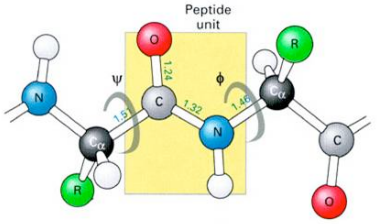




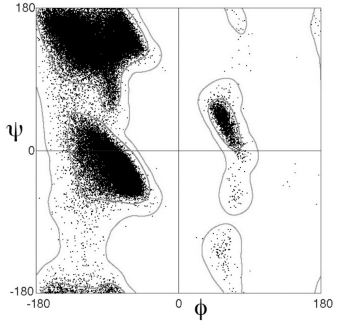
Covalent geometry

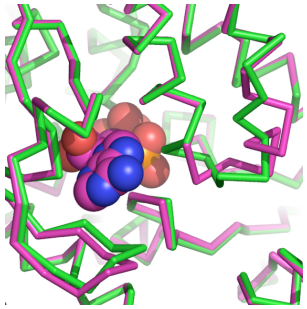


Images from PumMa web site (http://www.pumma.nl)

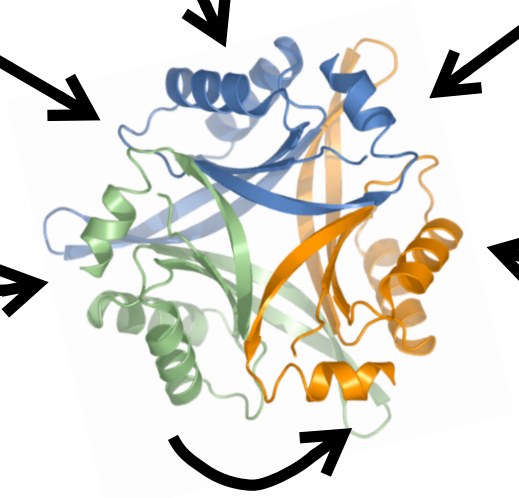


Mainchain distributions

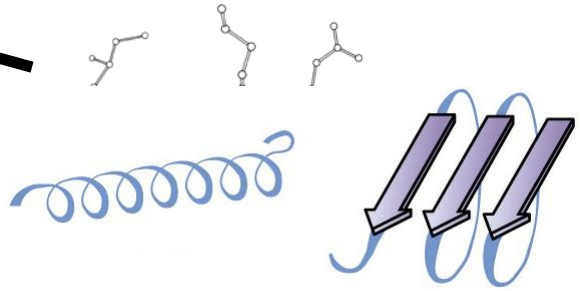




Related structures



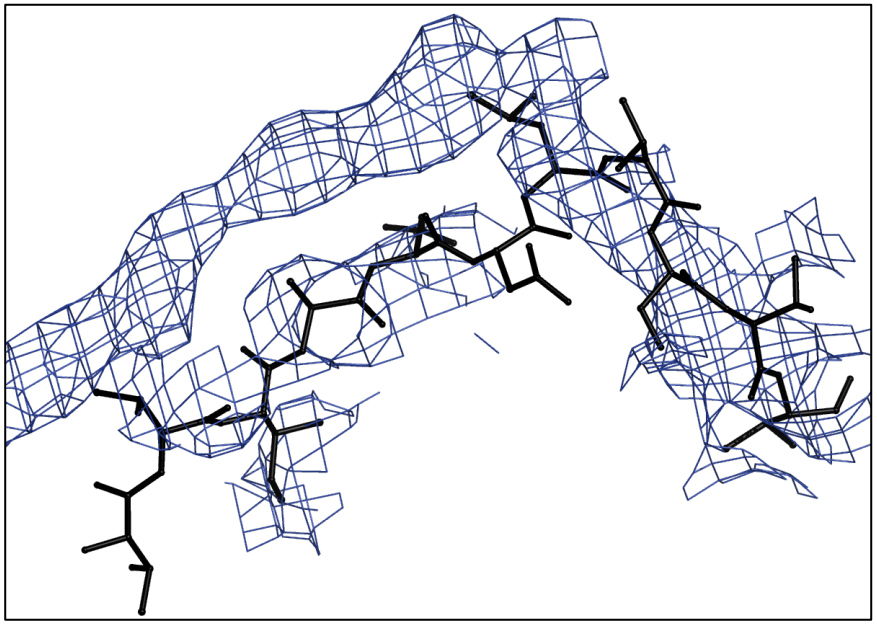
Internal symmetry



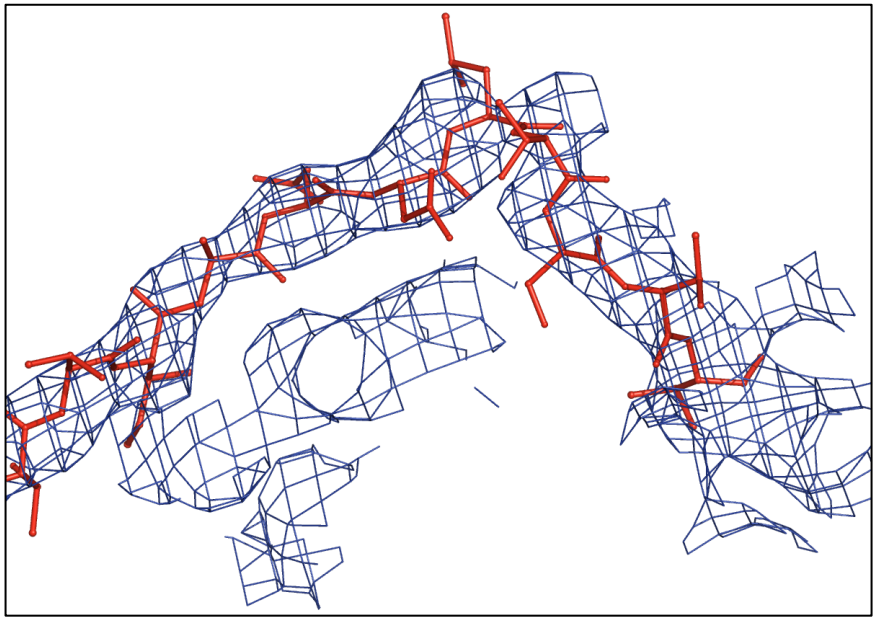
$$T_{RESTRAINTS} = T_{BOND} + T_{ANGLE} + \ldots + T_{NCS} + T_{RAMACHANDRAN} + T_{REFERENCE} + \ldots$$

Automated model refinement: *phenix.real_space_refine*

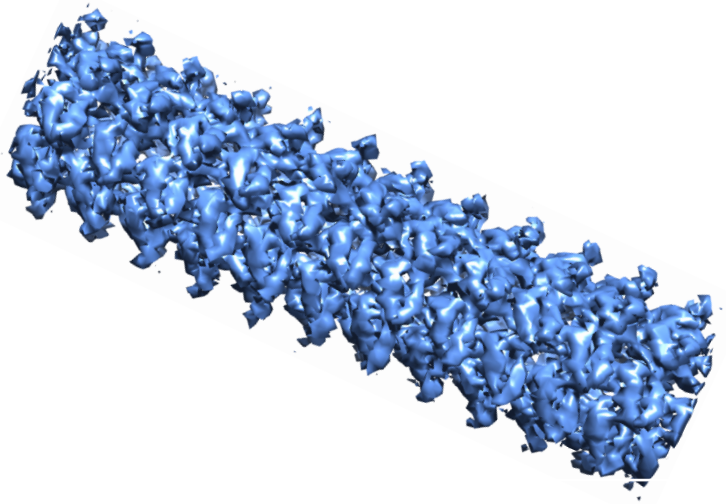
Start model before refinement

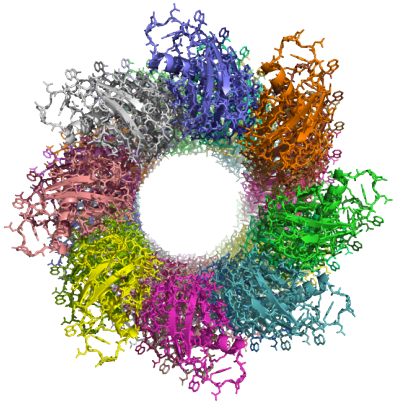


After *phenix.real_space_refine*



Examples: 3ZEE, resolution: 6.1 Å





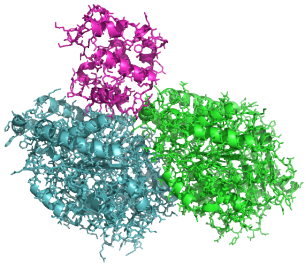
Residues/atoms: 4,116/32,830

Refinement: 45 min

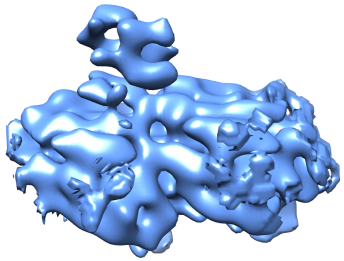
METRIC	Original	*Phenix*
Map CC	0.709	0.647
RMSD (bonds/angles)	0.04/4.05	0.01/1.23
Clashscore	18.34	18.59
Rama. outl., %	3.66	0
Rotamer outl., %	24.64	0
C-beta deviations	637	0

Validation

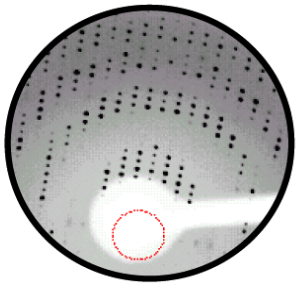
Model



Data



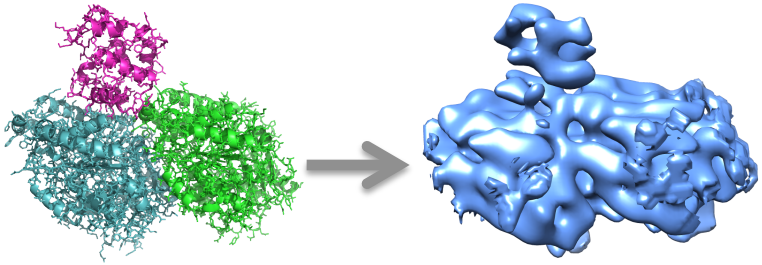
or



Cryo-EM

Diffraction

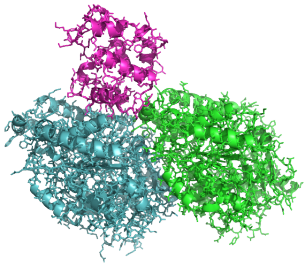
Model to data fit



Validation tools: Crystallography vs Cryo-EM

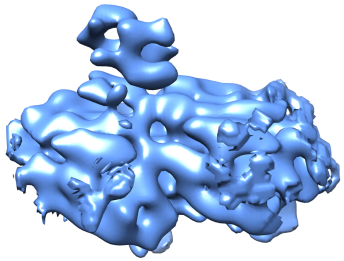
Exactly same

Model

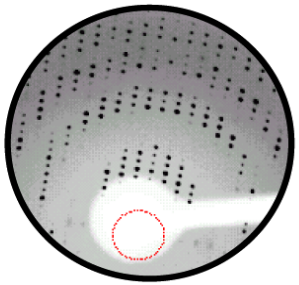


Different

Data



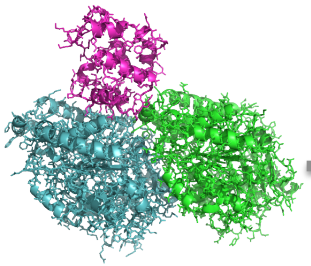
or

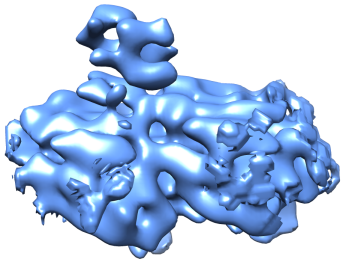


Cryo-EM

Diffraction

Model to data fit





Similar

Validation

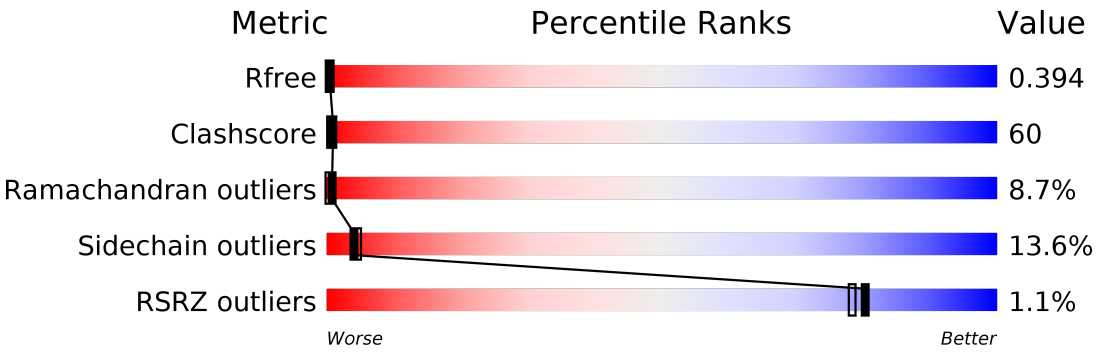
- Helps to save time later
- Helps to produce better models
- Helps to set correct expectations
- Minimize fraud or true mistakes

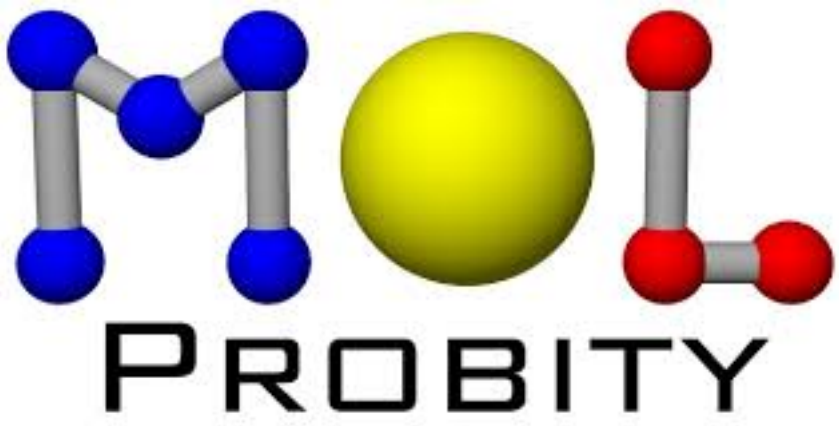
Acta Cryst D STRUCTURAL BIOLOGY
ISSN 2059-7983

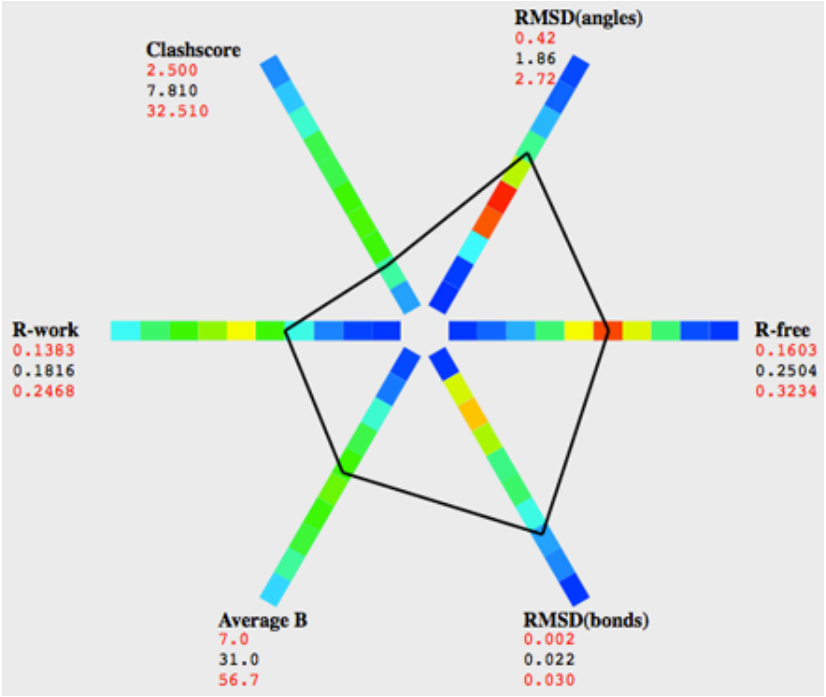
New tools for the analysis and validation of cryo-EM maps and atomic models

Pavel V. Afonine,[a,b]* Bruno P. Klaholz,[c] Nigel W. Moriarty,[a] Billy K. Poon,[a] Oleg V. Sobolev,[a] Thomas C. Terwilliger,[d,e] Paul D. Adams[a,f] and Alexandre Urzhumtsev[c,g]

Validation

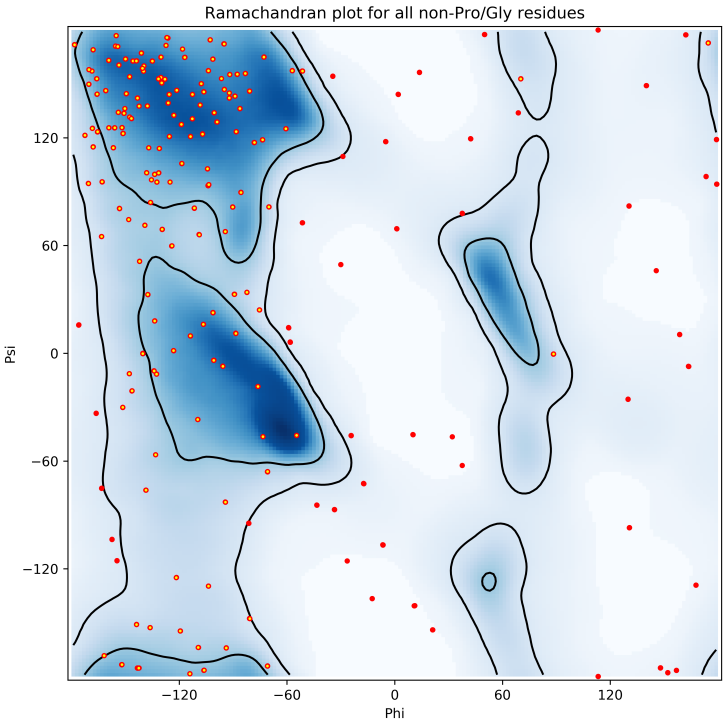

Metric
Percentile Ranks
Value
Rfree
0.394
Clashscore
60
Ramachandran outliers
8.7%
Sidechain outliers
13.6%
RSRZ outliers
1.1%
Worse
Better


PROBITY


RMSD(angles)
0.42
1.86
2.72
Clashscore
2.500
7.810
32.510
R-work
0.1383
0.1816
0.2468
R-free
0.1603
0.2504
0.3234
Average B
7.0
31.0
56.7
RMSD(bonds)
0.002
0.022
0.030

Example: Ramachandran plot outliers

3zx9



Clashscore: 245

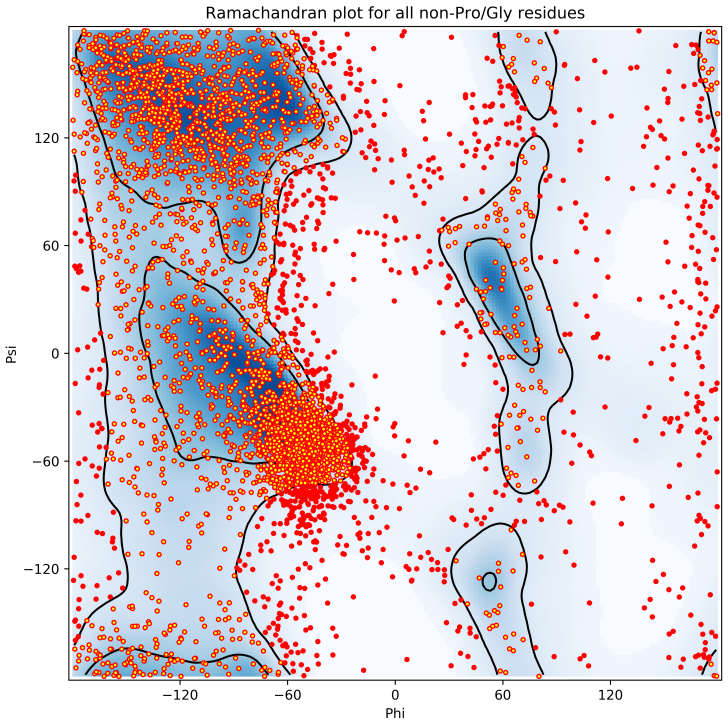
Rama outliers: 23%

Rotamer outliers: 17%

Year: 2011

Resolution: 17Å

5a9z



Clashscore: 197

Rama outliers: 25%

Rotamer outliers: 28%

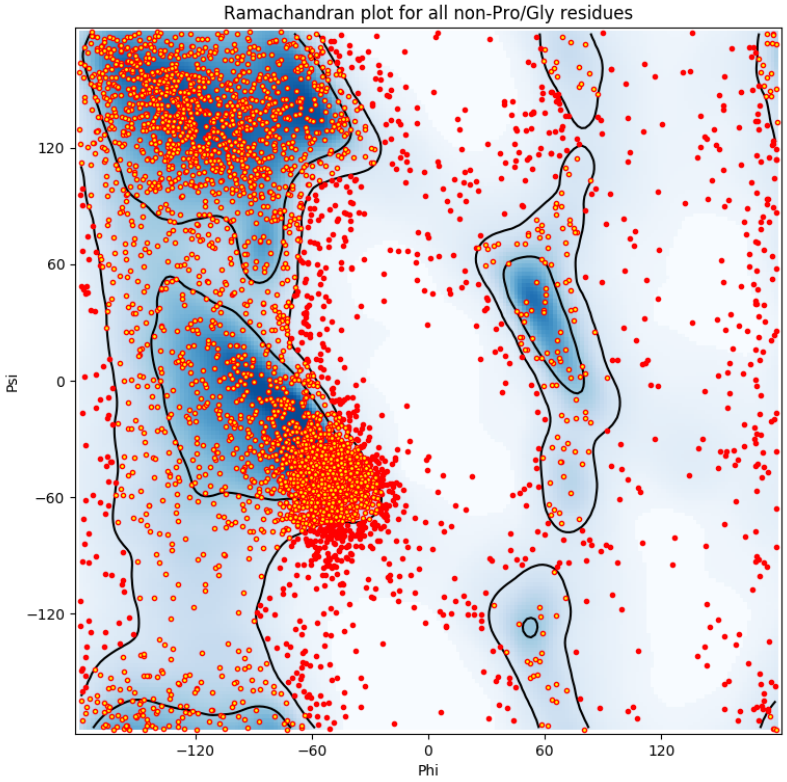
Year: 2015

Resolution: 4.7Å

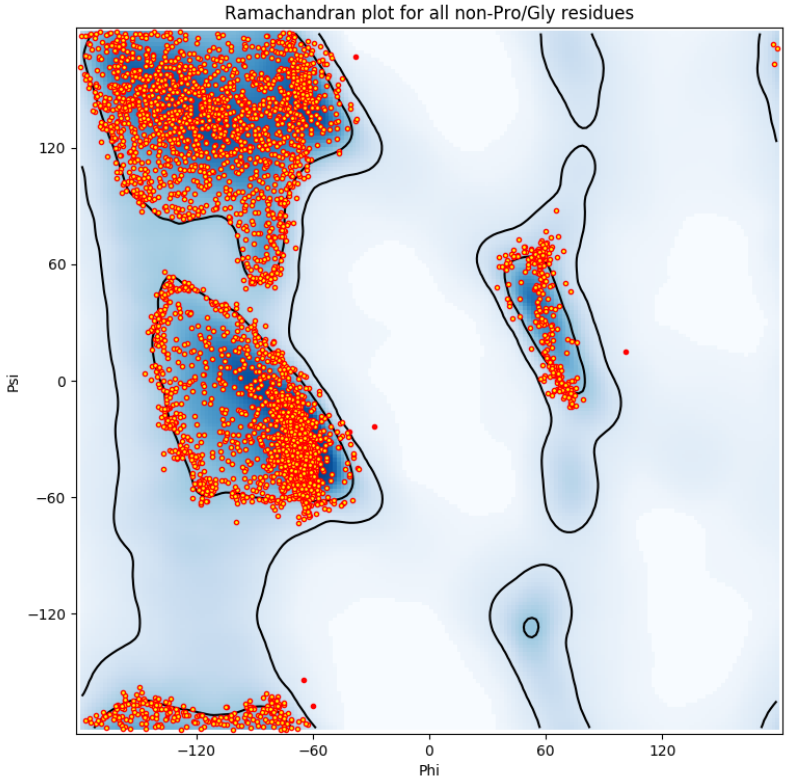
Ramachandran plot

PDB code: 5a9z

Original

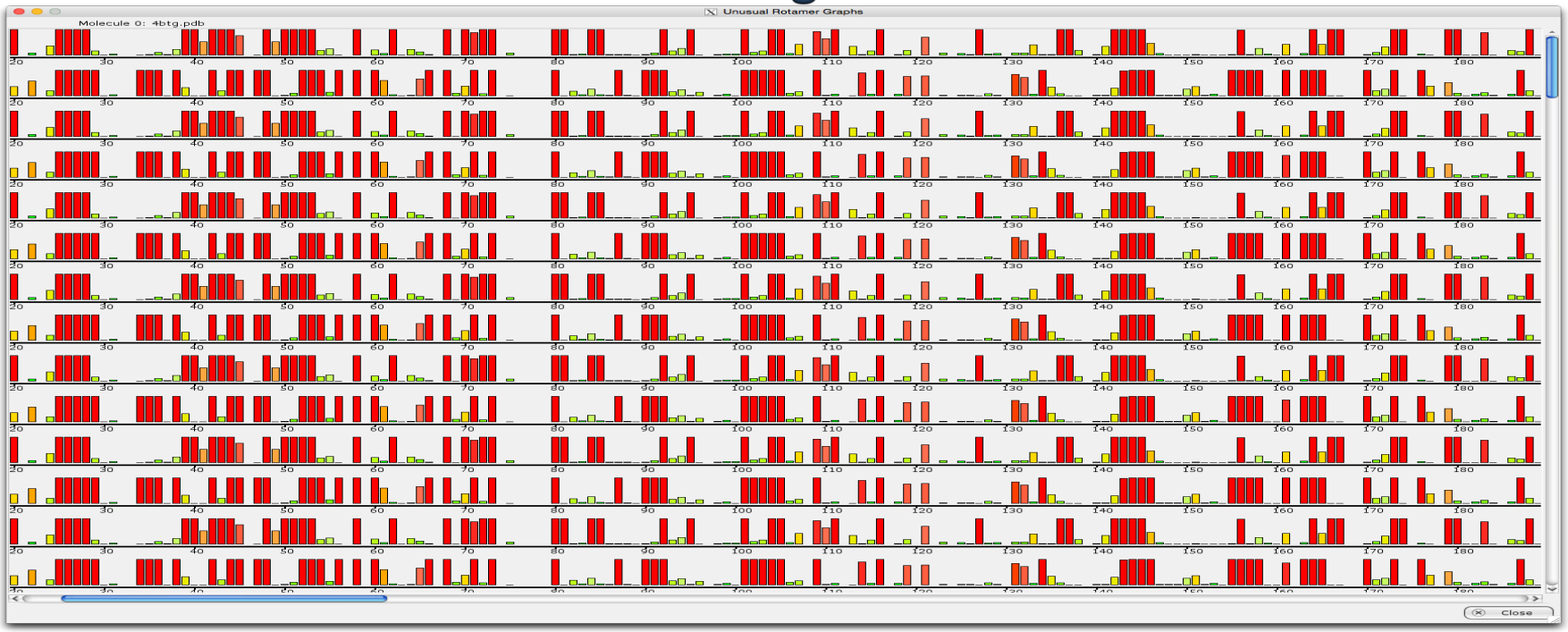


Refined with Ramachandran plot restraints



Example: side-chain rotamer outliers

4btg



Clashscore: 329

Rama outliers: 9%

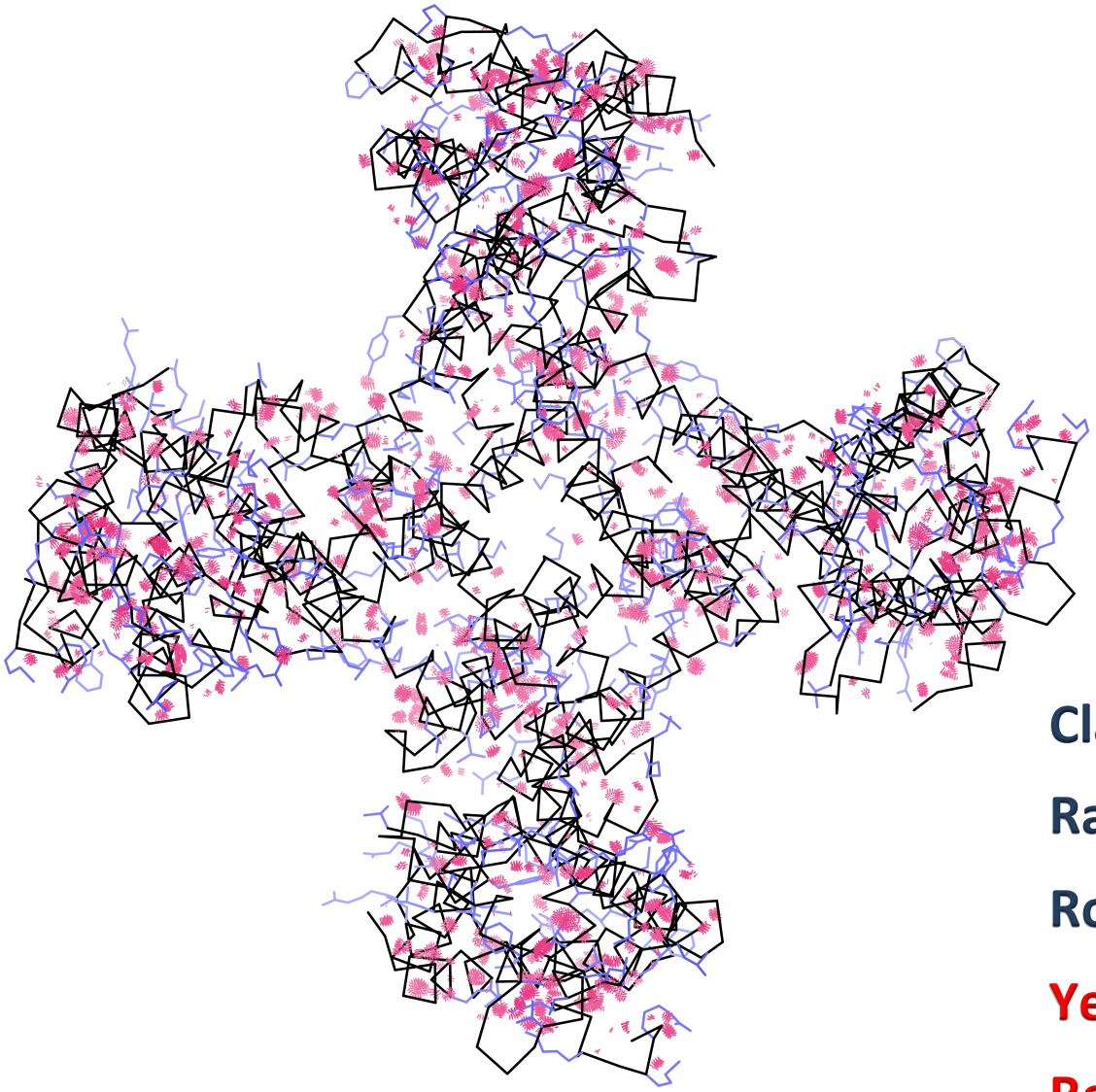
Rotamer outliers: 46%

Year: 2013

Resolution: 4.4Å

Example: steric clashes

3j5p



Clashscore: 78

Rama outliers: 0%

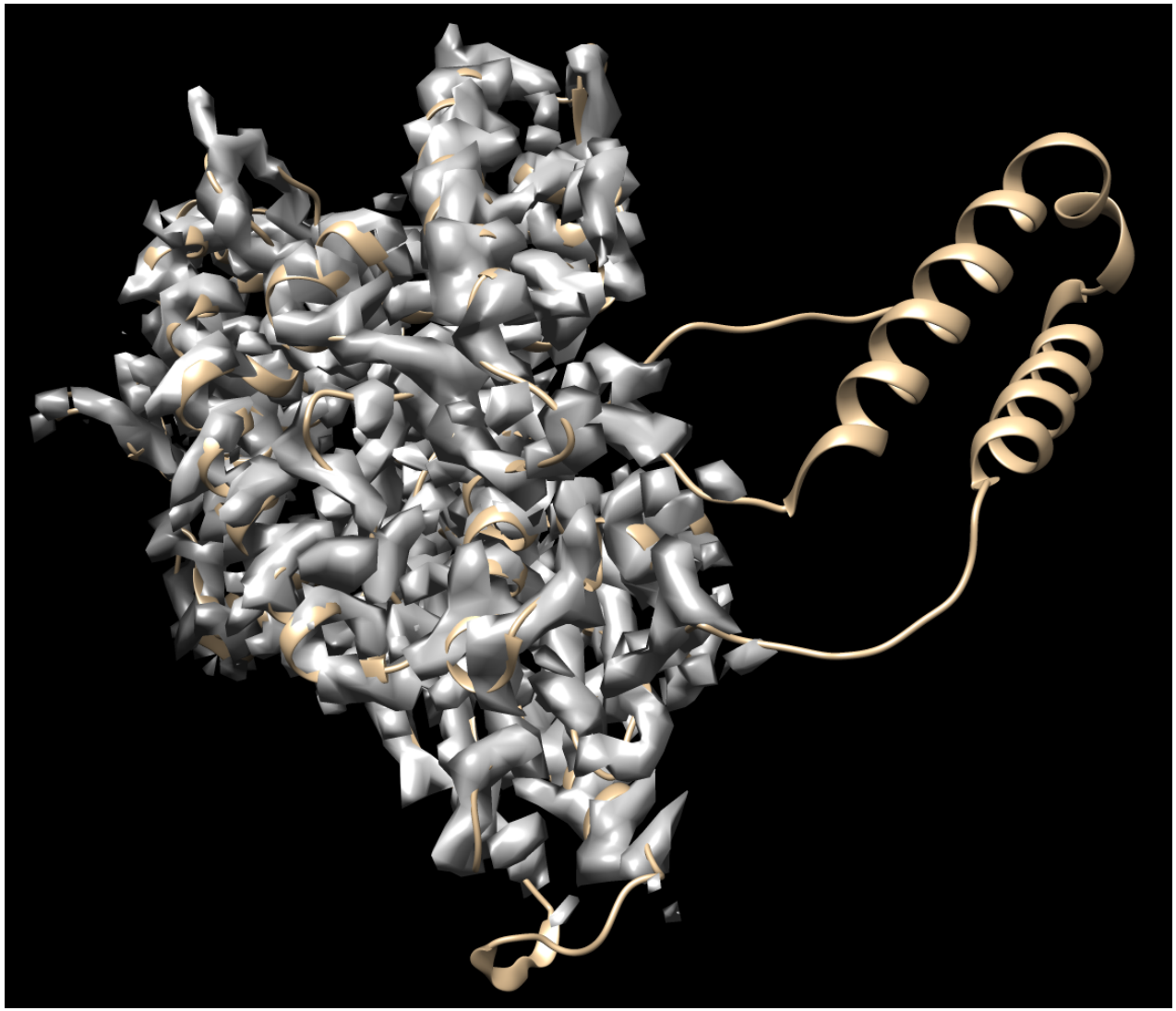
Rotamer outliers: 27%

Year: 2013

Resolution: 3.275Å

Validation: model-to-map fit

3j9e (emd_6240) | 3.3Å | CC= 0.85 | Year: 2015

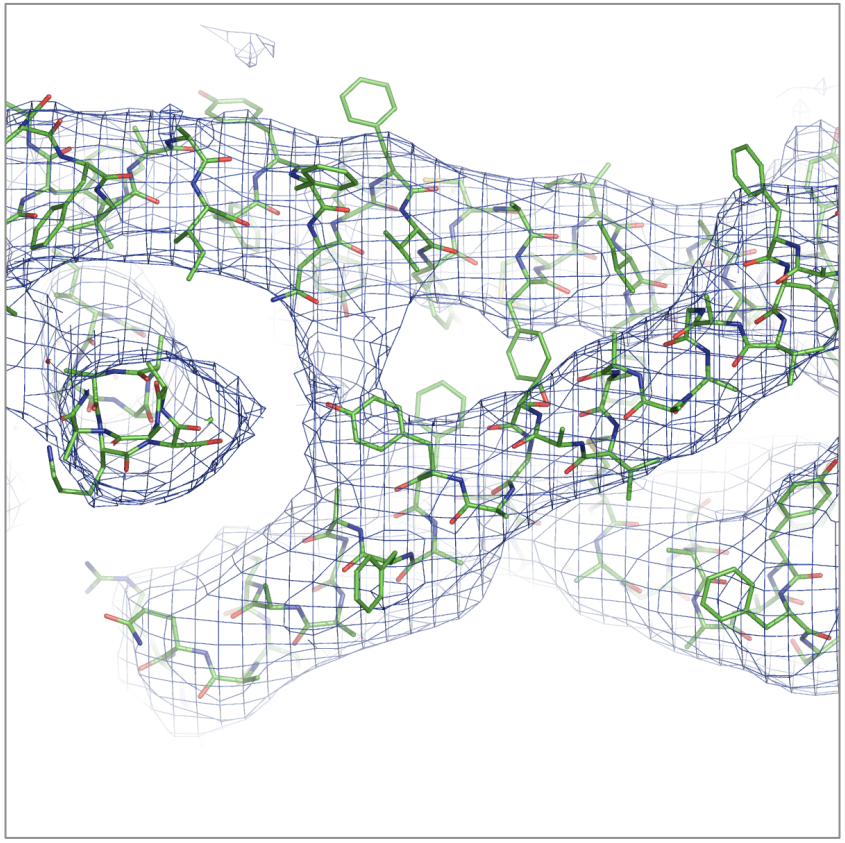


Validation: model-to-map fit

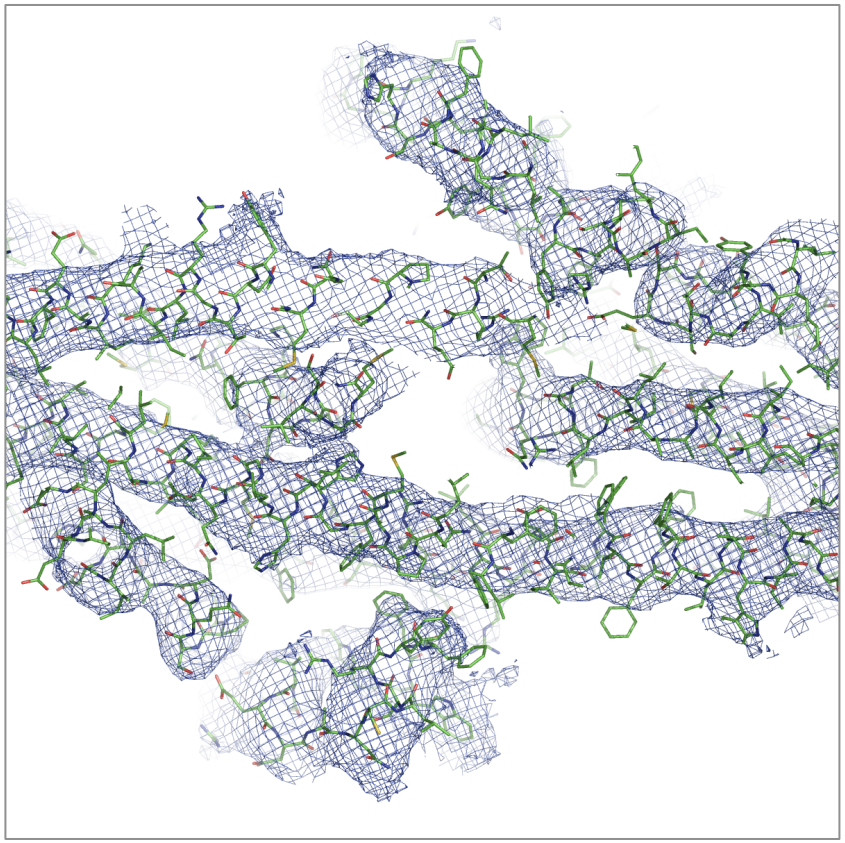
3a5x (emd_1641) | 4.0Å | CC <0

Data resolution

3j5q | 3.8Å



5tsi | 3.7Å



Model-map correlation coefficient (CC)

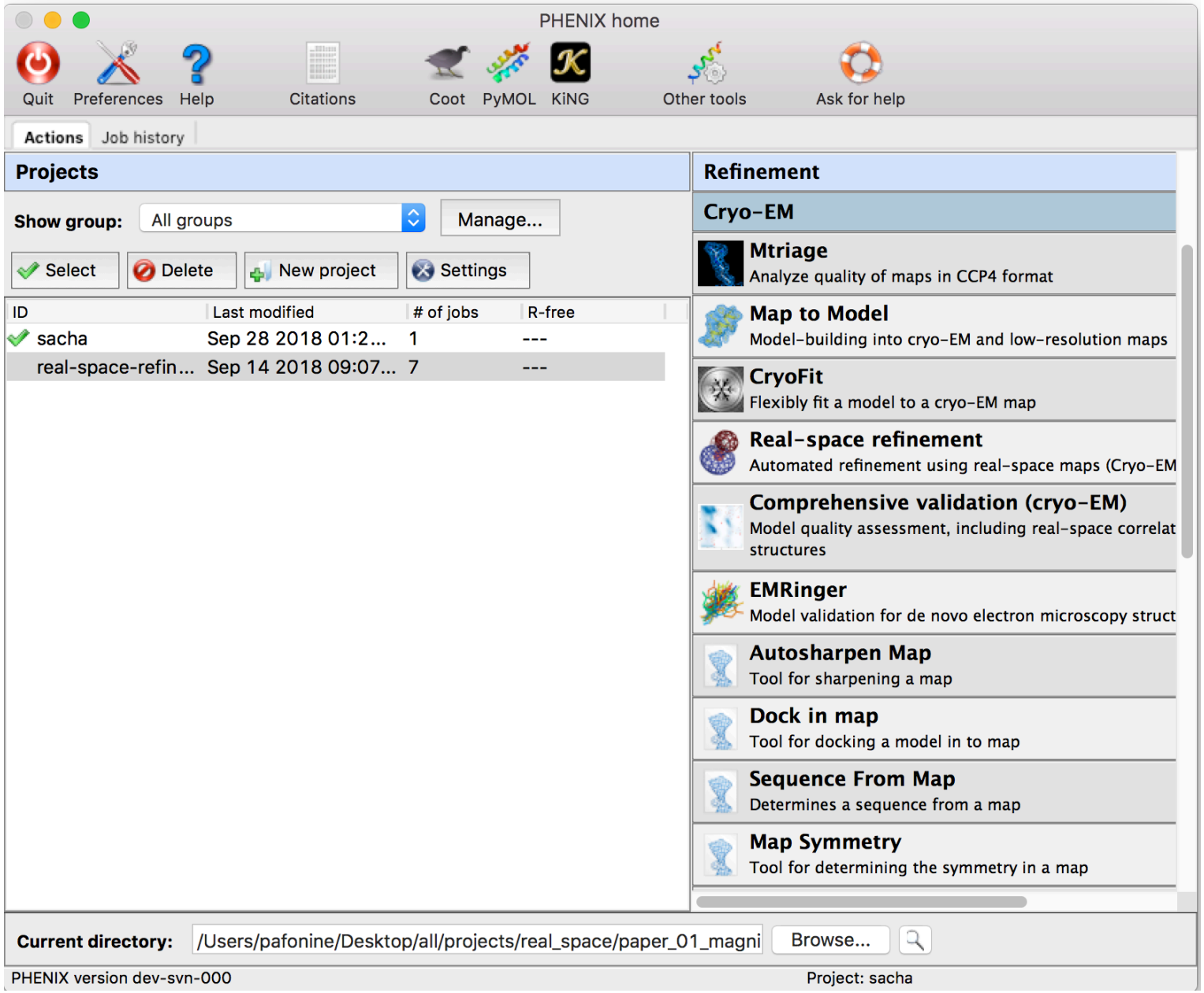
- Definition
  - With or w/o subtracting mean

$$CC(\rho_1,\rho_2)=\left(\sum_{\mathbf{n}}\left(\rho_1(\mathbf{n})\right)^2\right)^{-1/2}\left(\sum_{\mathbf{n}}\left(\rho_2(\mathbf{n})\right)^2\right)^{-1/2}\left(\sum_{\mathbf{n}}\rho_1(\mathbf{n})\rho_2(\mathbf{n})\right)$$

$$CC(\rho_1,\rho_2)=\left(\sum_{\mathbf{n}}\left(\rho_1(\mathbf{n})-\langle\rho_1\rangle\right)^2\right)^{-1/2}\left(\sum_{\mathbf{n}}\left(\rho_2(\mathbf{n})-\langle\rho_2\rangle\right)^2\right)^{-1/2}\left(\sum_{\mathbf{n}}\left(\rho_1(\mathbf{n})-\langle\rho_1\rangle\right)\left(\rho_2(\mathbf{n})-\langle\rho_2\rangle\right)\right)$$

- How model map is calculated
  - Approximation (e.g. N-gaussian)
  - Form-factors (electron, X-ray, neutron)
  - Fourier map
    - Box or sphere of Fourier map coefficients
- Region in the map used to calculate CC
  - Whole box
  - Mask around atoms
    - Atom radius

Phenix tools for cryo-EM



Resources

PHENIX Documentation

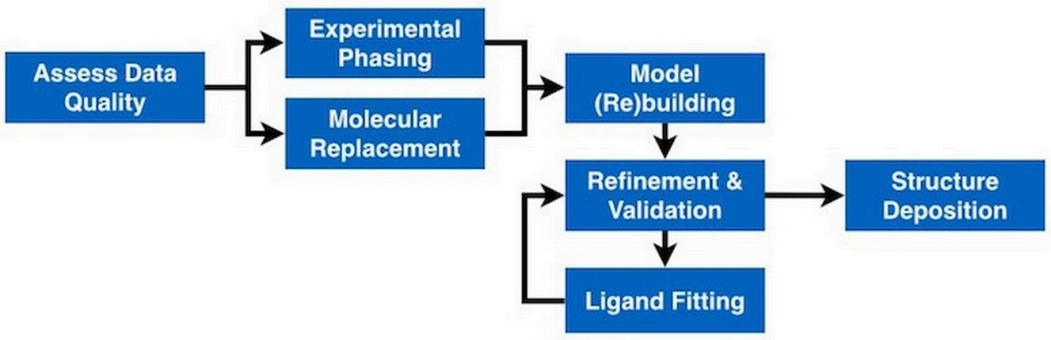
phenix-online.org/version_docs/1.9-1692/

Phenix *Python-based Hierarchical ENvironment for Integrated Xtallography*

Phenix Documentation - version 1.9-1692

Phenix programs and their functions
The Phenix graphical interface
Dictionary of crystallographic and other terms
FAQs: Frequently asked questions
How to install, setup and run Phenix
Complete Phenix reference documentation
Bibliography
Index

Crystallographic Structure Solution with Phenix



Phenix Documentation for X-ray Crystallography

Checking data quality | Experimental phasing | Molecular replacement | Model building | Structure refinement
Structure validation | Ligand fitting | Making geometry restraints | Structure deposition

Phenix Documentation for Neutron Crystallography

Structure refinement | Structure validation | Making geometry restraints | Structure deposition

Phenix Documentation for Electron Microscopy (EM)

Structure refinement | Convert map to structure factors | Extract box with map and model

Feedback

- **Feedback, questions, help**

  phenixbb@phenix-online.org
  bugs@phenix-online.org
  help@phenix-online.org

- **Reporting a bug or asking for help:**
  - **We can't help you if you don't help us to understand your problem**
  - **Do:**
    1) Make sure you can reproduce the problem using latest *Phenix* version
    2) Command and parameters used (series of GUI clicks that lead to problem)
    3) Input and output files
    4) Clearly explain the problem/question

PHENIX mailing list: www.phenix-online.org

Thanks !