

Model Refinement: X-ray vs cryo-EM

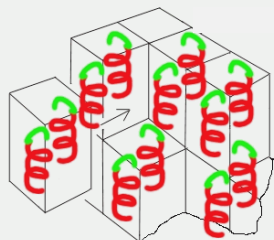
Pavel Afonine

Lawrence Berkeley National Laboratory (LBNL)

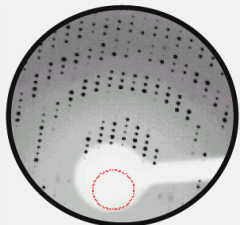
**September 26th, 2024
BNL**

Refinement

Crystallography



Initial model



Experimental
data

A priori
knowledge

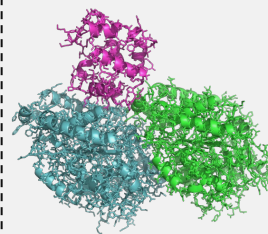
Score

Modify model
parameters

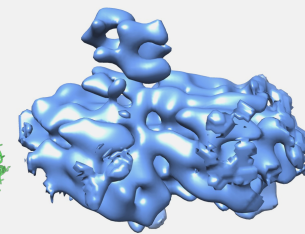
Improved
model

phenix.refine

Cryo-EM



Initial model



Experimental
data

A priori
knowledge

Score

Modify model
parameters

Improved
model

phenix.real_space_refine

Atomic model refinement: crystallography vs cryo-EM

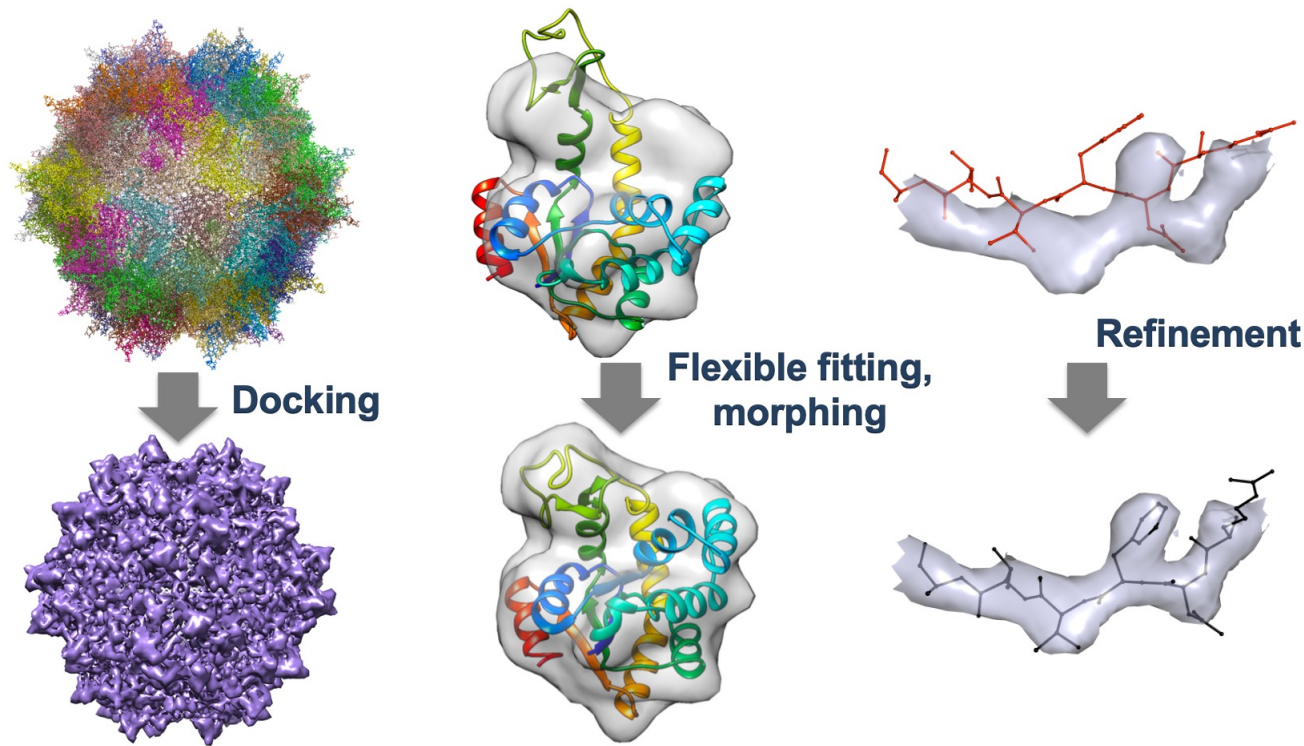
Crystallographic refinement

- 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

Not all model-to-data fitting is refinement



- Docking, flexible fitting, morphing are **not** refinement
- Refinement is to fine-tune an already fine atomic model
- Refinement does only small changes to the model (within *convergence radius of refinement*, $\sim 1\text{\AA}$)

Atomic model refinement: *phenix.real_space_refine*



STRUCTURAL
BIOLOGY

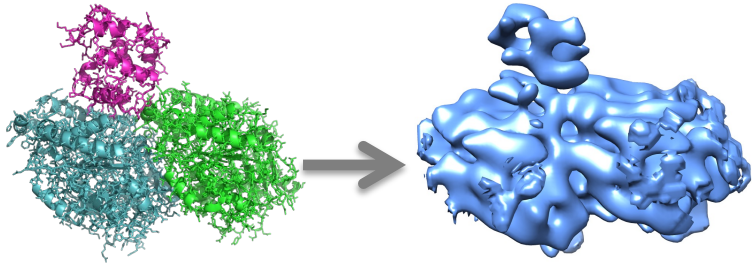
ISSN 2059-7983

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}

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

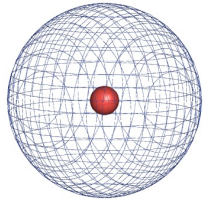
ρ_{calc} = model calculated map

- 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

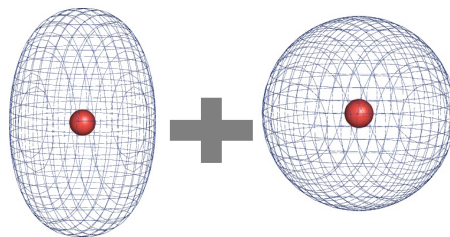
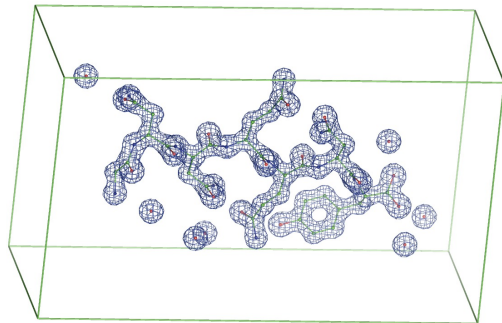
Model-to-map fit validation: CC_{MASK}

- 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
------	----	----	-----	---	---	--------	--------	--------	------	-------	---

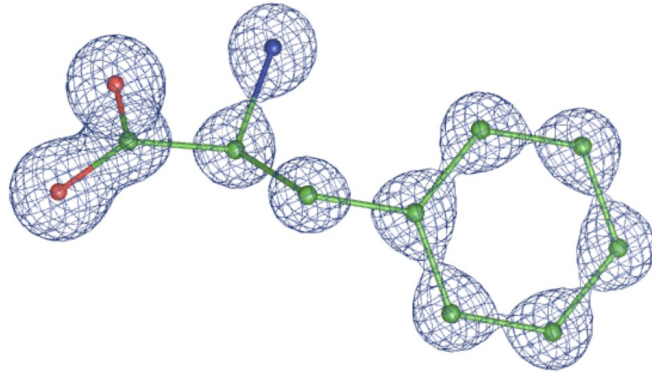


$$\rho_{MODEL}(\mathbf{r}) = \sum_{i=1}^{N_{atoms}} \rho_{atoms}(\mathbf{r})$$

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

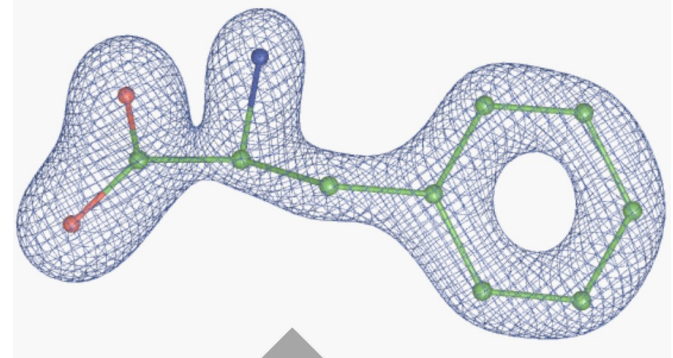
3Å model-calculated map

Exact model map



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

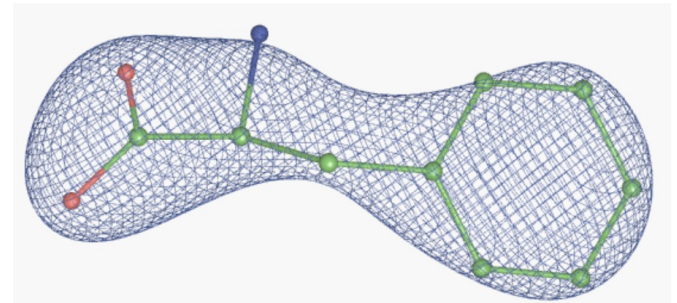
CC_{MASK}

A grey arrow pointing from the exact model map to the 3Å model-calculated map, with a large red 'X' drawn over it, indicating that this direct comparison is not the correct method for validation.

CC_{MASK}

A grey double-headed arrow pointing between the 3Å model-calculated map and the 3Å experimental map, indicating that the CC_{MASK} metric is used to compare them.

3Å experimental map



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

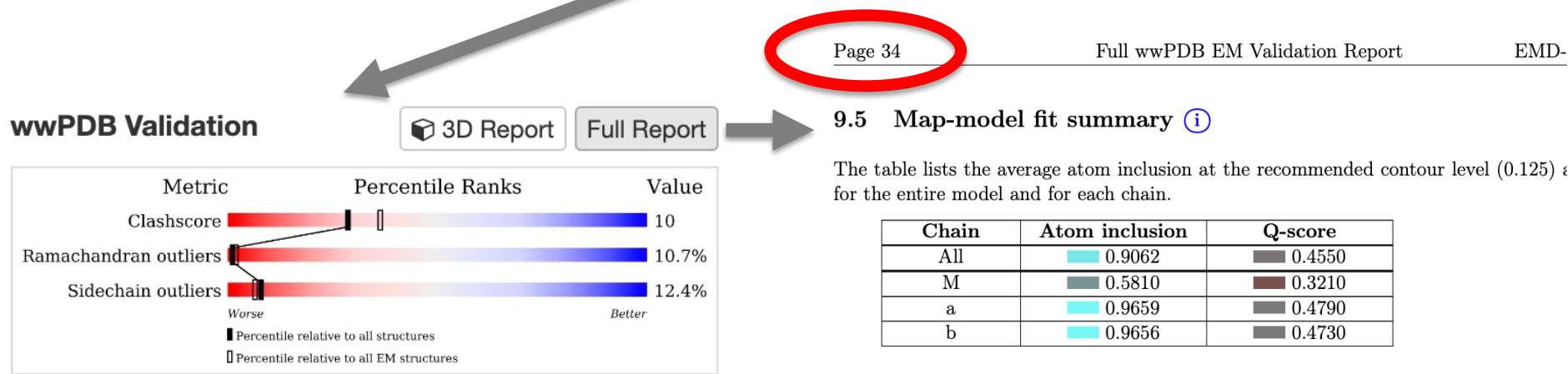
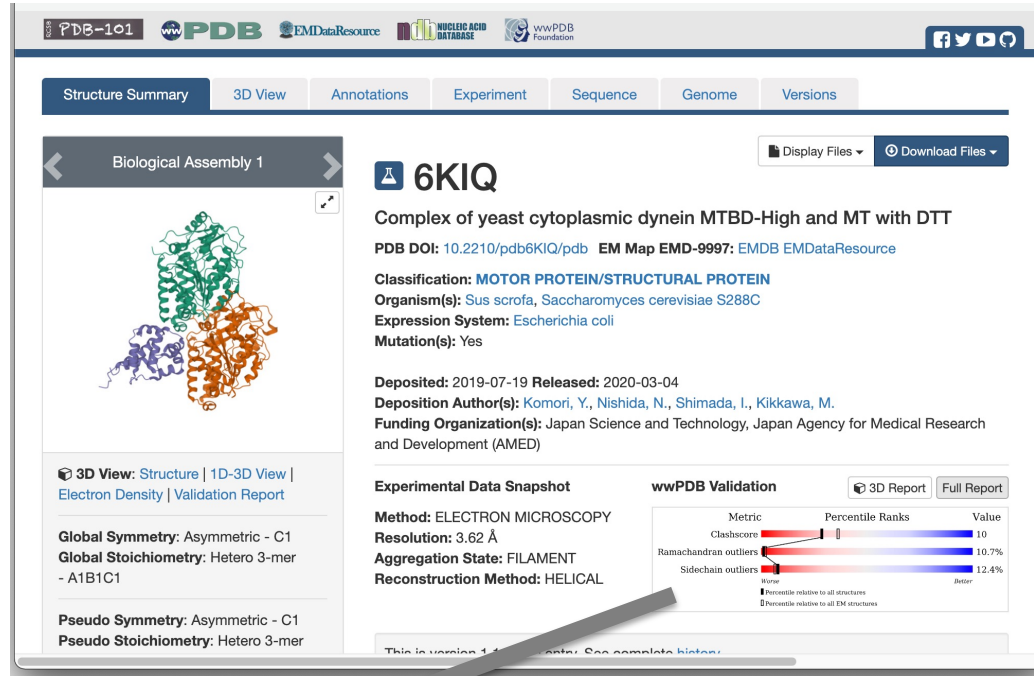
Atom inclusion

- **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 use shape of density
 - Does not account for missing atoms
 - Does not use map type
 - Unaware of alternative conformations

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
 - Does not use ADP, occupancy, atom type, ...
 - Does not use shape of density
 - Does not account for missing atoms
 - Alternative conformations are ***not*** handled
 - Anisotropic atoms are ***not*** handled

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



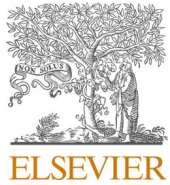
Model-to-map fit statistics is insufficient and very well hidden!

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

Variability refinement

Treasuring conformational changes



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

BBA - Biomembranes

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



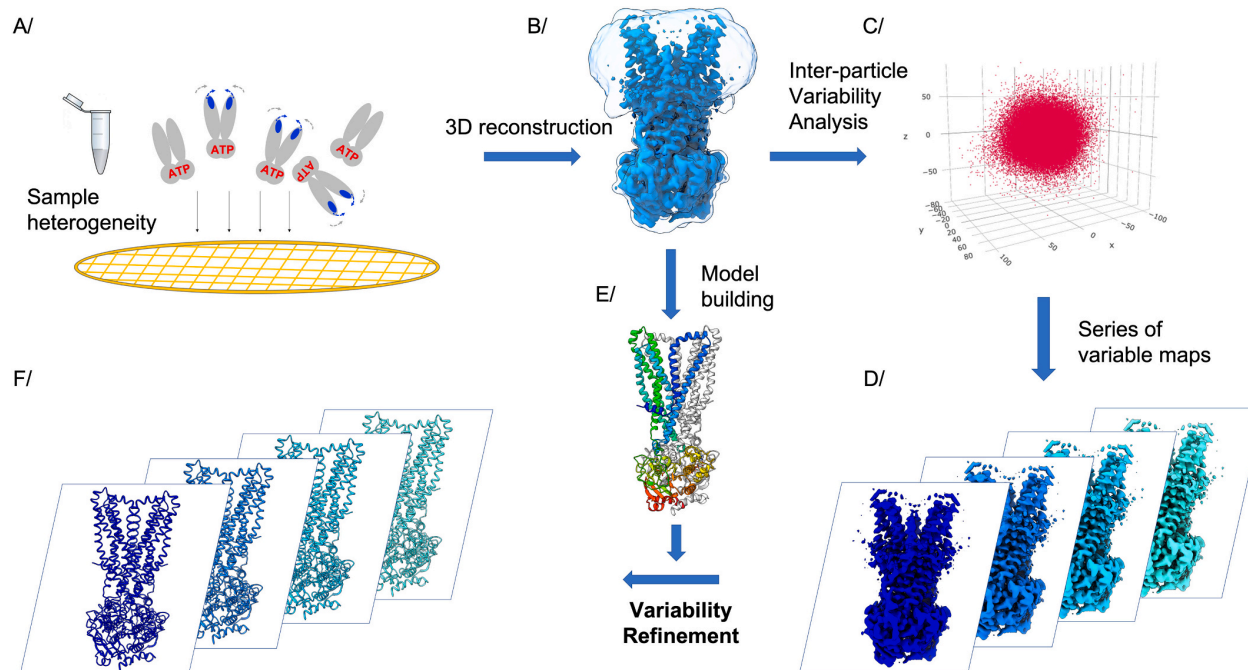
Review

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,*}

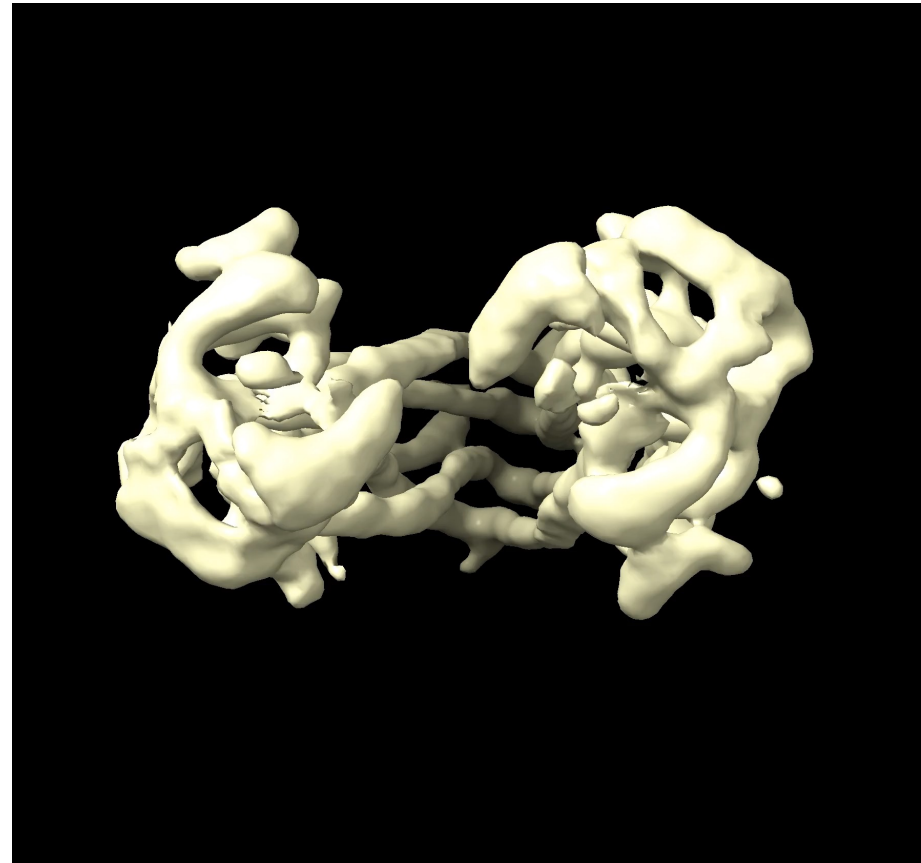
^a Molecular Biosciences and Integrated Bioimaging, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

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

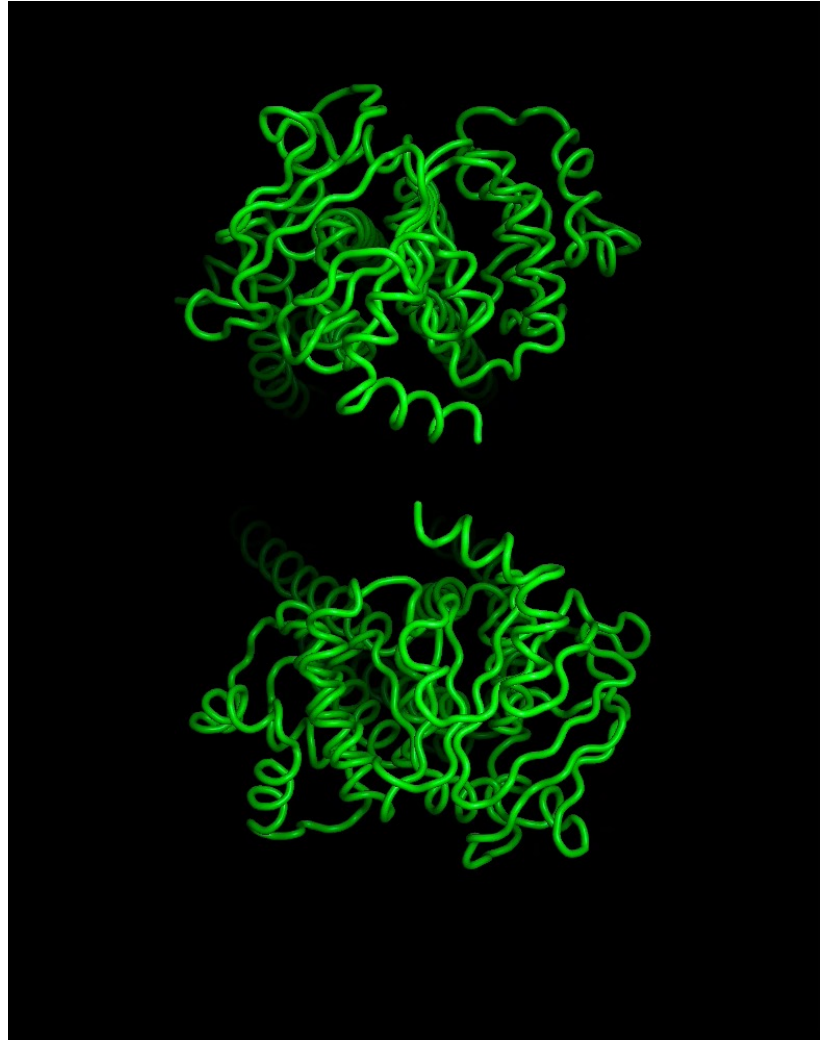
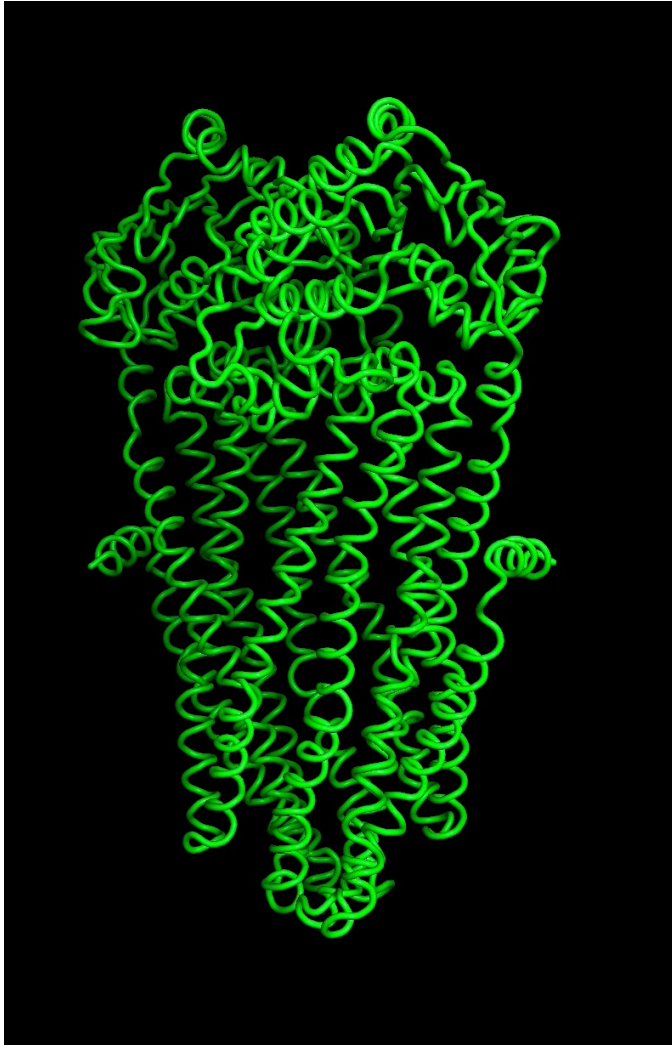


Maps

ABC transporter BmrA (unpublished!)



Refined ensembles of models



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

phenix.varref

map1.mrc ... mapN.mrc

model.pdb

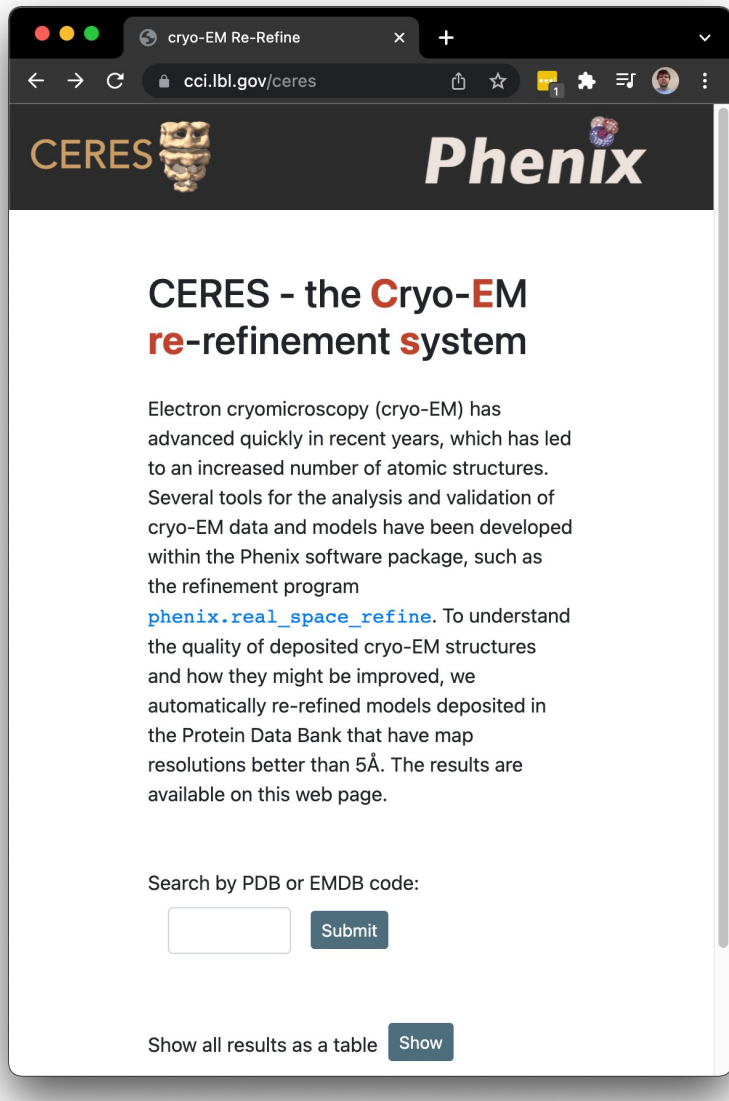
resolution=3

nproc=100

models_per_map=100

Output: ensemble of refined models that represents all maps

Automated re-refinement of deposited cryo-EM models



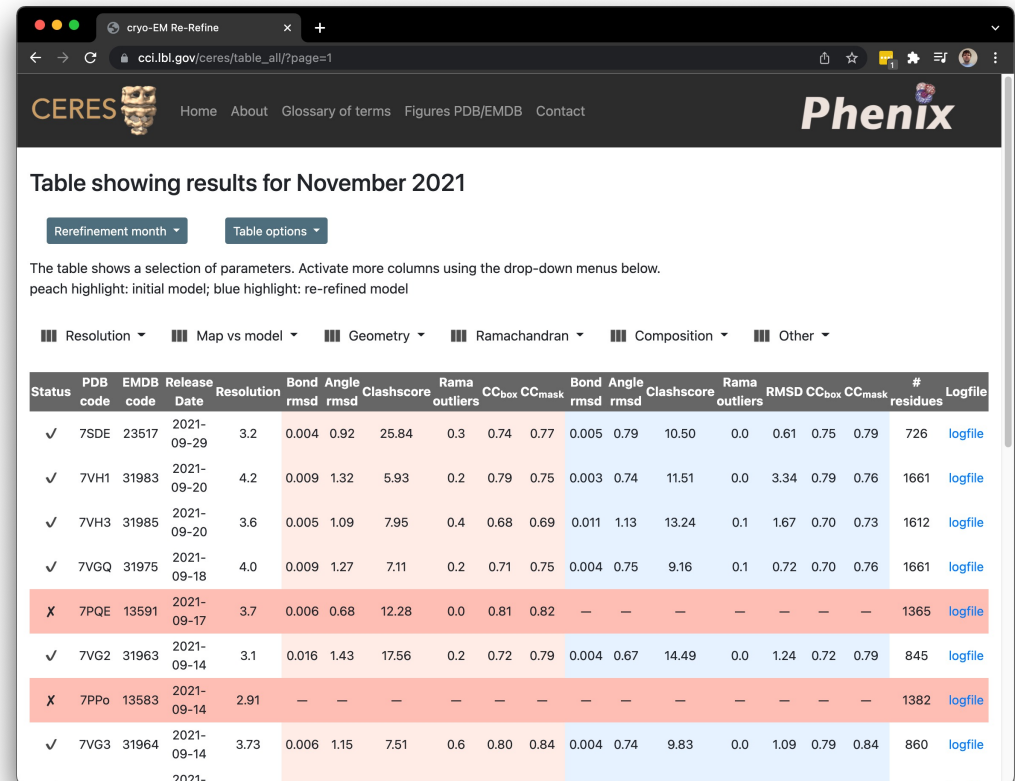
The screenshot shows the CERES website home page. The header features the CERES and Phenix logos. The main heading is "CERES - the Cryo-EM re-refinement system". The text describes the system's purpose: to improve the quality of deposited cryo-EM structures by automatically re-refining models. It mentions that the results are available on the website. At the bottom, there is a search bar for PDB or EMDB codes and a button to "Show all results as a table".

CERES - the Cryo-EM re-refinement system

Electron cryomicroscopy (cryo-EM) has advanced quickly in recent years, which has led to an increased number of atomic structures. Several tools for the analysis and validation of cryo-EM data and models have been developed within the Phenix software package, such as the refinement program [phenix.real_space_refine](#). To understand the quality of deposited cryo-EM structures and how they might be improved, we automatically re-refined models deposited in the Protein Data Bank that have map resolutions better than 5Å. The results are available on this web page.

Search by PDB or EMDB code:

Show all results as a table



The screenshot shows the CERES website table view for November 2021. The table displays a selection of parameters for various cryo-EM models, comparing the initial model (peach highlight) with the re-refined model (blue highlight). The table includes columns for Status, PDB code, EMDB code, Release Date, Resolution, Bond rmsd, Angle rmsd, Clashscore, Rama outliers, CC_{box}, CC_{mask}, and # residues. The table is filtered for November 2021.

Table showing results for November 2021

Rer refinement month: November 2021

The table shows a selection of parameters. Activate more columns using the drop-down menus below.
peach highlight: initial model; blue highlight: re-refined model

Resolution Map vs model Geometry Ramachandran Composition Other

Status	PDB code	EMDB code	Release Date	Resolution	Bond rmsd	Angle rmsd	Clashscore	Rama outliers	CC _{box}	CC _{mask}	Bond rmsd	Angle rmsd	Clashscore	Rama outliers	RMSD	CC _{box}	CC _{mask}	# residues	Logfile
✓	7SDE	23517	2021-09-29	3.2	0.004	0.92	25.84	0.3	0.74	0.77	0.005	0.79	10.50	0.0	0.61	0.75	0.79	726	logfile
✓	7VH1	31983	2021-09-20	4.2	0.009	1.32	5.93	0.2	0.79	0.75	0.003	0.74	11.51	0.0	3.34	0.79	0.76	1661	logfile
✓	7VH3	31985	2021-09-20	3.6	0.005	1.09	7.95	0.4	0.68	0.69	0.011	1.13	13.24	0.1	1.67	0.70	0.73	1612	logfile
✓	7VGQ	31975	2021-09-18	4.0	0.009	1.27	7.11	0.2	0.71	0.75	0.004	0.75	9.16	0.1	0.72	0.70	0.76	1661	logfile
X	7PQE	13591	2021-09-17	3.7	0.006	0.68	12.28	0.0	0.81	0.82	—	—	—	—	—	—	—	1365	logfile
✓	7VG2	31963	2021-09-14	3.1	0.016	1.43	17.56	0.2	0.72	0.79	0.004	0.67	14.49	0.0	1.24	0.72	0.79	845	logfile
X	7PPo	13583	2021-09-14	2.91	—	—	—	—	—	—	—	—	—	—	—	—	—	1382	logfile
✓	7VG3	31964	2021-09-14	3.73	0.006	1.15	7.51	0.6	0.80	0.84	0.004	0.74	9.83	0.0	1.09	0.79	0.84	860	logfile

- Developers: helps track the impact of new methods and tools
- Users: lets to see how their models can benefit from improved methods and tools