Phenix Tools for Validated Refinement of Atomic Models into maps (low-resolution, Cryo-EM, X-ray or neutron)

Pavel V. Afonine | **Youval Dar** | **Oleg V. Sobolev** | **Thomas C. Terwilliger** | **Alexandre Urzhumtsev** | **Paul D. Adams**

Lawrence Berkeley National Lab | **Los Alamos National Lab** | **Université de Lorraine & IGBMC, France** | **University of California, Berkeley**

**www.phenix-online.org | PAfonine@lbl.gov**

SUMMARY

The *phenix.real_space_refine* program implements both new *real-space refinement* methods and in-depth structure validation integrated into refinement workflow. The program is designed to fit atomic models into various kinds of maps (X-ray, cryo-EM, neutron) of broad range of quality (high- to low-resolution) and produce refined models with no or minimal geometry violations such as rotamer or Ramachandran plot outliers.

MOTIVATIONS

- **Rapidly increasing demand**
  - Cryo-EM structures in PDB by year (March 2015)
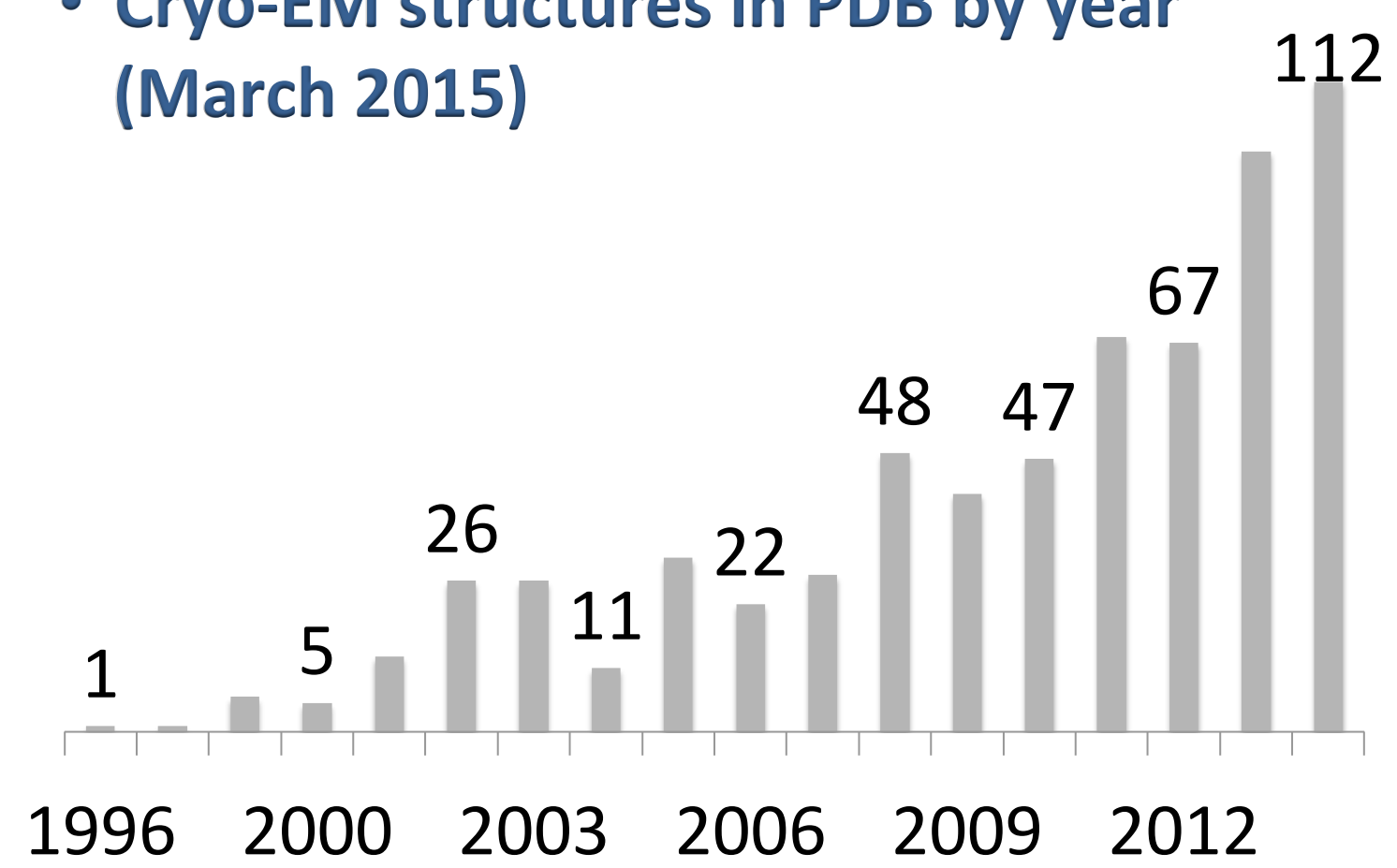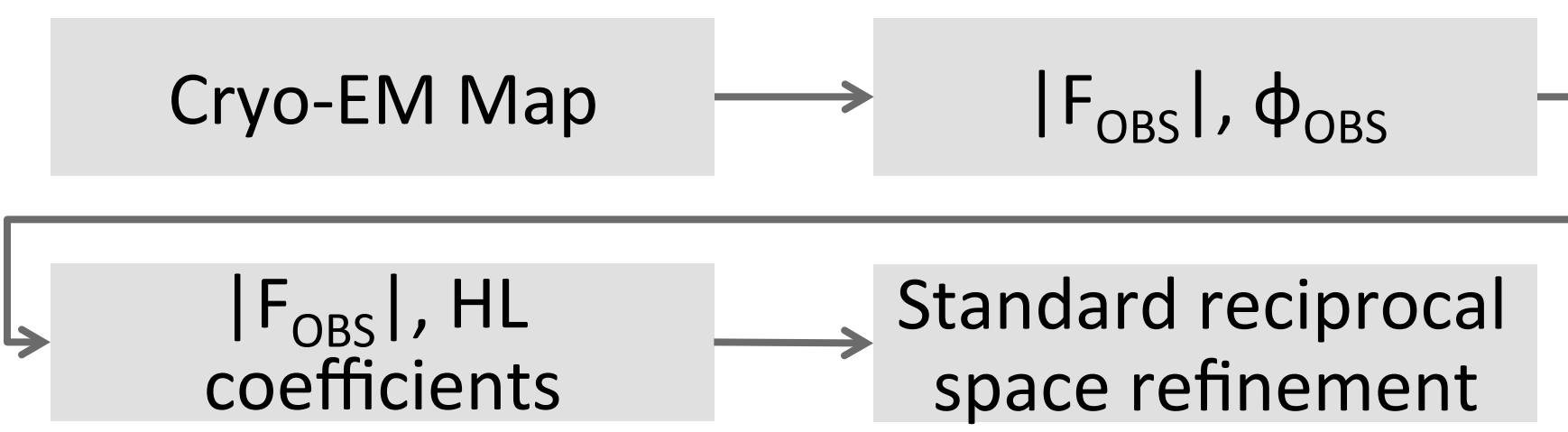
- **Inspired by community members**
  - User-request-driven development
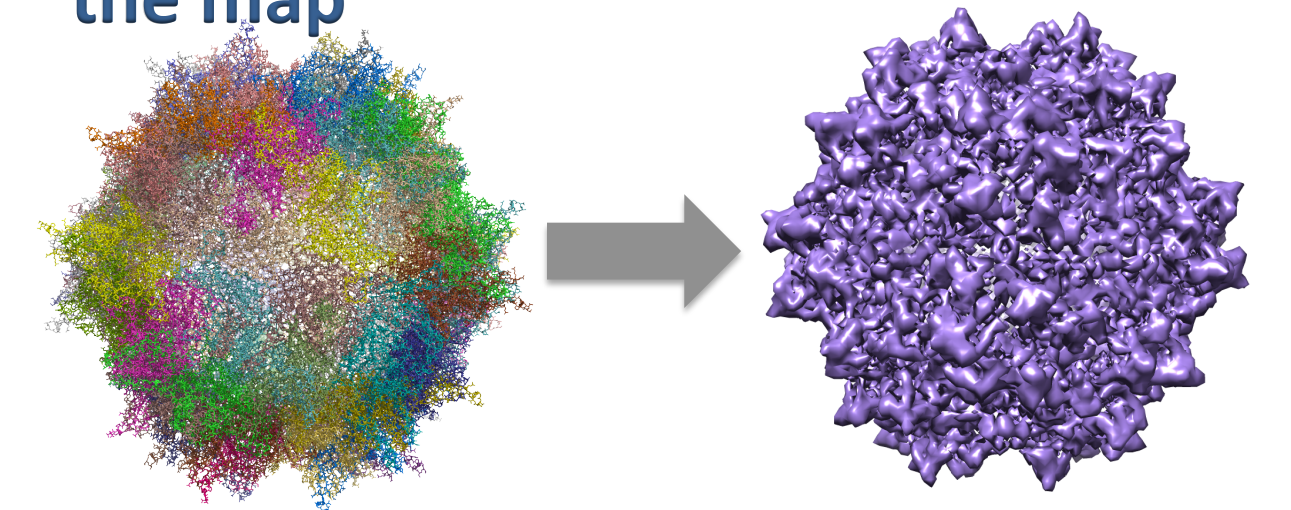 - Provided feedback and real-life data

  Adam Frost (UC San Francisco) | Gino Cingolani (Jefferson University, PA) | Valerie Pye (Cancer Research UK) | Wah Chiu & Corey Hryc (Baylor, TX) | Darcie Miller (St. Jude Research Hospital) | Anastasia Aksyuk (NIH) | Mavis Agbandje-Mckenna (University of Florida) | Stuart Howes (Eva Nogales Lab) … and many others

- **Lack of dedicated real space refinement software**
  - Often cryo-EM maps are being converted into reciprocal space to make use of current crystallographic refinement programs
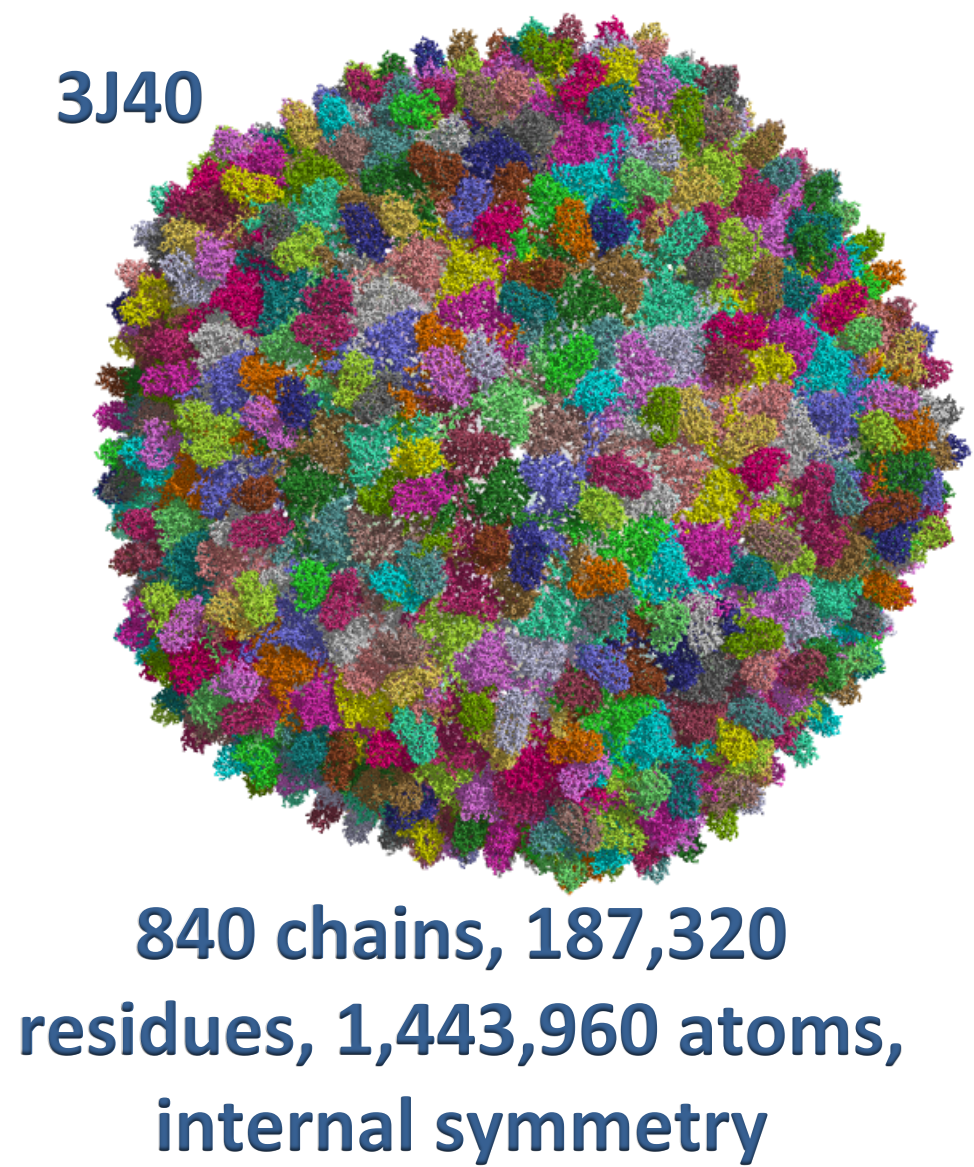
  Cryo-EM Map → $|F_{OBS}|, \phi_{OBS}$

  $|F_{OBS}|$, HL coefficients → Standard reciprocal space refinement

  Might not be the best approach

  - The cryo-EM map is closest to the experimental data: refinement should be done directly against the map

CHALLENGES

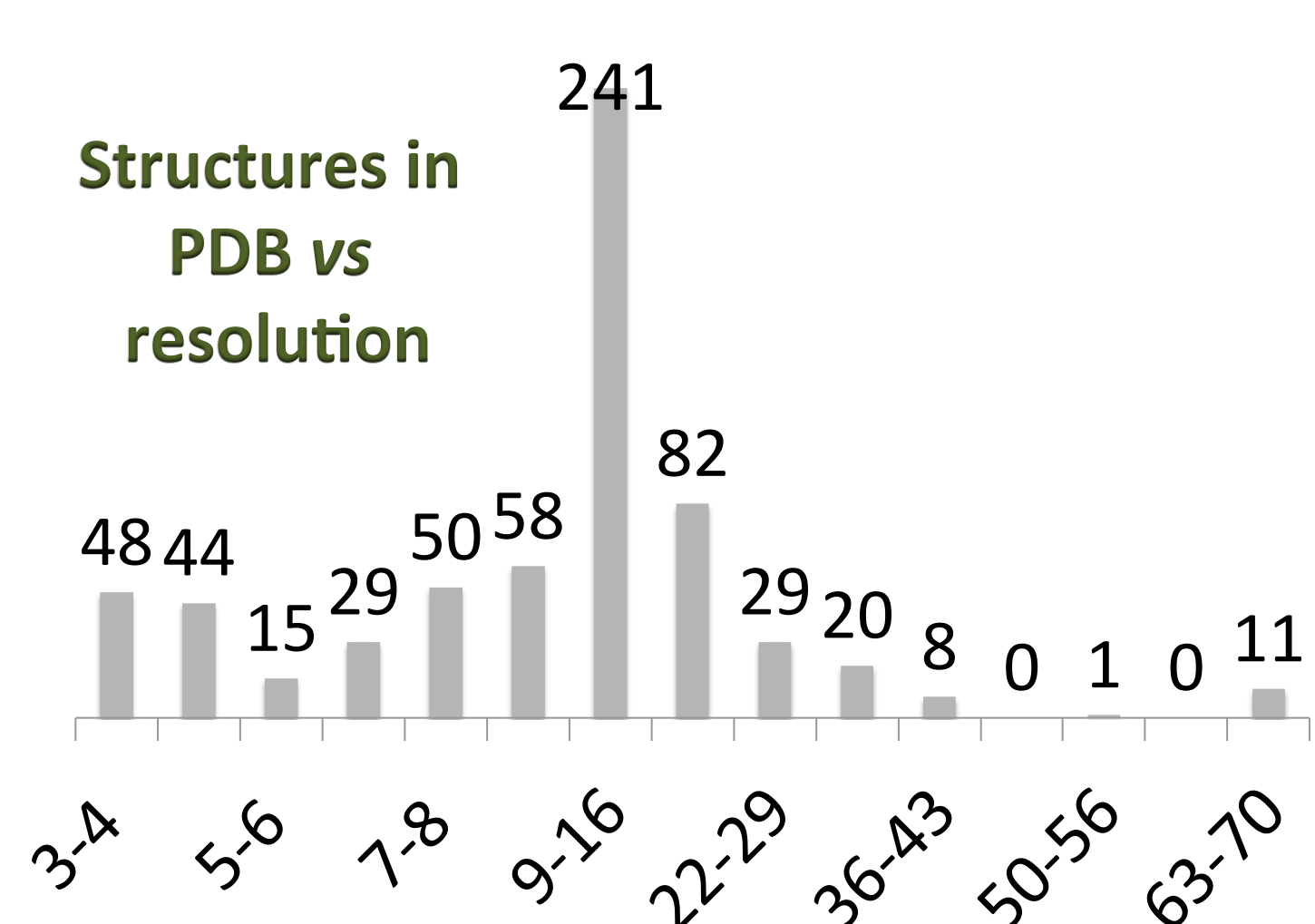- **Size: typically very large**
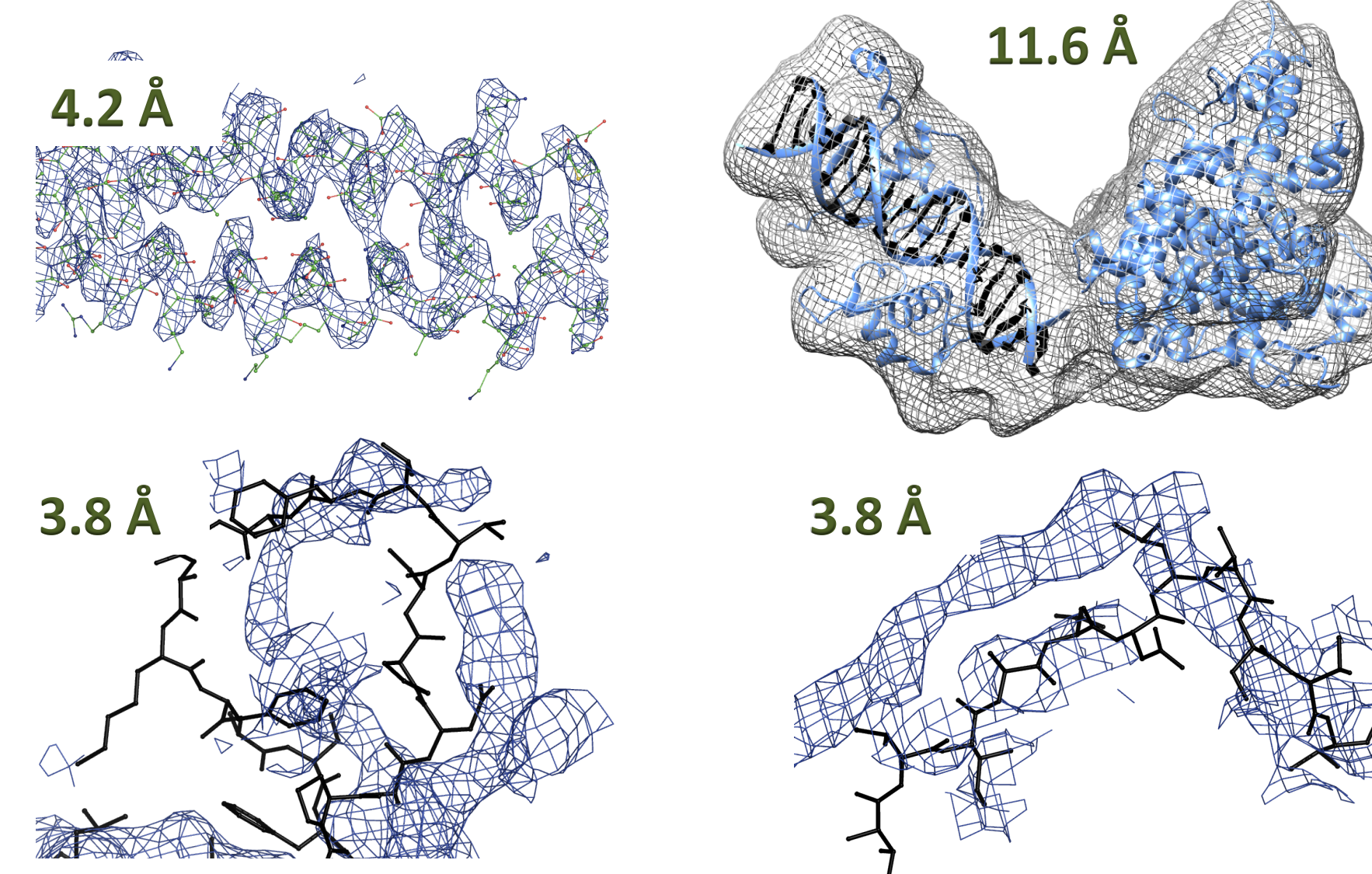
  3J40

  840 chains, 187,320 residues, 1,443,960 atoms, internal symmetry

- **Resolution: typically low or very low**

  Structures in PDB vs resolution

  241, 48, 44, 15, 29, 50, 58, 82, 29, 20, 8, 0, 1, 0, 11

  3-4 | 5-6 | 7-8 | 9-16 | 22-29 | 36-43 | 50-56 | 63-70

- **Maps vary in quality**

  4.2 Å, 11.6 Å, 3.8 Å, 3.8 Å

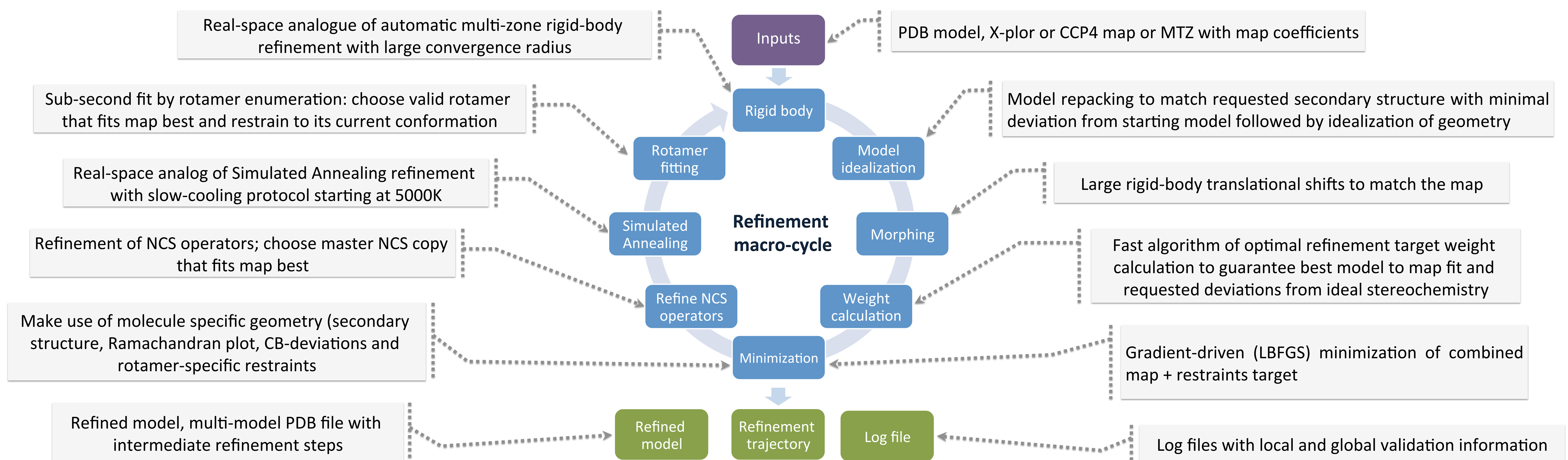- **Model geometry: from decent to very poor**
 - Rotamer outliers, more than:
    - 10% in 32% of models in PDB
    - 20% in 10% of models in PDB
  - Ramachandran plot outliers, more than:
    - 5% in 91% of models in PDB
    - 10% in 34% of models in PDB
  - RMSD covalent bonds, more than:
    - 0.05Å in 15% of models in PDB
      - Gold standard 0.01-0.02Å
  - RMSD covalent angles, more than:
    - 3° in 20% of models in PDB
      - Gold standard 1.5-2.5Å

REAL-SPACE REFINEMENT IN *Phenix*: STATE-OF-THE-ART

- Real-space analogue of automatic multi-zone rigid-body refinement with large convergence radius

- Sub-second fit by rotamer enumeration: choose valid rotamer that fits map best and restrain to its current conformation

- Real-space analog of Simulated Annealing refinement with slow-cooling protocol starting at 5000K

- Refinement of NCS operators; choose master NCS copy that fits map best

- Make use of molecule specific geometry (secondary structure, Ramachandran plot, CB-deviations and rotamer-specific restraints)

- Refined model, multi-model PDB file with intermediate refinement steps

Inputs

Rigid body

Rotamer fitting

Model idealization

Simulated Annealing

Morphing

Refine NCS operators

Weight calculation

Minimization

Refinement macro-cycle

Refined model

Refinement trajectory

Log file

- PDB model, X-plor or CCP4 map or MTZ with map coefficients

- Model repacking to match requested secondary structure with minimal deviation from starting model followed by idealization of geometry

- Large rigid-body translational shifts to match the map

- Fast algorithm of optimal refinement target weight calculation to guarantee best model to map fit and requested deviations from ideal stereochemistry

- Gradient-driven (LBFGS) minimization of combined map + restraints target

- Log files with local and global validation information

EXAMPLES OF USING *Phenix* REAL-SPACE REFINEMENT TO REFINE ATOMIC MODELS INTO CRYO-EM MAPS

- **Re-refinement of 3J5Q, resolution 3.8Å**
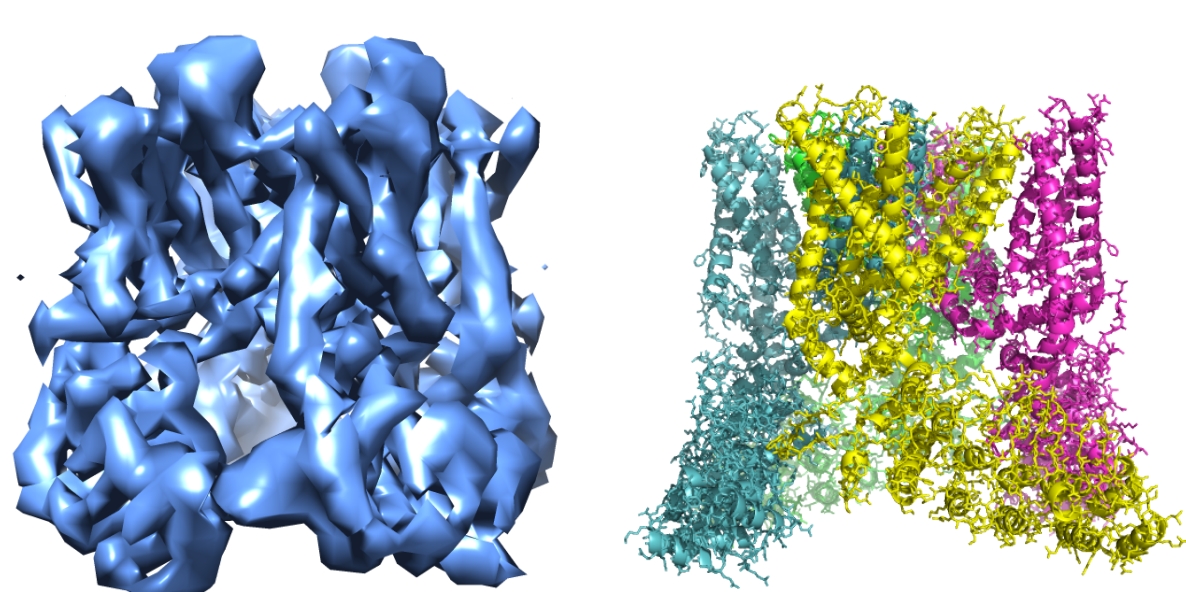
METRIC	Original	*Phenix*
Map CC	0.650	0.714
RMSD (bonds/angles)	0.01/1.34	0.01/1.31
Clashscore	100.9	32.84
Rama. outl., %	0.52	0
Rotamer outl., %	27.99	0
C-beta deviations	0	0

2,324 residues, refinement time: 20 minutes

- **Re-refinement of 3J6P, resolution 8.2Å**
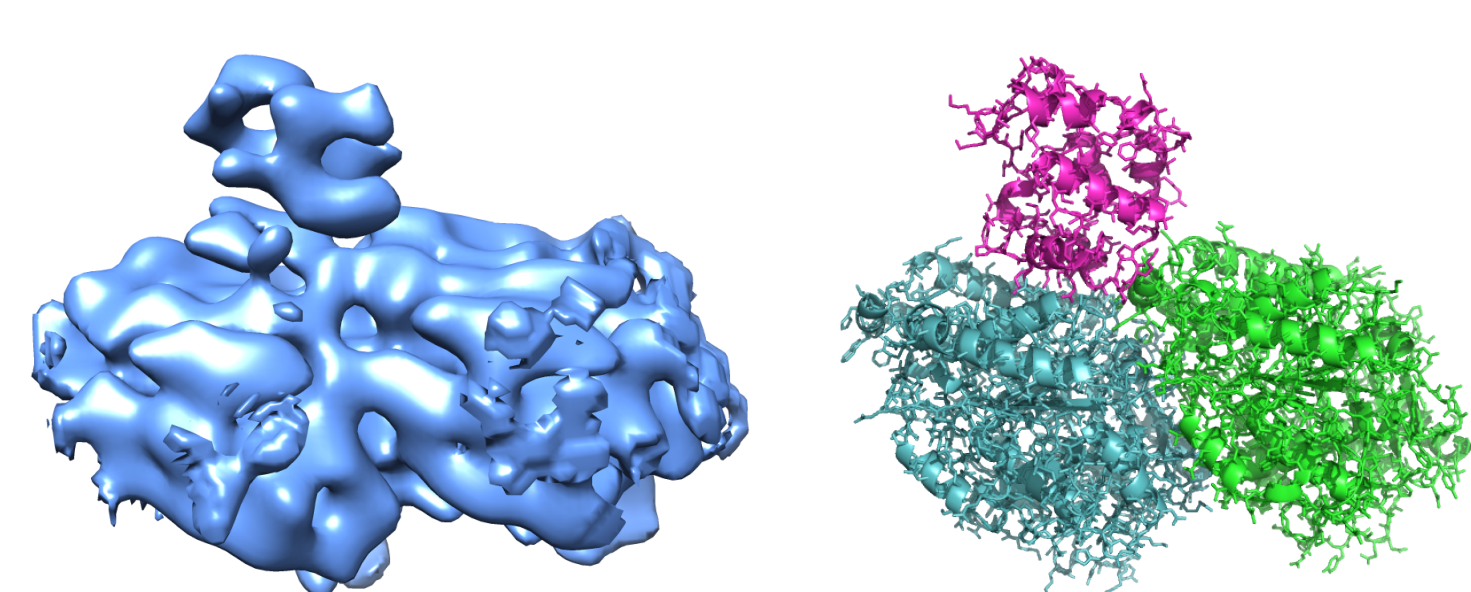
METRIC	Original	*Phenix*
Map CC	0.596	0.743
RMSD (bonds/angles)	0.03/2.34	0.00/1.11
Clashscore	92.37	34.73
Rama. outl., %	2.03	0.54
Rotamer outl., %	26.21	0
C-beta deviations	2	0

949 residues, refinement time: 15 minutes

- **Re-refinement of 4CI0, resolution 3.4Å**
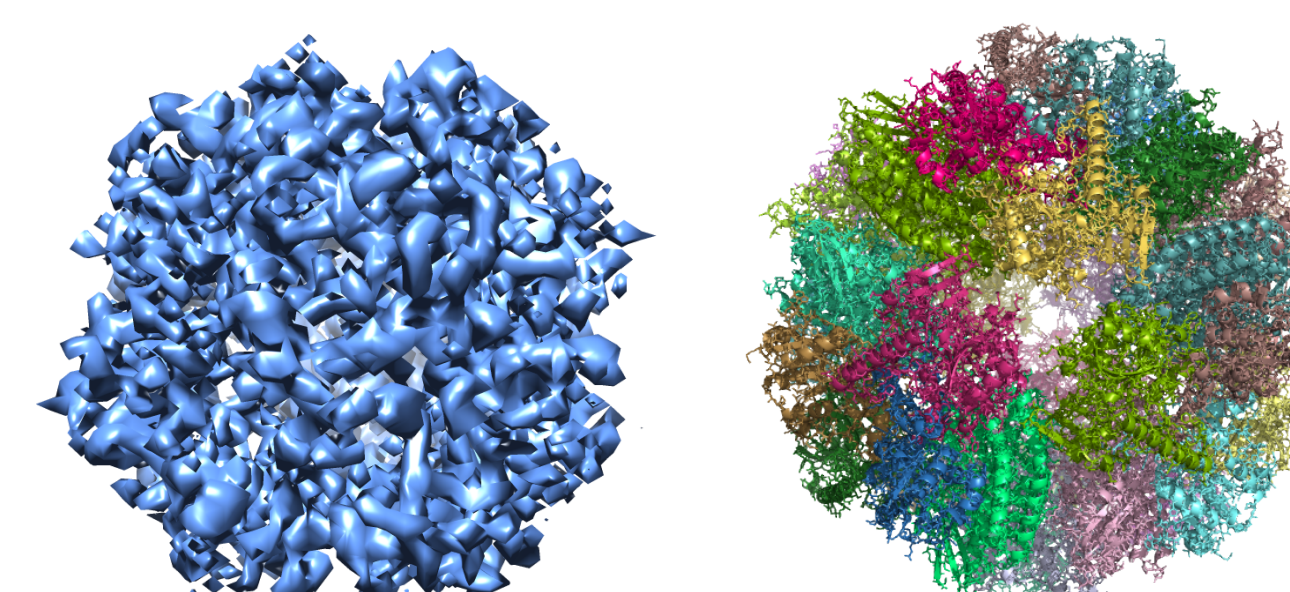
METRIC	Original	*Phenix*
Map CC	0.645	0.783
RMSD (bonds/angles)	0.02/2.05	0.01/1.21
Clashscore	117.1	18.79
Rama. outl., %	0.11	0.11
Rotamer outl., %	35.51	0
C-beta deviations	24	0

10,716 residues, refinement time: 173 minutes

- **Re-refinement of 4CI0, resolution 3.4Å**
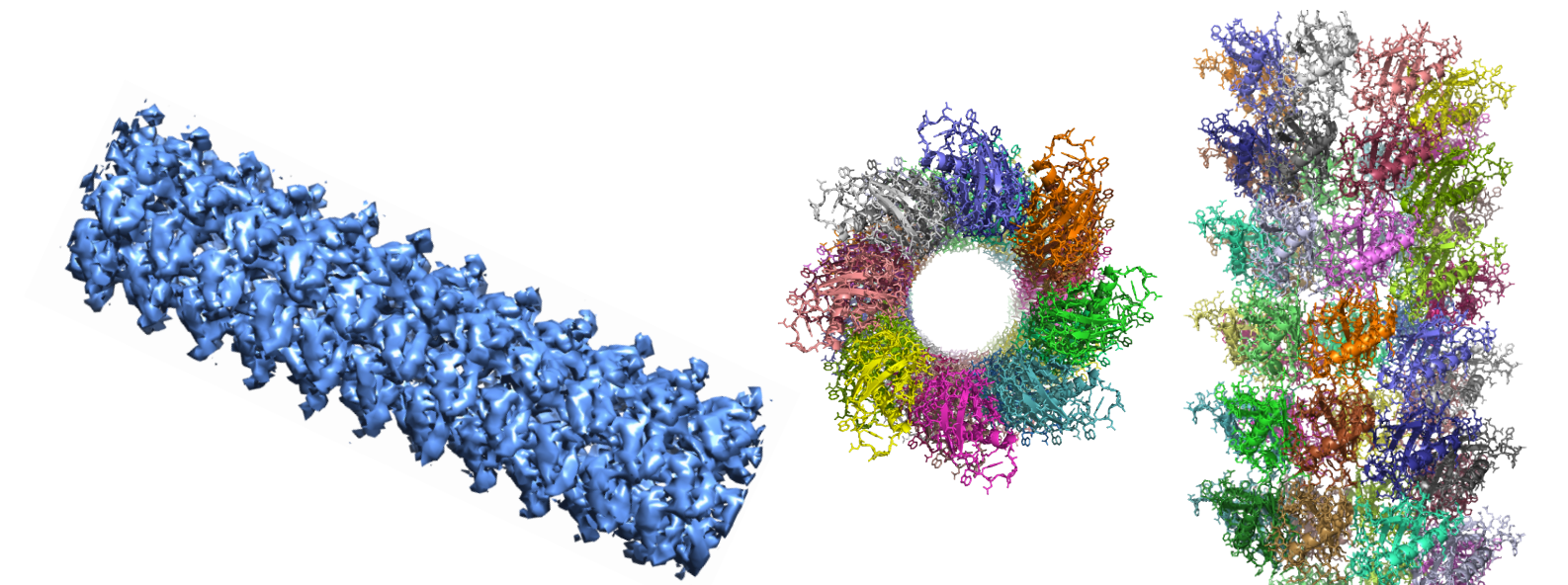
METRIC	Original	*Phenix*
Map CC	0.709	0.647
RMSD (bonds/angles)	0.04/4.05	0.01/1.23
Clashscore	18.34	18.59
Rama. outl., %	3.66	0
Rotamer outl., %	24.64	0
C-beta deviations	637	0

4,116 residues, refinement time: 45 minutes

Phenix

All tools and methods presented here are implemented in *Phenix* and available in most recent nightly build

More information:
`www.phenix-online.org`

Acknowledgements

April, 2015