

# *Model Refinement: cryo-EM*

**Pavel Afonine**



[phenix-online.org](http://phenix-online.org)



[lbl.gov](http://lbl.gov)



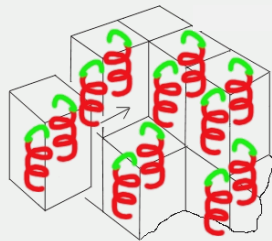
[qrefine.com](http://qrefine.com)

ACA, Denver, Colorado

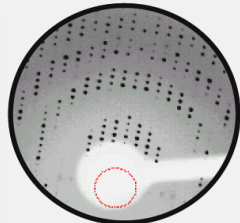
July 7<sup>th</sup> 2024

# Refinement in Phenix

## Crystallography



Initial model



Experimental  
data

*A priori*  
knowledge

Score

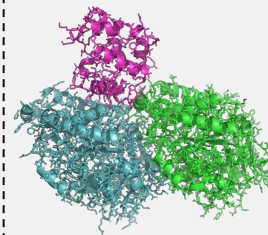
Modify model  
parameters

Improved  
model

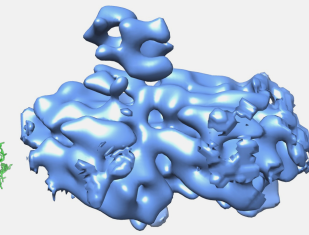
phenix.refine

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

# 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

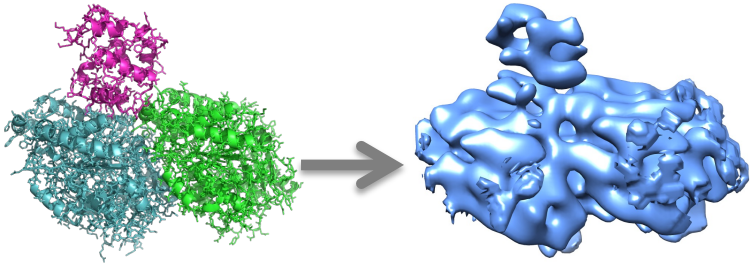
## Real-space refinement in *PHENIX* for cryo-EM and crystallography

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

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

# Model-to-map fit validation: $CC_{\text{MASK}}$

## Model to map fit



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

$\rho_{\text{obs}}$  = experimental map

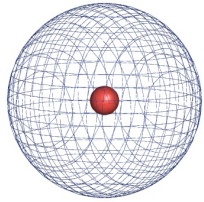
$\rho_{\text{calc}}$  = model calculated map

- 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
$CC_{\text{MASK}}$	Poor: < 0.3 So-so: 0.3-0.6 Good: > 0.6

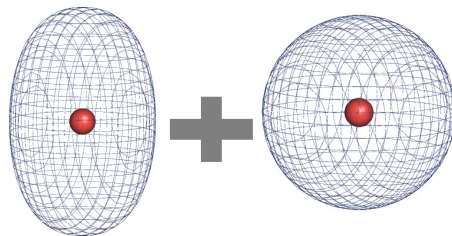
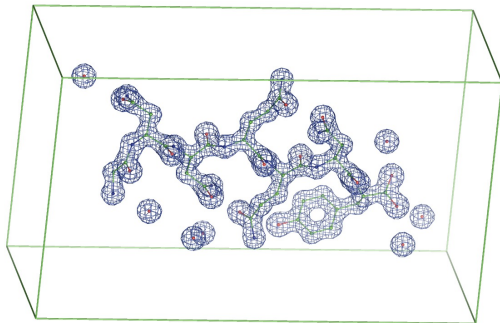
# Model-to-map fit validation: $CC_{\text{MASK}}$

- Gaussian IAM (Independent Atom Model)



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

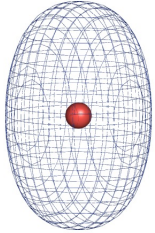
ATOM	25	CA	PRO	A	4	31.309	29.489	26.044	1.00	57.79	C
------	----	----	-----	---	---	--------	--------	--------	------	-------	---



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

# Model map

- Gaussian IAM (Independent Atom Model)
- Anisotropic:



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

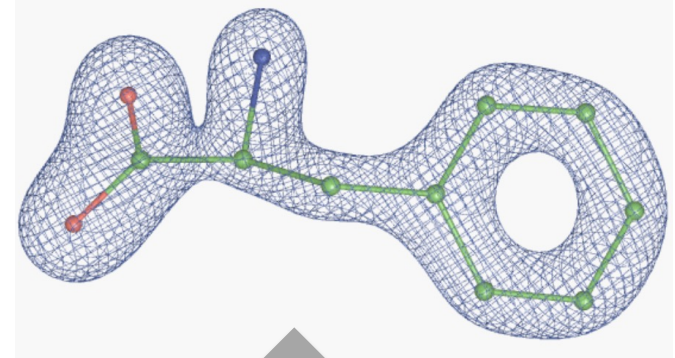
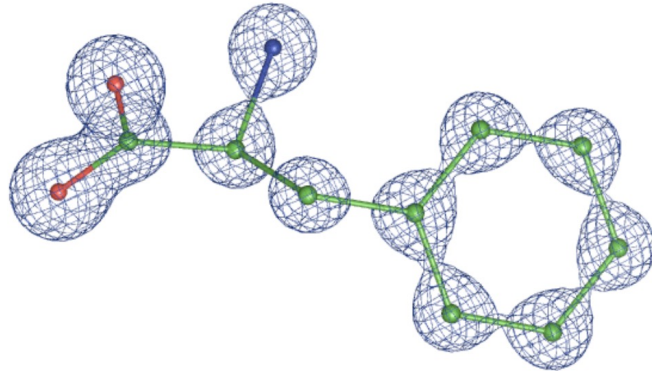
ATOM	25	CA	PRO	A	4	31.309	29.489	26.044	1.00	57.79	C	
ANISOU	25	CA	PRO	A	4	8443	7405	6110	2093	-24	-80	C



# Model-to-map fit validation: $CC_{\text{MASK}}$

## 3Å model-calculated map

## Exact model map



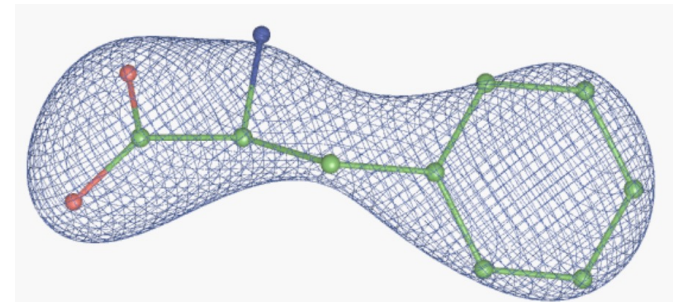
$CC_{\text{MASK}}$



$CC_{\text{MASK}}$



## 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**

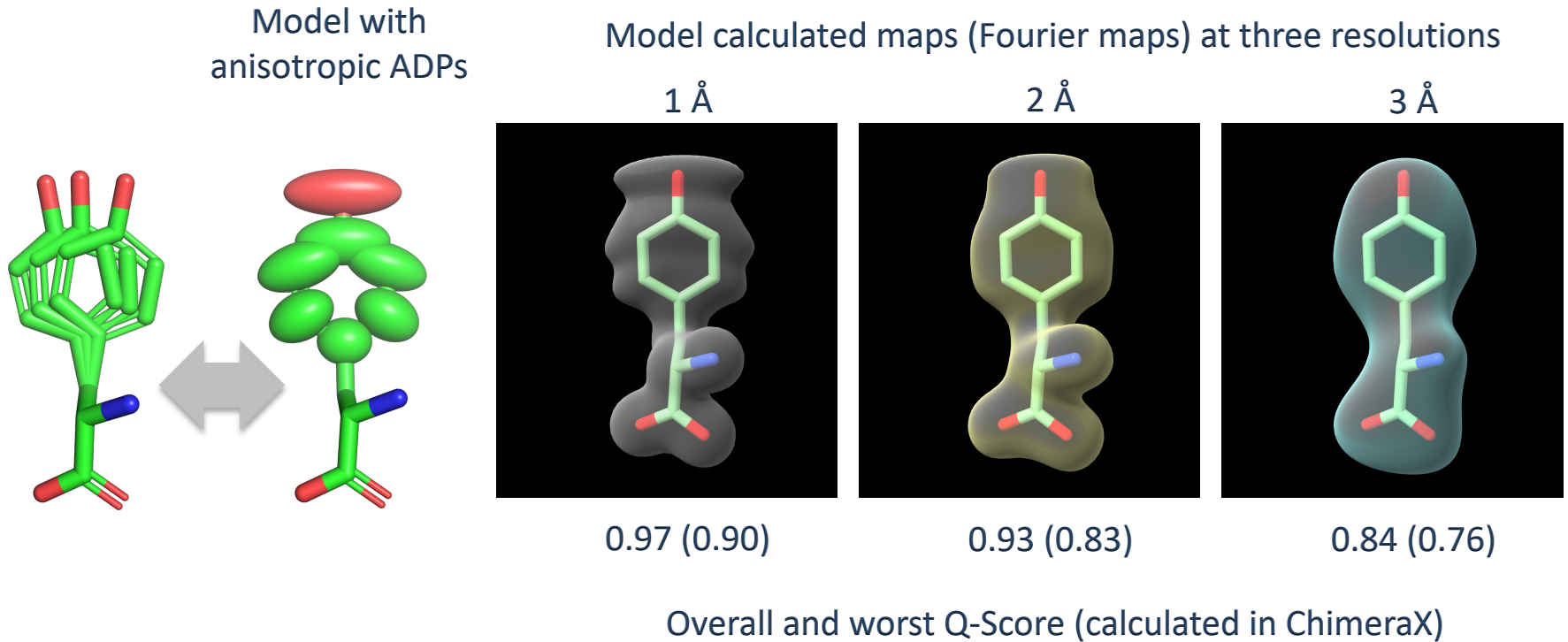
# Atom inclusion

- **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

Structure Summary | 3D View | Annotations | Experiment | Sequence | Genome | Versions

Biological Assembly 1

**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  
3D Report | Full Report

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

Global Symmetry: Asymmetric - C1  
Global Stoichiometry: Hetero 3-mer - A1B1C1  
Pseudo Symmetry: Asymmetric - C1  
Pseudo Stoichiometry: Hetero 3-mer

Page 34

Full wwPDB EM Validation Report

EMD-

wwPDB Validation

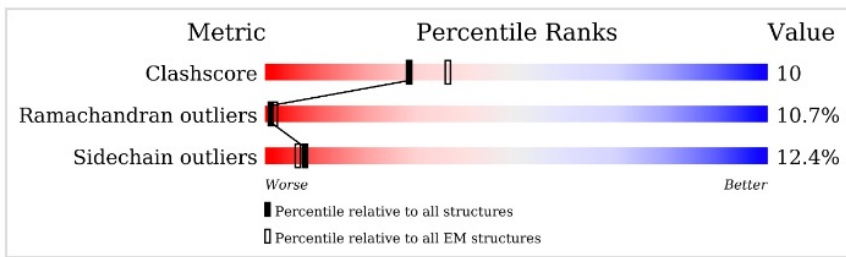
3D Report

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

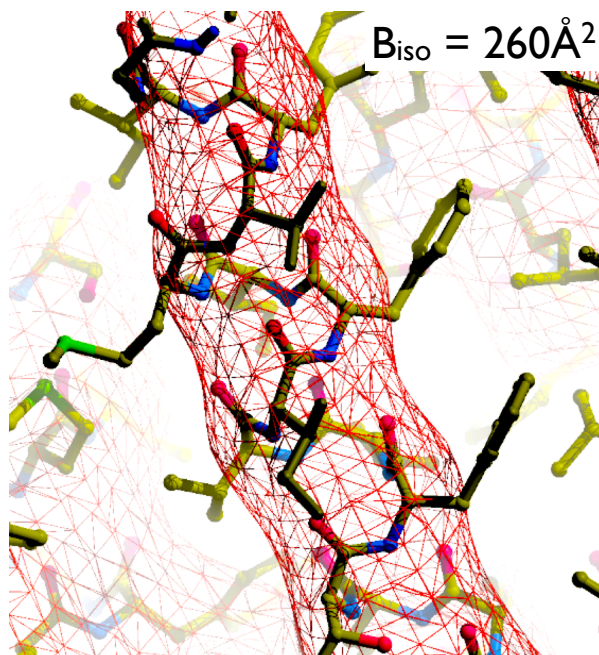


**Model-to-map fit statistics is insufficient and very well hidden!**

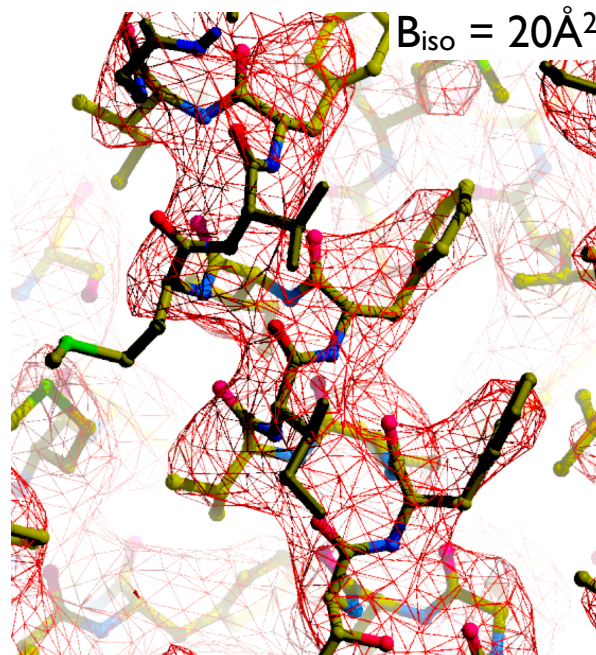
# Refinement: practical considerations

- 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

# Map sharpening and refinement



*Deposited Map*



*Autosharpened Map*

EMDB: 8414, PDB: 5tji

Fully automatic:

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



STRUCTURAL  
BIOLOGY

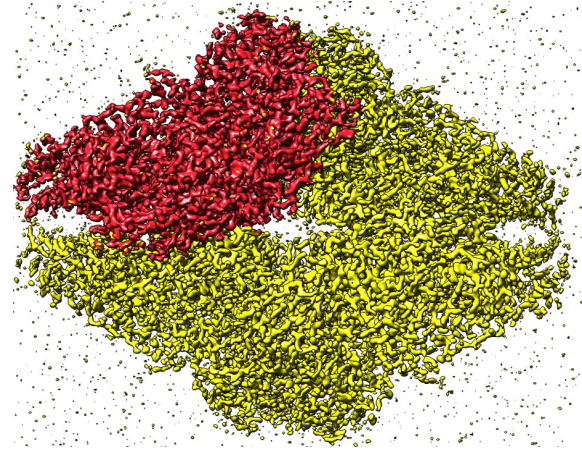
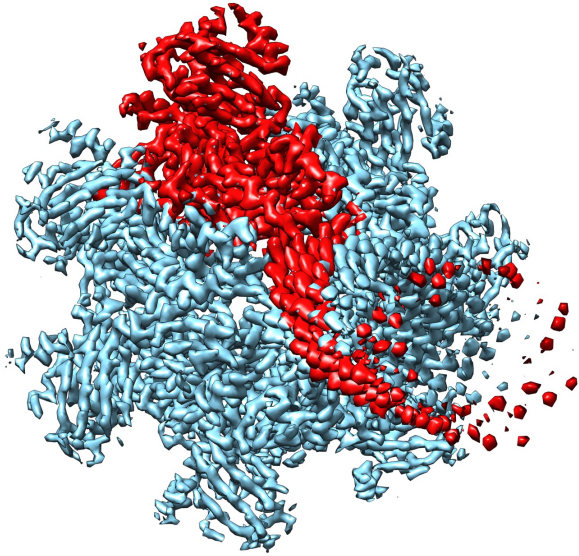
ISSN 2059-7983

Automated map sharpening by maximization of  
detail and connectivity

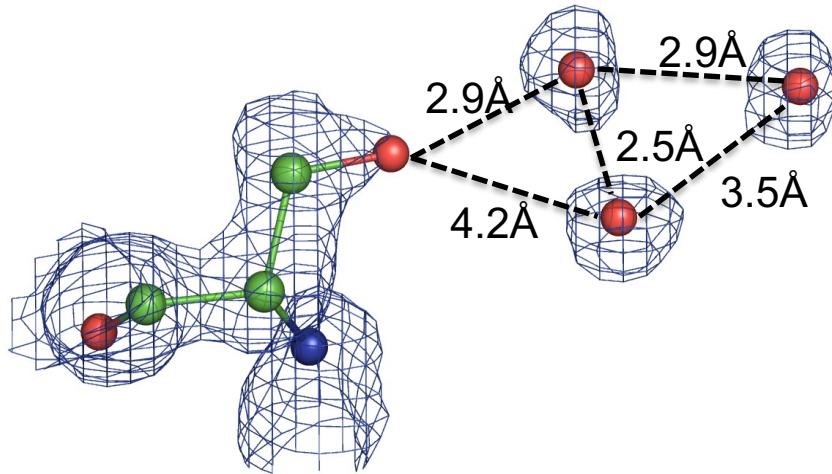
Thomas C. Terwilliger,<sup>a,b\*</sup> Oleg V. Sobolev,<sup>c</sup> Pavel V. Afonine<sup>c,d</sup> and  
Paul D. Adams<sup>d,e</sup>



# Map symmetry and refinement

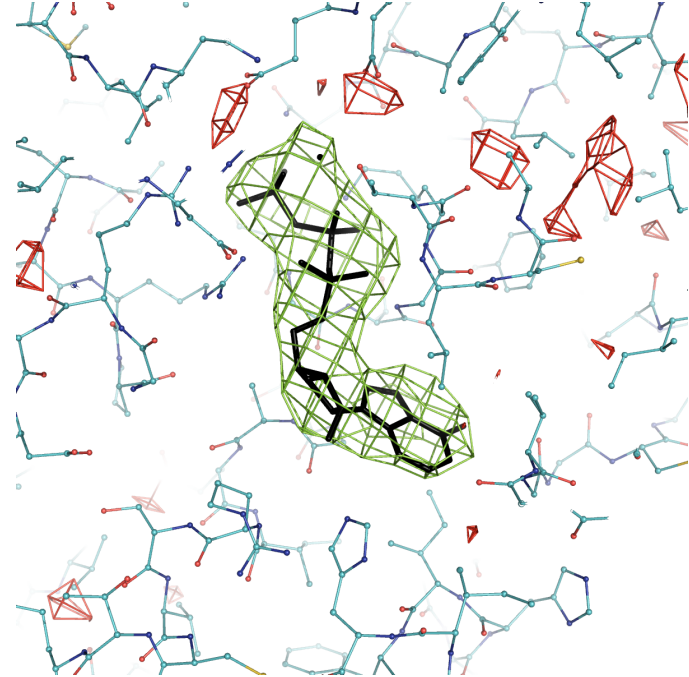
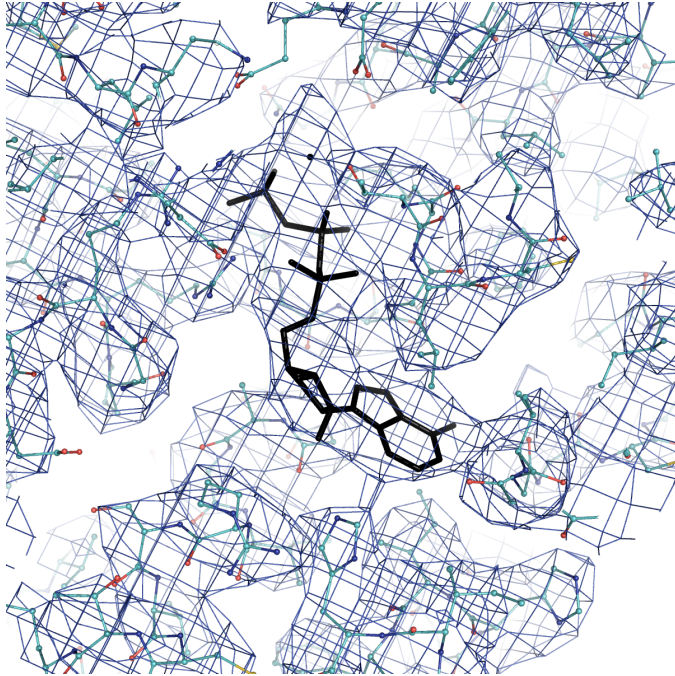


# Solvent building and refinement: phenix.douse



**Available in ChimeraX!**

# Maps and refinement



- Analogue of crystallographic Fo-Fc map
- Requires well-refined model (including B factors)

# Variability refinement

# Treasuring conformational changes



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

BBA - Biomembranes

journal homepage: [www.elsevier.com/locate/bbamem](https://www.elsevier.com/locate/bbamem)



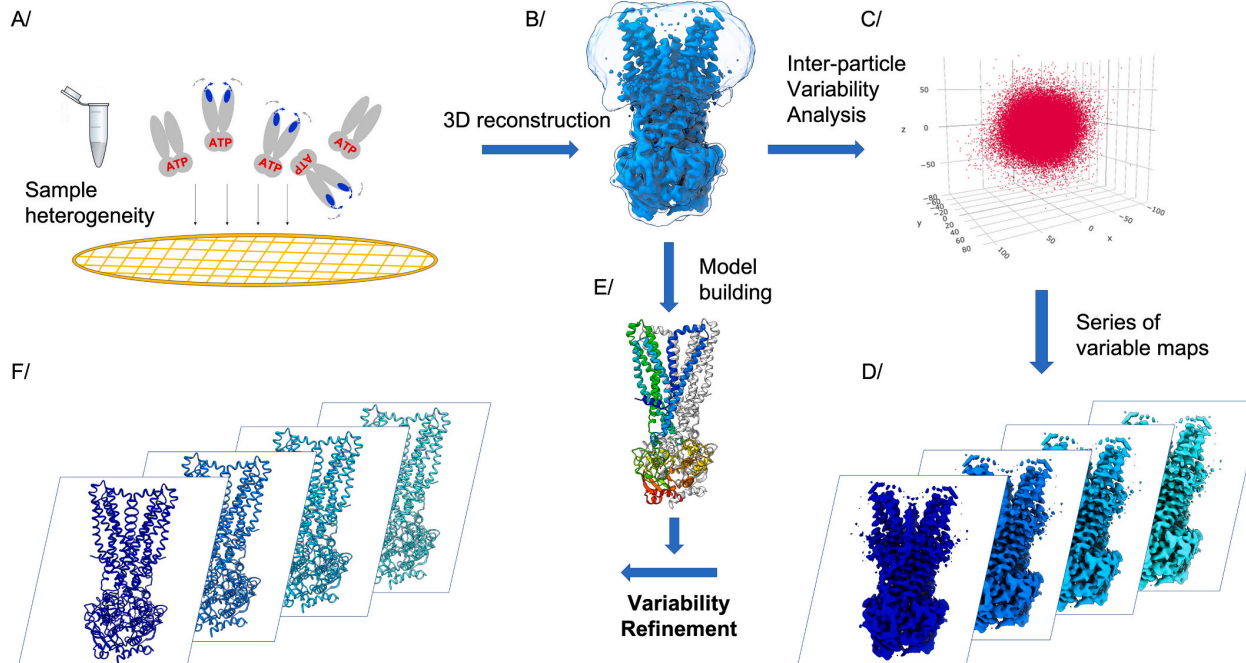
Review

## Conformational space exploration of cryo-EM structures by variability refinement

Pavel V. Afonine<sup>a,\*</sup>, Alexia Gobet<sup>b</sup>, Loïck Moissonnier<sup>b</sup>, Juliette Martin<sup>b</sup>, Billy K. Poon<sup>a</sup>, Vincent Chaptal<sup>b,\*</sup>

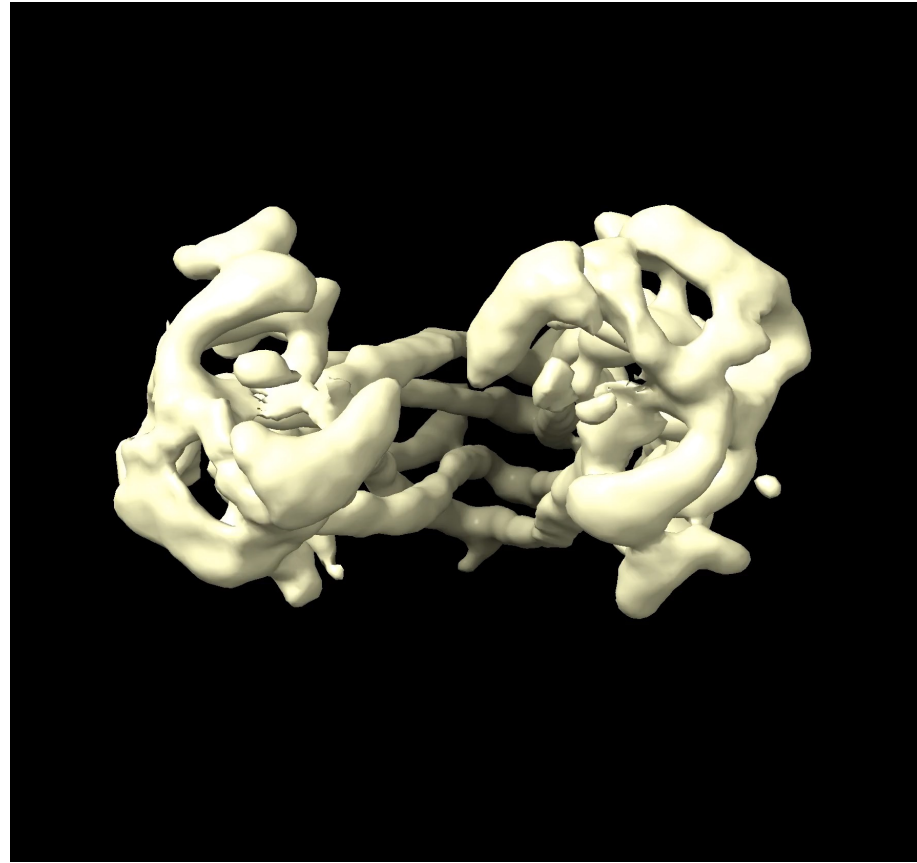
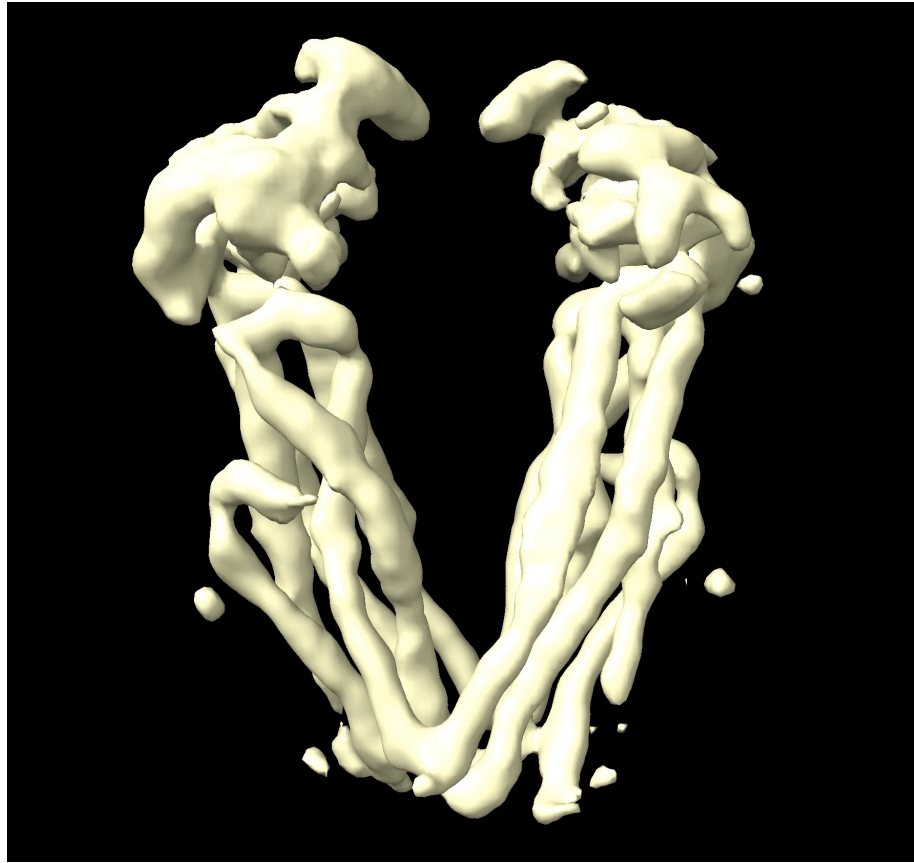
<sup>a</sup> Molecular Biosciences and Integrated Biomaging, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

<sup>b</sup> Molecular Microbiology and Structural Biochemistry, UMR5086 CNRS University Lyon1, 7 passage du Vercors, 69007 Lyon, France



# Maps

*ABC transporter BmrA (unpublished!)*



*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

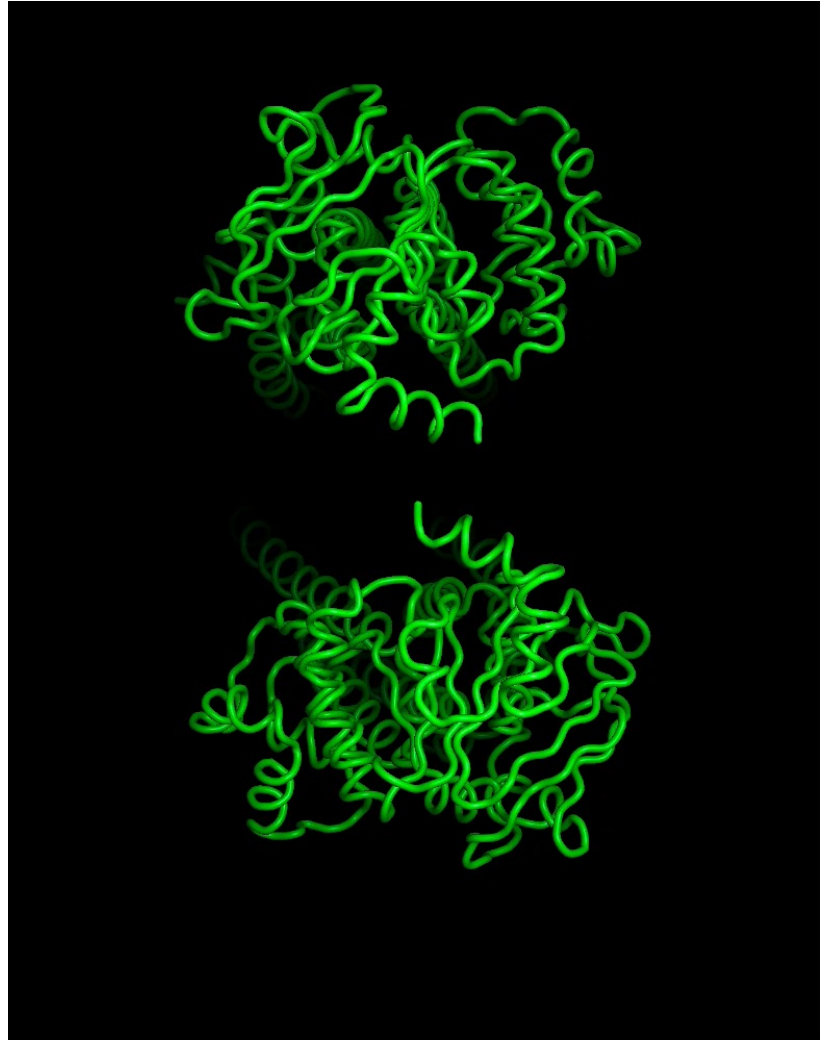
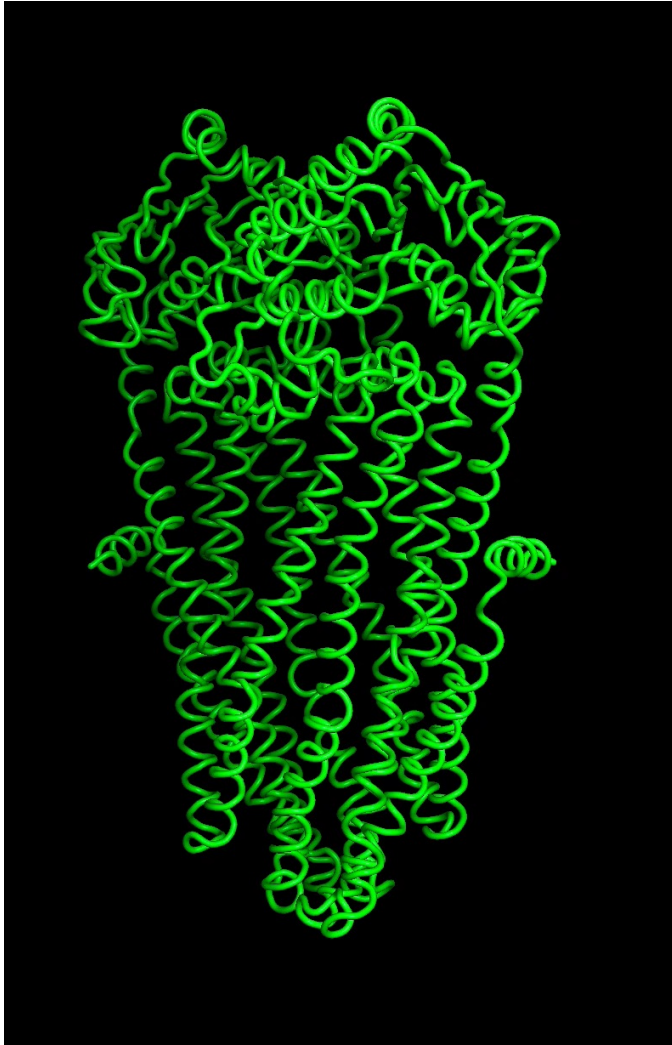
**Output:** ensemble of refined models that represents all maps

# Workflow

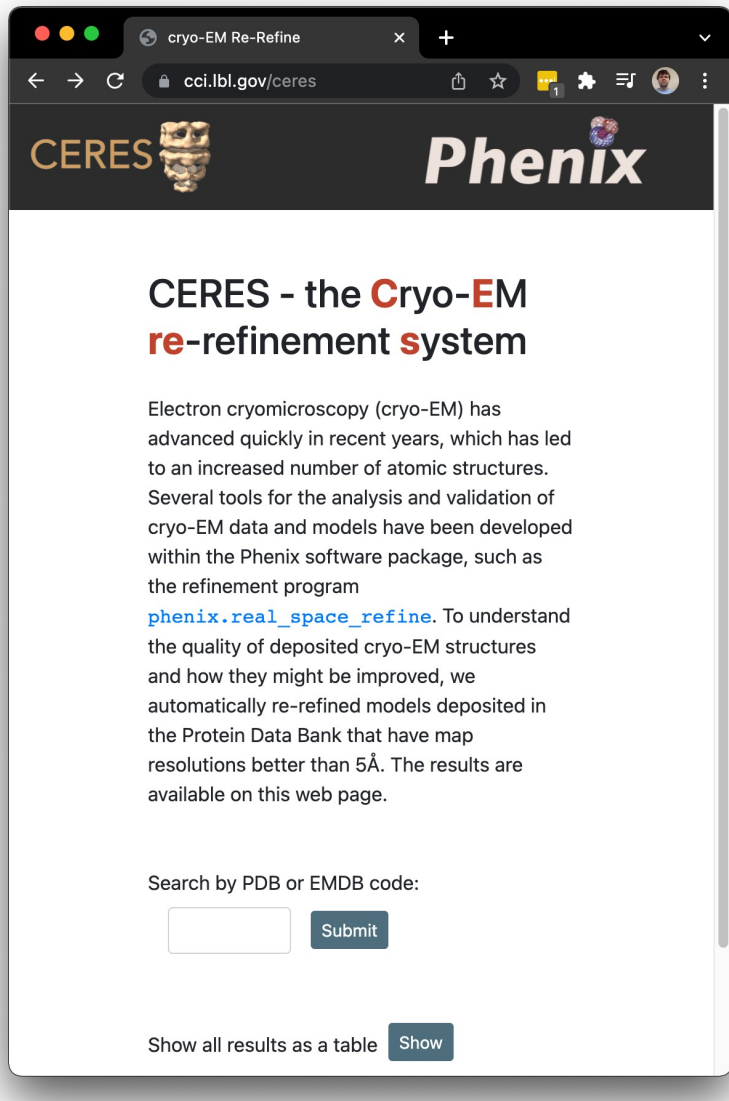
- Input model and maps
- Order maps by similarity using  $CC_{\text{box}}$
- Identify the map that is closest to input model (by  $CC_{\text{mask}}$ )
  - This is the starting point for the first refinement
  - Generate ensemble of 100 perturbed models (by MD)
  - Refine each model with *phenix.real\_space\_refine*
  - 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)



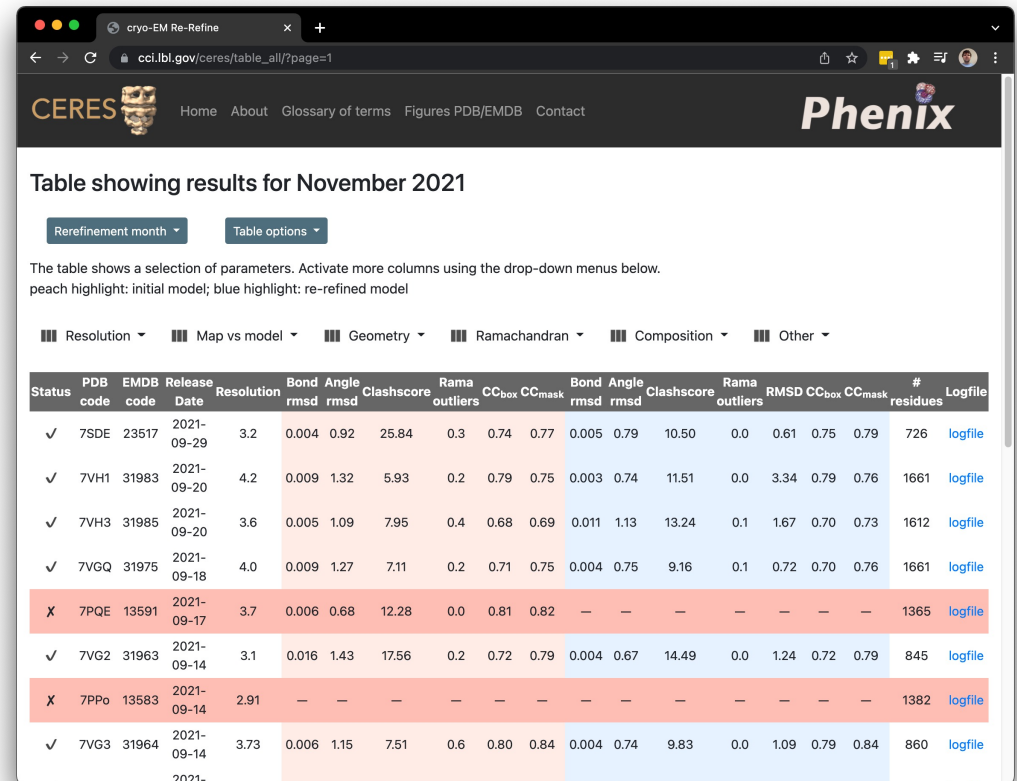
# Refined ensembles of models



# Automated re-refinement of deposited cryo-EM models



The screenshot shows the CERES website home page. The header features the CERES logo and the Phenix logo. The main heading is "CERES - the Cryo-EM re-refinement system". Below this, a paragraph explains that electron cryomicroscopy (cryo-EM) has advanced quickly, leading to more atomic structures. It mentions that several tools for analysis and validation have been developed within the Phenix software package, such as the refinement program [phenix.real\\_space\\_refine](#). The text states that to understand the quality of deposited cryo-EM structures and how they might be improved, models deposited in the Protein Data Bank are automatically re-refined. The results are available on this web page. At the bottom, there is a search bar with a "Submit" button and a "Show" button to display results as a table.



The screenshot shows a table of results for November 2021. The table has columns for Status, PDB code, EMDB code, Release Date, Resolution, Bond rmsd, Angle rmsd, Clashscore, Rama outliers, CC<sub>box</sub>, CC<sub>mask</sub>, Bond rmsd, Angle rmsd, Clashscore, Rama outliers, RMSD, CC<sub>box</sub>, CC<sub>mask</sub>, # residues, and Logfile. The table is filtered for 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.

Status	PDB code	EMDB code	Release Date	Resolution	Bond rmsd	Angle rmsd	Clashscore	Rama outliers	CC <sub>box</sub>	CC <sub>mask</sub>	Bond rmsd	Angle rmsd	Clashscore	Rama outliers	RMSD	CC <sub>box</sub>	CC <sub>mask</sub>	# 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	<a href="#">logfile</a>
✓	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	<a href="#">logfile</a>
✓	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	<a href="#">logfile</a>
✓	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	<a href="#">logfile</a>
X	7PQE	13591	2021-09-17	3.7	0.006	0.68	12.28	0.0	0.81	0.82	—	—	—	—	—	—	—	1365	<a href="#">logfile</a>
✓	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	<a href="#">logfile</a>
X	7PPo	13583	2021-09-14	2.91	—	—	—	—	—	—	—	—	—	—	—	—	—	1382	<a href="#">logfile</a>
✓	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	<a href="#">logfile</a>

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