

Model Refinement: cryo-EM

Pavel Afonine



phenix-online.org



lbl.gov



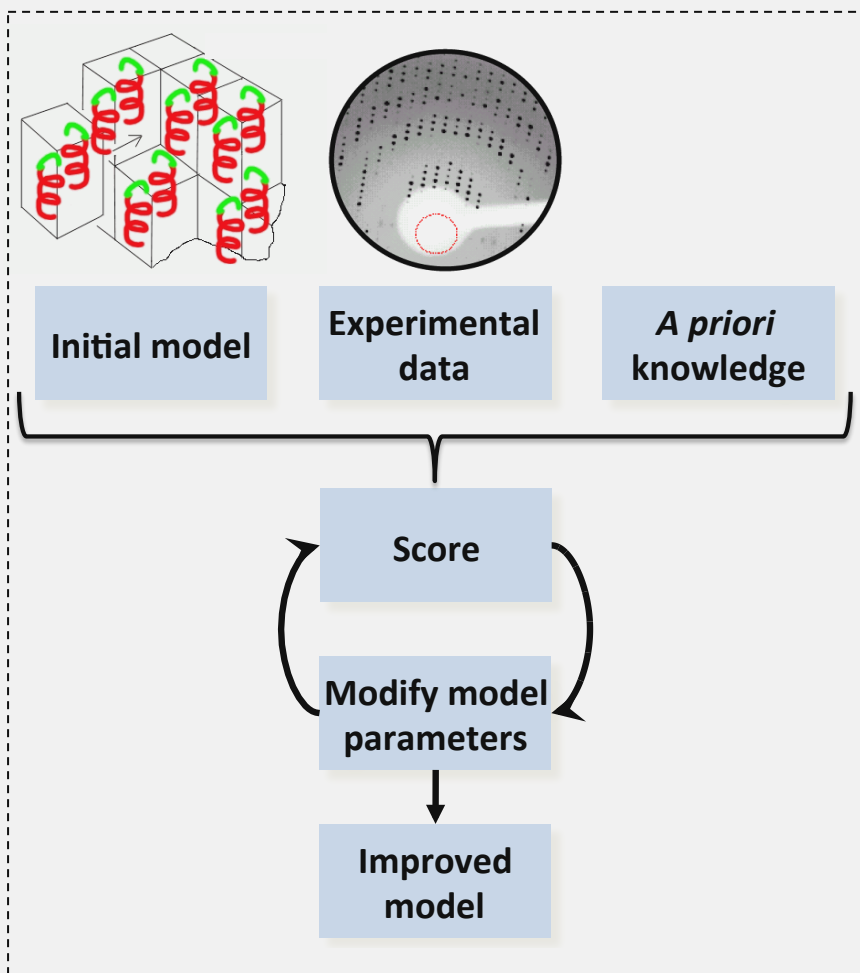
qrefine.com

Hosted by the Oklahoma COBRE in Structural Biology

March 19th 2024

Refinement in Phenix

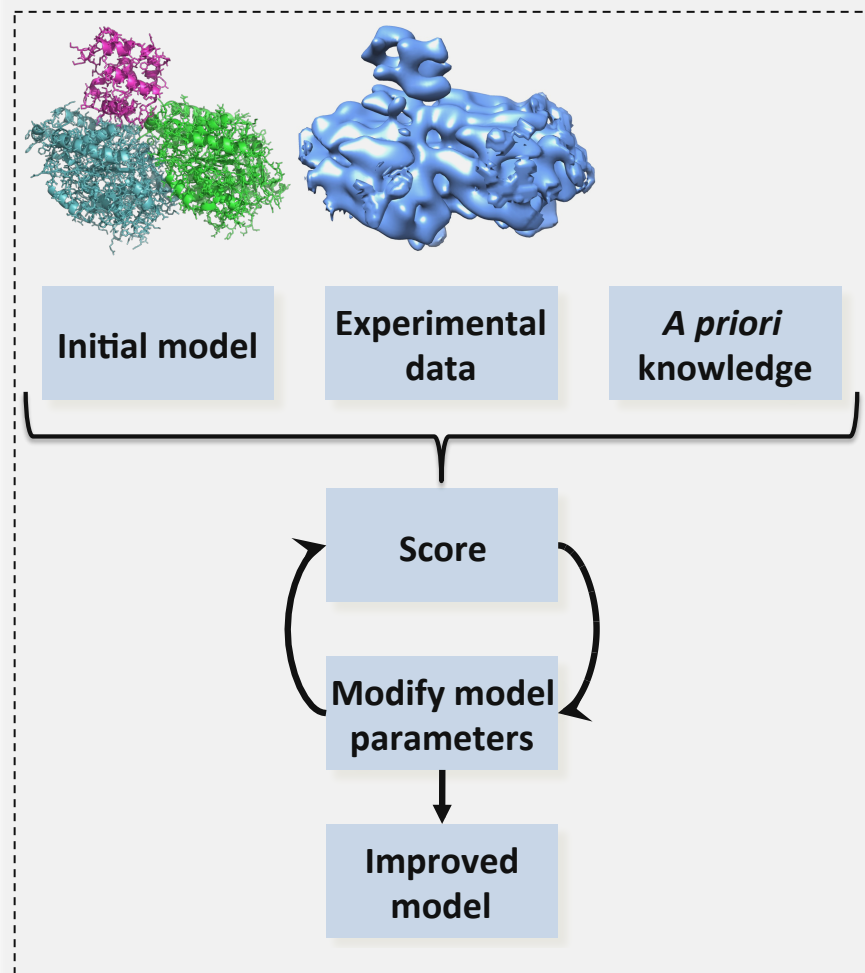
Crystallography



phenix.refine

Available since 2005

Cryo-EM



phenix.real_space_refine

Available since 2013

Atomic model refinement: crystallography vs cryo-EM

Crystallographic refinement

- Improving model improves map
 - (2mFo-DFc, Model phase), (mFo-DFc, Model phase)
 - Better model leads to better map
 - Better map leads to more model built
 - Improving model in one place lets build more model elsewhere in the unit cell
 - Refine all model parameters (XYZ, B) from start to end of structure solution
 - Build solvent (ordered water) early
- Experimental data never changed
- Data / restraints weight is global and time expensive to find best value
- Whole model needs to be refined

Cryo-EM refinement

- Changing model does not change map
 - Build solvent (water) last
 - Get as complete and accurate model as possible before refining B factors and occupancies
- Experimental data changes a lot during the process (filtering, boxing, using maps with implied symmetry or not, etc.)
 - What map to use in refinement?
 - Refined B factors depend on map used
- Data / restraints weight can be local and is always optimal
- Boxed parts of the model can be refined

Atomic model refinement: *phenix.real_space_refine*



STRUCTURAL
BIOLOGY

ISSN 2059-7983

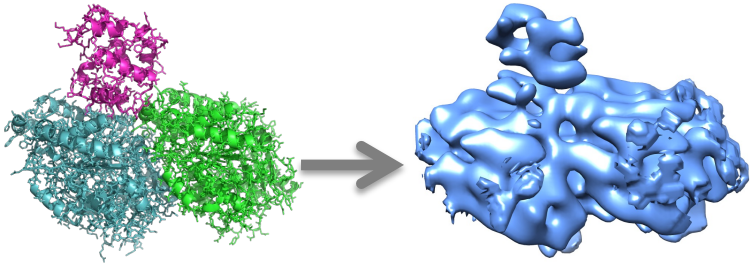
Real-space refinement in *PHENIX* for cryo-EM and crystallography

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}

How we evaluate refinement progress (model-to-map fit) or what's the analogue of crystallographic R-factor?

Model-to-map fit validation: CC_{MASK}

Model to map fit



$$CC_{\text{MASK}} = \frac{\sum \rho_{\text{obs}} \rho_{\text{calc}}}{(\sum \rho_{\text{obs}}^2 \sum \rho_{\text{calc}}^2)^{1/2}}$$

ρ_{obs} = experimental map

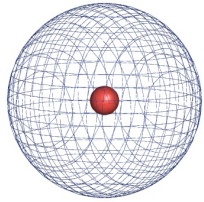
ρ_{calc} = model calculated map

- Easy interpretation: -1: anticorrelation, 0: no correlation, 1: perfect correlation
- Uses all atomic model parameters (XYZ, B-factors, occ, atom type)
- Not specific to map type (any map: x-ray, neutron, electron, cryo-EM, ...)
- Can be calculated locally (per atom, residue, chain, molecule, whole box, ...)
 - Local resolution can be trivially taken into account

| Metric | Expected value |
|--------------------|----------------|
| CC_{MASK} | Poor: < 0.3 |
| | So-so: 0.3-0.6 |
| | Good: > 0.6 |

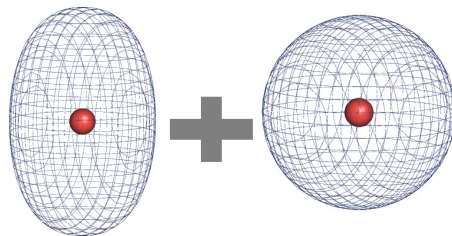
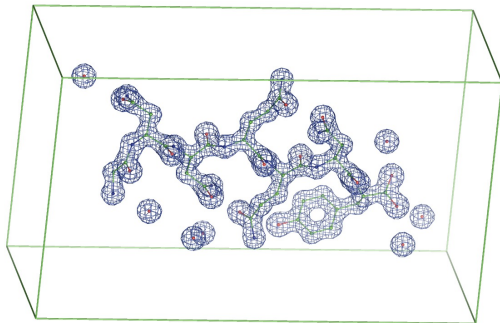
Model-to-map fit validation: CC_{MASK}

- Gaussian IAM (Independent Atom Model)



$$\rho_{atom}(\mathbf{r}, \mathbf{r}_0, B, q) = q \sum_{k=1}^5 a_k \left(\frac{4\pi}{b_k + B} \right)^{3/2} \exp\left(-\frac{4\pi^2 |\mathbf{r} - \mathbf{r}_0|^2}{b_k + B} \right)$$

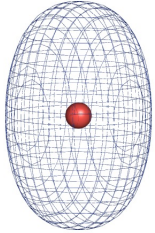
| | | | | | | | | | | | |
|------|----|----|-----|---|---|--------|--------|--------|------|-------|---|
| ATOM | 25 | CA | PRO | A | 4 | 31.309 | 29.489 | 26.044 | 1.00 | 57.79 | C |
|------|----|----|-----|---|---|--------|--------|--------|------|-------|---|



$$\rho_{\text{MODEL}}(\mathbf{r}) = \sum_{i=1}^{N_{\text{atoms}}} \rho_{\text{atoms}}(\mathbf{r})$$

Model map

- Gaussian IAM (Independent Atom Model)
- Anisotropic:



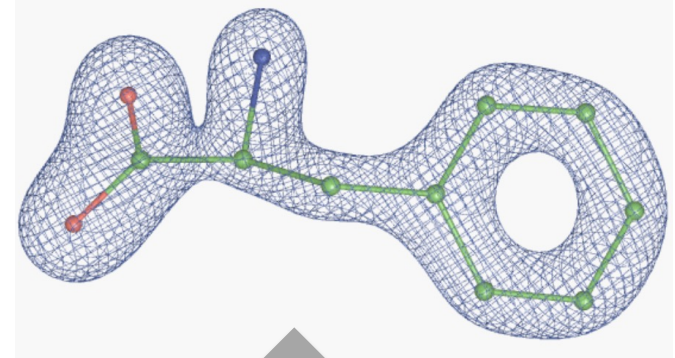
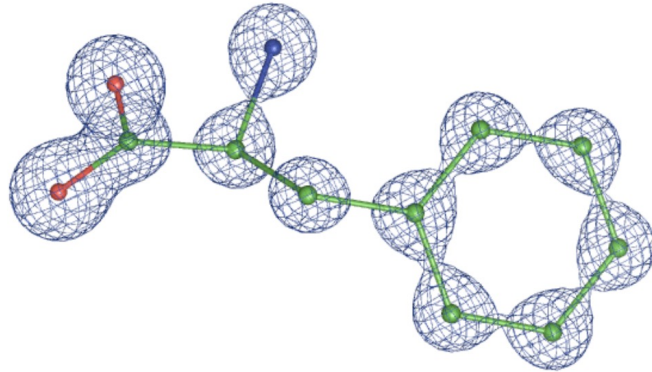
$$\rho_{atom}(\mathbf{r}, \mathbf{U}, q) = q \sum_{j=1}^5 \frac{q a_j (4\pi)^{3/2}}{|8\pi^2 \mathbf{U}_{cart} + b_j \mathbf{I}|^{1/2}} \exp\left(-4\pi^2 (\mathbf{r} - \mathbf{r}_0)^T \mathbf{A}^T [8\pi^2 \mathbf{U}_{cart} + b_j \mathbf{I}]^{-1} \mathbf{A} (\mathbf{r} - \mathbf{r}_0)\right)$$

| | | | | | | | | | | | | |
|--------|----|----|-----|---|---|--------|--------|--------|------|-------|-----|---|
| ATOM | 25 | CA | PRO | A | 4 | 31.309 | 29.489 | 26.044 | 1.00 | 57.79 | C | |
| ANISOU | 25 | CA | PRO | A | 4 | 8443 | 7405 | 6110 | 2093 | -24 | -80 | C |

Model-to-map fit validation: CC_{MASK}

3Å model-calculated map

Exact model map



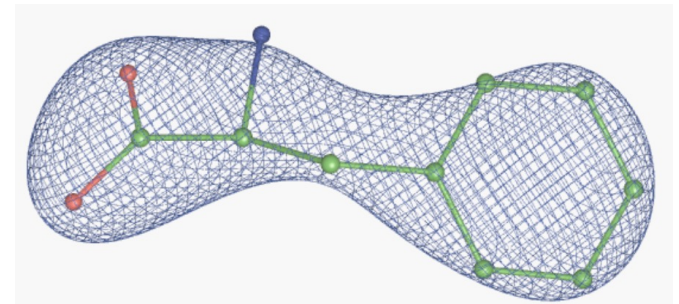
CC_{MASK}



CC_{MASK}



3Å experimental map



$$\rho_{\text{MODEL}}(\mathbf{r}) = \sum_{i=1}^{N_{\text{atoms}}} \rho_{\text{atoms}}(\mathbf{r})$$

- FT exact model map
- Remove terms up to specified resolution
- FT back to real space to get a Fourier image = “Model map”

Other popular model-to-map fit metrics and reasons why they are not as good as CCmask

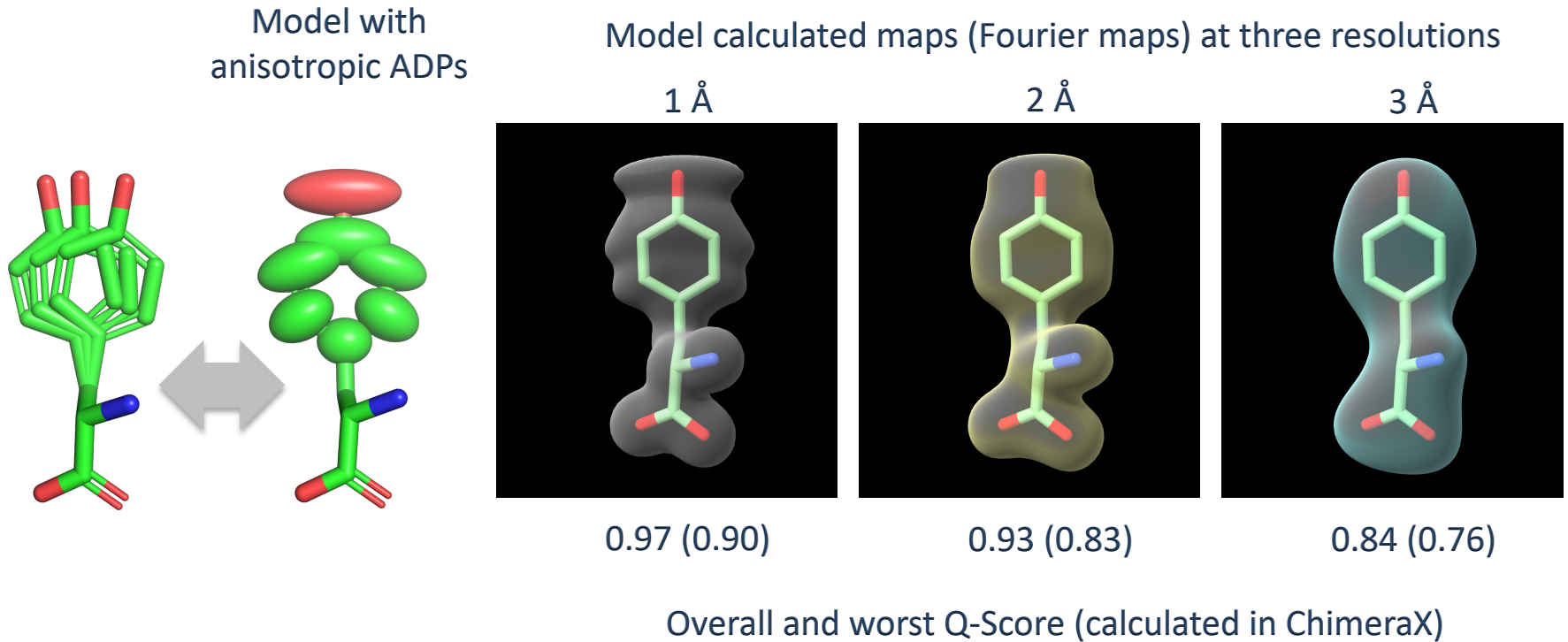
Atom inclusion

- **Atom inclusion:** fraction of atoms inside molecular envelope contoured at a given level
 - Contouring threshold: Arbitrarily? What is optimal level?
 - No use of atomic model parameters such as ADP, occupancy, atom type, ...
 - Does not compare shape of density:
 - How SER placed into PHE density is going to score?
 - How water O placed into Mg peak will score?
 - Does not account for missing atoms
 - Does not use map type (x-ray, neutron, electron)
 - Partially occupied atoms (alternative conformations):
 - Chosen level for fully occupied atoms needs to be scaled by occupancy for partially occupied atoms

Q-Score

- **Q-score:** measure the resolvability of individual atoms in a cryo-EM map, using an atomic model fitted to or built into the map
 - No use of atomic model parameters such as ADP, occupancy, atom type, ...
 - Shape of density:
 - How SER placed into PHE density is going to score?
 - How water O placed into Mg peak will score?
 - Does not account for missing atoms (it shouldn't given the definition)
 - Alternative conformations are **not** handled
 - How anisotropic atoms are **not** handled
 - Does not use map type (x-ray, neutron, electron)

Example: Q-Score for exact (model-generated) map



- Why Q-Score is not perfect (=1) given these are exact model-generated maps?
- Why it varies with the resolution?

Validation reports (RCSB): only Q-score and atom inclusion

6KIQ
Complex of yeast cytoplasmic dynein MTBD-High and MT with DTT

PDB DOI: 10.2210/pdb6KIQ/pdb EM Map EMD-9997: EMDB EMDataResource

Classification: **MOTOR PROTEIN/STRUCTURAL PROTEIN**
Organism(s): *Sus scrofa*, *Saccharomyces cerevisiae* S288C
Expression System: *Escherichia coli*
Mutation(s): Yes

Deposited: 2019-07-19 Released: 2020-03-04
Deposition Author(s): Komori, Y., Nishida, N., Shimada, I., Kikkawa, M.
Funding Organization(s): Japan Science and Technology, Japan Agency for Medical Research and Development (AMED)

Experimental Data Snapshot
Method: ELECTRON MICROSCOPY
Resolution: 3.62 Å
Aggregation State: FILAMENT
Reconstruction Method: HELICAL

wwPDB Validation
Metric Percentile Ranks Value
Clashscore 10
Ramachandran outliers 10.7%
Sidechain outliers 12.4%

Page 34

Full wwPDB EM Validation Report

EMD-

wwPDB Validation

3D Report

Full Report

9.5 Map-model fit summary ⓘ

The table lists the average atom inclusion at the recommended contour level (0.125) for the entire model and for each chain.

| Chain | Atom inclusion | Q-score |
|-------|----------------|---------|
| All | 0.9062 | 0.4550 |
| M | 0.5810 | 0.3210 |
| a | 0.9659 | 0.4790 |
| b | 0.9656 | 0.4730 |

Model-to-map fit statistics is insufficient and very well hidden!

Refinement: practical considerations

- Final stages
 - Refine B-factors (Atomic Displacement Parameters)
 - Group B factor or individual
 - Refine occupancies
 - Use Hydrogen atoms (and keep them in the final model!)
 - Add water (phenix.douse: command line and GUI):
 - Also available in ChimeraX

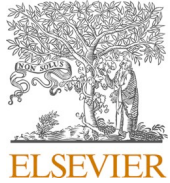
mmCIF

- 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 too large to fit into PDB file format
 - *Phenix* provides full support for mmCIF I/O

| | |
|--|---|
|  | letters to the editor |
|  STRUCTURAL BIOLOGY ISSN 2059-7983 | 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 | |

Variability refinement

Treasuring conformational changes



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

BBA - Biomembranes

journal homepage: www.elsevier.com/locate/bbamem



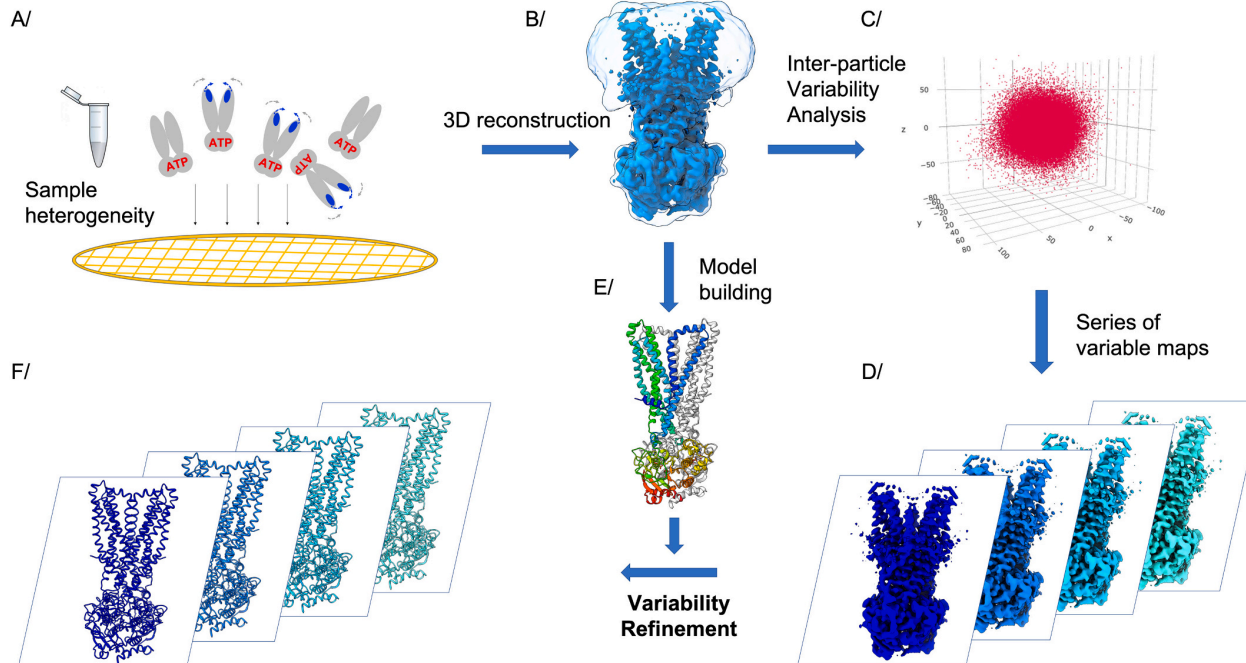
Review

Conformational space exploration of cryo-EM structures by variability refinement

Pavel V. Afonine^{a,*}, Alexia Gobet^b, Loïck Moissonnier^b, Juliette Martin^b, Billy K. Poon^a, Vincent Chaptal^{b,*}

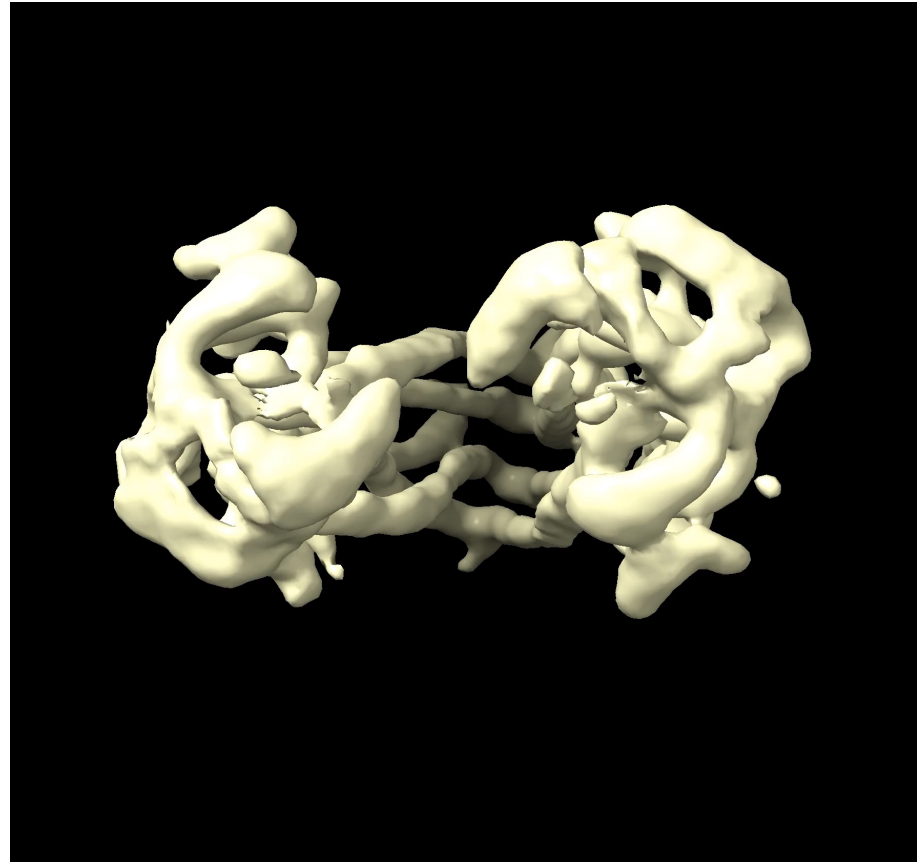
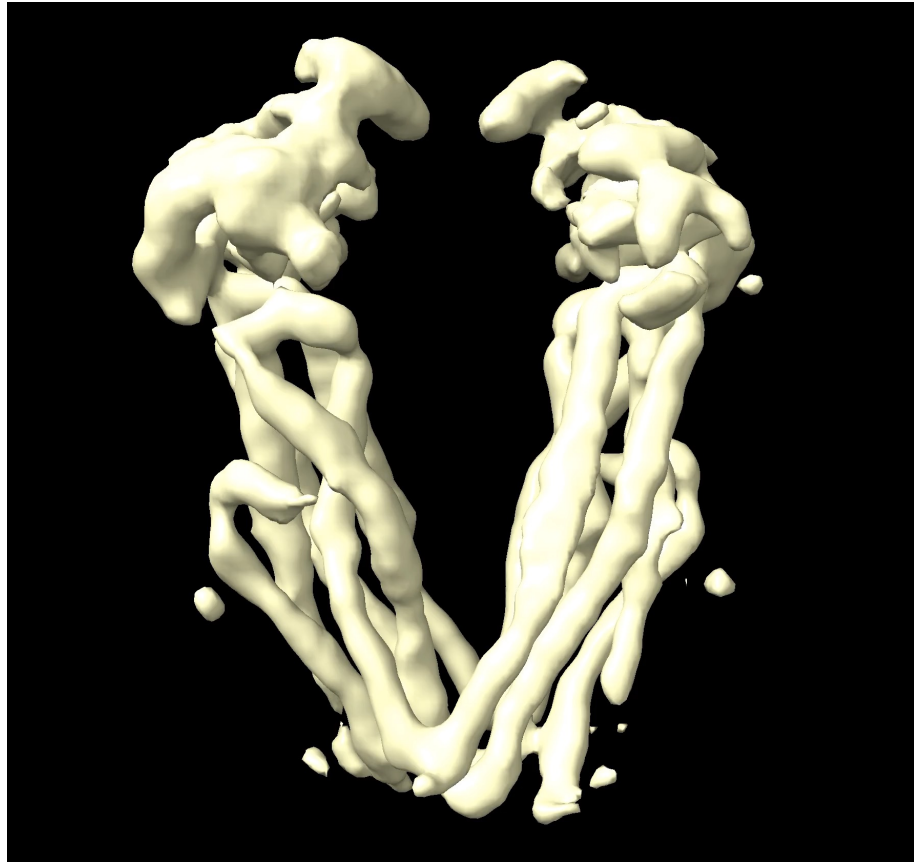
^a Molecular Biosciences and Integrated Bioimaging, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

^b Molecular Microbiology and Structural Biochemistry, UMR5086 CNRS University Lyon1, 7 passage du Vercors, 69007 Lyon, France



Maps

ABC transporter BmrA (unpublished!)



phenix.varref – new Phenix tool to represent ensemble of maps with ensemble of atomic models

phenix.varref

map1.mrc ... mapN.mrc

model.pdb

resolution=3

nproc=100

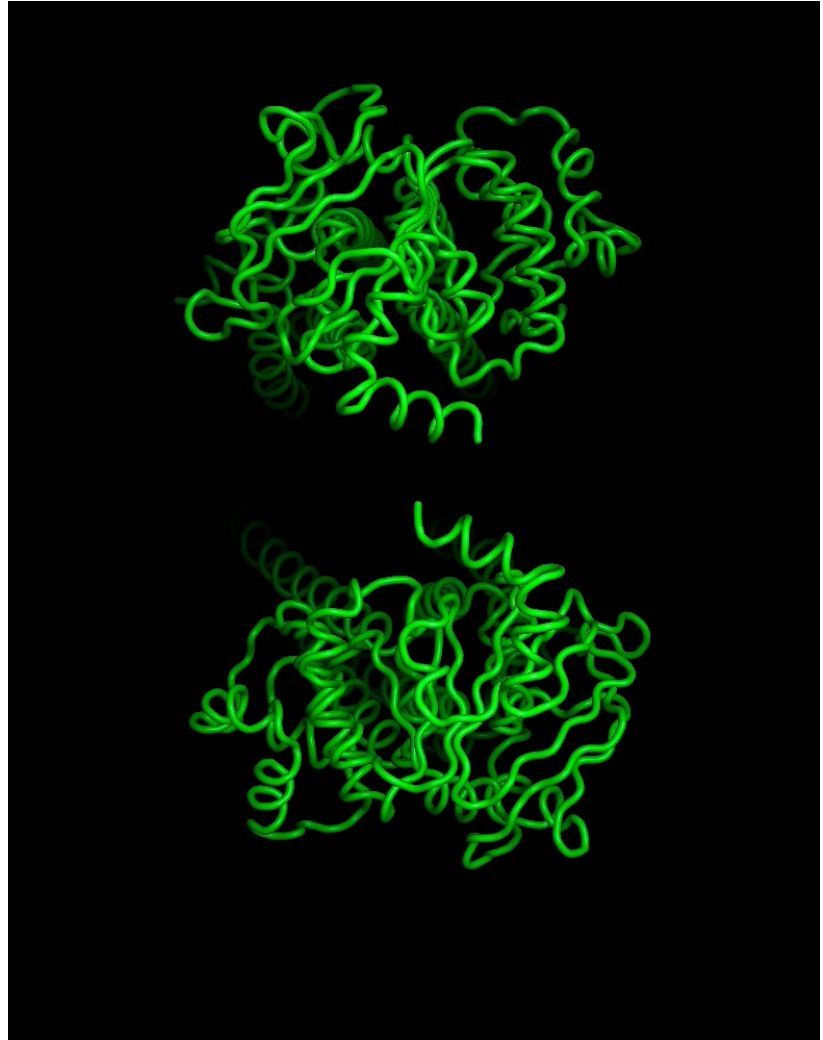
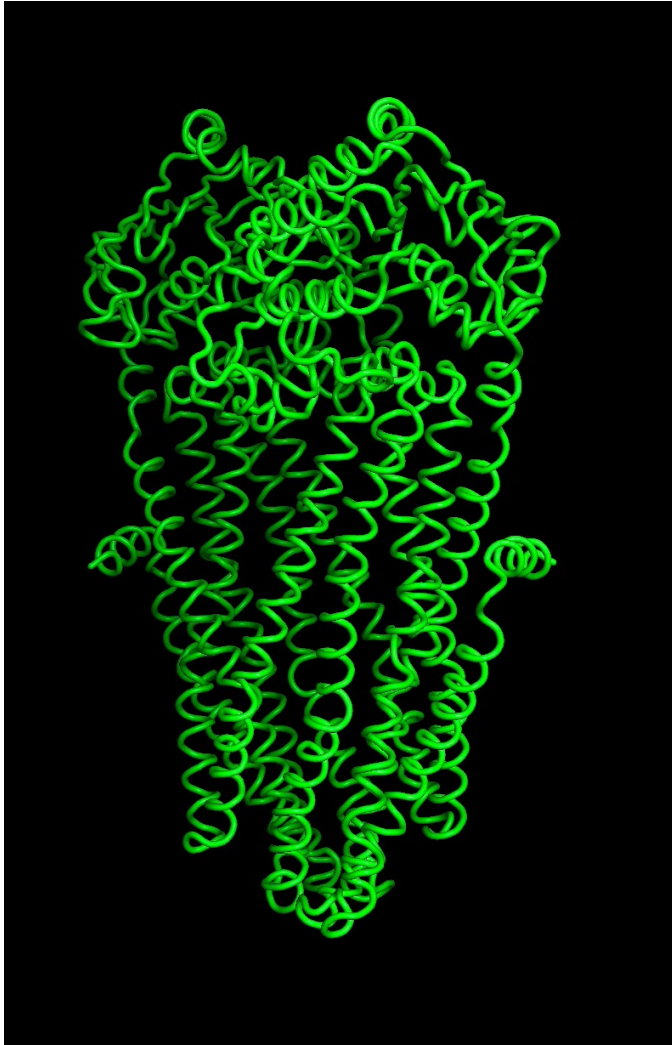
models_per_map=100

Output: ensemble of refined models that represents all maps

Workflow

- Input model and maps
- Order maps by similarity using CC_{box}
- Identify the map that is closest to input model (by CC_{mask})
 - This is the starting point for the first refinement
 - Generate ensemble of 100 perturbed models (by MD)
 - Refine each model with *phenix.real_space_refine*
 - Combine all refined models to yield overall best fitting model
- Refine ensemble of refined models against the next closest map
 - Combined all refined models to yield overall best fitting model
- ...and so on for all maps.
- Result:
 - N models corresponding to N maps
 - 100 models per map (can be used to estimate uncertainty)

Refined ensembles of models



Future

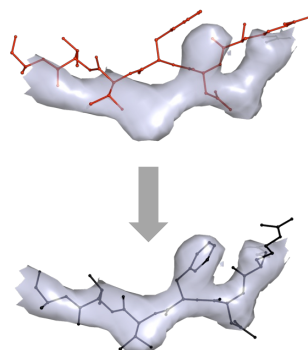
QM derived restraints

T

Optimize
consensus
between model-
to-data fit and...
common sense

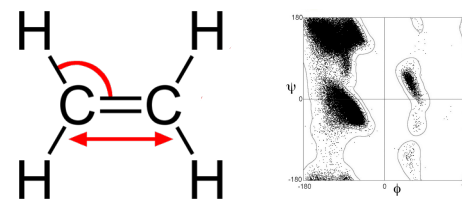
=

T_{DATA}



+

W * T_{RESTRAINTS}



~~Bonds, angles, planes,
torsions, chirality, non-
bonded repulsion~~

**Replace with
energies/gradients
from QM
calculations**

Q|R: Quantum Refinement project

Blend of expertise and background

Crystallography
methods



Software
development



QM
expertise

www.qrefine.com



Quantum Refinement

<http://qrefine.com> qrefine@googlegroups.com

Q|R @ GitHub

Pending.AI (Australia)



Mark
Waller



Holger
Kruse

Shanghai University

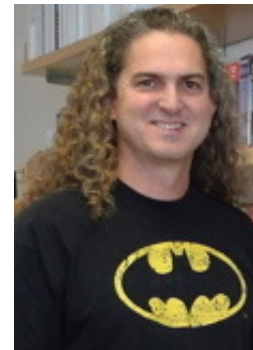


Malgorzata
Biczysko

LBLN (USA)



Pavel
Afonine



Nigel
Moriarty


Q|R: History of progress

 STRUCTURAL BIOLOGY
ISSN 2059-7983

Q|R: quantum-based refinement

2016

Min Zheng,^{a,b} Jeffrey R. Reimers,^{a,c} Mark P. Waller^{a,b*} and Pavel V. Afonine^{a,d*}

 STRUCTURAL BIOLOGY
ISSN 2059-7983

Solving the scalability issue in quantum-based refinement: Q|R#1

2017


Min Zheng,^{a,b} Nigel W. Moriarty,^c Yanting Xu,^a Jeffrey R. Reimers,^{a,d} Pavel V. Afonine^{a,c*} and Mark P. Waller^{a*}

 STRUCTURAL BIOLOGY
ISSN 2059-7983

Including crystallographic symmetry in quantum-based refinement: Q|R#2

2019

Min Zheng,^{a,b} Malgorzata Biczysko,^a Yanting Xu,^a Nigel W. Moriarty,^c Holger Kruse,^d Alexandre Urzhumtsev,^{e,f} Mark P. Waller^{g*} and Pavel V. Afonine^{c*}

 STRUCTURAL BIOLOGY
ISSN 2059-7983

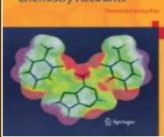
Real-space quantum-based refinement for cryo-EM: Q|R#3

2020

Lum Wang,^a Holger Kruse,^b Oleg V. Sobolev,^c Nigel W. Moriarty,^c Mark P. Waller,^{d*} Pavel V. Afonine^c and Malgorzata Biczysko^{a*}


Optimal clustering for quantum refinement of biomolecular structures: Q|R#4

2023



Research | Published: 25 September 2023

Theoretical Chemistry Accounts

 Quantum Refinement
<http://qrefine.com> qrefine@googlegroups.com

Open source on GitHub

Since 2016

QM-based refinement of bio-macromolecules

2010-2014

Impossible

Ufimtsev, Martinez, *et al.*
GPU accelerated software
(Terachem)

www.petachem.com



2017

Possible but very
limited

Q|R Team
Divide-and-conquer using
chemical intuition

www.qrefine.com



Happening around now

Unlimited
(in principle)

Isayev Lab
AI/ML trained QM
potentials

www.olexandrisayev.com



+

