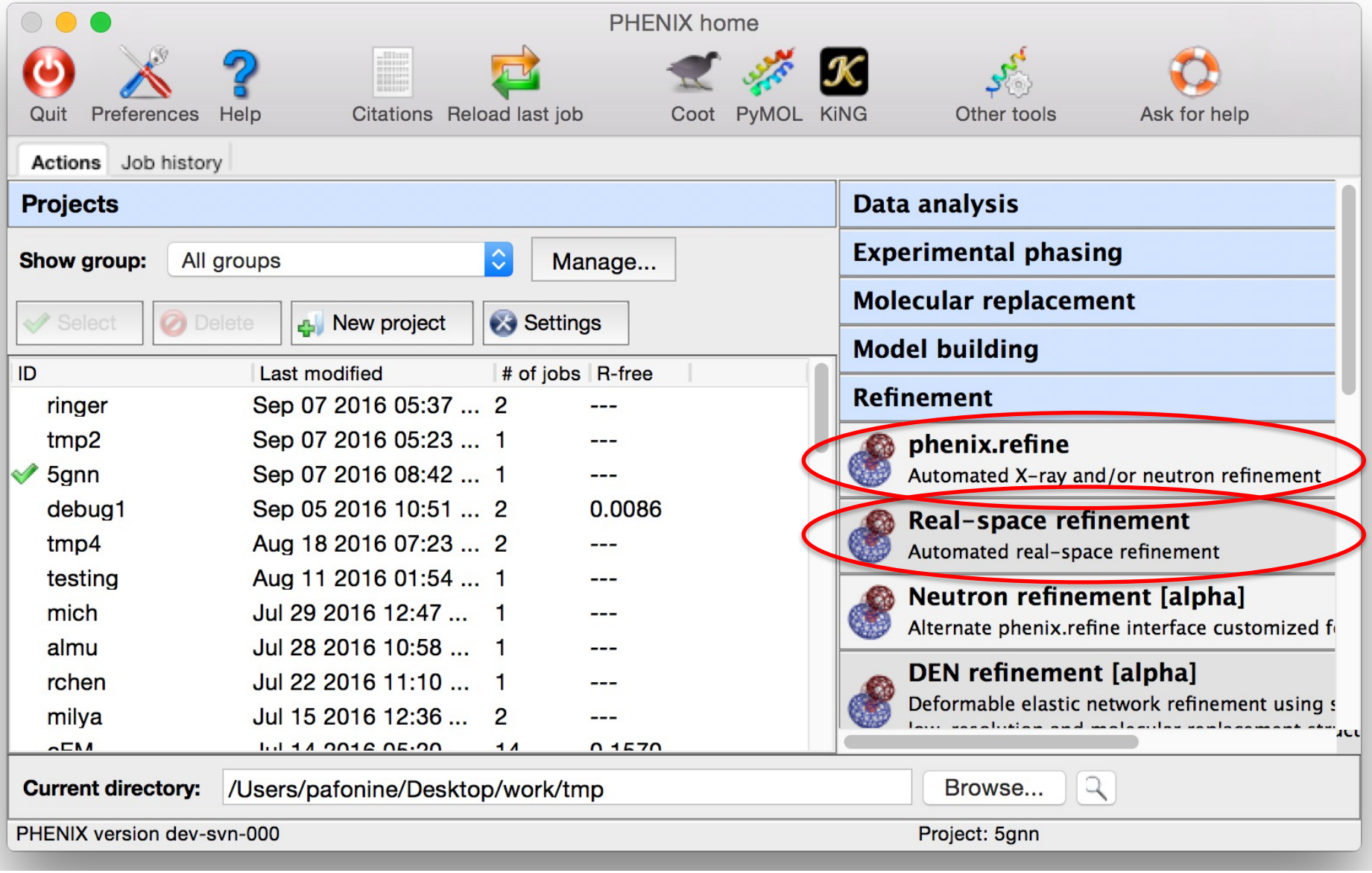
Model Refinement

Pavel Afonine

Lawrence Berkeley National Lab, California, USA

July 29, 2022
ACA
Portland, Oregon

Refinement tools in *Phenix*



Reading

RESEARCH PAPERS

Acta Cryst. (2018). D**74**, 531-544
https://doi.org/10.1107/S2059798318006551
Cited by 672

Part of *CCP-EM Spring Symposium 2017*

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

P. V. Afonine, B. K. Poon, R. J. Read, O. V. Sobolev, T. C. Terwilliger, A. Urzhumtsev and P. D. Adams

RESEARCH PAPERS

OPEN ACCESS

Acta Cryst. (2012). D**68**, 352-367
https://doi.org/10.1107/S0907444912001308
Cited by 2576

Part of *CCP4 Study Weekend 2011*

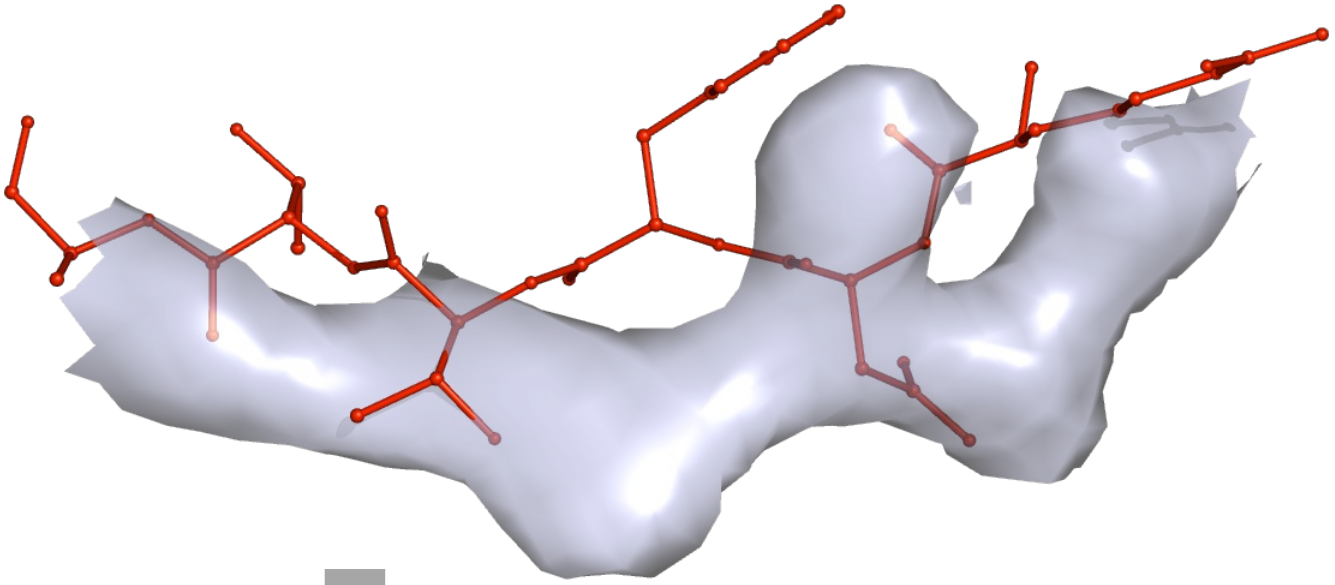
Towards automated crystallographic structure refinement with *phenix.refine*

P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart and P. D. Adams

phenix.refine is a program within the *PHENIX* package that supports crystallographic structure

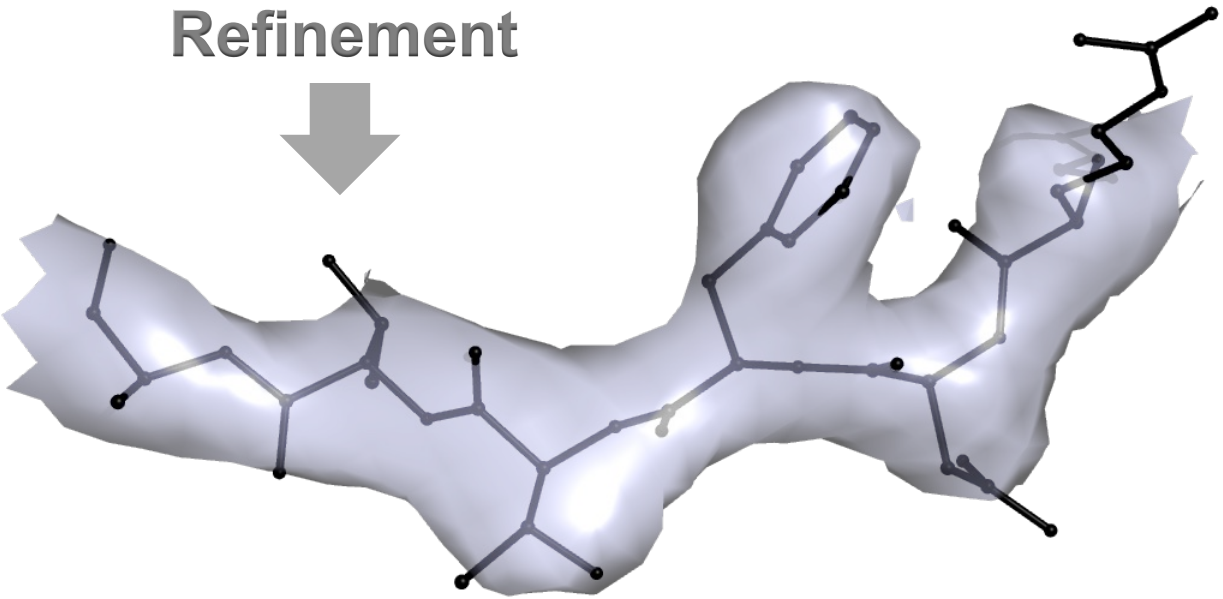
Structure refinement in a nutshell

Initial (poor) model



Refinement

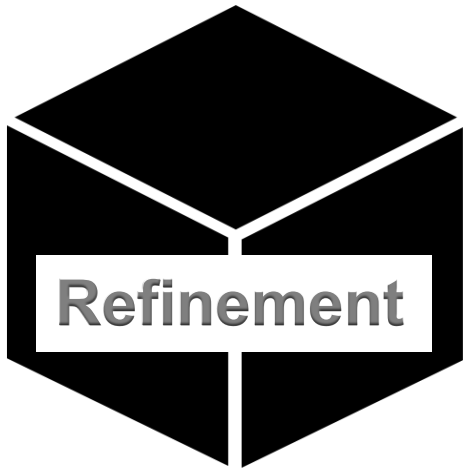
Improved (refined) model



Structure refinement: black box?

Model (PDB, mmCIF)

Data (MTZ)



Refined (better) model

Structure refinement: black box?

- Does it always work?
- Is it always as easy as giving a ‘black box’ input files and hoping it produces a good model?

Structure refinement: black box?

- Does it always work?
- Is it always as easy as giving a ‘black box’ input files an hoping it produces a good model?

  - No. Because:
    - Parametrizing refinement is too complex
    - Default settings suit most common scenario

Structure refinement: lots of jargon

Reference model?

TLS?

Rotamer fixing?

AltLocs?

Group B vs individual?

Local minima?

ADP?

Clashes?

NCS?

IAS?

tNCS?

Grid search?

Weights?

CDL?

SA?

Minimization?

Rama plot restraints?

f' & f''?

Hydrogens?

Restraints?

Bulk-Solvent?

Rama-Z?

Anisotropy?

Rigid body?

Twinning?

SS restraints?

NQH flips?

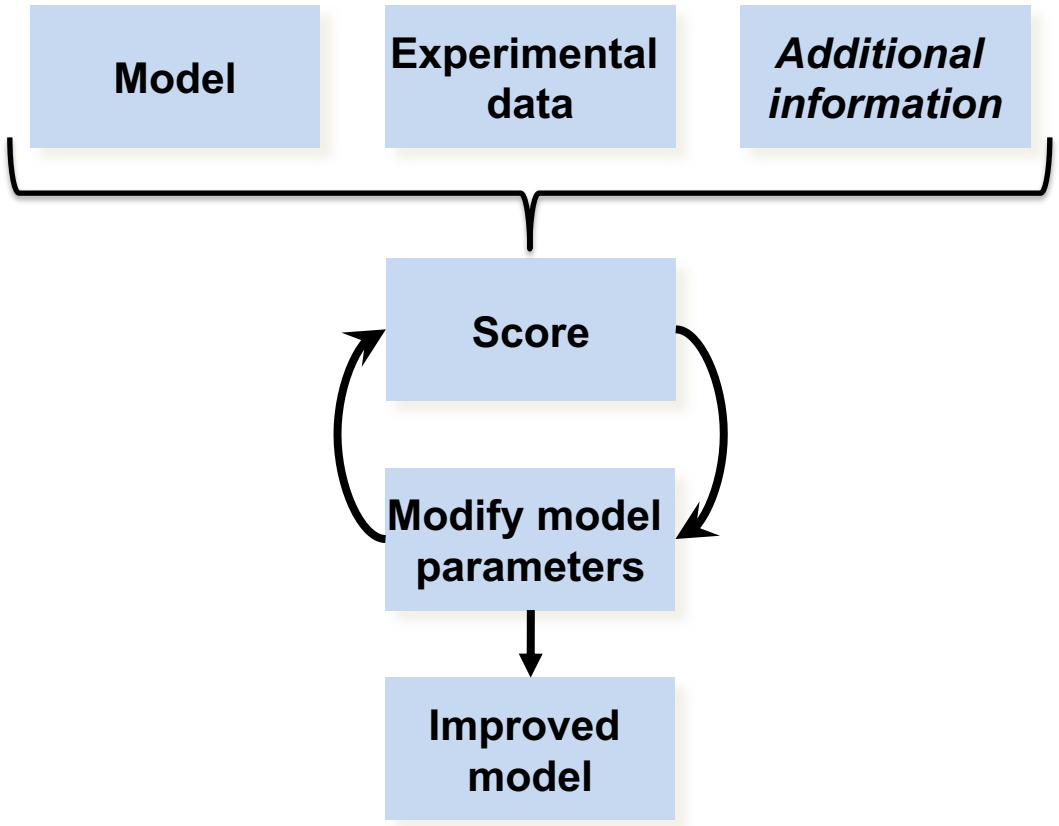
Structure refinement: default doesn't cover all!

- Default refinement is parameterized and optimized to work with typical model and data
  - Low-, high-, ultra-low-, ultra-high- resolution data need special attention
  - Incomplete models need special attention
  - Specialized fine-tuning of complete models may be needed
  - AlphaFold predicted models may need special treatement

Structure refinement: decision-making variables

- **Crystal**
  - Mosaicity
  - Twining
  - Solvent content
  - Symmetry
- **Data**
  - Resolution
  - Errors
  - Completeness
  - Processing
- **Model**
  - Stage
  - Source
  - Parameterization
  - Fit to data

Refinement in a nutshell (II)



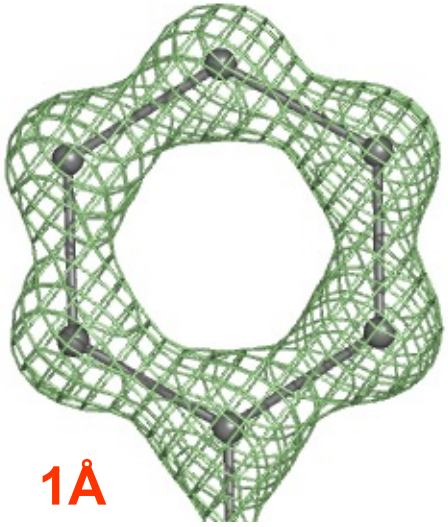
Refinement – optimization process of fitting model parameters to experimental data

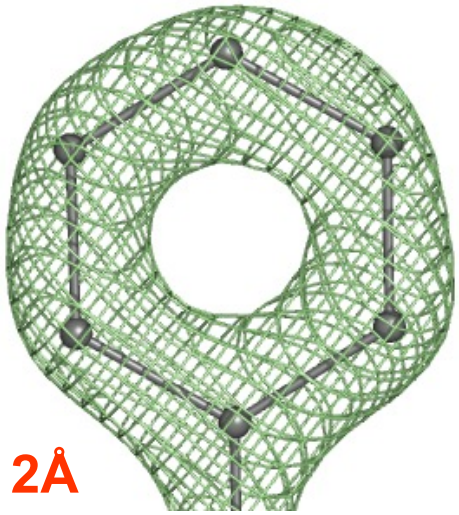
Restraints in refinement

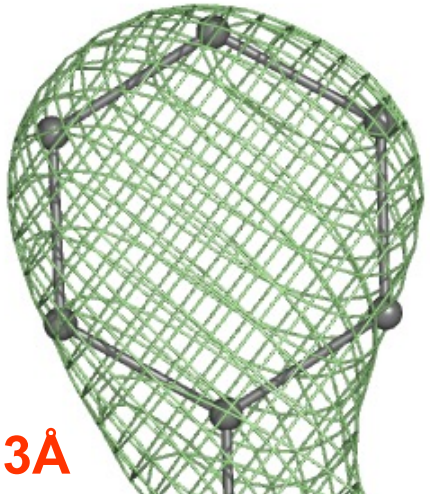
$$T = T_{\mathrm{DATA}}(F_{\mathrm{OBS}}, F_{\mathrm{MODEL}}) + wT_{\mathrm{RESTRAINTS}}$$

$T_{RESTRAINTS} = T_{BOND} + T_{ANGLE} + T_{DIHEDRAL} + T_{PLANARITY} + T_{NONBONDED} + T_{CHIRALITY}$

- Restraints (extra information) are needed because data are not perfect (errors, resolution, etc)







Structure refinement: decision-making variables

- **Crystal**
  - Mosaicity
  - Twining
  - Solvent content
  - Symmetry
- **Data**
  - Resolution
  - Errors
  - Completeness
  - Processing
- **Model**
  - Stage
  - Parameterization
  - Fit to data

- **Weight** (w) in

$$T = T_{\mathrm{DATA}}(F_{\mathrm{OBS}}, F_{\mathrm{MODEL}}) + wT_{\mathrm{RESTRAINTS}}$$

is additional complication

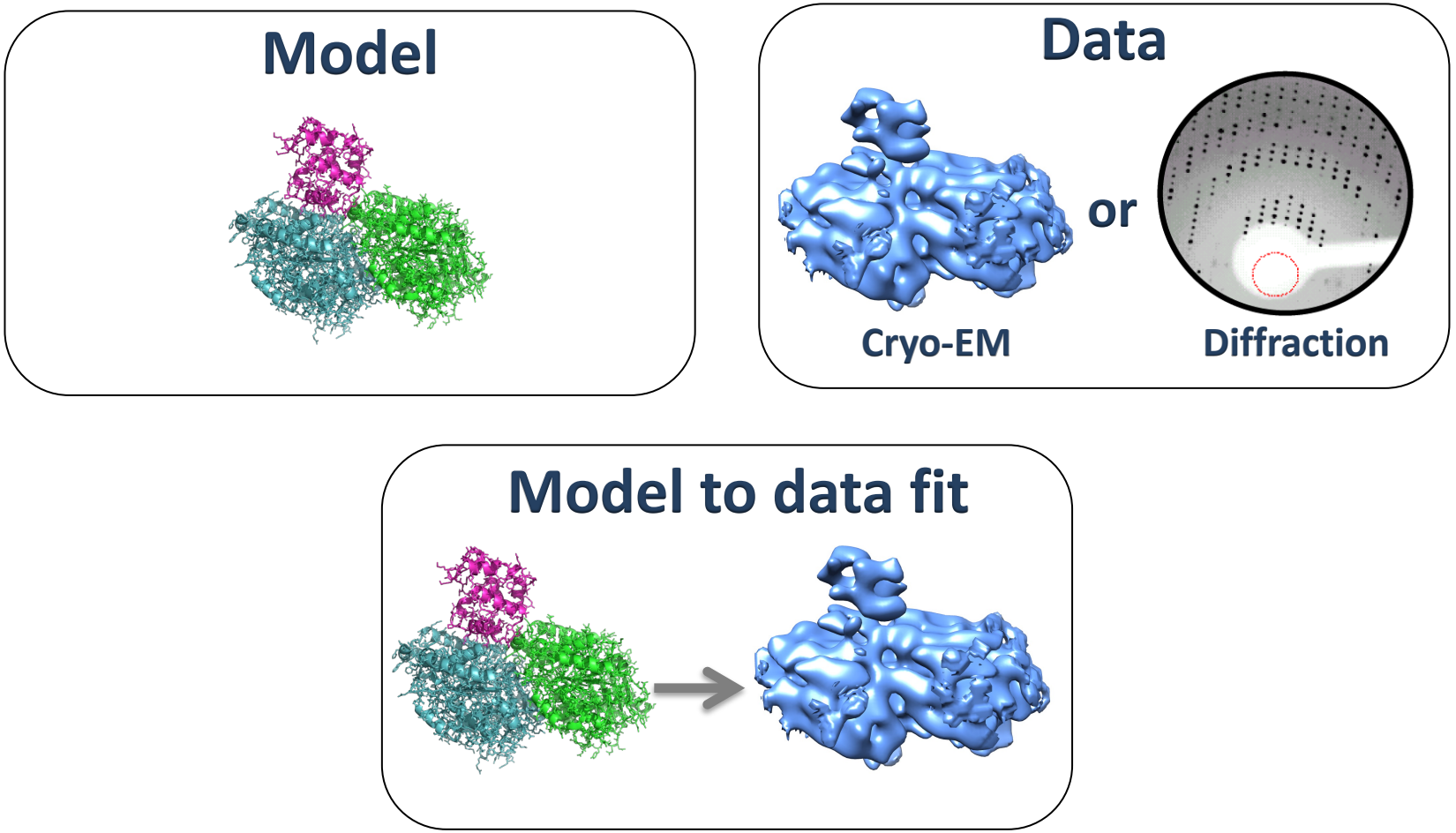
Bottom line

- Refinement is hard (read impractical) to automate fully
- User involvement is needed

How you know…

- … refinement worked ?
- … you did it correctly ?
- … the model you got is good enough to publish ?

Validation



Validation = checking model, data and model-to-data fit are all make sense and obey to prior expectations

Summary

- Refinement is highly automated but user decision-making is almost always needed
- Knowing about model and data guides decision-making
- Use model parameterization that is appropriate for the data quality and stage of refinement
- Use optimization method that is appropriate for the refinement stage
- Do validation throughout the process of structure solution

Feedback, asking for help, etc…

- Feedback, questions

  phenixbb@phenix-online.org
  bugs@phenix-online.org
  help@phenix-online.org

- Reporting a problem, bug or asking for help:

  - We can't help you if you don't help us to understand your problem

  - Send us:

    1) PHENIX version;
    2) Command and parameters used;
    3) Input and output files;
    4) Clearly explain the problem.

Subscribe to PHENIX mailing list: www.phenix-online.org

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