Refinement (Cryo-EM)

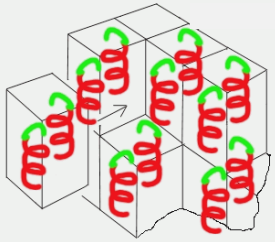
Pavel Afonine

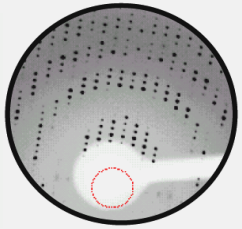
Lawrence Berkeley National Lab, California, USA

July 29, 2022
ACA
Portland, Oregon

Refinement

Crystallography





Initial model

Experimental data

A priori knowledge

Score

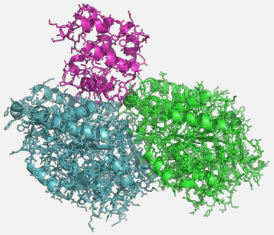
Modify model parameters

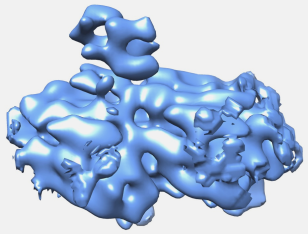
Improved model

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: *phenix.real_space_refine*

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]

Atomic model refinement: crystallography vs cryo-EM

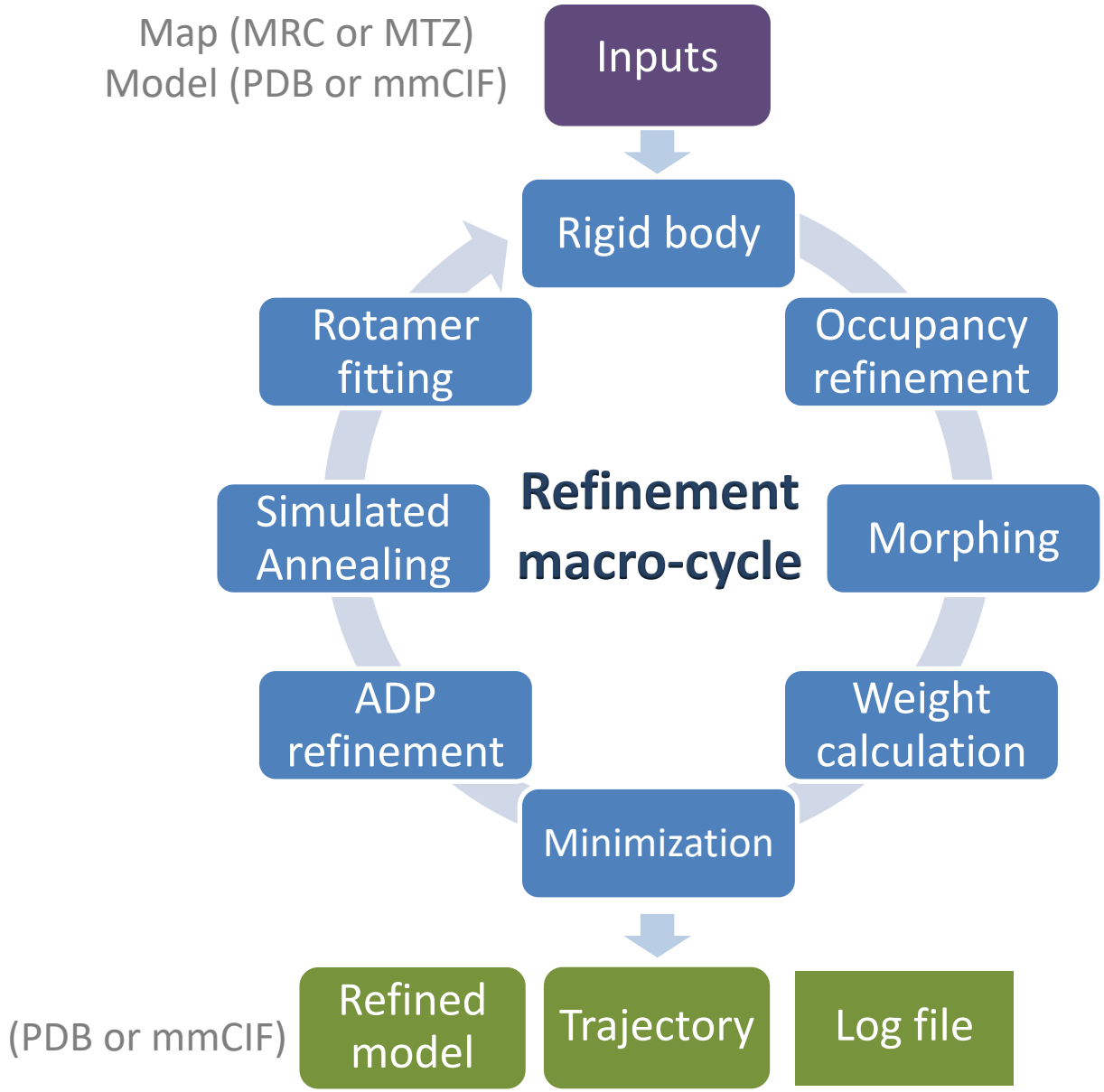
Crystallographic refinement

- 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

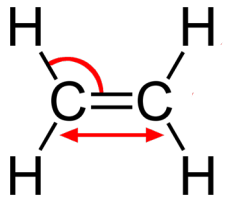
- Changing model does not change map
  - Build solvent (water) last
  - Get as complete and accurate model as possible before refining B factors and occupancies
- Experimental data changes a lot during the process (filtering, boxing, using maps with implied symmetry or not, etc.)
  - What map to use 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

Refinement protocol

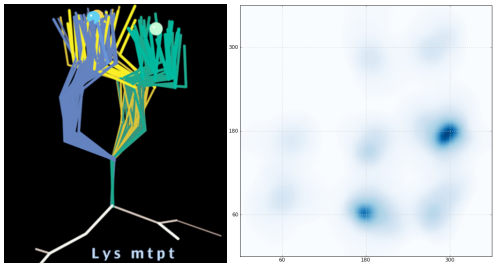


Restraints

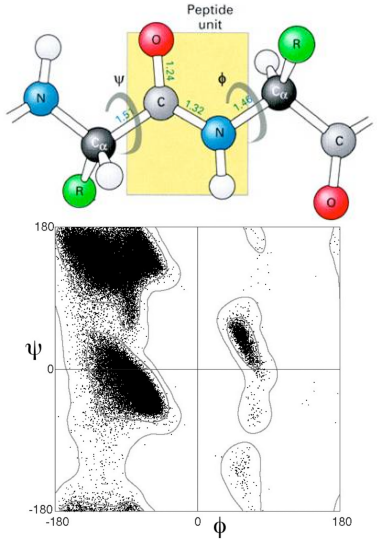
Covalent geometry

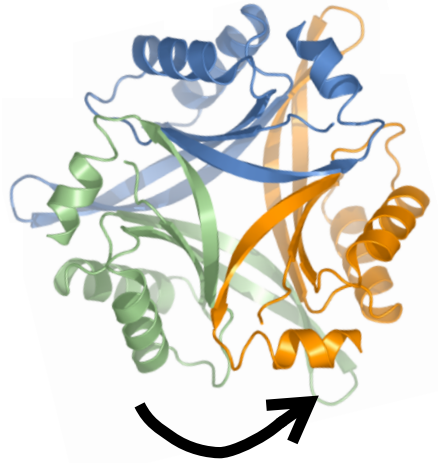


Side chain distributions

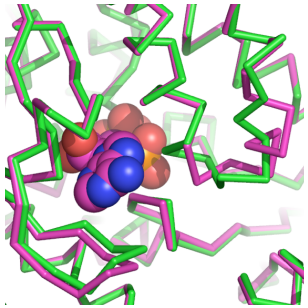


Main chain distributions

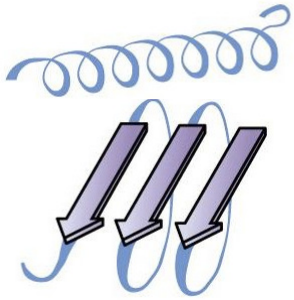




Internal symmetry (NCS)

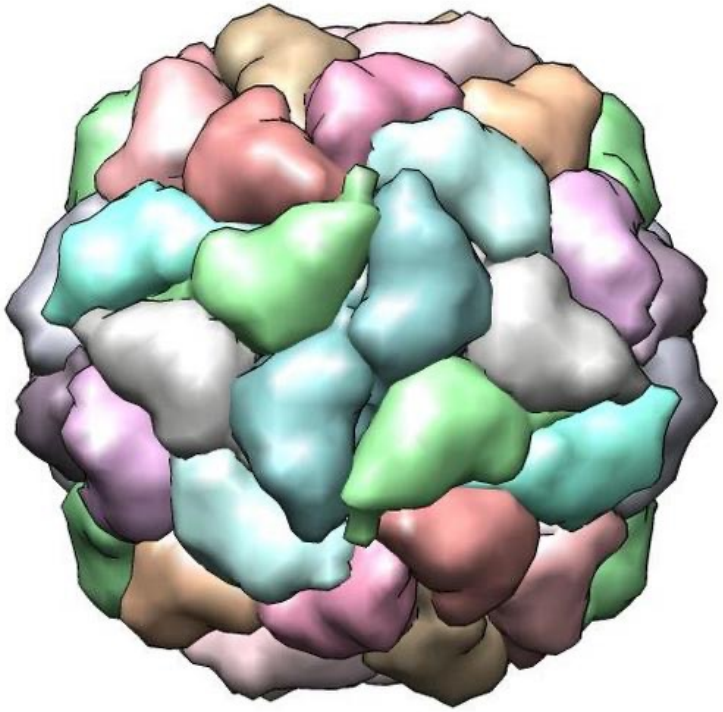


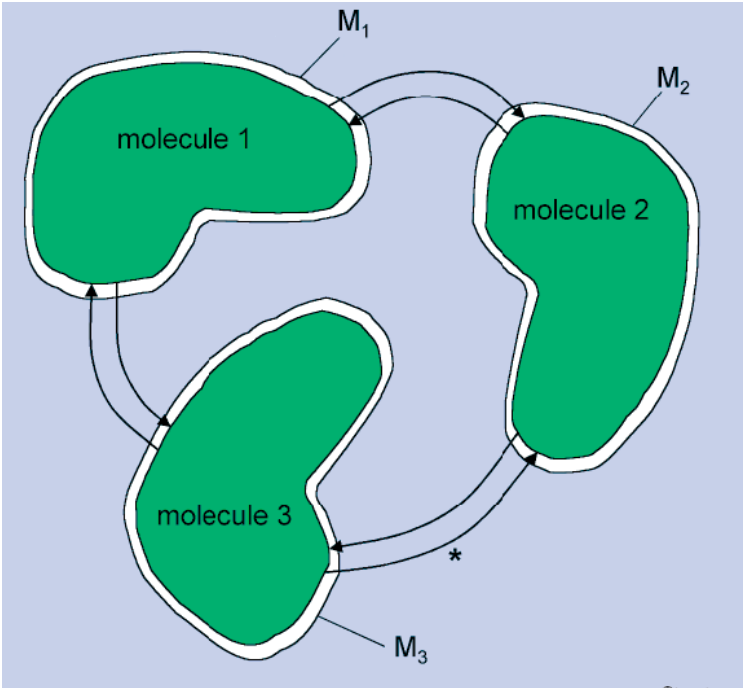
Similar (homologous) structures (reference model restraints)



Secondary structure

NCS (molecular symmetry)





Source: Internet

- **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

Real-space refinement – B-factors (new)

- Much faster than in previous versions
- Isotropic group or individual
- Symmetry-aware
  - Strict symmetry: B-factors are identical across symmetry copies
- Can use multiprocessing
- Data / restraints weight is always optimal
  - Ensures best model-to-map fit and physically meaningful B-factors

- Refine B-factors at the very last step
  - Because refined model does not feed back to the map
- Refined B-factor values depend on the map used (original *vs* sharpened or else manipulated)

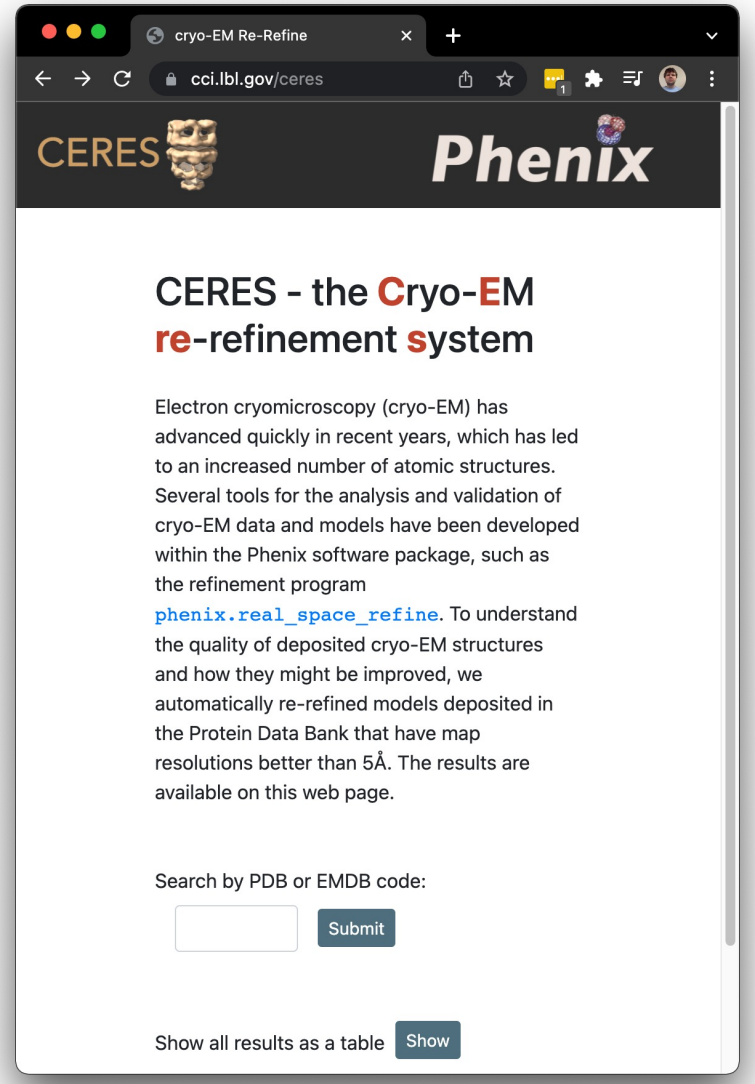
Real-space refinement – occupancy (new)

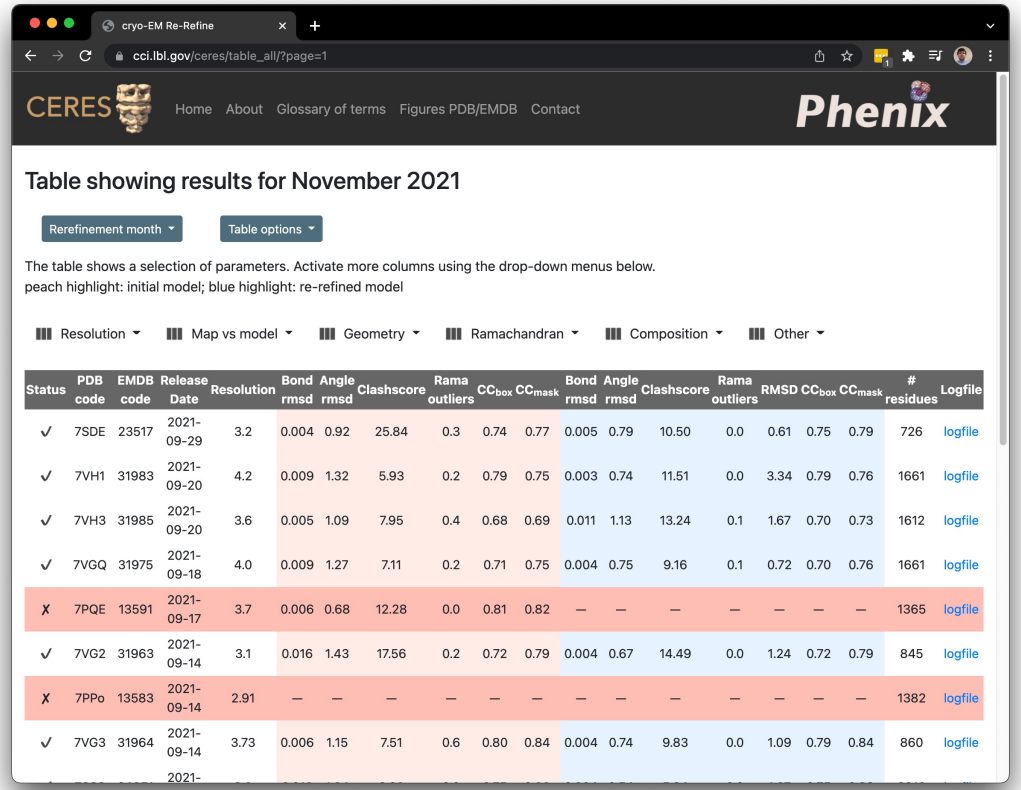
- Symmetry-aware
  - Strict symmetry: B-factors are identical across symmetry copies
- Can use multiprocessing
- Rules are similar to *phenix.refine*:
  - Look for (phenix-online.org):

    13 typical occupancy refinement scenarios and available options in phenix.refine

- Refine occupancies at the very last step
  - Because refined model does not feed back to the map

Automated re-refinement of deposited cryo-EM models

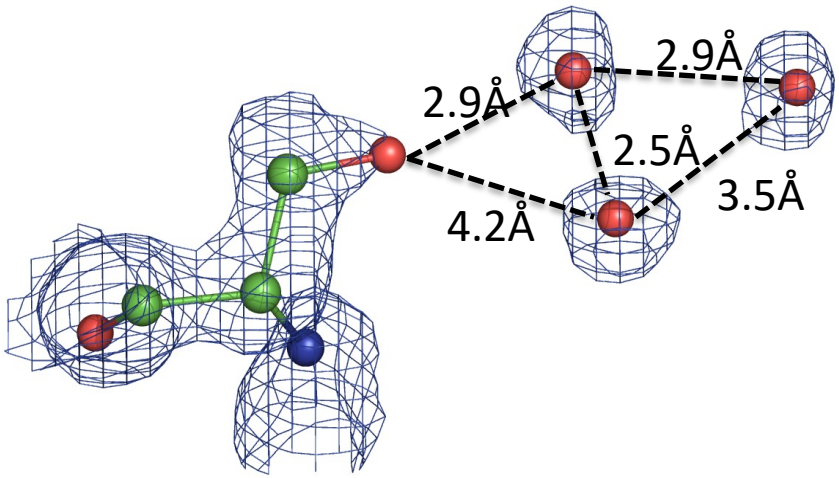




Status	PDB code	EMDB code	Release Date	Resolution	Bond rmsd	Angle rmsd	Clashscore	Rama outliers	CC_{box}	CC_{mask}	Bond rmsd	Angle rmsd	Clashscore	Rama outliers	RMSD	CC_{box}	CC_{mask}	# residues	Logfile
✓	7SDE	23517	2021-09-29	3.2	0.004	0.92	25.84	0.3	0.74	0.77	0.005	0.79	10.50	0.0	0.61	0.75	0.79	726	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
✗	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
✗	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

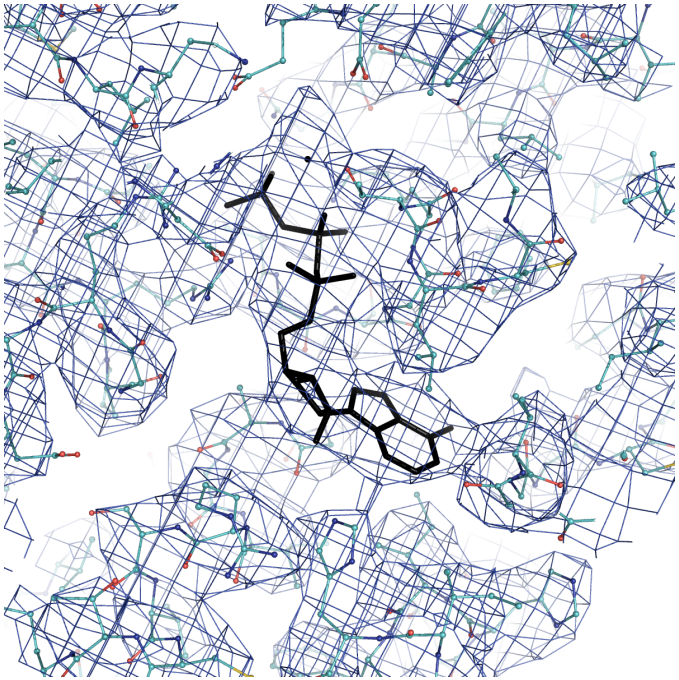
- Developers: helps track the impact of new methods and tools
- Users: lets to see how their models can benefit from improved methods and tools

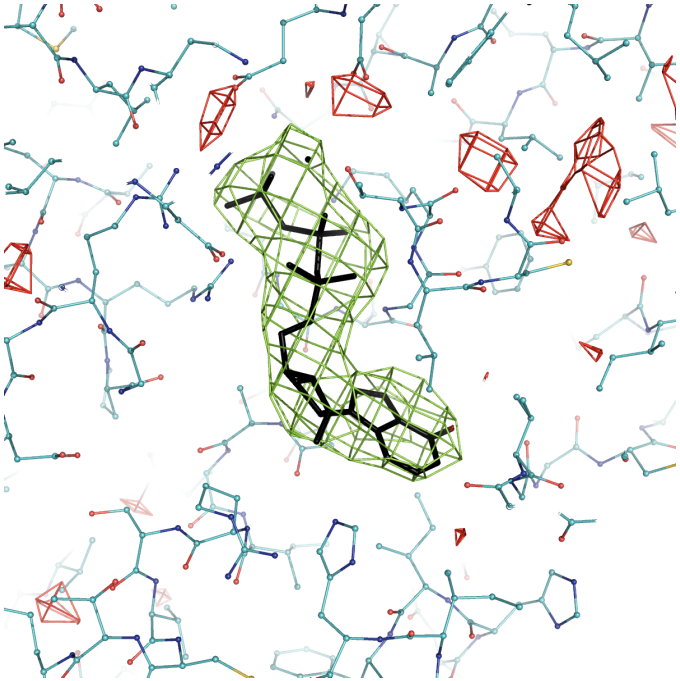
Automated water building: *phenix.douse*



Difference maps

5L4g, EMDB 4002





User support

- **Feedback, questions, help**

  Mailing list (all, developers and users): phenixbb@phenix-online.org
  Bug reports (developers only): bugs@phenix-online.org
  Ask for help (developers only): 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
  - All data sent to us kept confidentially

The Phenix Project

Lawrence Berkeley Laboratory

Paul Adams, Pavel Afonine, Dorothee Liebschner, Nigel Moriarty, Billy Poon, Christopher Schlicksup, Oleg Sobolev

Los Alamos National Laboratory
New Mexico Consortium

Tom Terwilliger, Li-Wei Hung

UTHealth

Matt Baker, Corey Hyrc

University of Cambridge

Randy Read, Airlie McCoy, Tristan Croll, Claudia Millán Nebot, Rob Oeffner

Duke University

Jane & David Richardson, Christopher Williams, Vincent Chen

An NIH/NIGMS funded Program Project

Liebschner D, *et al.*, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. Acta Cryst. 2019 **D75**:861–877