

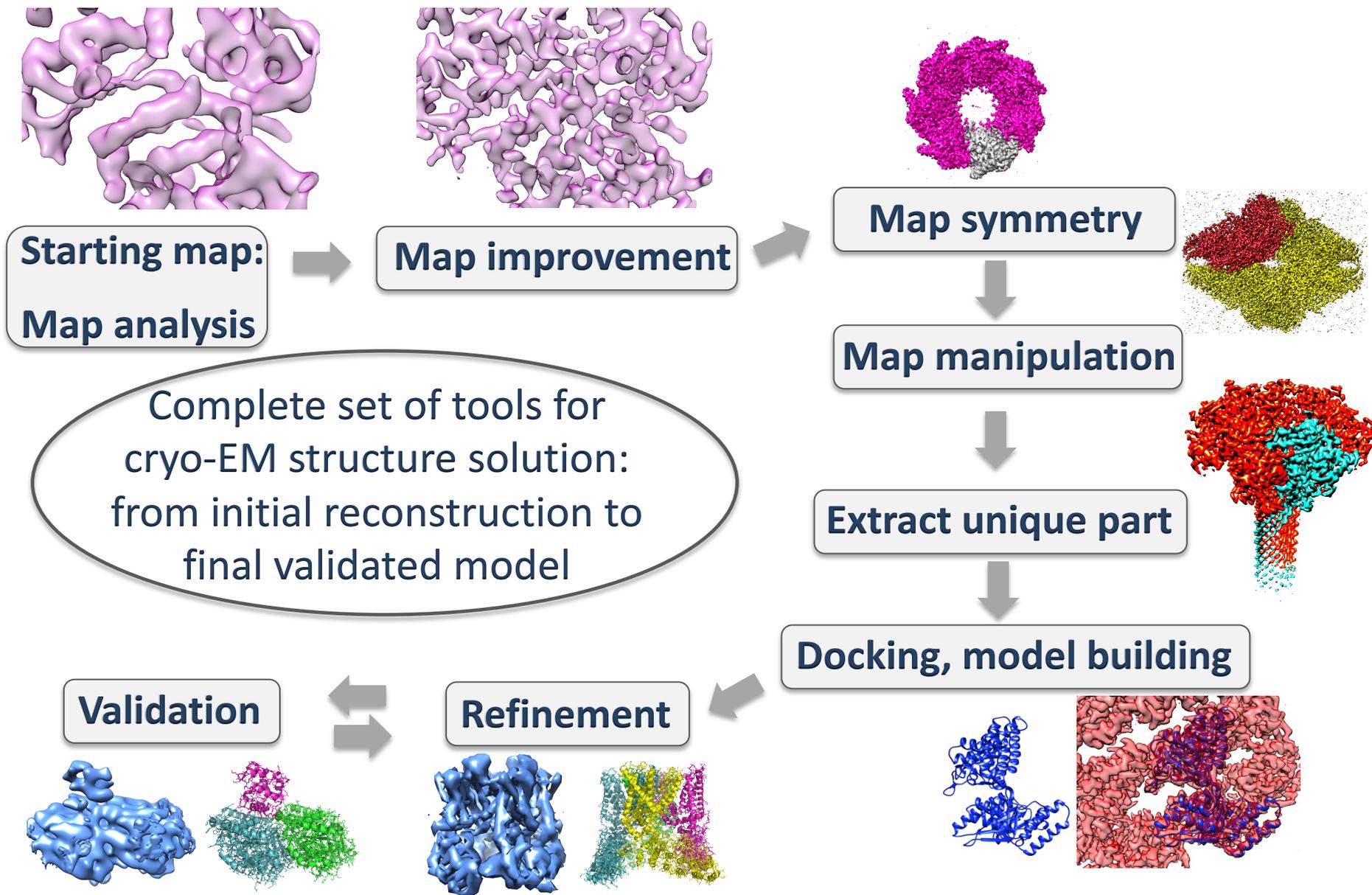
***Phenix* Tools for Cryo-EM: Refinement**

Pavel Afonine

LBNL, Berkeley, California, USA

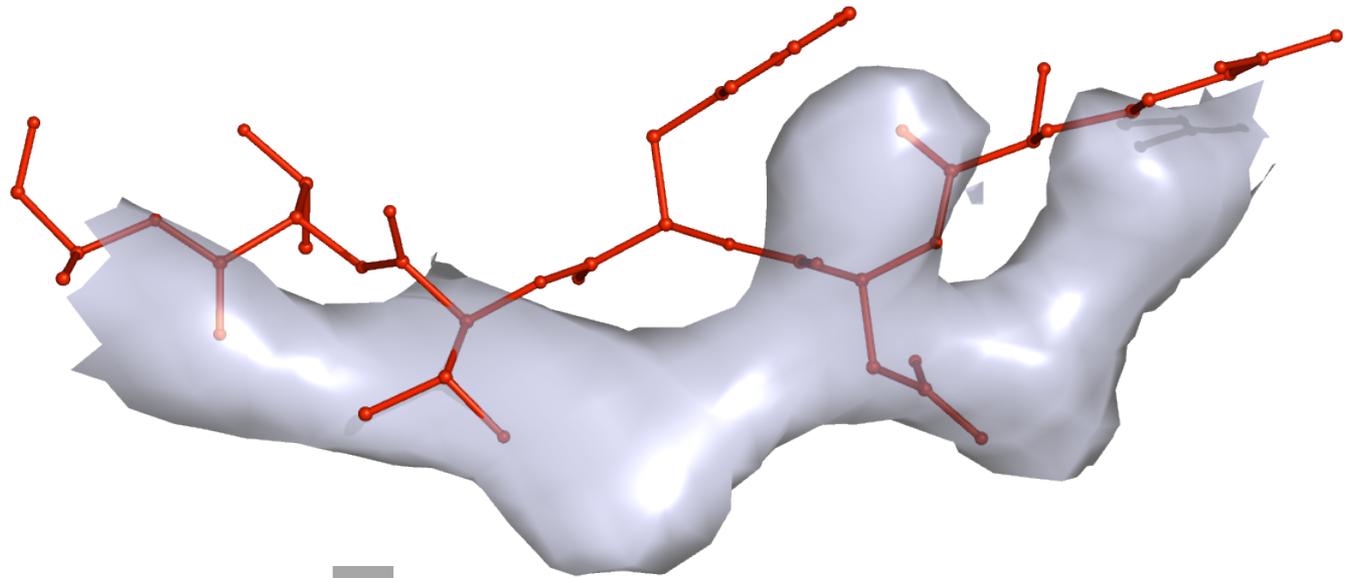
January 30, 2020

Cryo-EM tools in *Phenix*



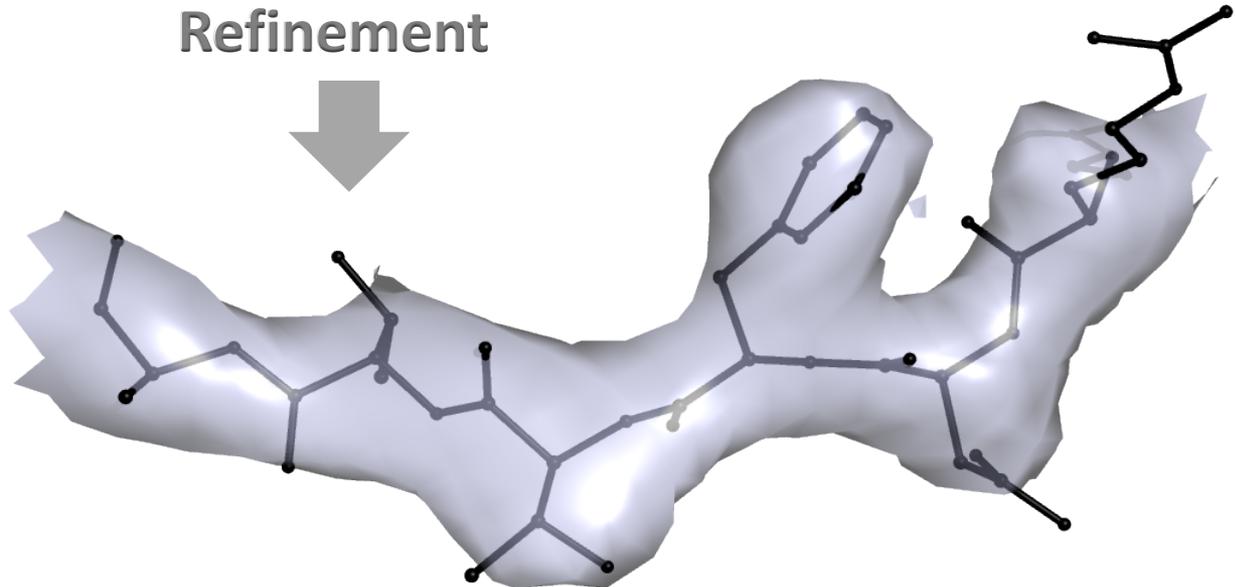
Structure refinement

Initial (poor)
model



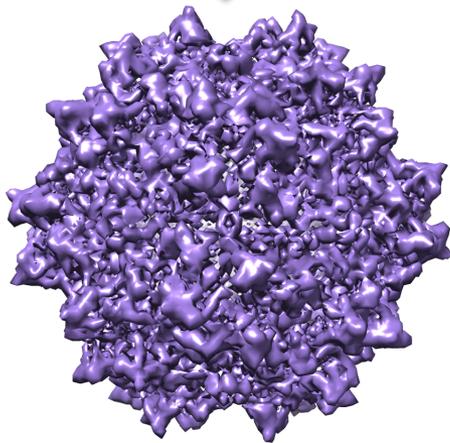
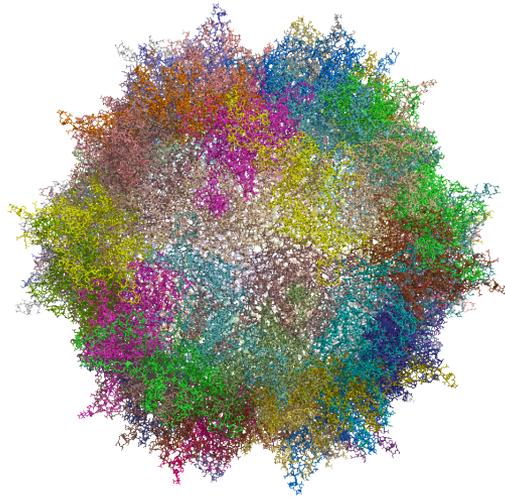
Refinement

Improved
(refined)
model

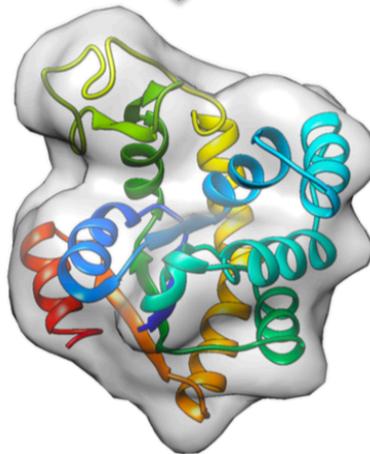
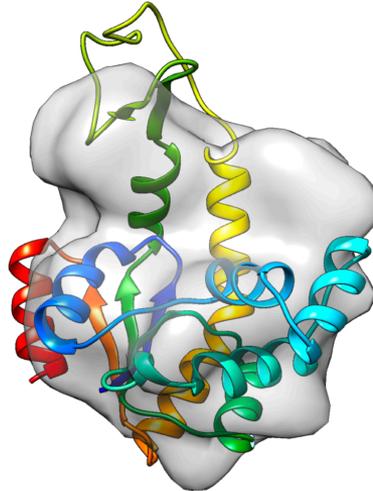


Scenarios for model-to-map fitting

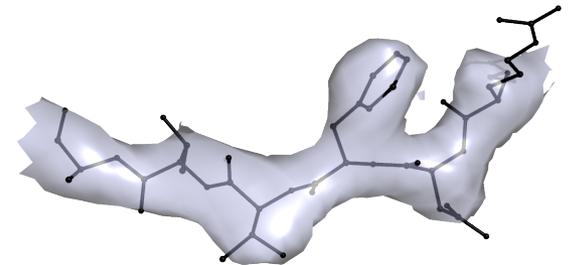
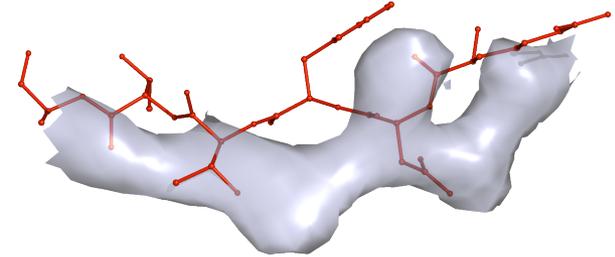
Docking



Flexible fitting, morphing

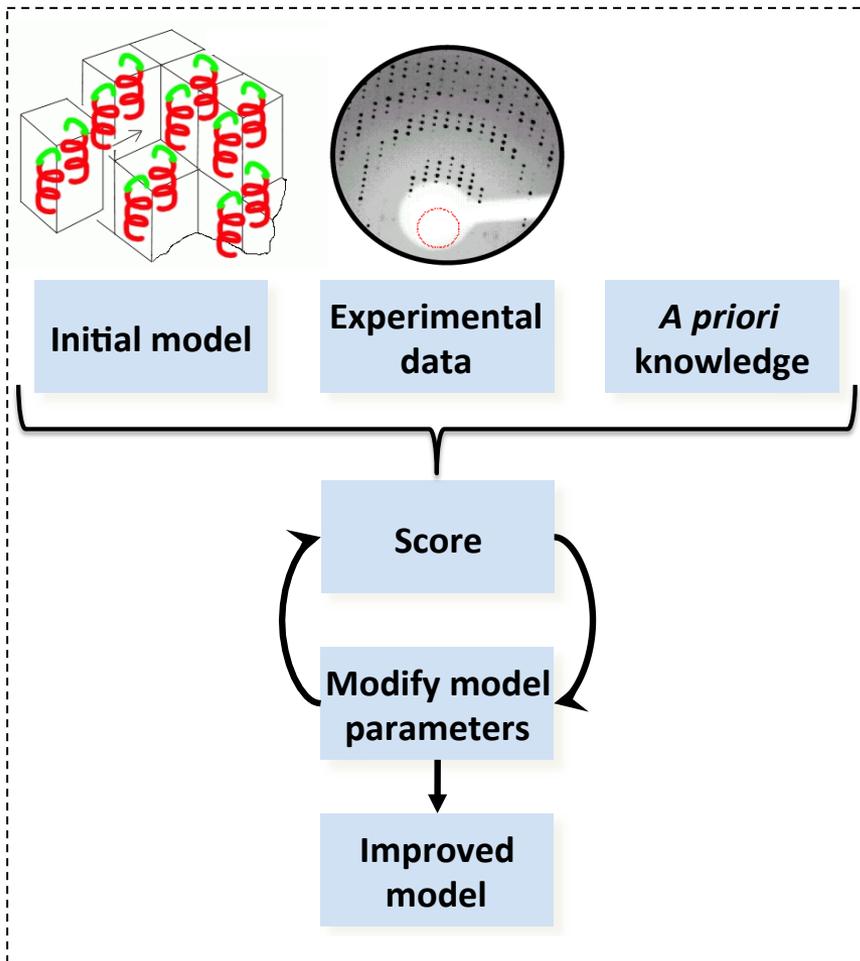


Refinement



Refinement tools in *Phenix*

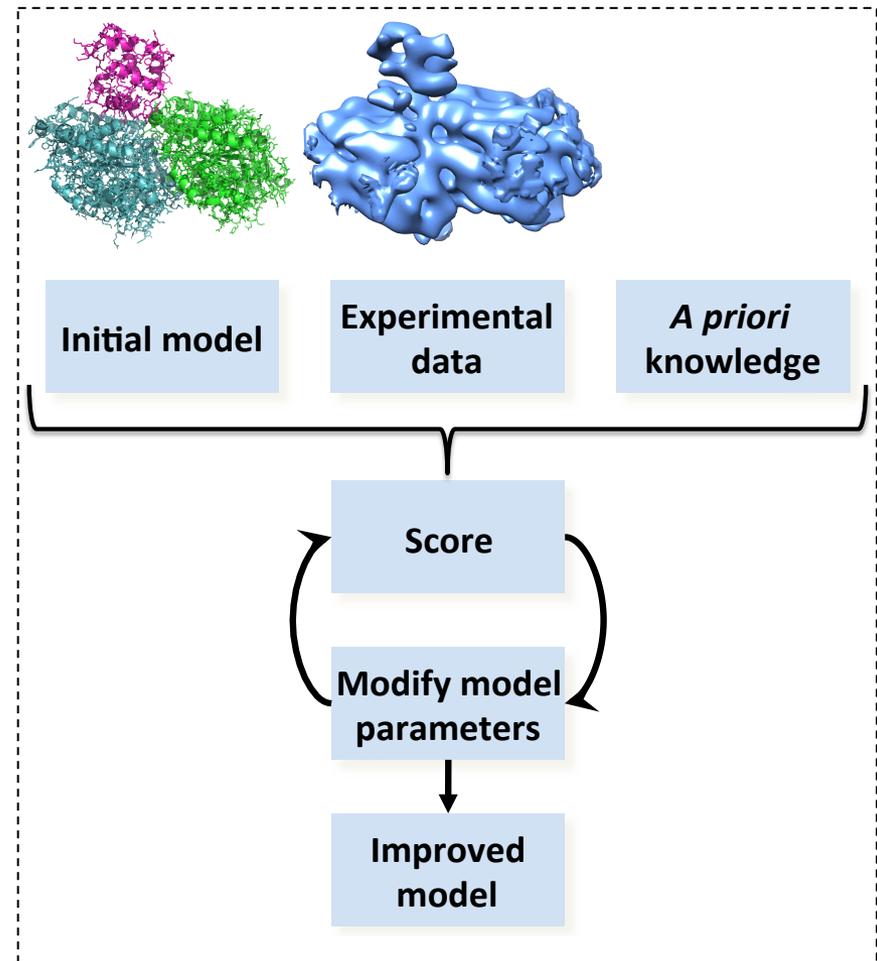
Crystallography



phenix.refine

Available since 2005

Cryo-EM



phenix.real_space_refine

Available since 2013

Refinement tools in *Phenix*

PHENIX home

Quit Preferences Help Citations Reload last job Coot PyMOL KiNG Other tools Ask for help

Actions Job history

Projects

Show group: All groups Manage...

Select Delete New project Settings

ID	Last modified	# of jobs	R-free
ringer	Sep 07 2016 05:37 ...	2	---
tmp2	Sep 07 2016 05:23 ...	1	---
✓ 5gnn	Sep 07 2016 08:42 ...	1	---
debug1	Sep 05 2016 10:51 ...	2	0.0086
tmp4	Aug 18 2016 07:23 ...	2	---
testing	Aug 11 2016 01:54 ...	1	---
mich	Jul 29 2016 12:47 ...	1	---
almu	Jul 28 2016 10:58 ...	1	---
rchen	Jul 22 2016 11:10 ...	1	---
milya	Jul 15 2016 12:36 ...	2	---
SEM	Jul 14 2016 05:20 ...	14	0.1570

Data analysis

Experimental phasing

Molecular replacement

Model building

Refinement

-  **phenix.refine**
Automated X-ray and/or neutron refinement
-  **Real-space refinement**
Automated real-space refinement
-  **Neutron refinement [alpha]**
Alternate phenix.refine interface customized for neutron data
-  **DEN refinement [alpha]**
Deformable elastic network refinement using small-angle scattering data and molecular replacement structure

Current directory: /Users/pafonine/Desktop/work/tmp Browse...

PHENIX version dev-svn-000 Project: 5gnn

Real-space refinement with *phenix.real_space_refine*

- No Fourier space involved
- No structure factors
- No R-factors
- Model refined directly into the map



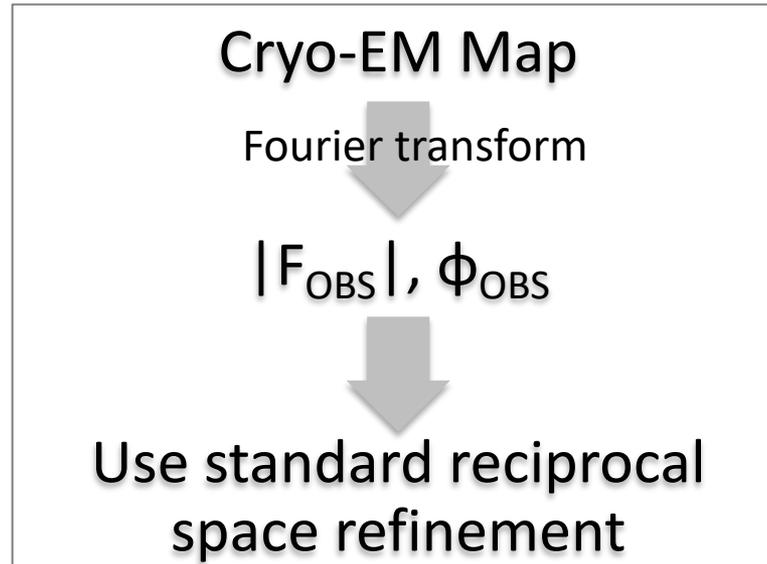
ISSN 2059-7983

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}

Repurposing crystallographic software for cryo-EM

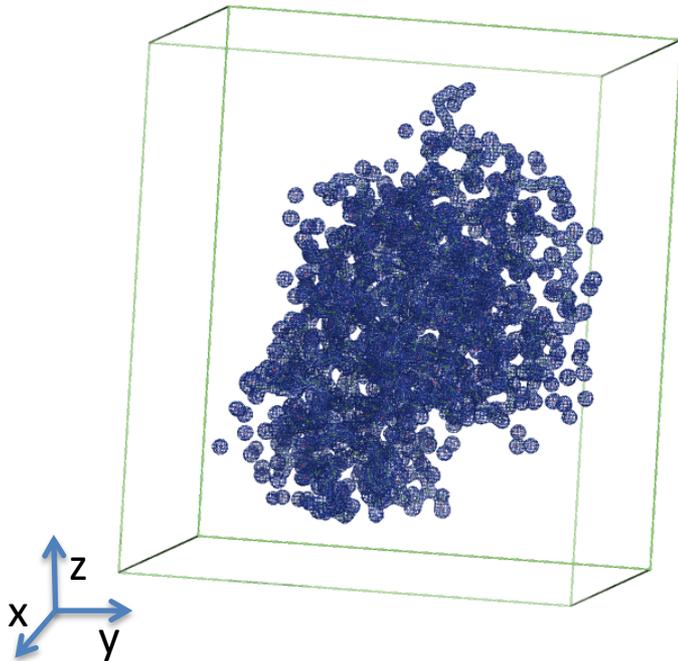
Legacy procedure



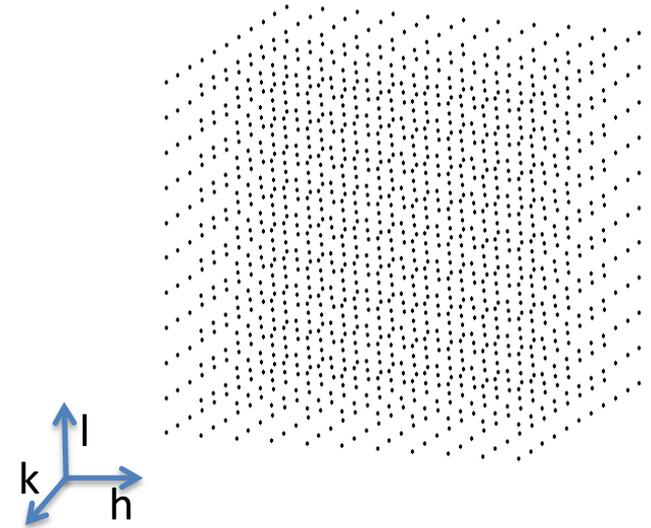
- Is conversion map to structure factors lossless?
- Are crystal bulk-solvent and anisotropic scaling still applicable?
- R factors will largely depend on masking used. What is R_{free} ?
- 2mFo-DFc and mFo-DFc maps?
- Form-factors ?

Map to structure factors conversion

Map in real space (ρ)



Map in Fourier space (F)

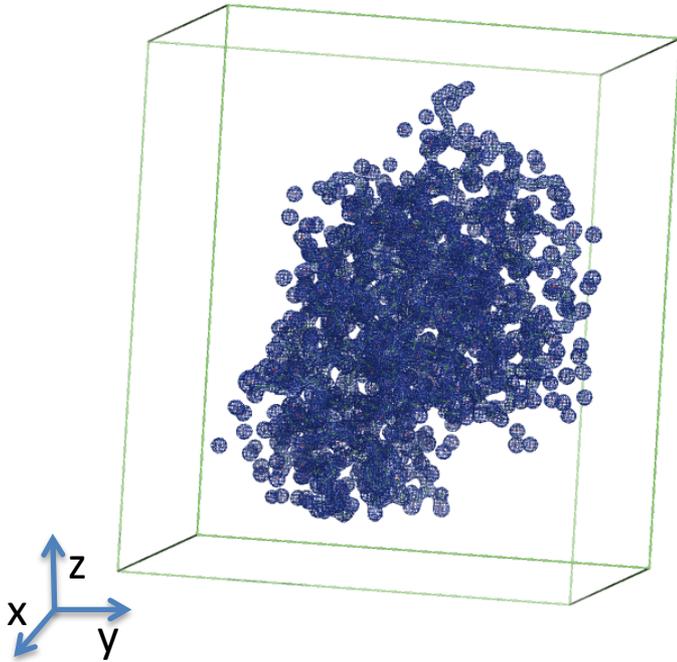


$$\rho(\mathbf{r}) = \frac{1}{V_{cell}} \sum_{h_{min}}^{h_{max}} \sum_{k_{min}}^{k_{max}} \sum_{l_{min}}^{l_{max}} F(\mathbf{s}) \exp(-2\pi i \mathbf{s} \mathbf{r})$$

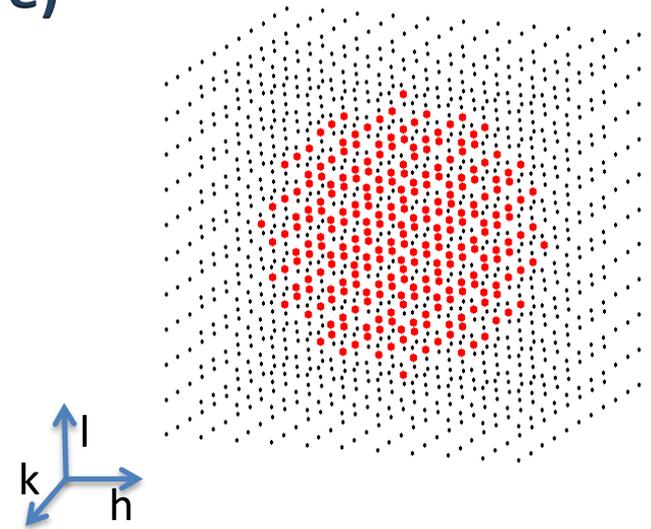
- Lossless conversion: $\rho \leftrightarrow F$

Map to structure factors conversion

Map in real space (ρ)



Map in Fourier space (F)
truncated by resolution (red sphere)



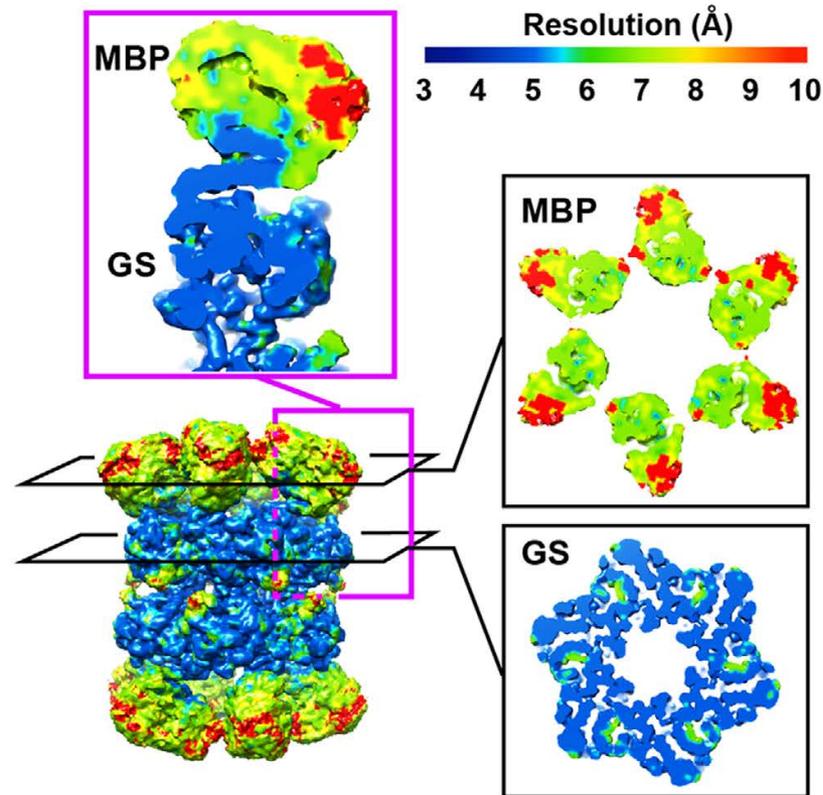
- **Not** a lossless conversion: $\rho \leftrightarrow F$

Real-space refinement

- **Calculations are faster**
 - Large models
 - Quick turn-around when model building or development
- **Local targets**
 - Easy to make parallel
 - Employ methods with large convergence radii
- **Weight between data and restraints**
 - Can always be optimized
 - Can vary across reconstruction volume

Real-space refinement: local data/restraints weight

- PDB code: 5LDF



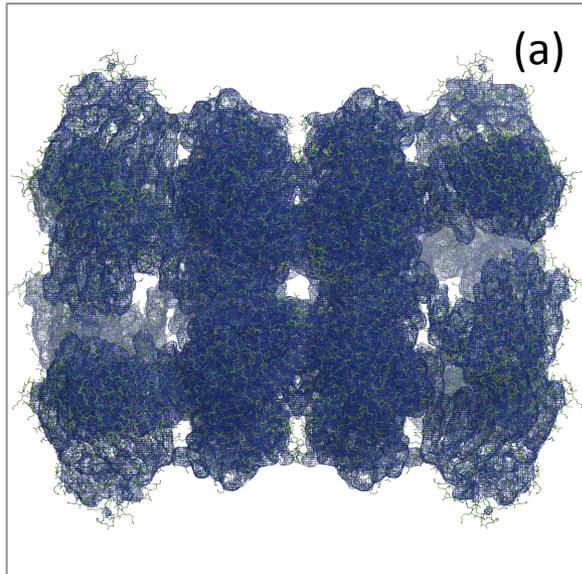
Coscia et al (2016). Sci. Reports, 6, 30909

Resolution (map quality) varies across the volume

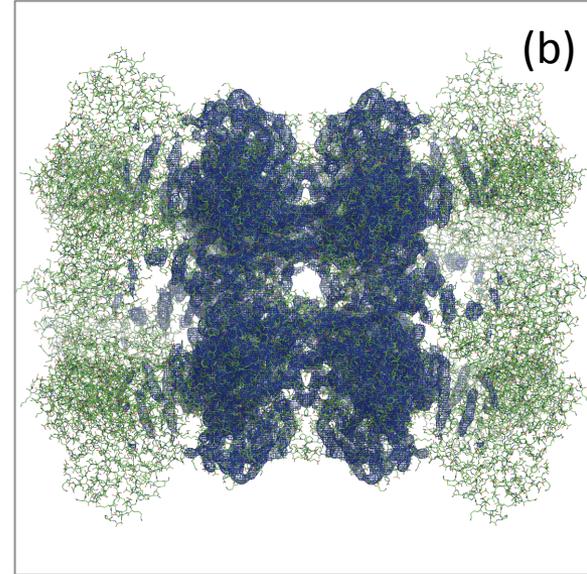
Real-space refinement: local data/restraints weight

- 5LDF

Low contouring level



High contouring level



$$T = T_{\text{DATA}}(\rho_{\text{OBS}}, \rho_{\text{MODEL}}) + wT_{\text{RESTRAINTS}}$$

- **Contribution of restraints can be weighted by local map quality**
 - Poorer resolved regions may be restrained stronger
 - Better resolved regions may use less restraints and more data
 - Local optimal weight can be quickly obtained in real space

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**

Refinement target

- Least-squares (map similarity) or cross-correlation:

$$LS = \sum_{\text{all grid points}} (\rho_{obs} - k * \rho_{calc})^2$$

$$CC = \frac{\sum_n \rho_{obs} \rho_{calc}}{\sum_n \rho_{obs}^2 \sum_n \rho_{calc}^2}$$

ρ_{obs} = experimental map

ρ_{calc} = model calculated map

- Accurate (matches shape of model-calculated map with experimental map)
- Very slow to calculate
 - Not used in *phenix.real_space_refine*

Refinement target

- Atom-centered:

$$T = - \sum_{atoms} \rho_{obs}(x_{atom}, y_{atom}, z_{atom})$$

$x_{atom}, y_{atom}, z_{atom}$ = coordinates of atom center

- Less accurate
- Very fast to calculate (more than 100 times faster than LS or CC)
 - Used in *phenix.real_space_refine*

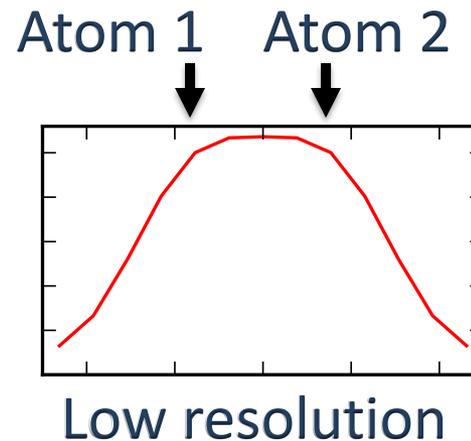
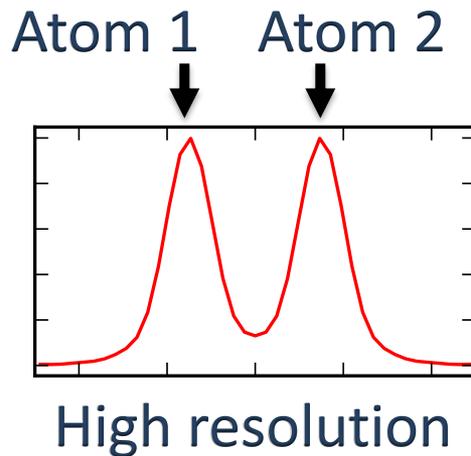
Refinement target

- Atom-centered:

$$T = - \sum_{atoms} \rho_{obs}(x_{atom}, y_{atom}, z_{atom})$$

$x_{atom}, y_{atom}, z_{atom}$ = coordinates of atom center

- Why it is less accurate?

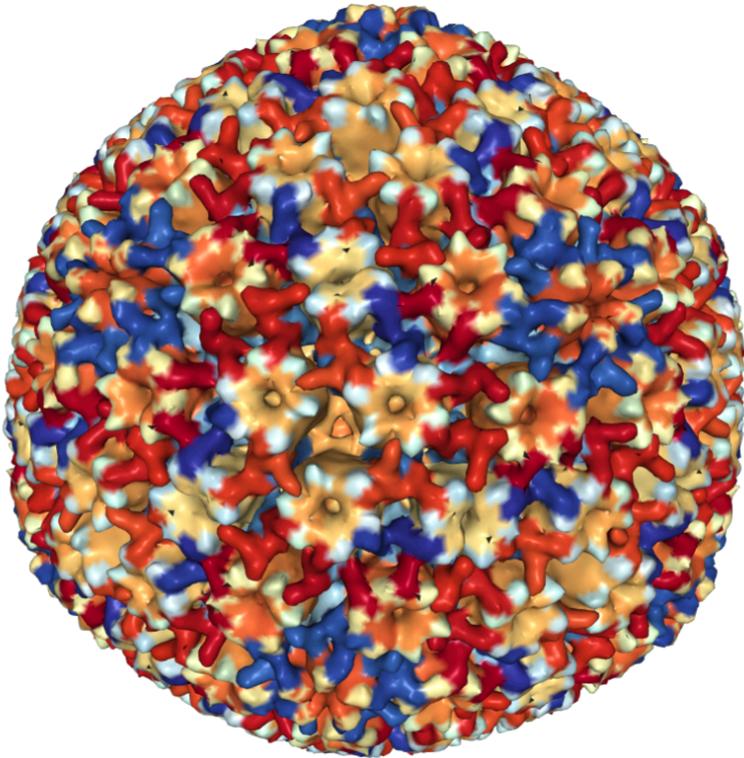


Moving atoms to nearest peaks \neq making correct model

Target calculation example

- PDB: 5VKU

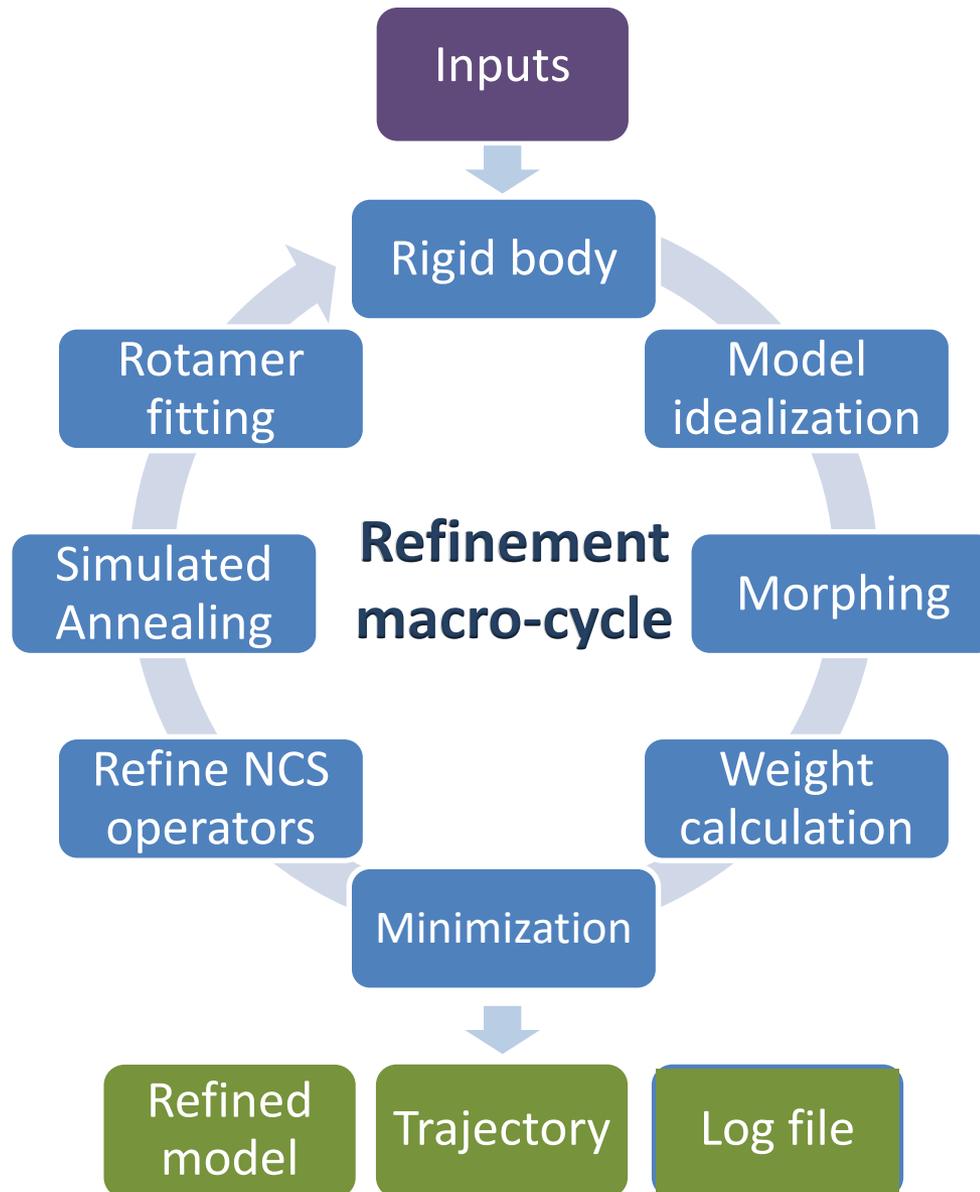
3720 chains | 1,872,060 residues | 14,917,620 atoms



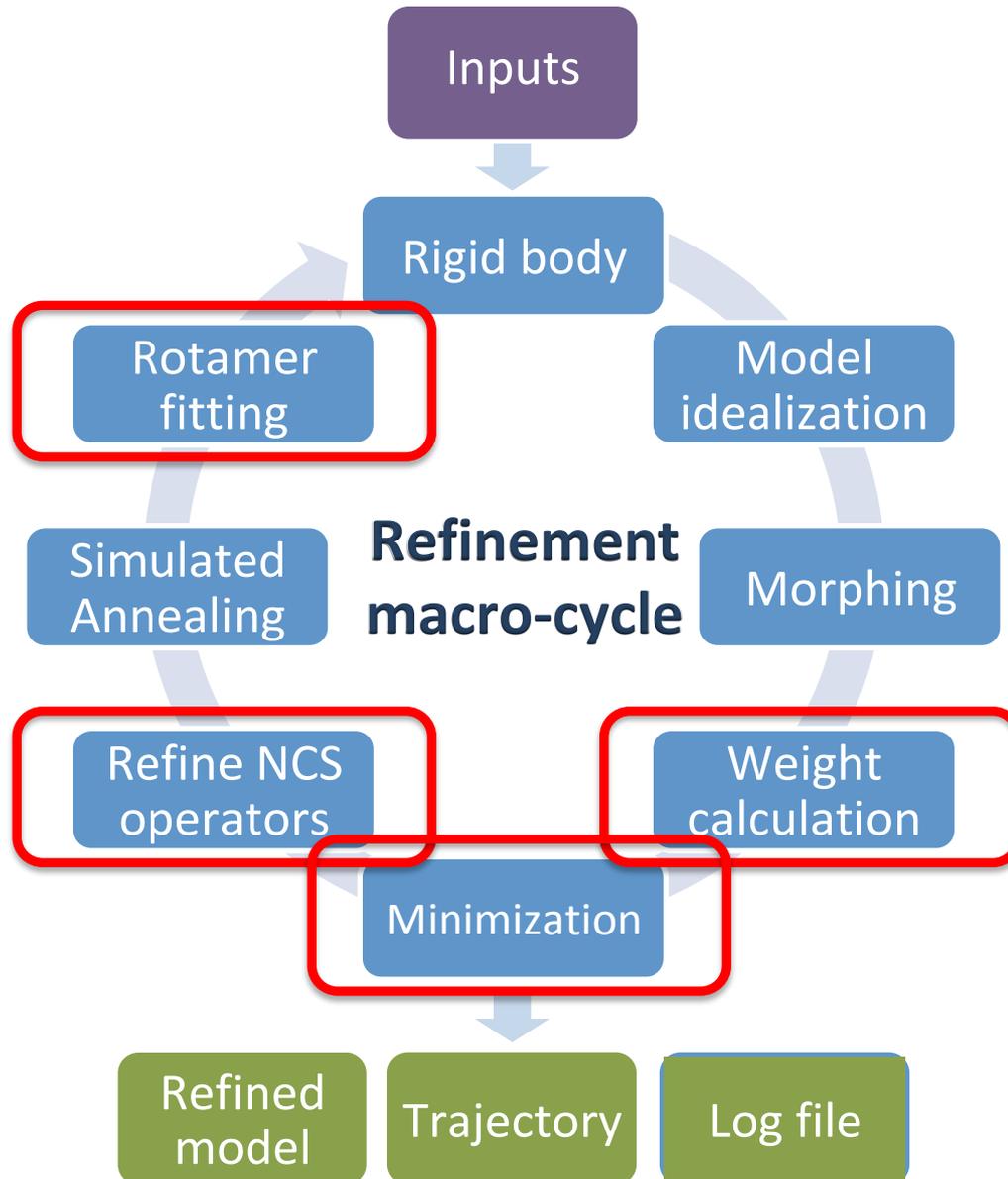
- Calculate one set of ρ_{calc} – **never finished on my laptop**
- Calculate real-space refinement target – **several seconds**

$$T = - \sum_{atoms} \rho(x_{atom}, y_{atom}, z_{atom})$$

Automated model refinement: *phenix.real_space_refine*

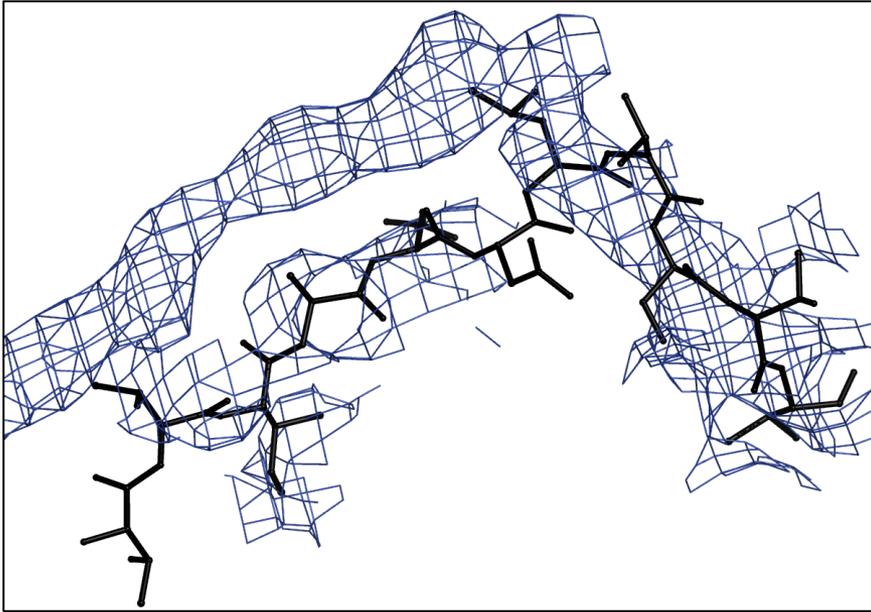


Automated model refinement: *phenix.real_space_refine*

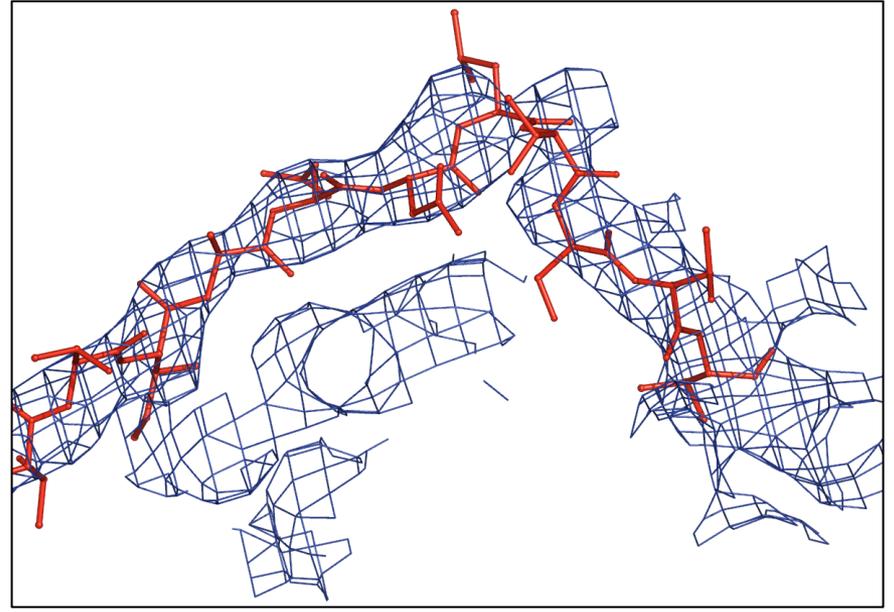


Morphing

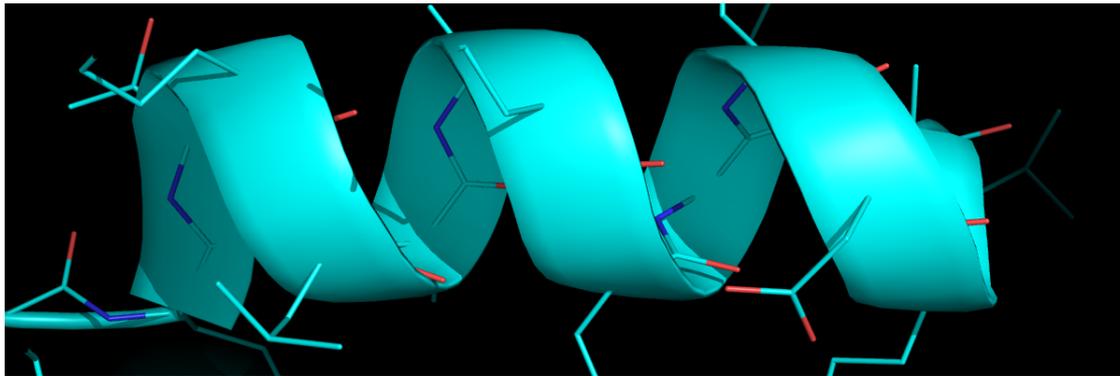
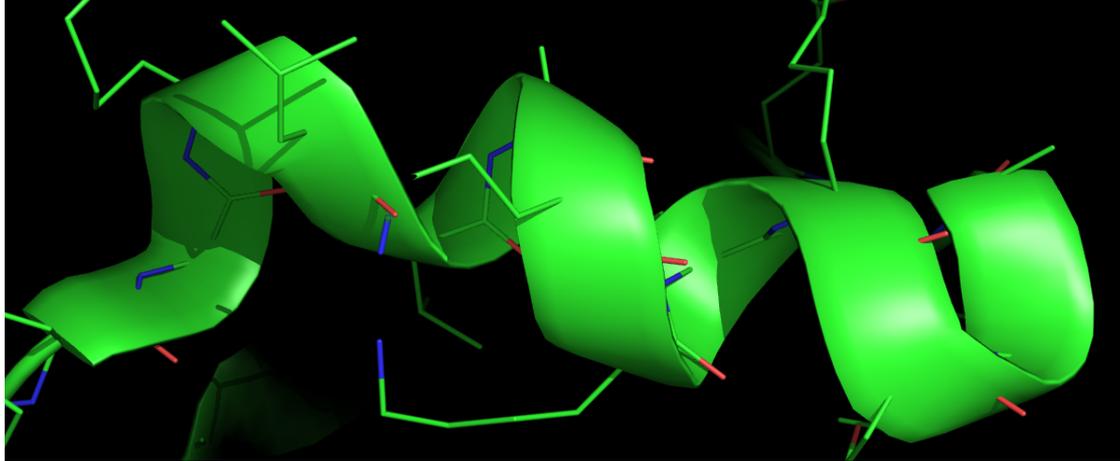
Start model before refinement



After *phenix.real_space_refine*



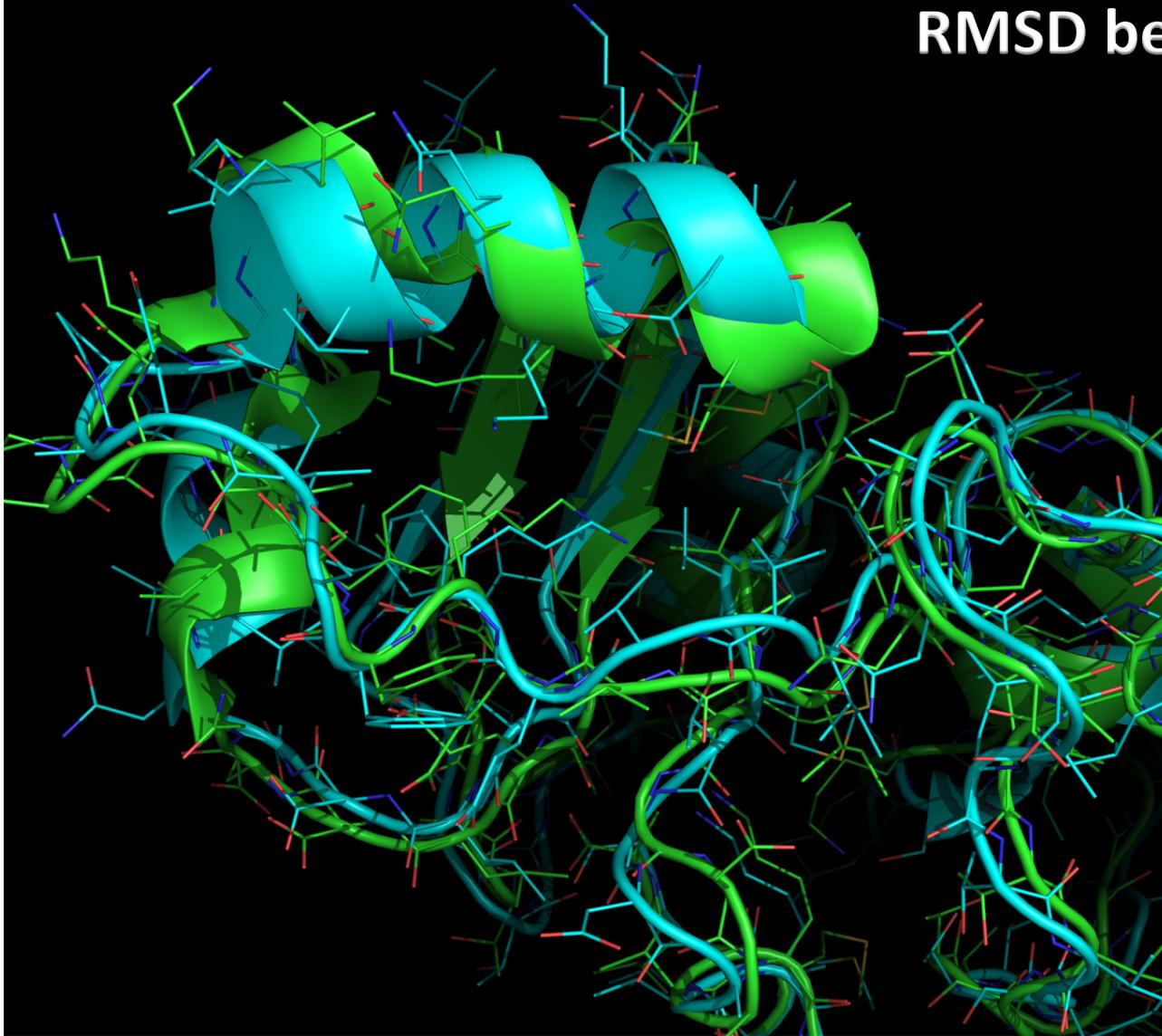
Model regularization



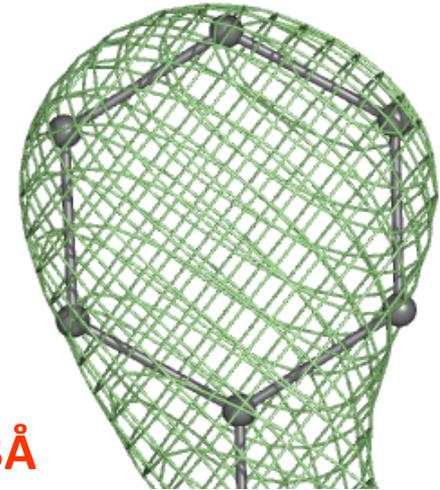
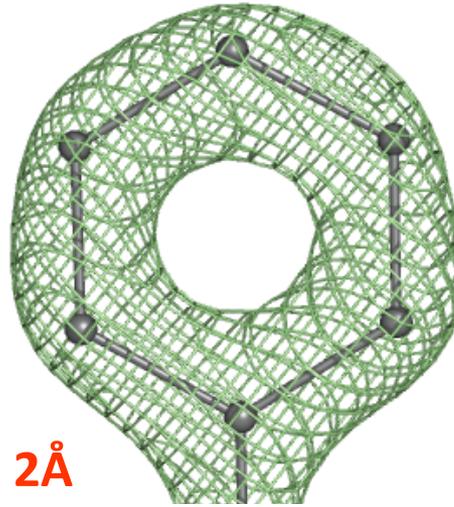
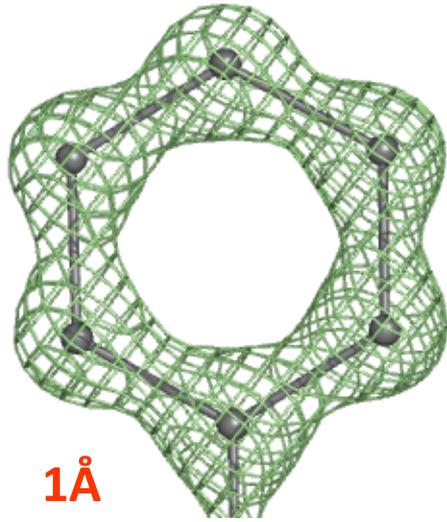
Model regularization

Before and after regularization

RMSD between two models
less than 1.5Å



Restraints



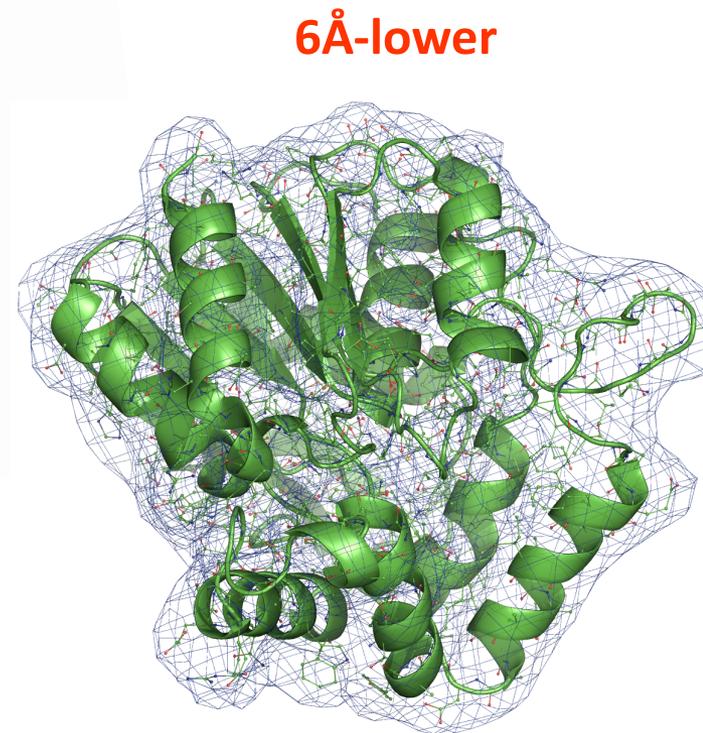
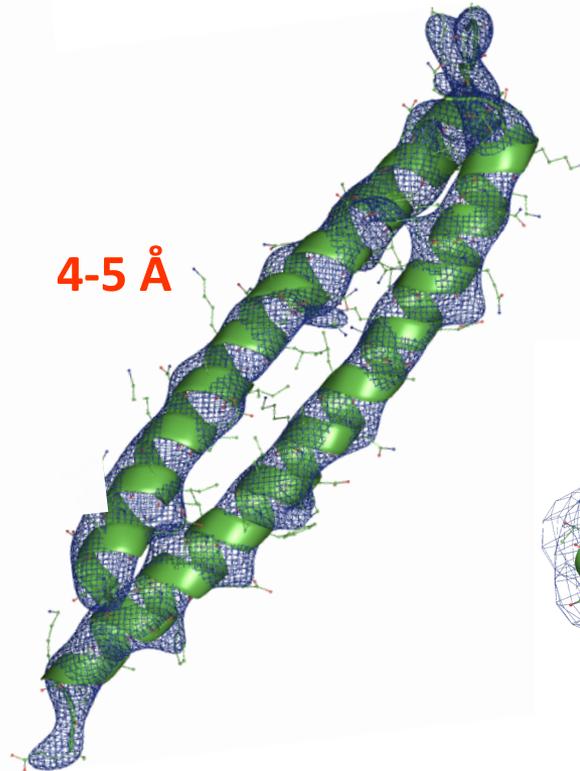
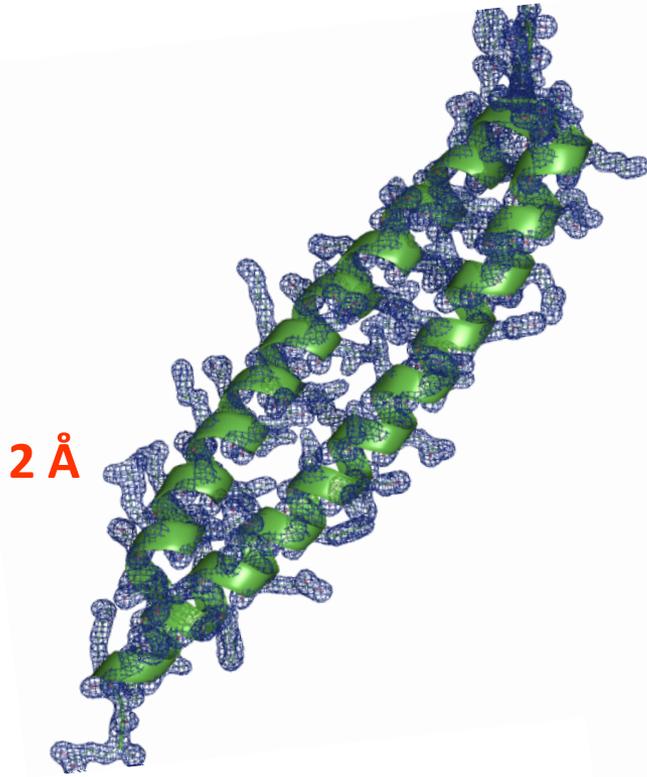
- Lower the resolution, less detailed the map
- Need extra information to keep correct geometry during refinement

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

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

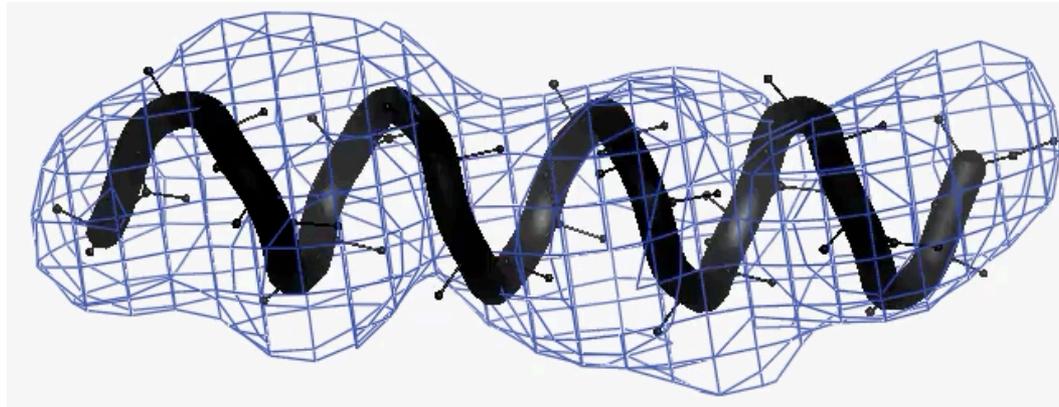
Restraints

- Low resolution map is not sufficient to maintain secondary



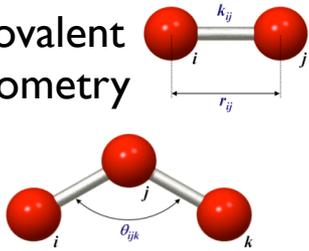
Restraints

- Example: refinement of a perfect α -helix into low-res map
 - Using standard restraints on covalent geometry isn't sufficient
 - Model geometry deteriorates as result of refinement



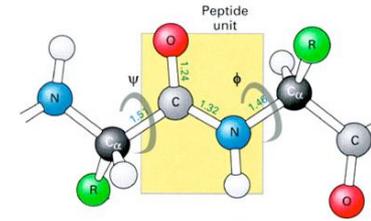
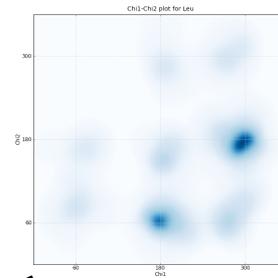
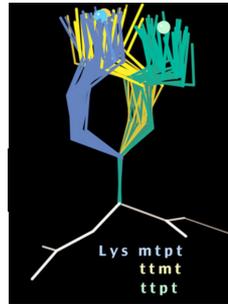
Restraints

Covalent geometry

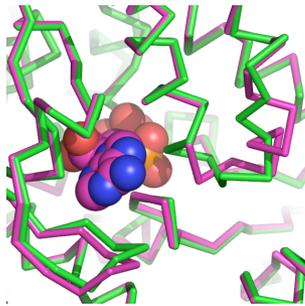
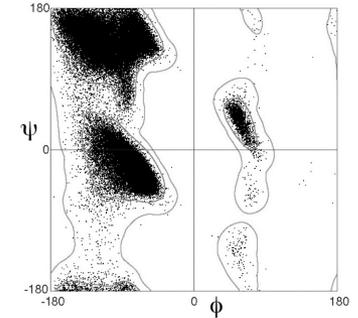


Images from PumMa web site (<http://www.pumma.nl>)

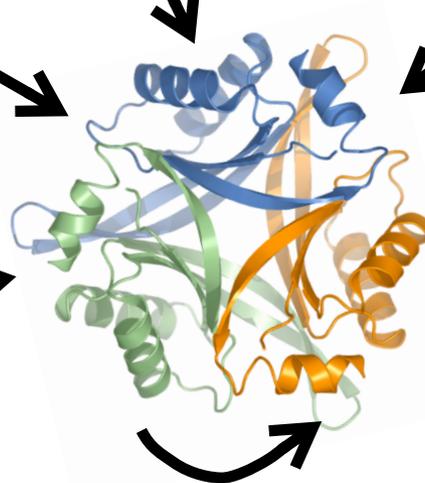
Sidechain distributions



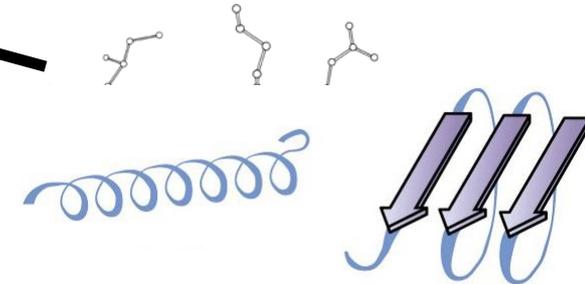
Mainchain distributions



Related structures

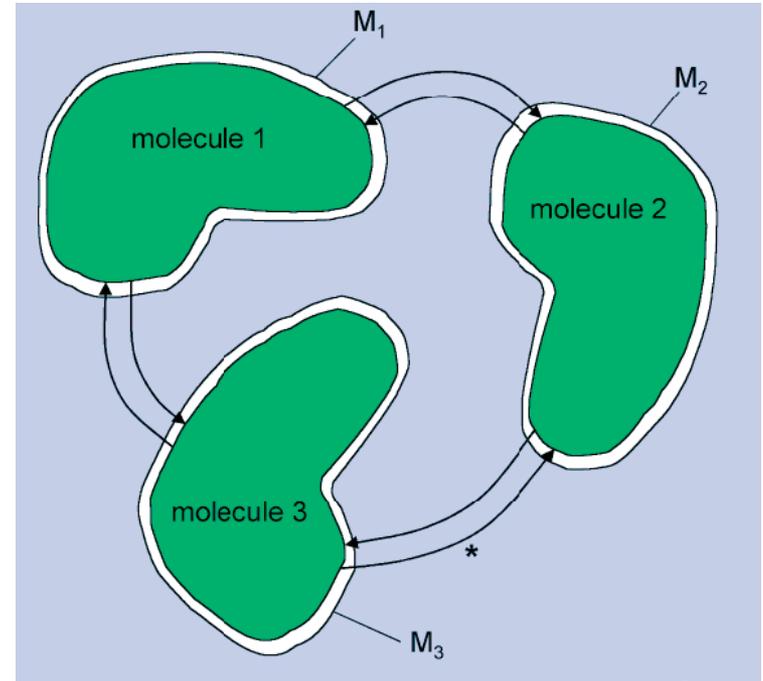
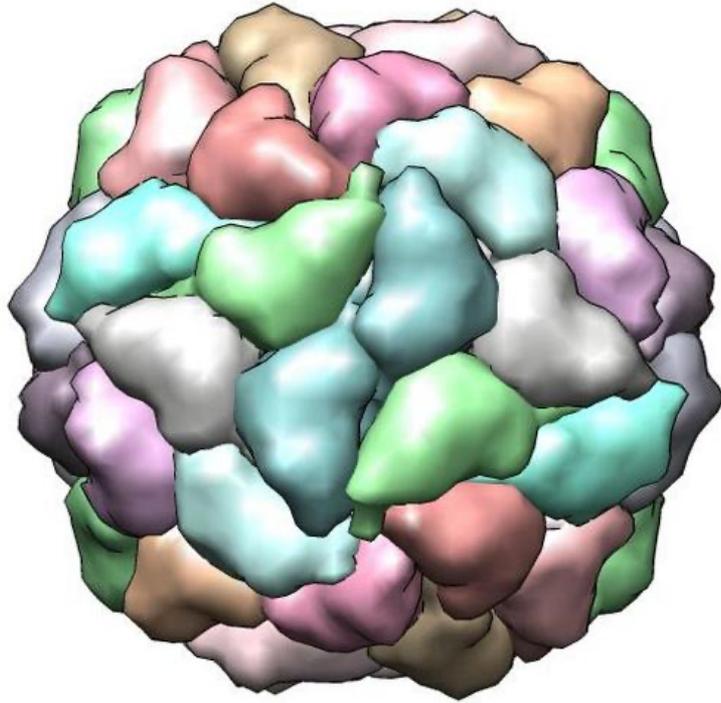


Internal symmetry



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

NCS (internal symmetry): constraints vs restraints



- **Constraints:** molecules 1, 2 and 3 are required to be identical
- **Restraints:** molecules 1, 2 and 3 are required to be similar but not necessarily identical
- *phenix.real_space_refine*: only NCS constraints available

Refinement: practical considerations and misc. facts

- Running with all defaults is ok in most cases
- Minimal required inputs:
 - Model, map and map resolution
 - Map resolution is only used to calculate CC for your information and does not affect refinement results in any way
- Need to adjust parameters
 - Something isn't quite right with the refined model
 - Input model is poor and needs many adjustments to fit the map
- Refinement at 3-3.5 Å and lower, always use:
 - Ramachandran plot restraints
 - Secondary structure restraints
 - Reference model restraints (if quality homology model is available)
 - "NCS" (molecular symmetry)

Refinement: practical considerations and misc facts

- NCS:
 - **Symmetry related copies:**
 - Can be found automatically as part of refinement or using *phenix.simple_ncs_from_pdb* tool
 - Can be specified manually
 - Automatic determination relies on model quality
 - Always check automatically detected NCS copies!
 - **Symmetrized map:**
 - Always use NCS constraints
 - **Symmetry was not used in reconstruction, resolution:**
 - Better than 2.0Å: don't use NCS
 - 2-3.5 (or 4) Å: use NCS restraints

Refinement: practical considerations and misc facts

- Secondary structure (SS) annotation
 - SS information is encoded as HELIX/SHEET records in PDB file or equivalent in mmCIF file
 - *phenix.secondary_structure_restraints* can create SS annotation
 - No software can annotate SS fully reliably and correctly!
- Secondary structure (SS) restraints:
 - Always use at 3Å and worse
 - Better than 3Å: use if needed (model has geometry violations)
 - Setting up SS restraints:
 - Use one of available SS annotation tools to get initial draft
 - Before using in refinement, manually validate and edit initial annotation to make it as accurate as possible
 - Incorrect SS annotation will propagate into refined model

Refinement: practical considerations and misc facts

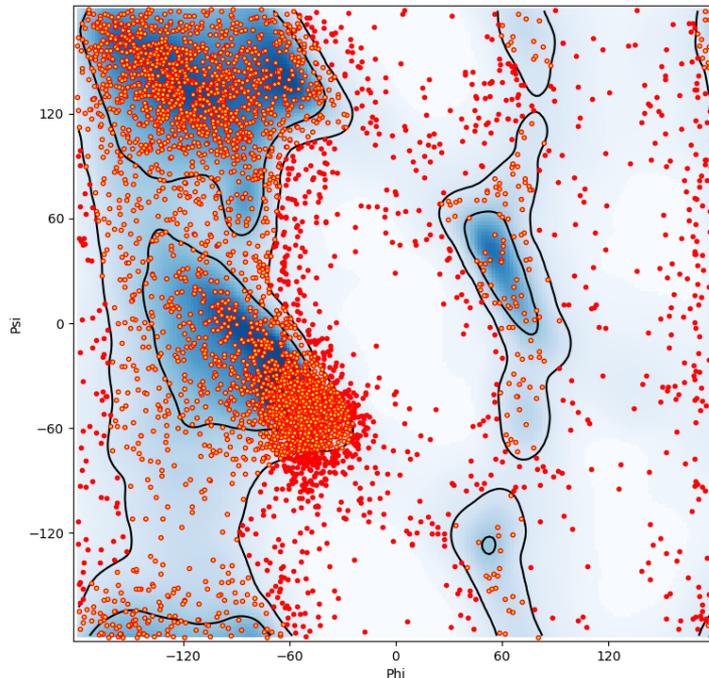
- Ramachandran plot restraints
 - Always use at 3Å and worse
 - Better than 3Å: use if needed (model has geometry violations)
 - Don't use to fix outliers. Fix outliers first, then use Ramachandran plot restraints to stop re-occurring outliers..
 - Check Ramachandran plot regularly (see Validation topic)

Refinement: practical considerations and misc facts

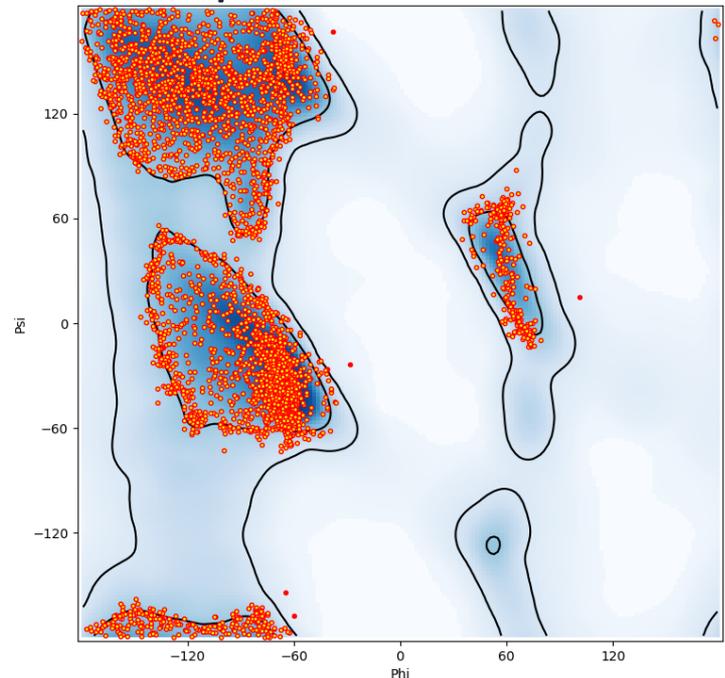
- Ramachandran plot restraints
 - Don't use to fix outliers. Fix outliers first, then use Ramachandran plot restraints to prevent re-occurring outliers..

PDB code: 5a9z

Original



Refined with Ramachandran plot restraints

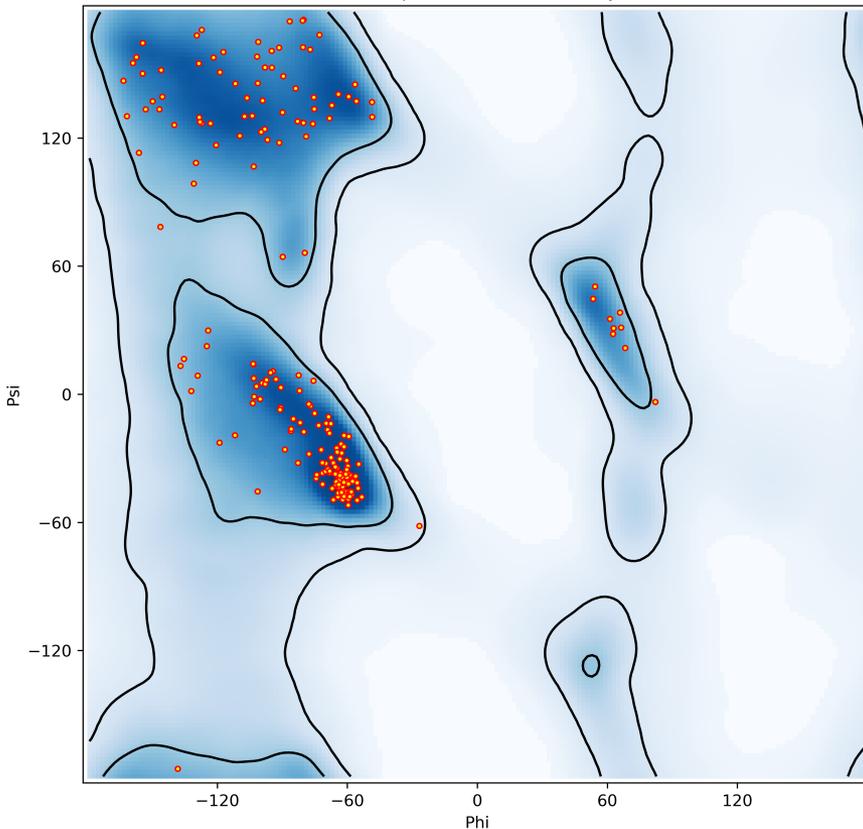


Bad idea to use Ramachandran plot restraints in this case. Fix outliers first!

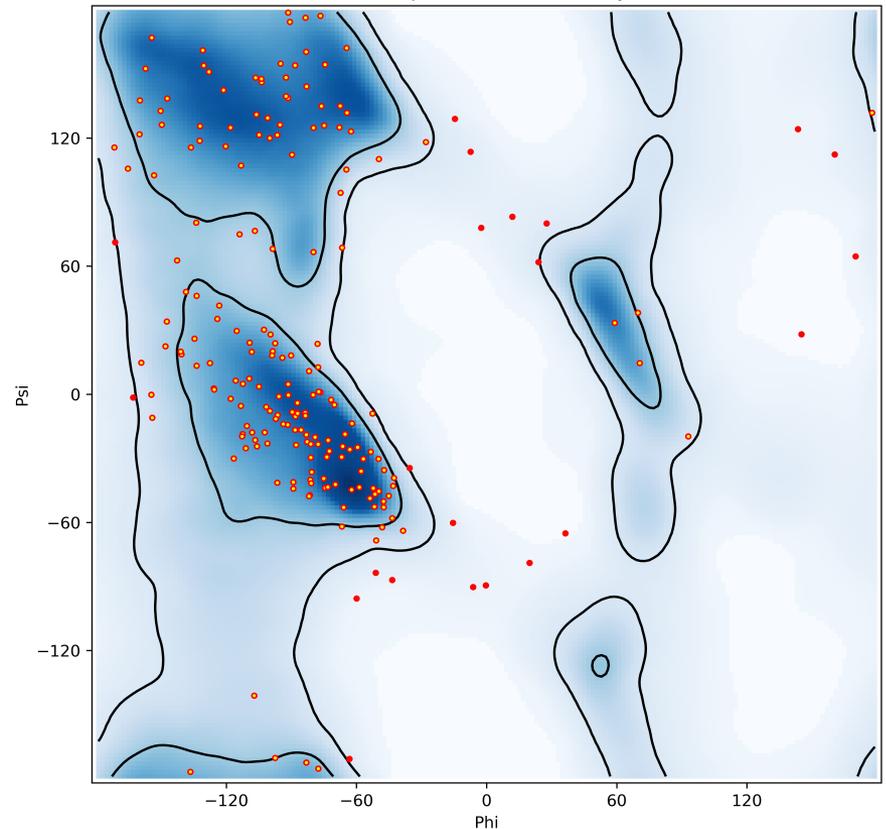
Refinement: practical considerations and misc facts

- Ramachandran plot restraints
 - Use to stop outliers from occurring

Before refinement



After refinement



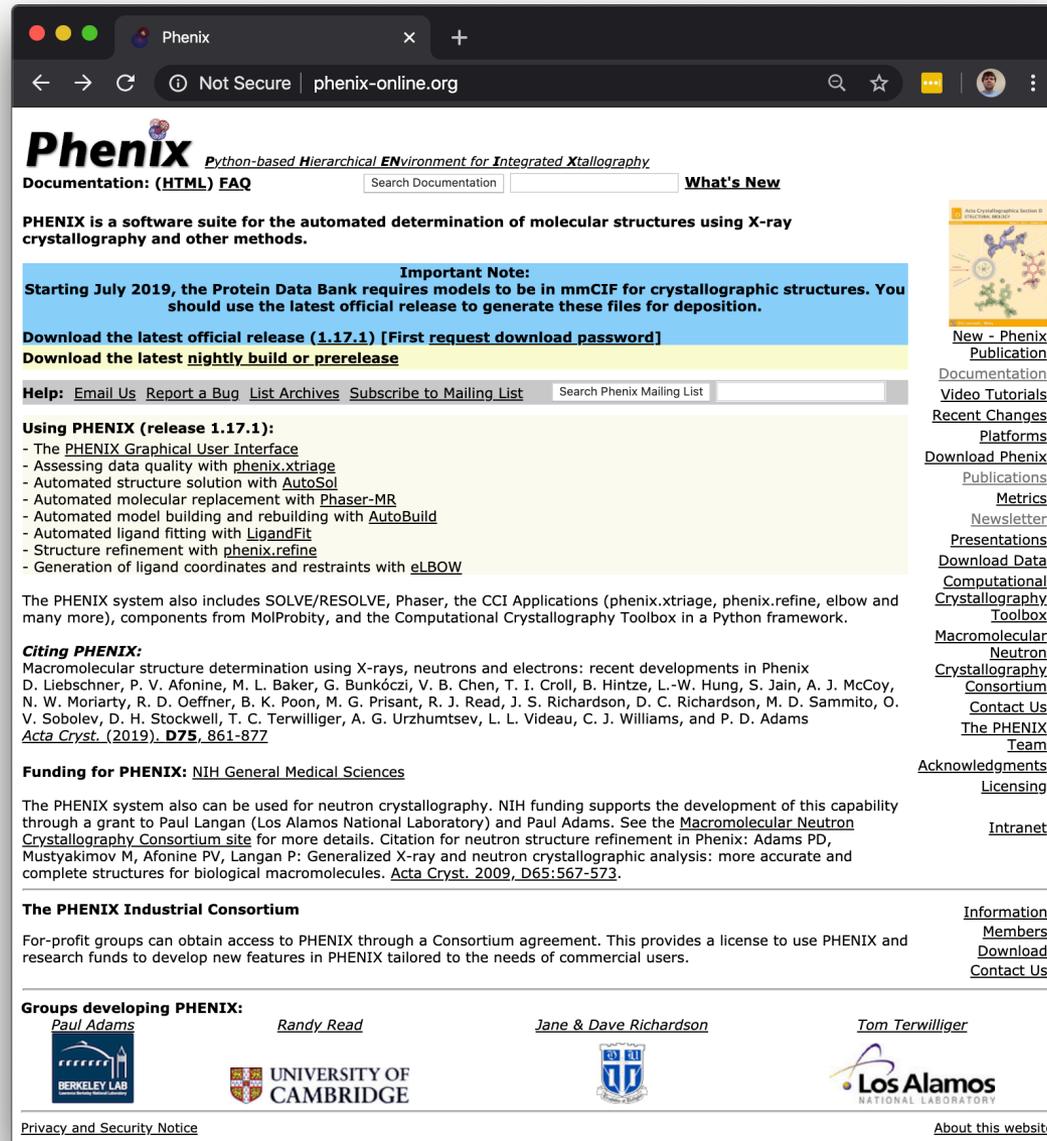
Good idea to use Ramachandran plot restraints!

Refinement: practical considerations and misc facts

- mmCIF file format for atomic models
 - **Mandatory use for crystallographic models since July 2019**
 - PDB formatted files are not accepted any more
 - Some cryo-EM models may be large enough to not fit into PDB file format
 - *Phenix* provides full support for mmCIF I/O

	letters to the editor
 <p>STRUCTURAL BIOLOGY</p> <p>ISSN 2059-7983</p>	Announcing mandatory submission of PDBx/mmCIF format files for crystallographic depositions to the Protein Data Bank (PDB)
Received 21 February 2019 Accepted 3 April 2019	Paul D. Adams,^{a,b} Pavel V. Afonine,^a Kumaran Baskaran,^c Helen M. Berman,^d John Berrisford,^e Gerard Bricogne,^f David G. Brown,^g Stephen K. Burley,^{d,h,i,*} Minyu Chen,^j Zukang Feng,^d Claus Flensburg,^f Aleksandras Gutmanas,^e Jeffrey C. Hoch,^{k,*} Yasuyo Ikegawa,^j Yumiko Kengaku,^j Eugene Krissinel,^l Genji Kurisu,^{j,*} Yuhe Liang,^d Dorothee Liebschner,^a Lora Mak,^e John L. Markley,^{c,*} Nigel W. Moriarty,^a Garib N. Murshudov,^m Martin Noble,ⁿ Ezra Peisach,^d Irina Persikova,^d Billy K. Poon,^a Oleg V. Sobolev,^a Eldon L. Ulrich,^c Sameer Velankar,^{e,*} Clemens Vornrhein,^f John Westbrook,^d Marcin Wojdyr,^{f,l} Masashi Yokochi^j and Jasmine Y. Young^d
Edited by R. J. Read, University of Cambridge, England	

Resources



The screenshot shows the Phenix website homepage. At the top, there's a navigation bar with the Phenix logo and the tagline "Python-based Hierarchical ENvironment for Integrated Xtallography". Below this, there are links for "Documentation: (HTML) FAQ", a search box for documentation, and "What's New". A prominent blue box contains an "Important Note" about the Protein Data Bank requirements starting in July 2019. Below that, there are links to download the latest official release (1.17.1) and the latest nightly build or prerelease. A "Help" section includes links for "Email Us", "Report a Bug", "List Archives", and "Subscribe to Mailing List". A "Using PHENIX (release 1.17.1):" section lists various tools and features like PHENIX Graphical User Interface, phenix.xtriage, AutoSol, Phaser-MR, AutoBuild, LigandFit, phenix.refine, and eLBOW. A paragraph describes the system's components, including SOLVE/RESOLVE, Phaser, CCI Applications, MolProbity, and the Computational Crystallography Toolbox. A "Citing PHENIX:" section provides a list of authors and a citation from Acta Cryst. (2019), D75, 861-877. "Funding for PHENIX:" is attributed to NIH General Medical Sciences. A paragraph explains the grant from Paul Langan and Paul Adams, with a citation from Acta Cryst. 2009, D65:567-573. The "The PHENIX Industrial Consortium" section mentions access for for-profit groups. At the bottom, "Groups developing PHENIX:" lists Paul Adams (Berkeley Lab), Randy Read (University of Cambridge), Jane & Dave Richardson (University of Cambridge), and Tom Terwilliger (Los Alamos National Laboratory). Footer links include "Privacy and Security Notice" and "About this website".

Phenix Python-based Hierarchical ENvironment for Integrated Xtallography
Documentation: (HTML) FAQ Search Documentation What's New

PHENIX is a software suite for the automated determination of molecular structures using X-ray crystallography and other methods.

Important Note:
Starting July 2019, the Protein Data Bank requires models to be in mmCIF for crystallographic structures. You should use the latest official release to generate these files for deposition.

Download the latest official release (1.17.1) [First request download password]
Download the latest nightly build or prerelease

Help: Email Us Report a Bug List Archives Subscribe to Mailing List Search Phenix Mailing List

Using PHENIX (release 1.17.1):

- The PHENIX Graphical User Interface
- Assessing data quality with phenix.xtriage
- Automated structure solution with AutoSol
- Automated molecular replacement with Phaser-MR
- Automated model building and rebuilding with AutoBuild
- Automated ligand fitting with LigandFit
- Structure refinement with phenix.refine
- Generation of ligand coordinates and restraints with eLBOW

The PHENIX system also includes SOLVE/RESOLVE, Phaser, the CCI Applications (phenix.xtriage, phenix.refine, elbow and many more), components from MolProbity, and the Computational Crystallography Toolbox in a Python framework.

Citing PHENIX:
Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L.-W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, and P. D. Adams *Acta Cryst.* (2019), **D75**, 861-877

Funding for PHENIX: NIH General Medical Sciences

The PHENIX system also can be used for neutron crystallography. NIH funding supports the development of this capability through a grant to Paul Langan (Los Alamos National Laboratory) and Paul Adams. See the [Macromolecular Neutron Crystallography Consortium site](#) for more details. Citation for neutron structure refinement in Phenix: Adams PD, Mustyakimov M, Afonine PV, Langan P: Generalized X-ray and neutron crystallographic analysis: more accurate and complete structures for biological macromolecules. *Acta Cryst.* 2009, **D65**:567-573.

The PHENIX Industrial Consortium

For-profit groups can obtain access to PHENIX through a Consortium agreement. This provides a license to use PHENIX and research funds to develop new features in PHENIX tailored to the needs of commercial users.

Groups developing PHENIX:

- Paul Adams
- Randy Read
- Jane & Dave Richardson
- Tom Terwilliger

BERKELEY LAB UNIVERSITY OF CAMBRIDGE Los Alamos NATIONAL LABORATORY

Privacy and Security Notice About this website

Phenix paper
Video tutorials
Documentation
Recent Changes
Platforms
Download Phenix Publications
Metrics
Newsletter
Presentations
Download Data
Computational Crystallography Toolbox
Macromolecular Neutron Crystallography Consortium
Contact Us
The PHENIX Team
Acknowledgments
Licensing
Intranet
Information
Members
Download
Contact Us

Resources

PHENIX Documentation

phenix-online.org/documentation/

Phenix Python-based Hierarchical Environment for Integrated Xtallography

Phenix Documentation - 1.17.1-3660

- [Phenix programs and their functions](#)
- [The Phenix graphical interface](#)
- [Overview of video tutorials](#)
- [Tutorials and Examples](#)
- [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

```
graph TD; DQA[Data Quality Assessment] --> EP[Experimental Phasing]; DQA --> MR[Molecular Replacement]; EP --> DM[Density Modification]; MR --> DM; DM --> MRB[Model (Re)building]; MRB --> RV[Refinement & Validation]; RV --> LF[Ligand Fitting]; LF --> MRB; RV --> SD[Structure Deposition];
```

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](#) | [All](#)

Phenix Documentation for Neutron Crystallography

- [Structure refinement](#) | [Structure validation](#) | [Making geometry restraints](#) | [Structure deposition](#) | [All](#)

Cryo-EM Structure Solution with Phenix

```
graph TD; ASM[Auto-sharpen Map] --> DMI[Dock Model into Map]; ASM --> AMI[Autobuild Model into Map]; DMI --> RSR[Real Space Refinement]; AMI --> RSR; RSR --> VMM[Validation of Model and Map]; VMM --> SD[Structure Deposition];
```

Phenix Documentation for Electron Microscopy (EM)

- [Structure refinement](#) | [Map statistics \(resolution, etc\)](#) | [Auto-sharpen a map](#) | [Dock a model into a map](#) | [Rapid model-building](#) | [Guess sequences from map](#) | [Flexibly fit a model to a map](#) | [Build a model](#) | [Making geometry restraints](#) | [Compare CA/P in two models](#) | [Identify symmetry in a map](#) | [Combine best parts of focused maps](#) | [Extract box with map and model](#) | [Convert map to structure factors](#) | [Segment a map](#) | [Adjust a CA/CB model](#) | [Fix register errors](#) | [All](#)

User support

- 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
 - Make sure the problem still exist using the latest *Phenix* version
 - Send us all inputs (files, non-default parameters) and tell us steps that lead to the problem

Subscribe to *Phenix* mailing list: www.phenix-online.org