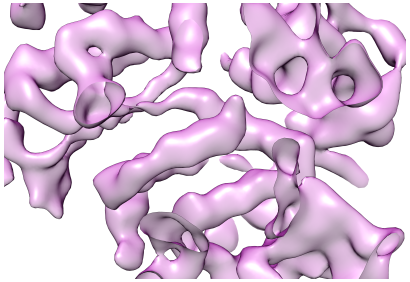
Phenix Tools for Cryo-EM: Refinement

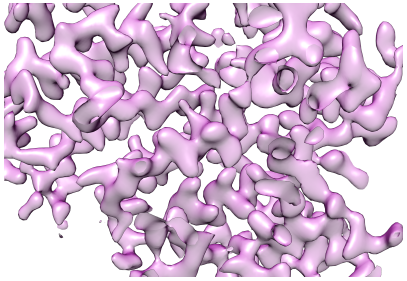
Pavel Afonine

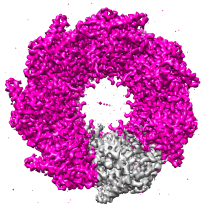
LBNL, Berkeley, California, USA

January 30, 2020

Cryo-EM tools in *Phenix*





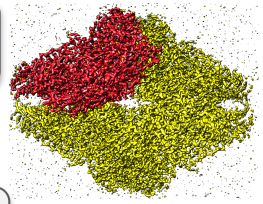


Starting map:

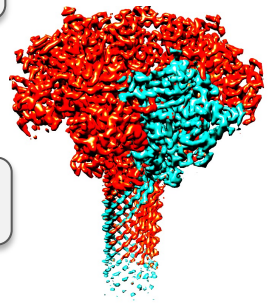
Map analysis

Map improvement

Map symmetry

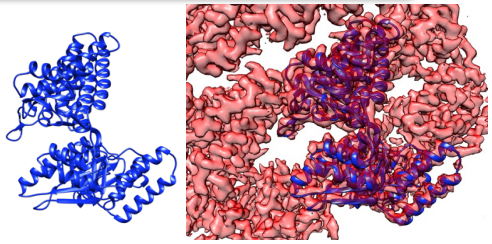


Map manipulation

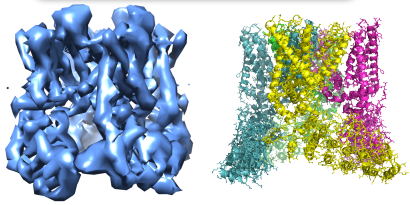


Extract unique part

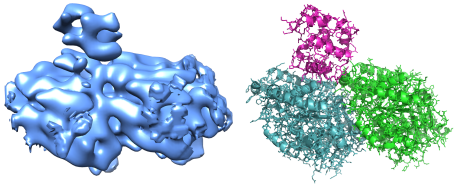
Docking, model building



Refinement



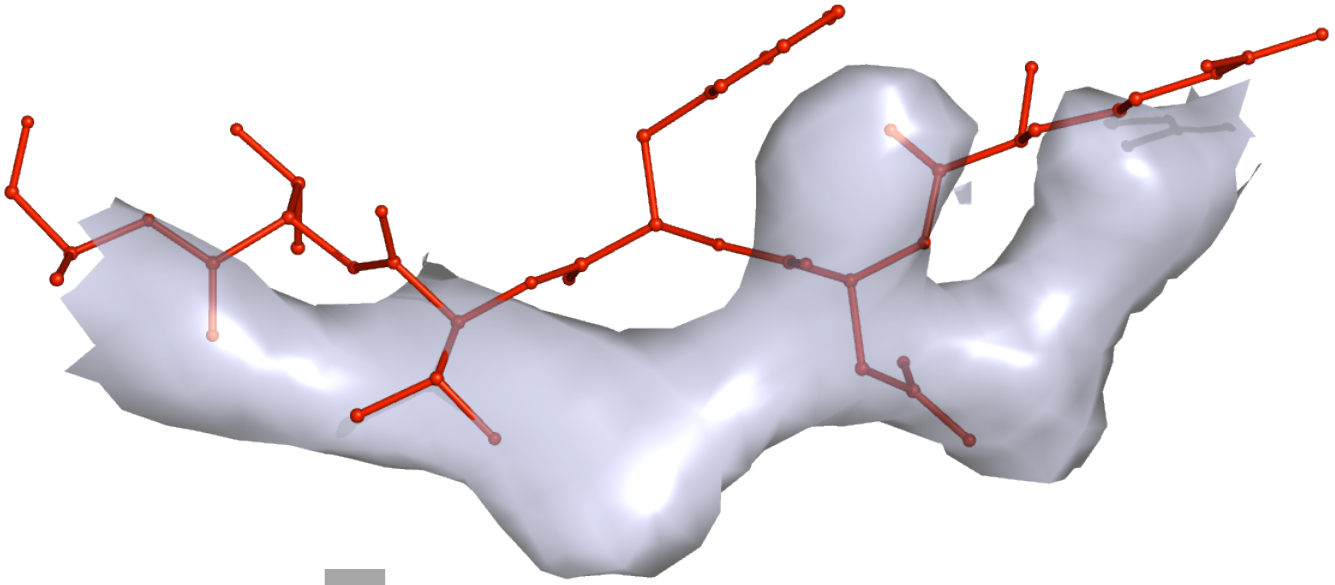
Validation



Complete set of tools for cryo-EM structure solution: from initial reconstruction to final validated model

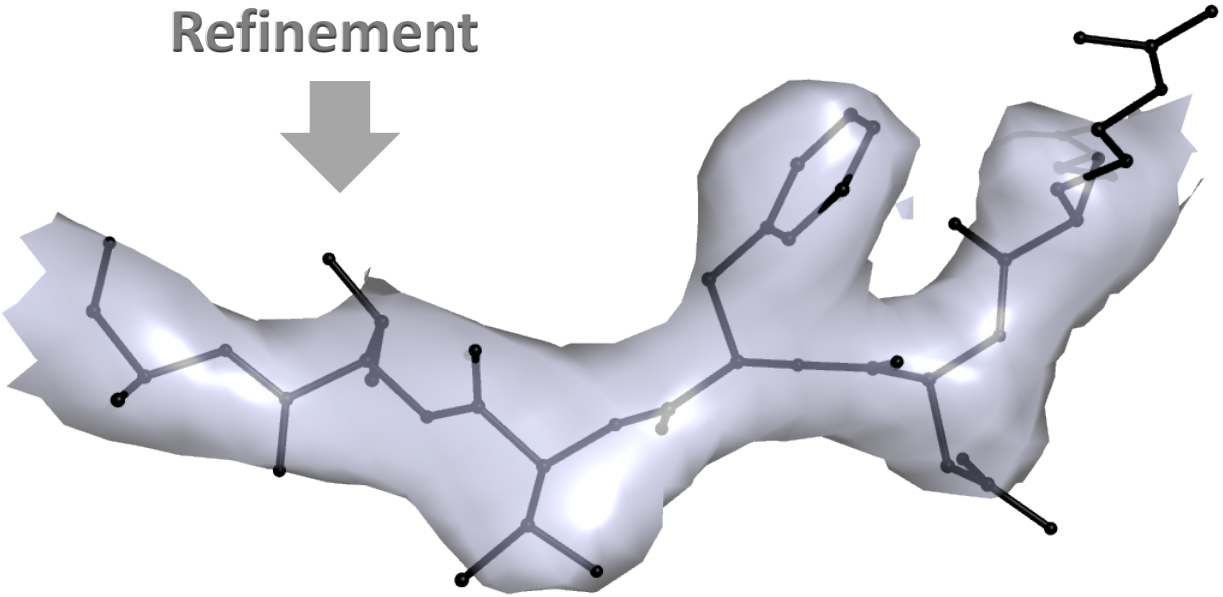
Structure refinement

Initial (poor) model



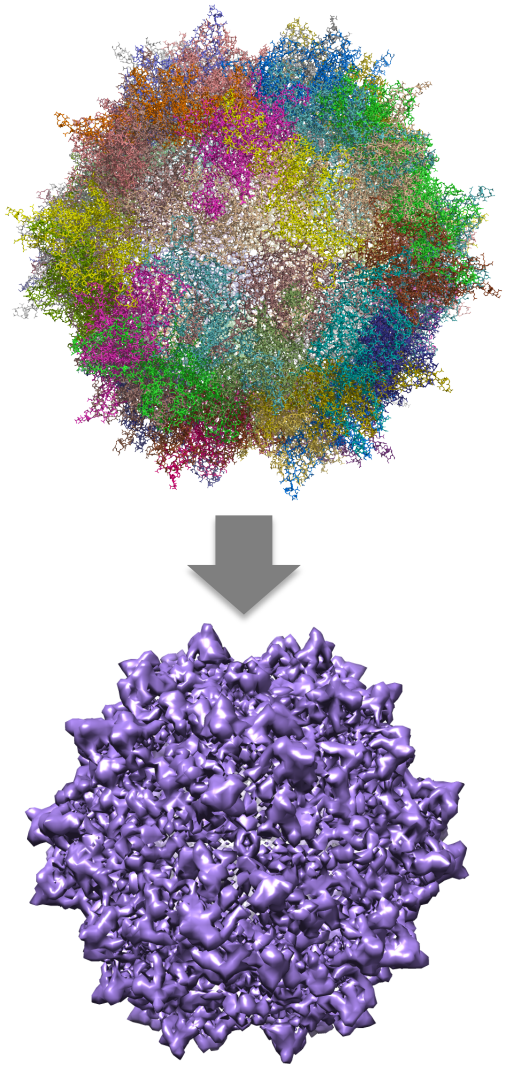
Refinement

Improved (refined) model

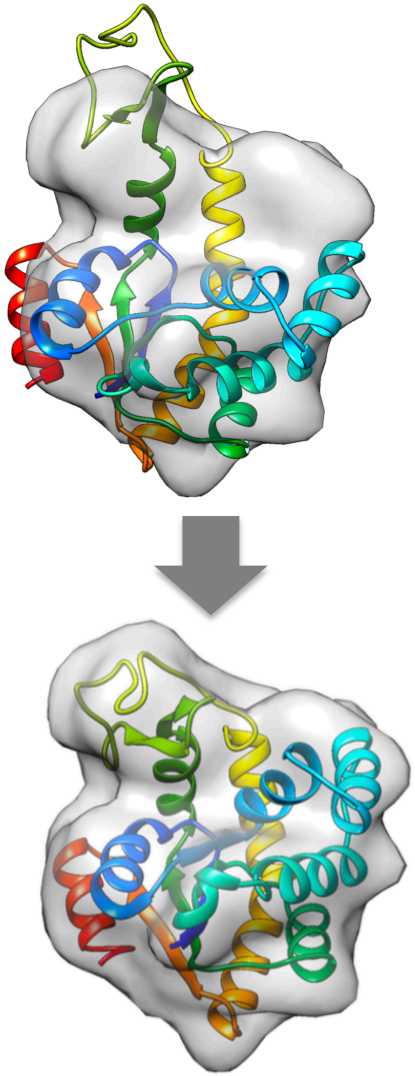


Scenarios for model-to-map fitting

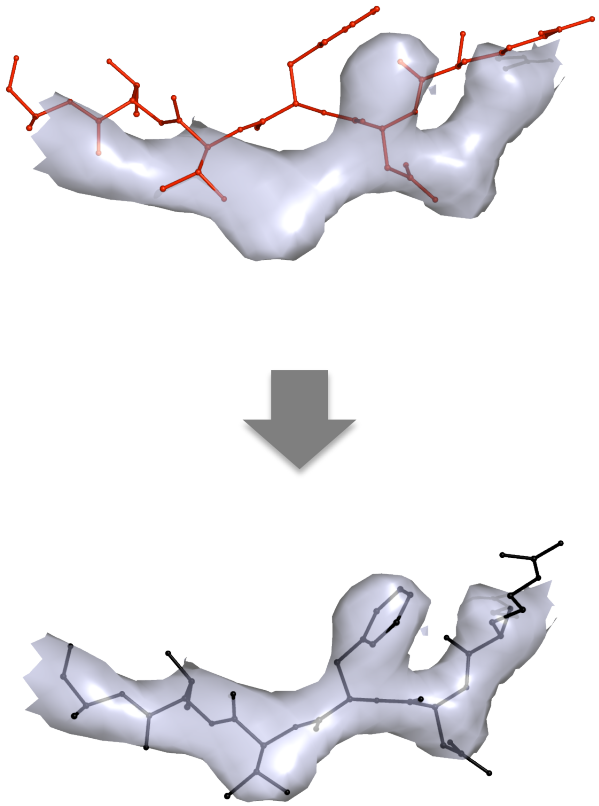
Docking



Flexible fitting, morphing

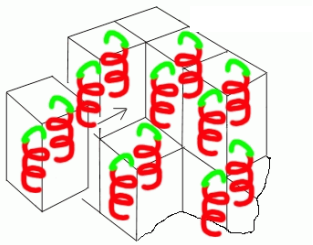


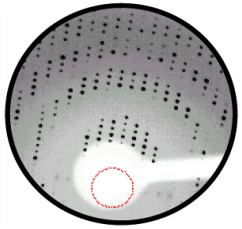
Refinement



Refinement tools in *Phenix*

Crystallography





Initial model

Experimental data

A priori knowledge

Score

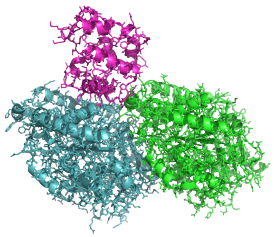
Modify model parameters

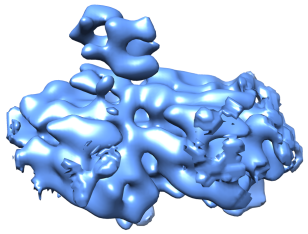
Improved model

phenix.refine

Available since 2005

Cryo-EM





Initial model

Experimental data

A priori knowledge

Score

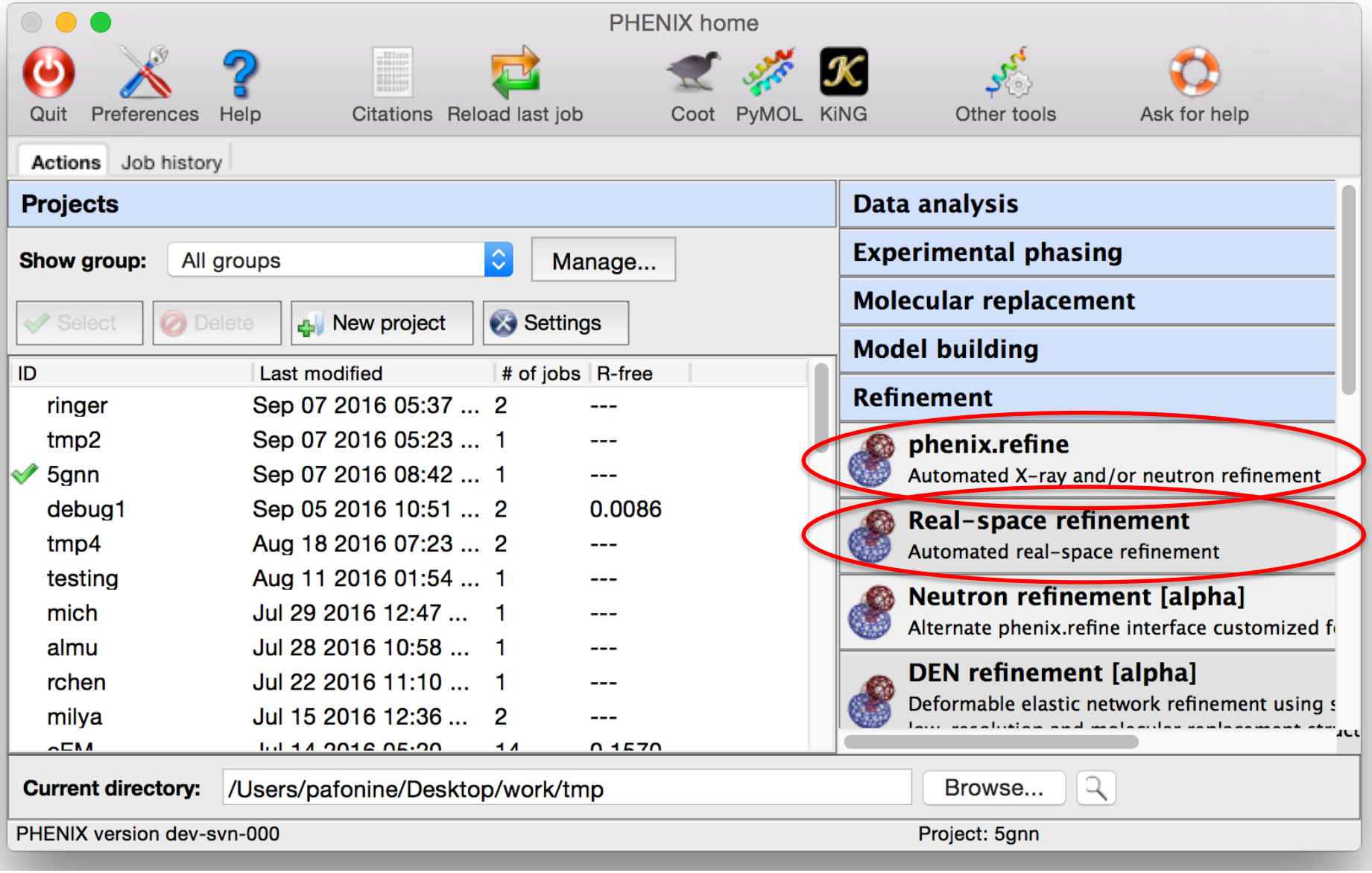
Modify model parameters

Improved model

phenix.real_space_refine

Available since 2013

Refinement tools in *Phenix*



Real-space refinement with *phenix.real_space_refine*

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

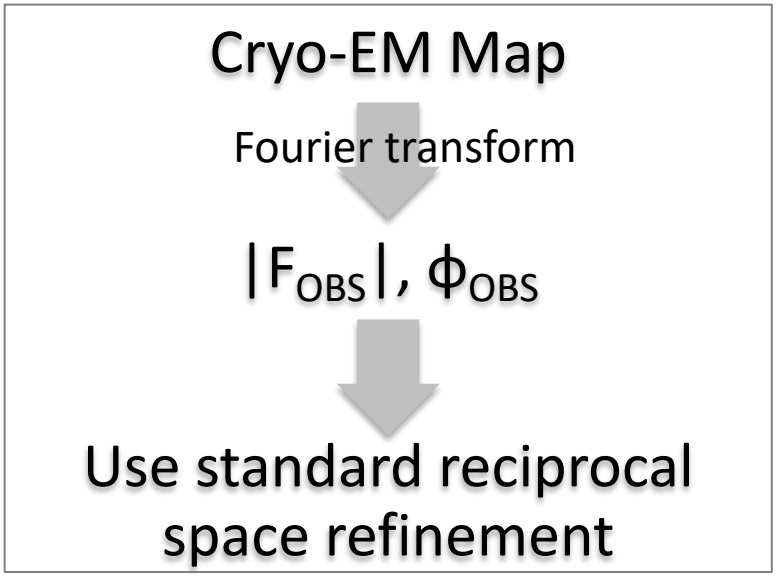
Acta Cryst D STRUCTURAL BIOLOGY
ISSN 2059-7983

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]

Repurposing crystallographic software for cryo-EM

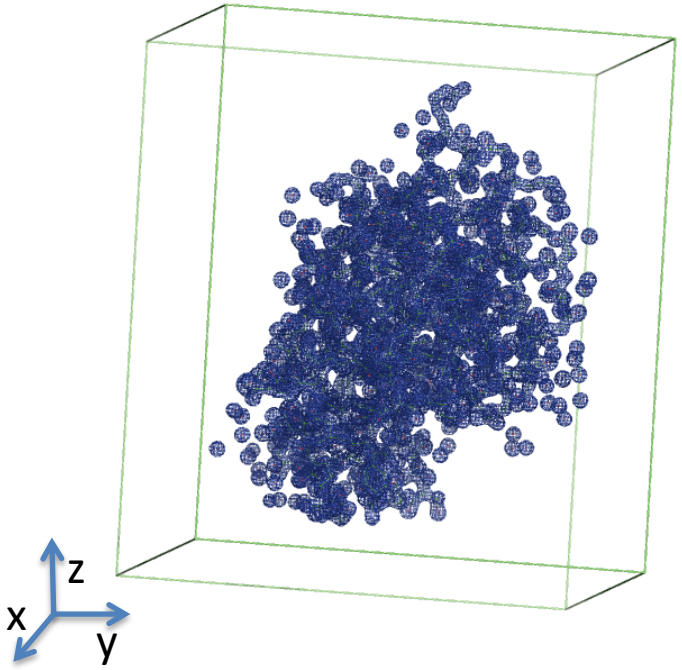
Legacy procedure



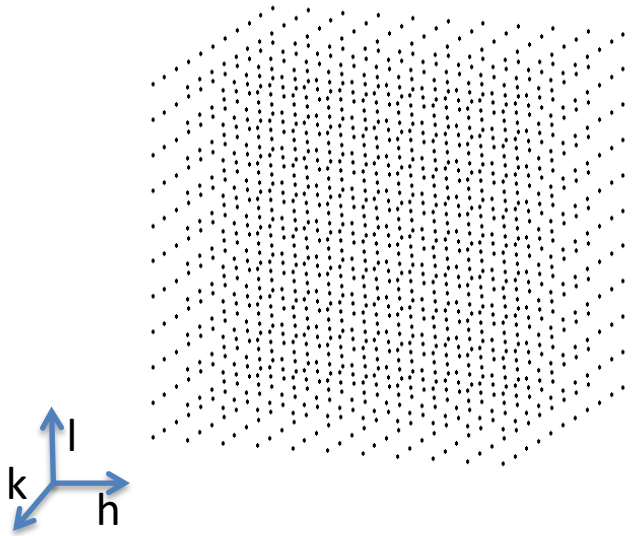
- Is conversion map to structure factors lossless?
- Are crystal bulk-solvent and anisotropic scaling still applicable?
- *R* factors will largely depend on masking used. What is R_{free}?
- 2mFo-DFc and mFo-DFc maps?
- Form-factors ?

Map to structure factors conversion

Map in real space (ρ)



Map in Fourier space (F)

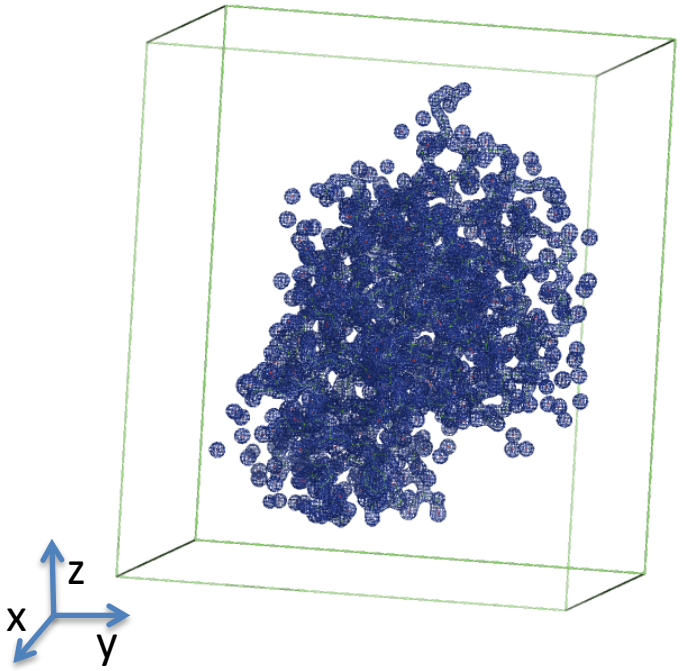


$$\rho(\mathbf{r}) = \frac{1}{V_{cell}} \sum_{h_{min}}^{h_{max}} \sum_{k_{min}}^{k_{max}} \sum_{l_{min}}^{l_{max}} F(\mathbf{s}) \exp(-2\pi i \mathbf{s}\mathbf{r})$$

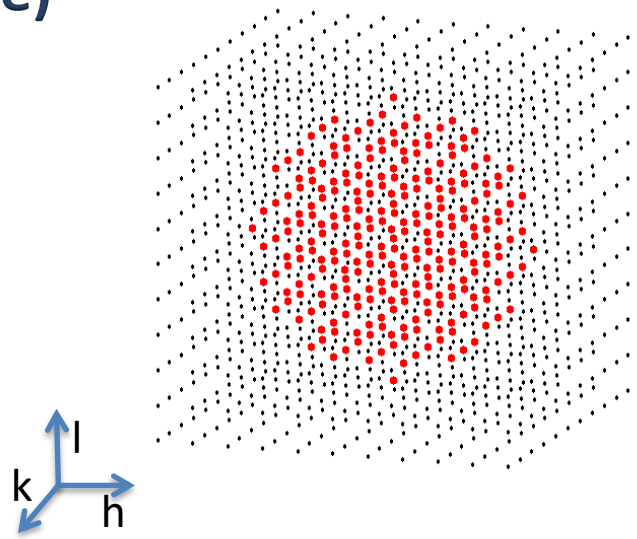
- **Lossless conversion: ρ <> F**

Map to structure factors conversion

Map in real space (ρ)



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



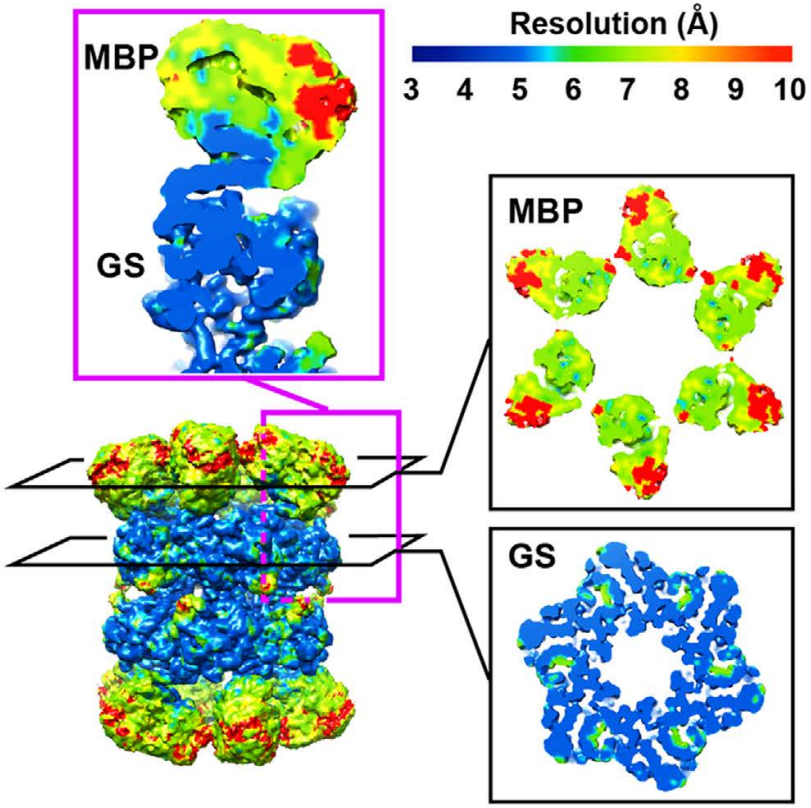
- **Not** a lossless conversion: $\rho <> F$

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**



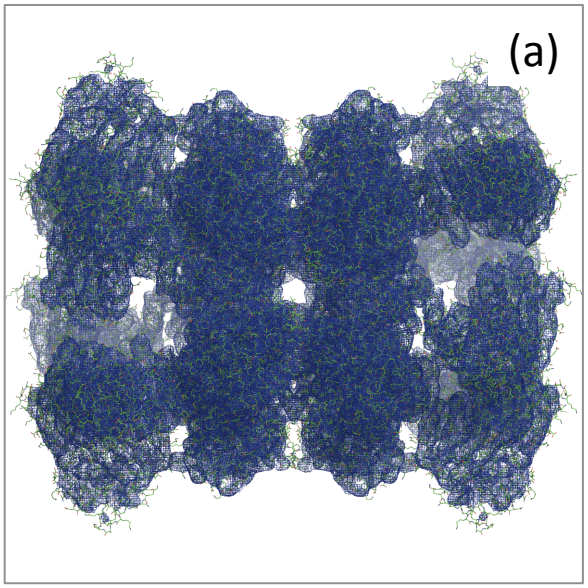
Coscia et al (2016). Sci. Reports, 6, 30909

Resolution (map quality) varies across the volume

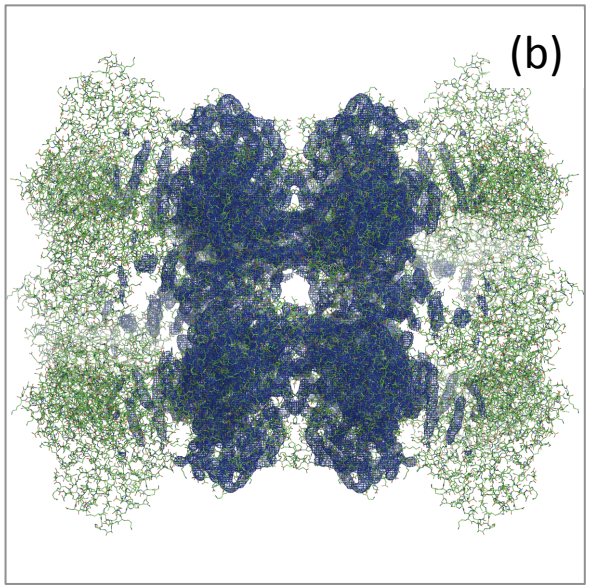
Real-space refinement: local data/restraints weight

- **5LDF**

Low contouring level



High contouring level



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

- **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:**

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

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*

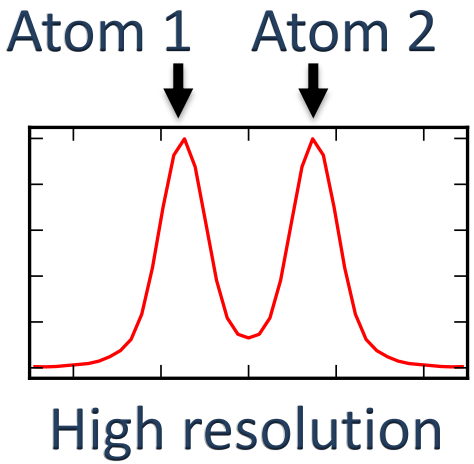
Refinement target

- **Atom-centered:**

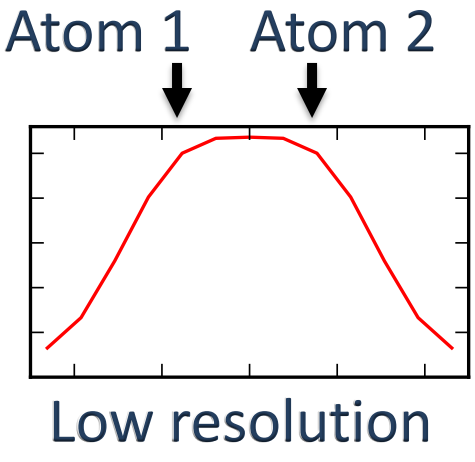
$$T = -\sum_{atoms} \rho_{obs}(x_{atom}, y_{atom}, z_{atom})$$

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

- Why it is less accurate?



High resolution



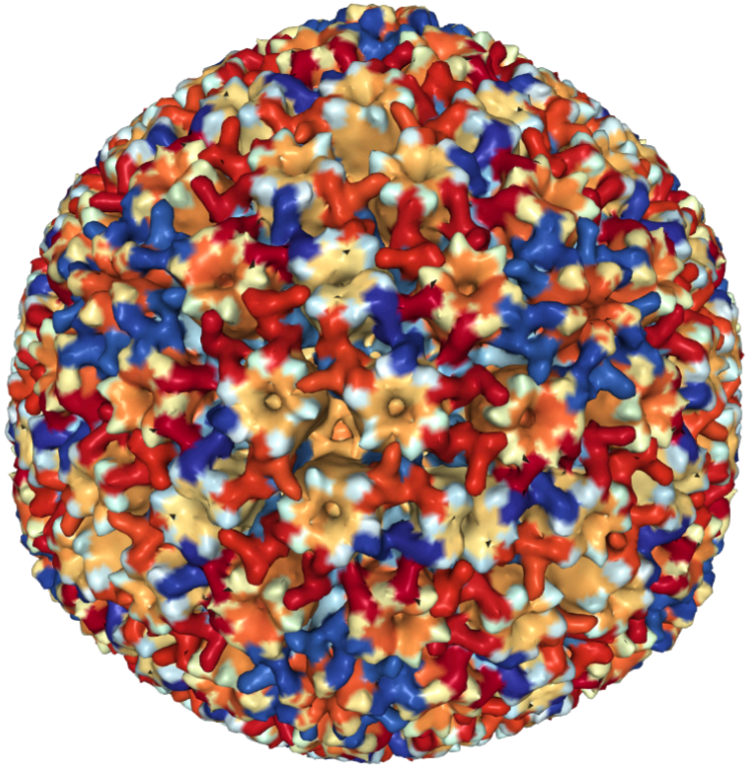
Low resolution

Moving atoms to nearest peaks ≠ making correct model

Target calculation example

- **PDB: 5VKU**

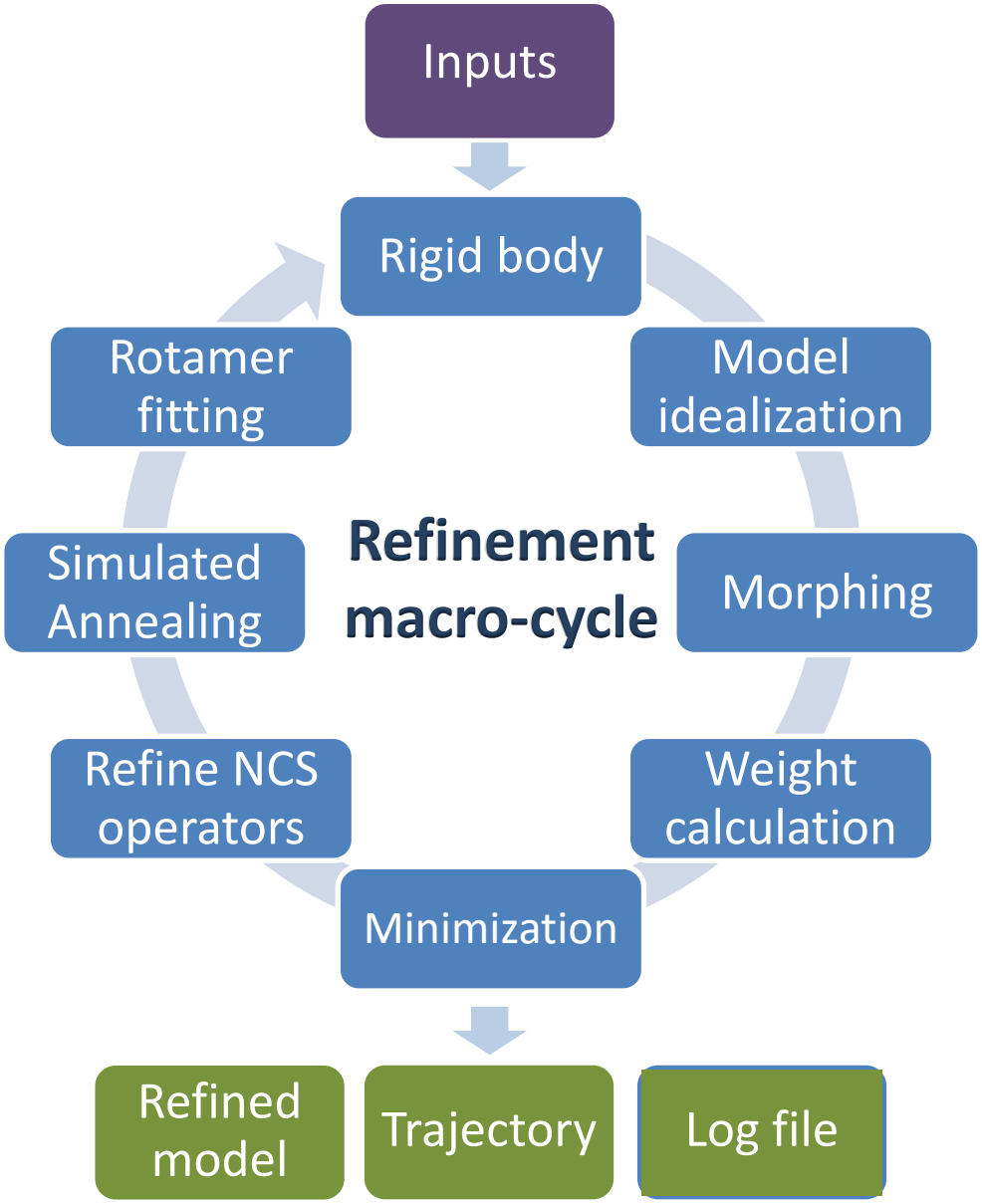
  3720 chains | 1,872,060 residues | 14,917,620 atoms



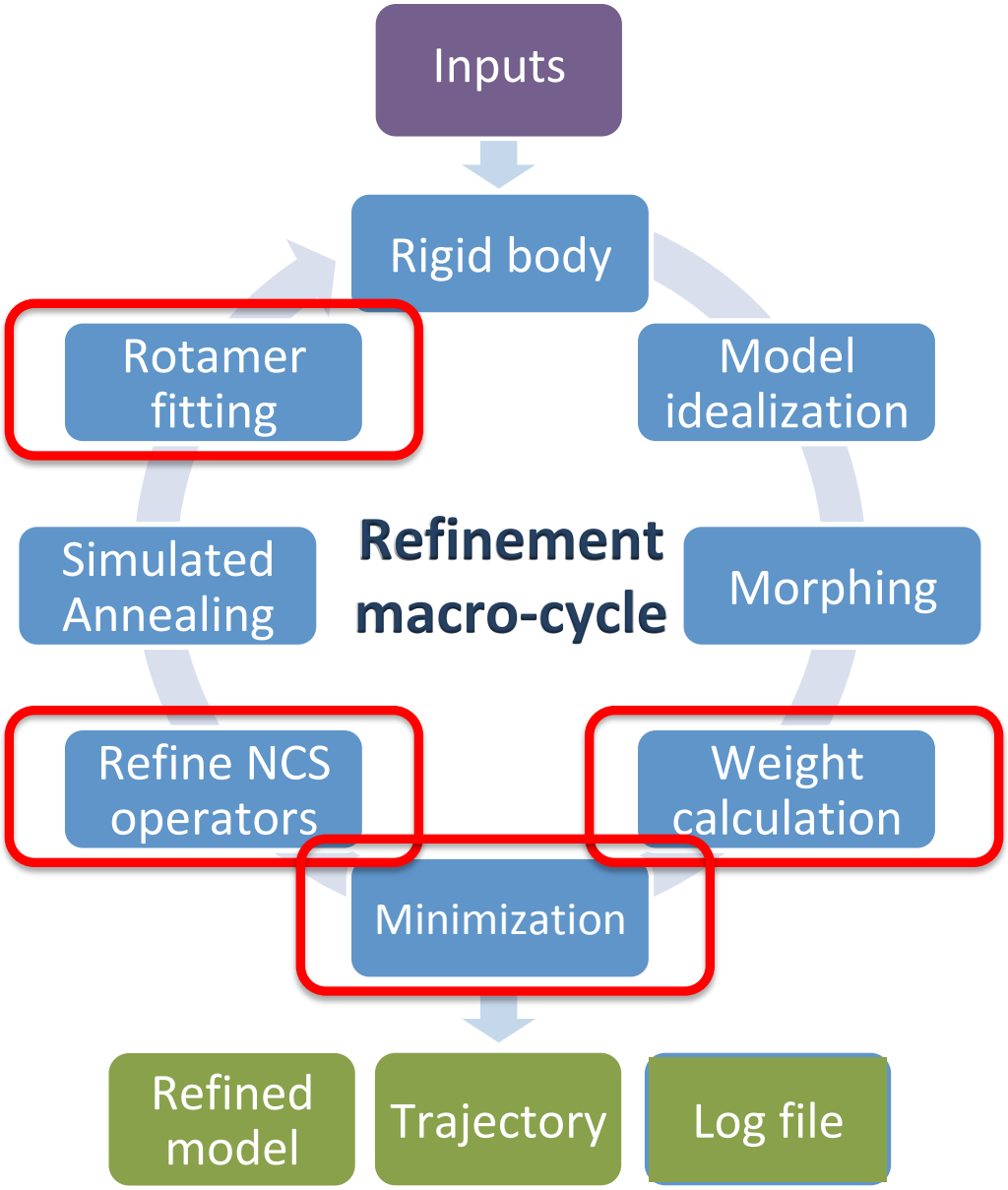
- **Calculate one set of ρ_{calc} – never finished on my laptop**
- **Calculate real-space refinement target – several seconds**

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

Automated model refinement: *phenix.real_space_refine*

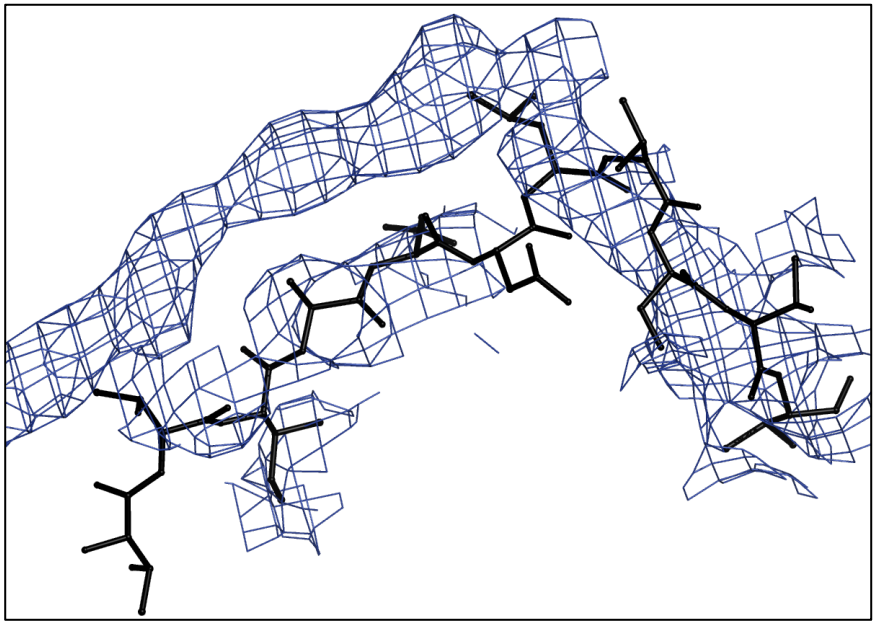


Automated model refinement: *phenix.real_space_refine*

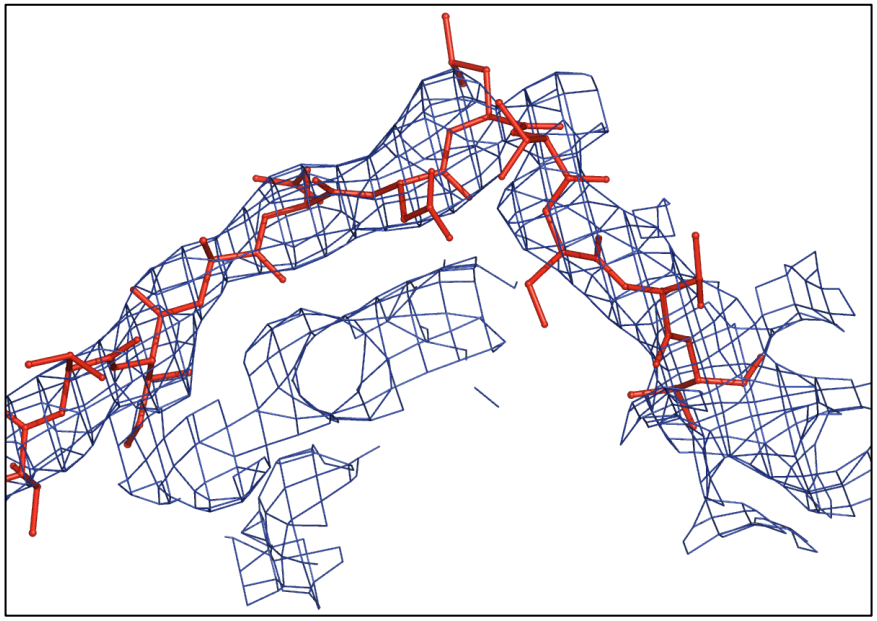


Morphing

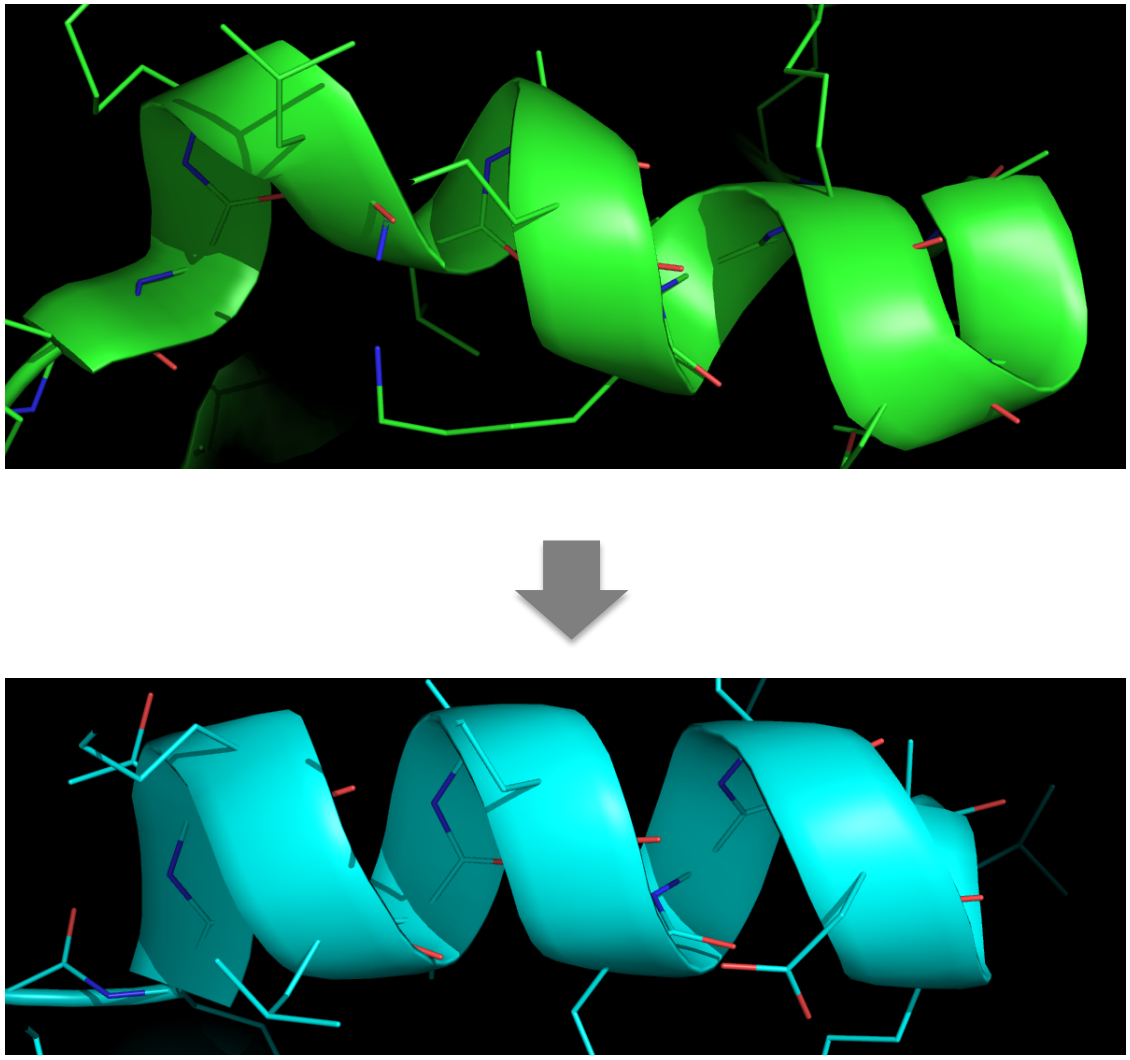
Start model before refinement



After *phenix.real_space_refine*



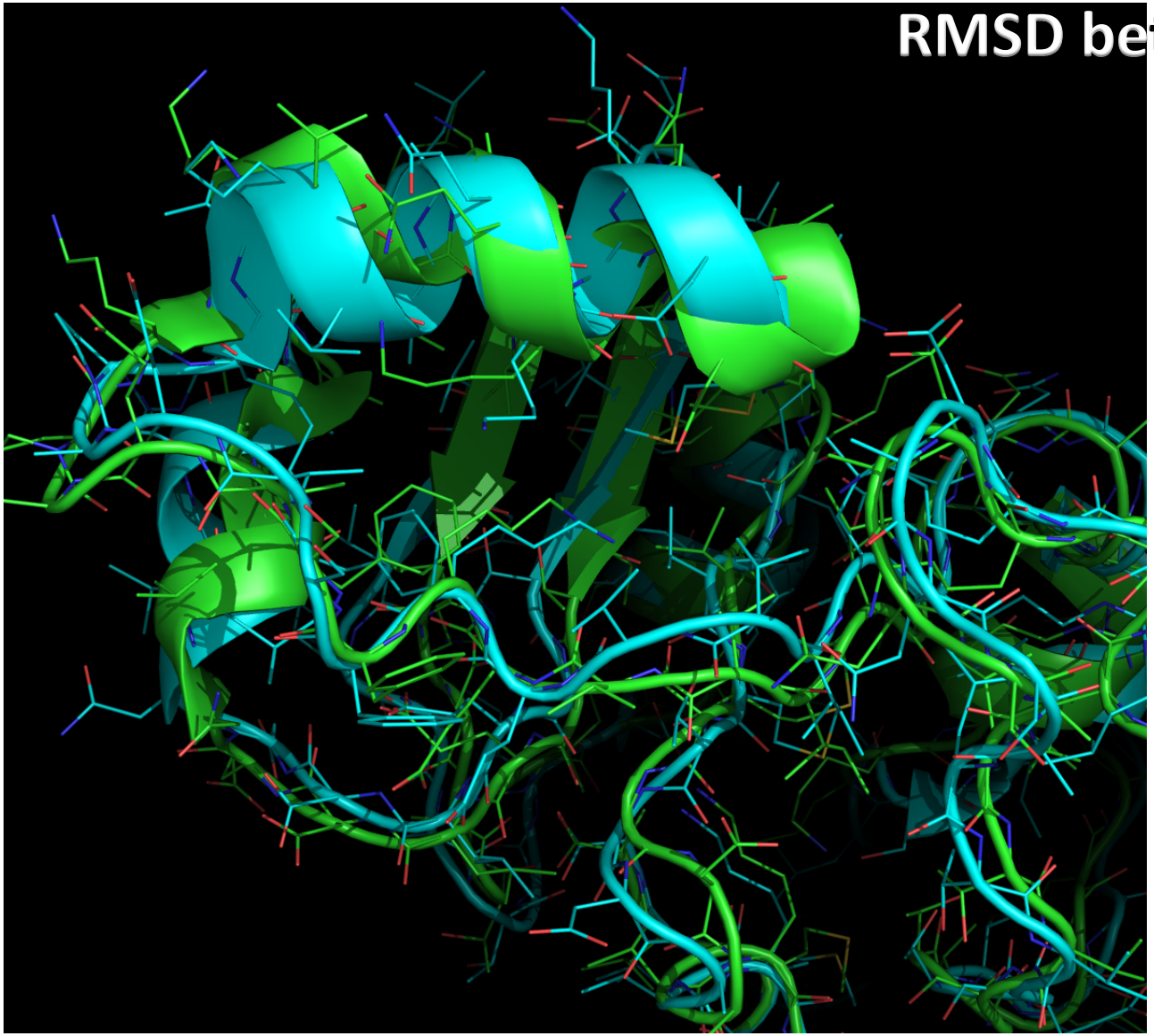
Model regularization



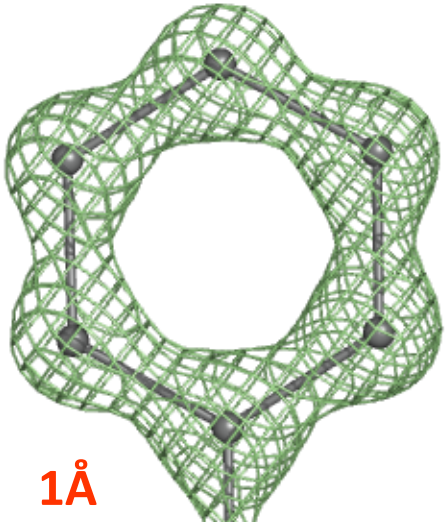
Model regularization

Before and after regilarization

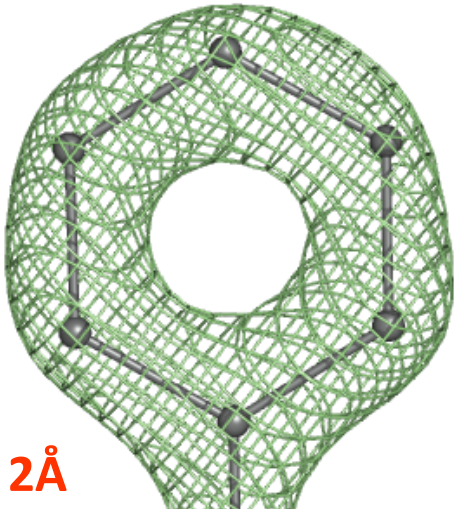
RMSD between two models less than 1.5Å



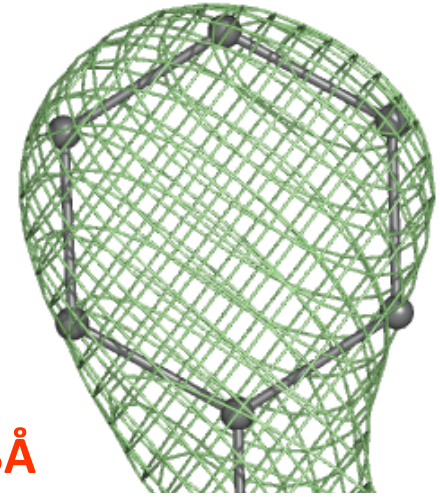
Restraints



1Å



2Å



3Å

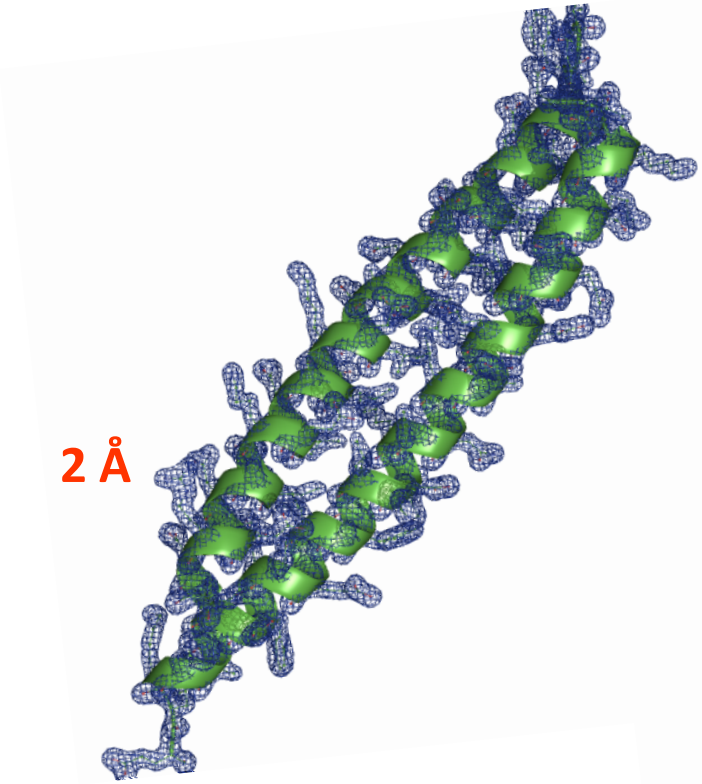
- Lower the resolution, less detailed the map
  - Need extra information to keep correct geometry during refinement

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

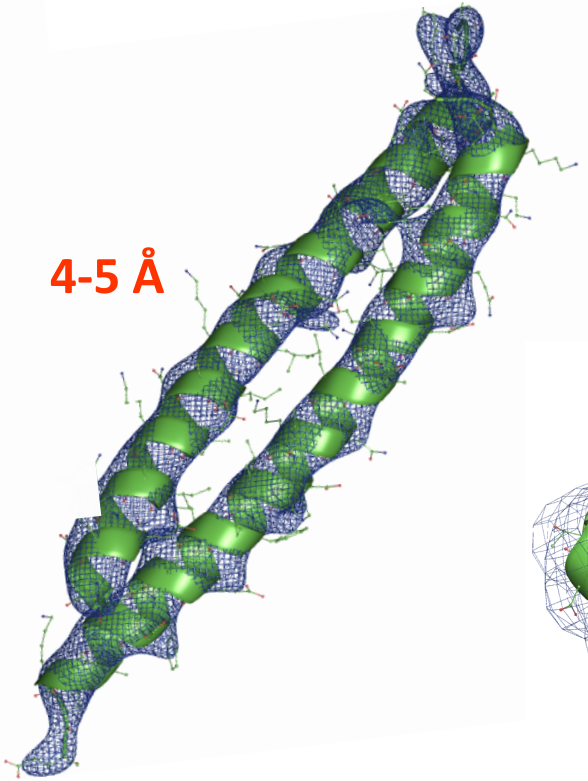
$$T_{\text{RESTRAINTS}} = T_{\text{BOND}} + T_{\text{ANGLE}} + T_{\text{DIHEDRAL}} + T_{\text{PLANARITY}} + T_{\text{NONBONDED}} + T_{\text{CHIRALITY}}$$

Restraints

- Low resolution map is not sufficient to maintain secondary

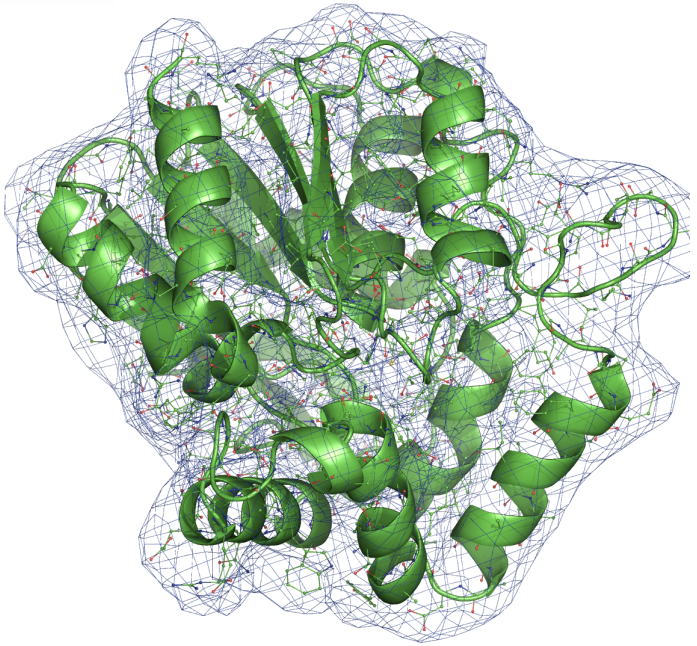


2 Å



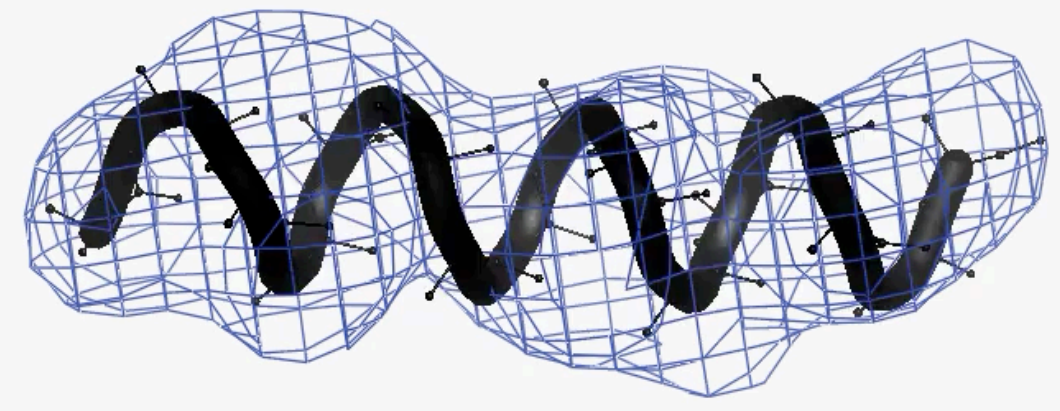
4-5 Å

6Å-lower



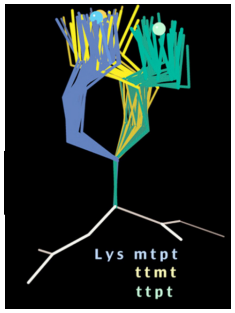
Restraints

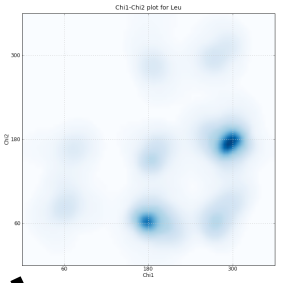
- Example: refinement of a perfect α-helix into low-res map
  - Using standard restraints on covalent geometry isn't sufficient
    - Model geometry deteriorates as result of refinement



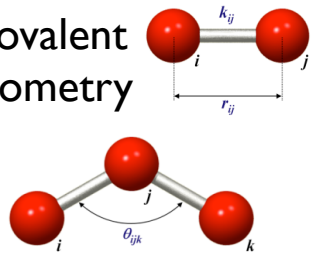
Restraints

Sidechain distributions

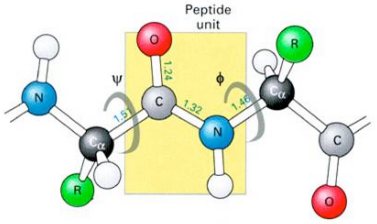




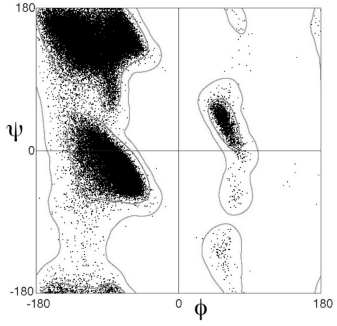
Covalent geometry

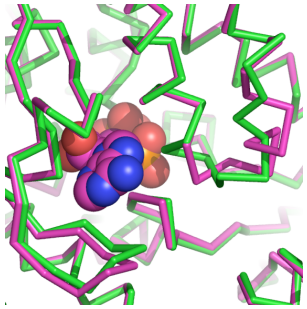


Images from PumMa web site (http://www.pumma.nl)

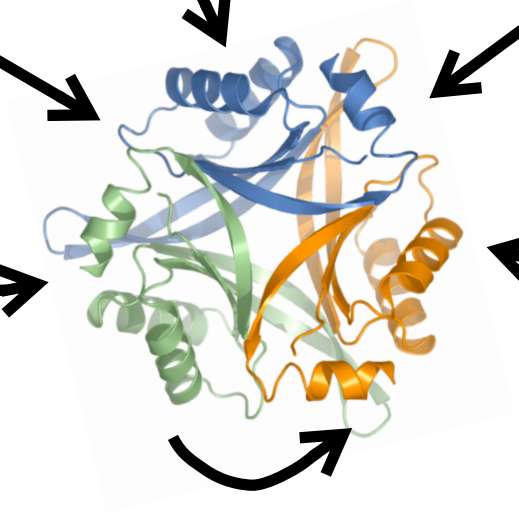


Mainchain distributions

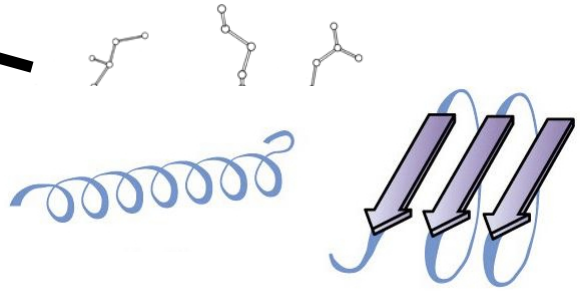




Related structures

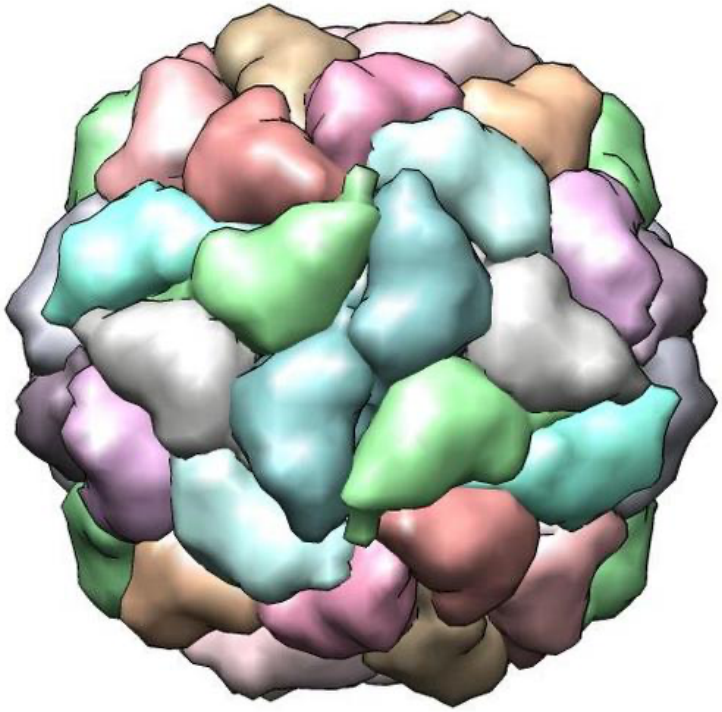


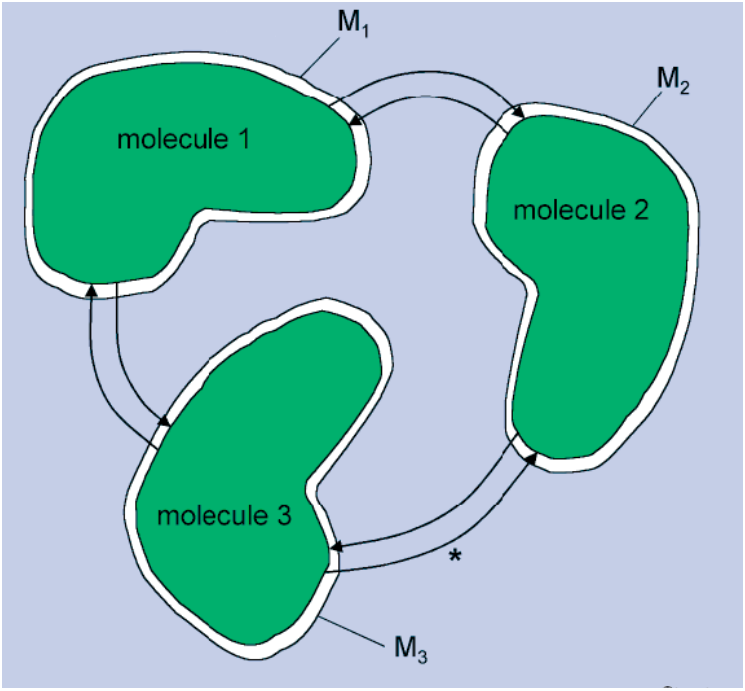
Internal symmetry



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

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

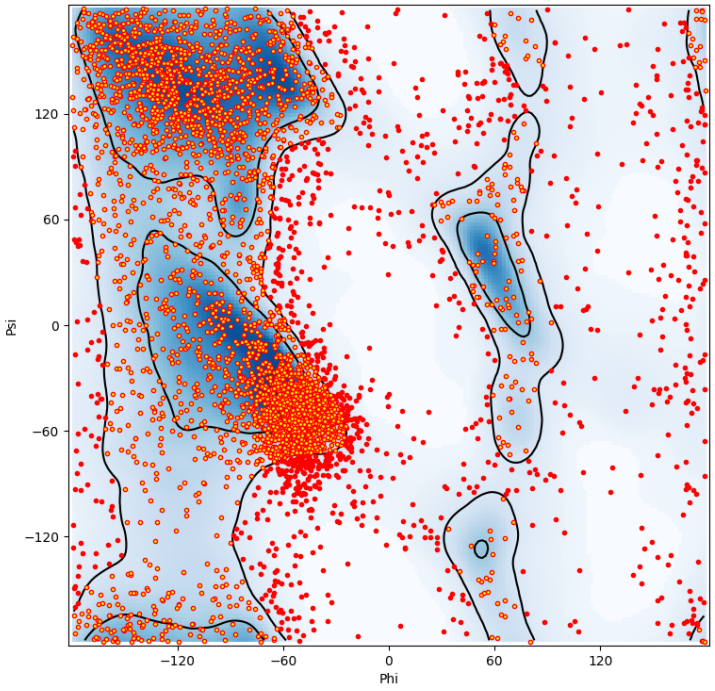
- Ramachandran plot restraints
  - Always use at 3Å and worse
  - Better than 3Å: use if needed (model has geometry violations)
  - Don’t use to fix outliers. Fix outliers first, then use Ramachandran plot restraints to stop re-occurring outliers..
  - Check Ramachandran plot regularly (see Validation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

Refinement: practical considerations and misc facts

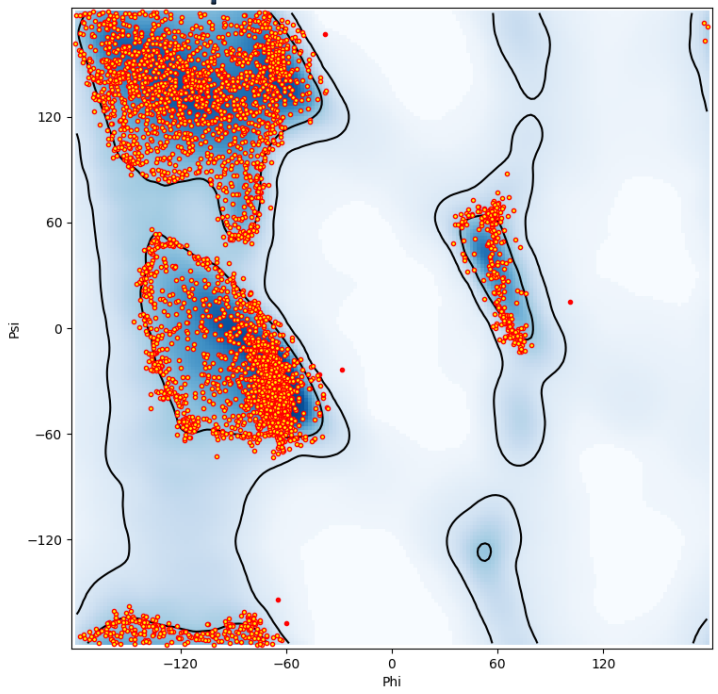
- Ramachandran plot restraints
  - Don’t use to fix outliers. Fix outliers first, then use Ramachandran plot restraints to prevent re-occurring outliers..

PDB code: 5a9z

Original



Refined with Ramachandran plot restraints

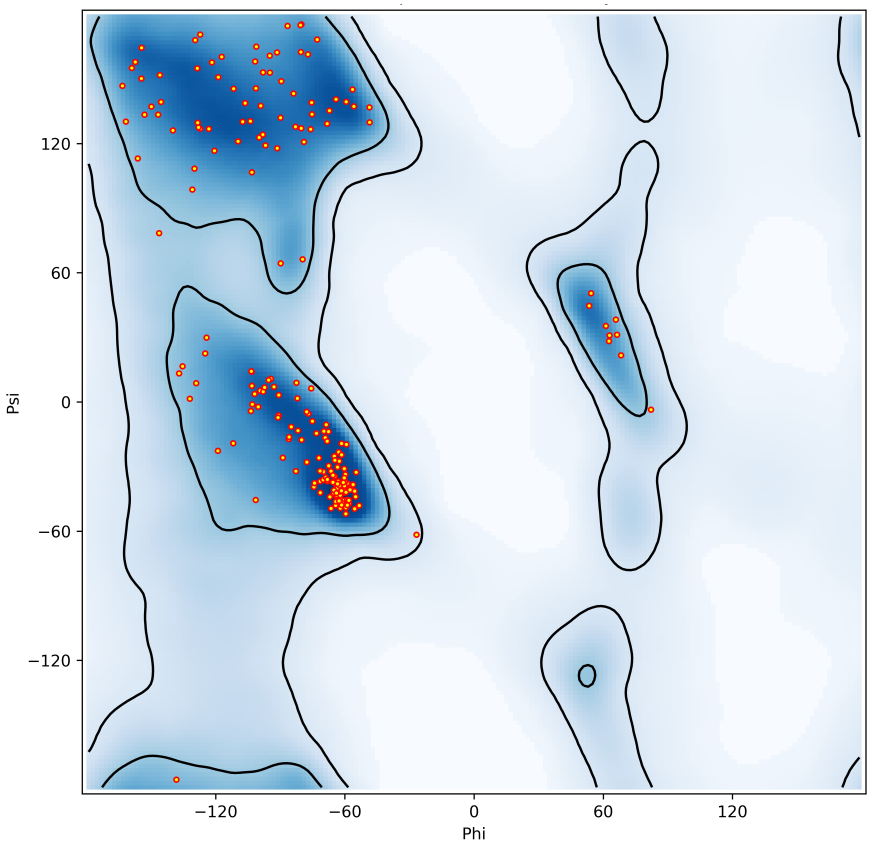


Bad idea to use Ramachandran plot restraints in this case. Fix outliers first!

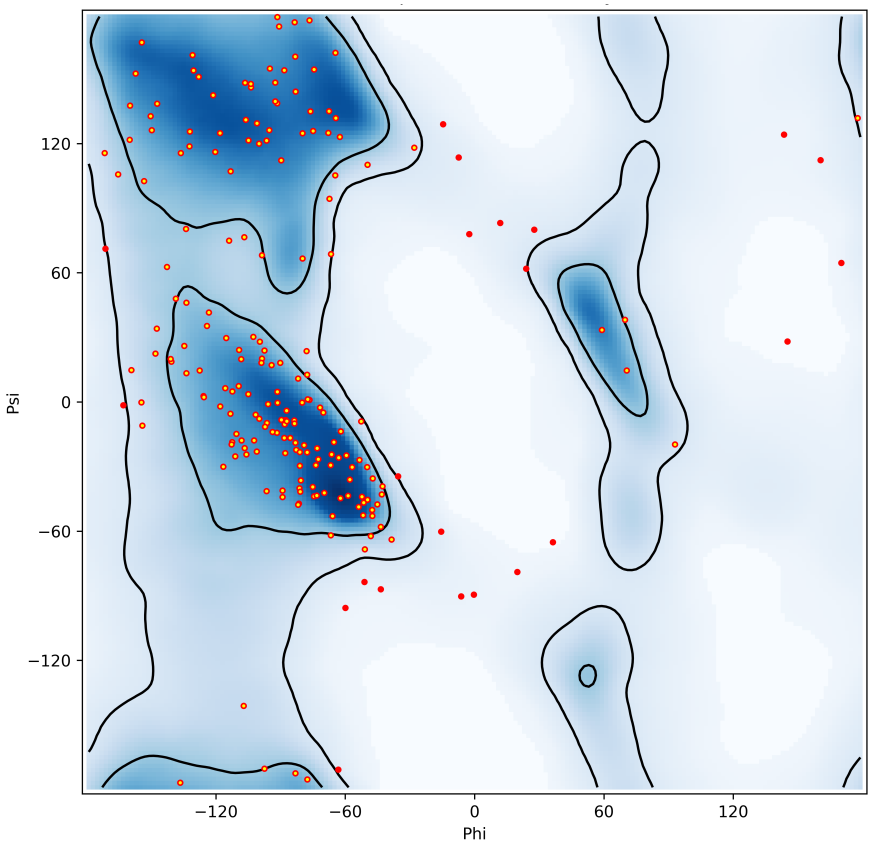
Refinement: practical considerations and misc facts

- Ramachandran plot restraints
  - Use to stop outliers from occurring

Before refinement



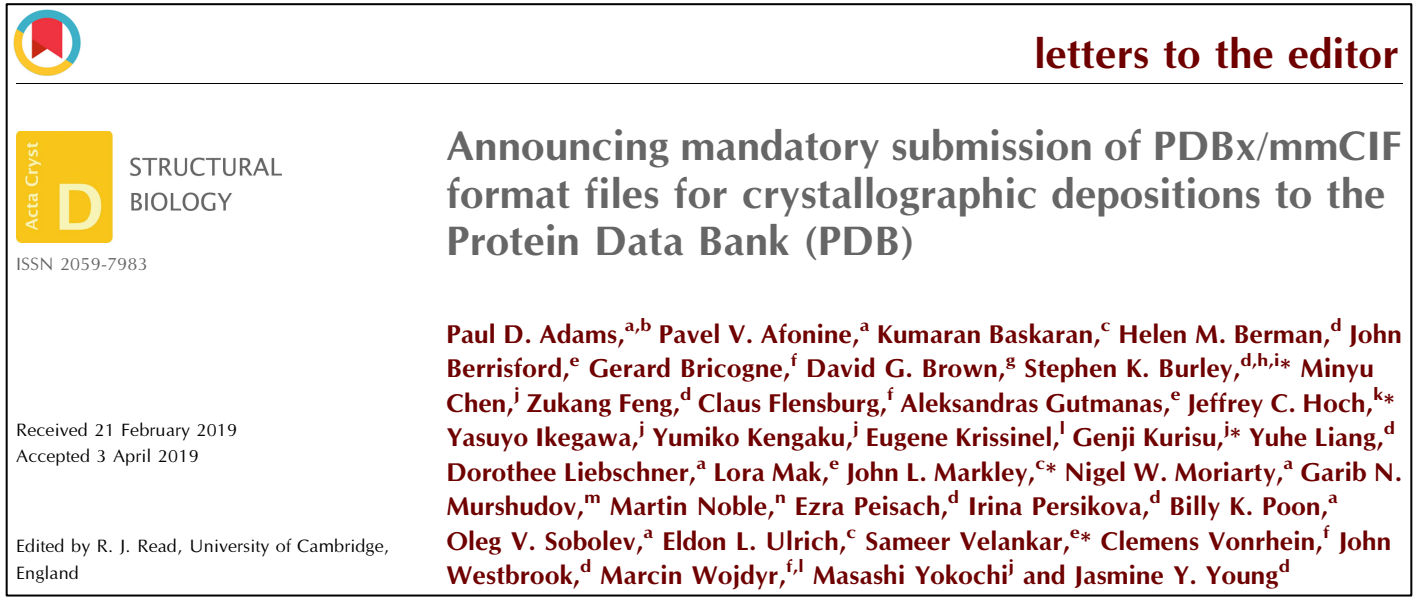
After refinement



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

Acta Cryst D STRUCTURAL BIOLOGY

ISSN 2059-7983

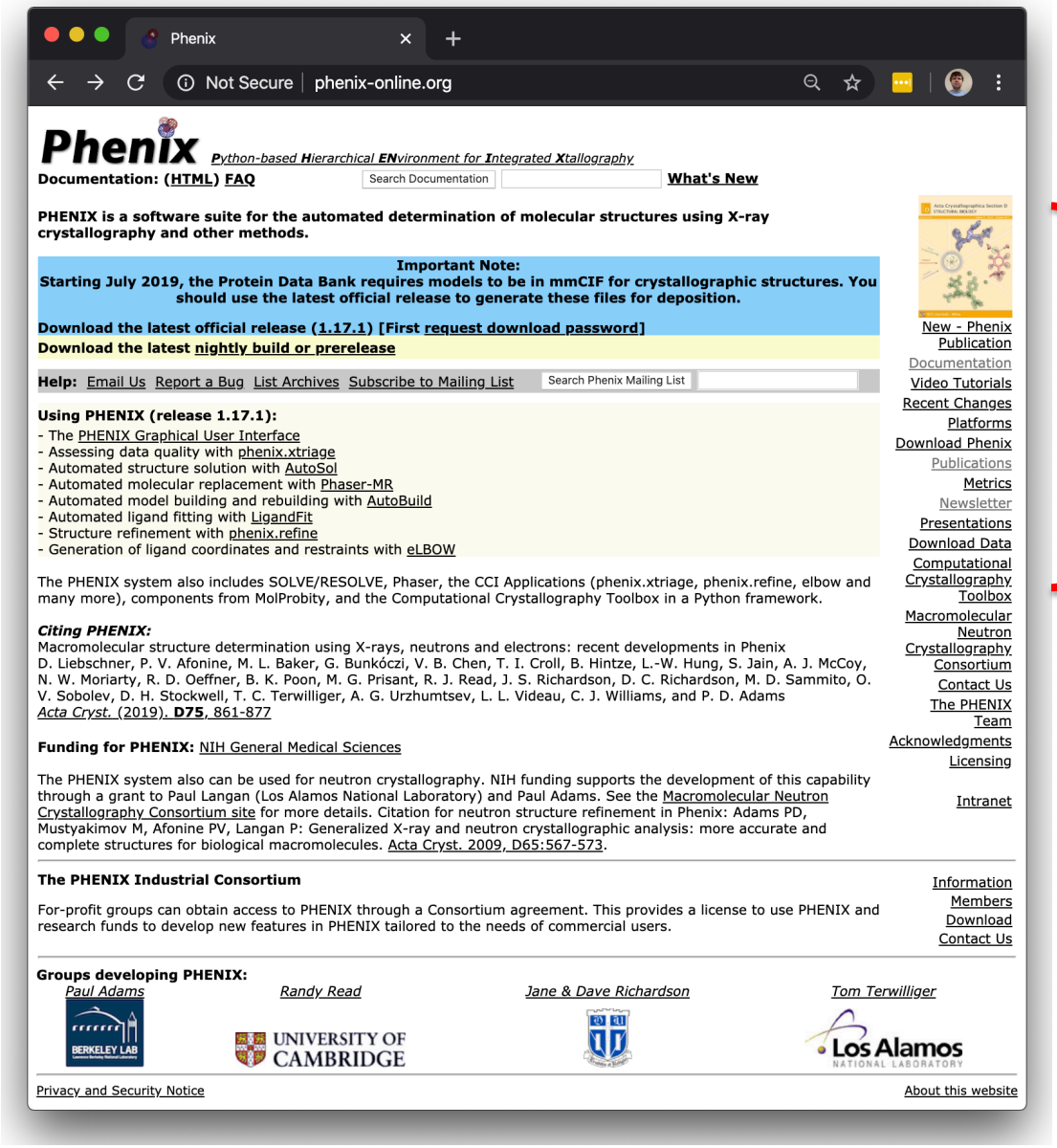
Announcing mandatory submission of PDBx/mmCIF format files for crystallographic depositions to the Protein Data Bank (PDB)

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 Vonrhein,[f] John Westbrook,[d] Marcin Wojdyr,[f,l] Masashi Yokochi[j] and Jasmine Y. Young[d]

Received 21 February 2019
Accepted 3 April 2019

Edited by R. J. Read, University of Cambridge, England

Resources



Phenix paper

Video tutorials

Documentation

Relevant papers

Bi-annual newsletters

Slides from workshops

Resources

PHENIX Documentation

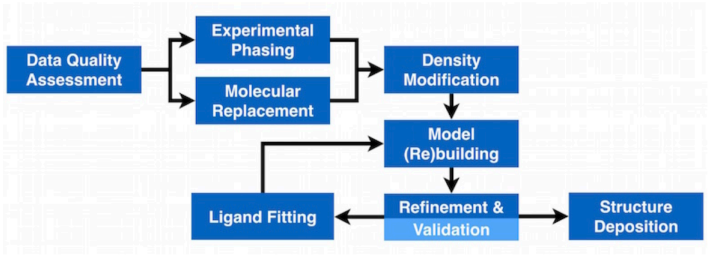
Not Secure | phenix-online.org/documentation/

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



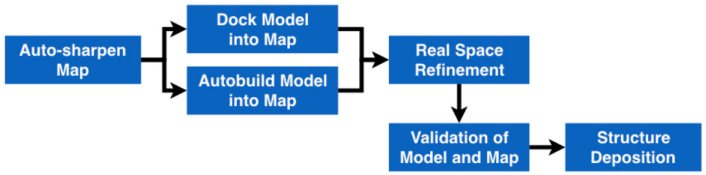
Phenix Documentation for X-ray Crystallography

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



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 parameyters) and tell us steps that lead to the problem

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