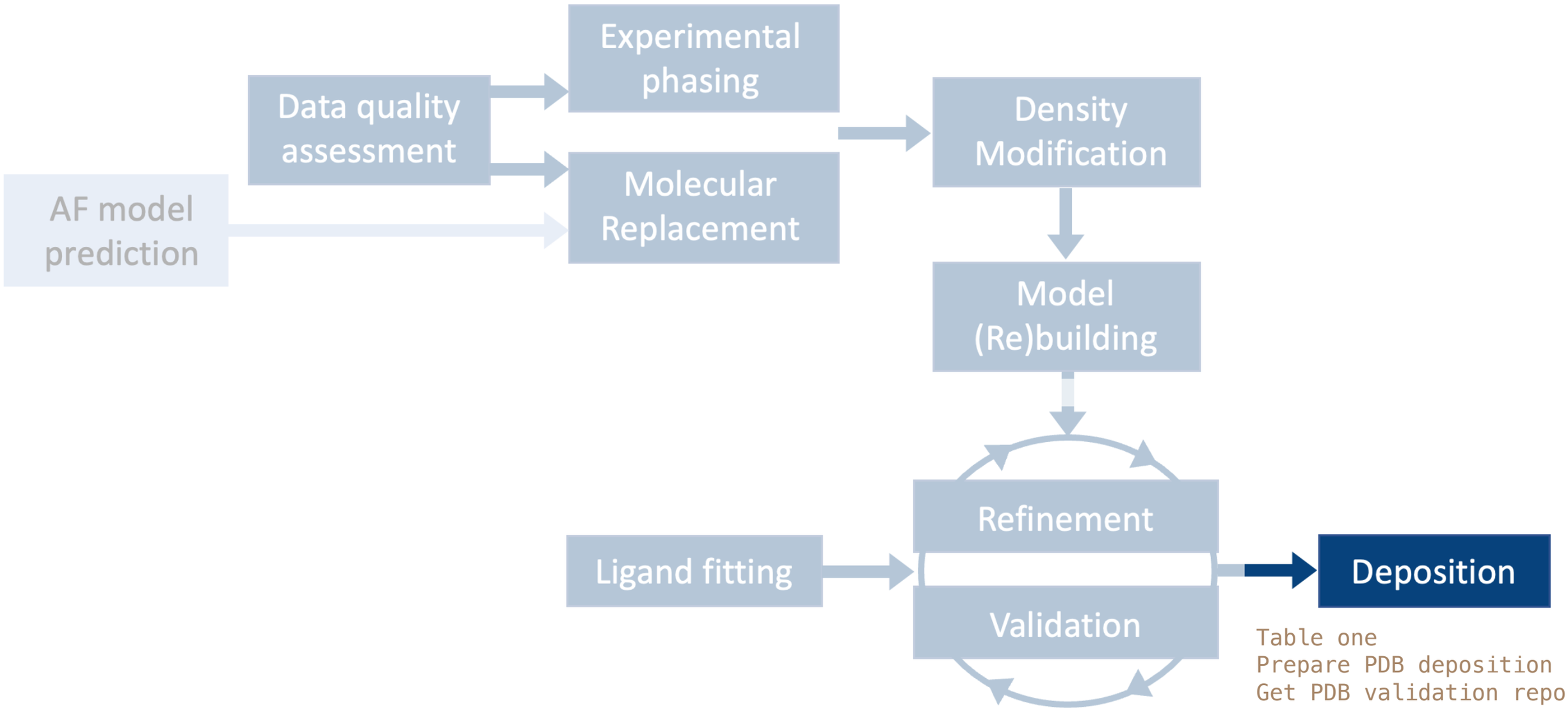
CBMS workbench, June 6 2024

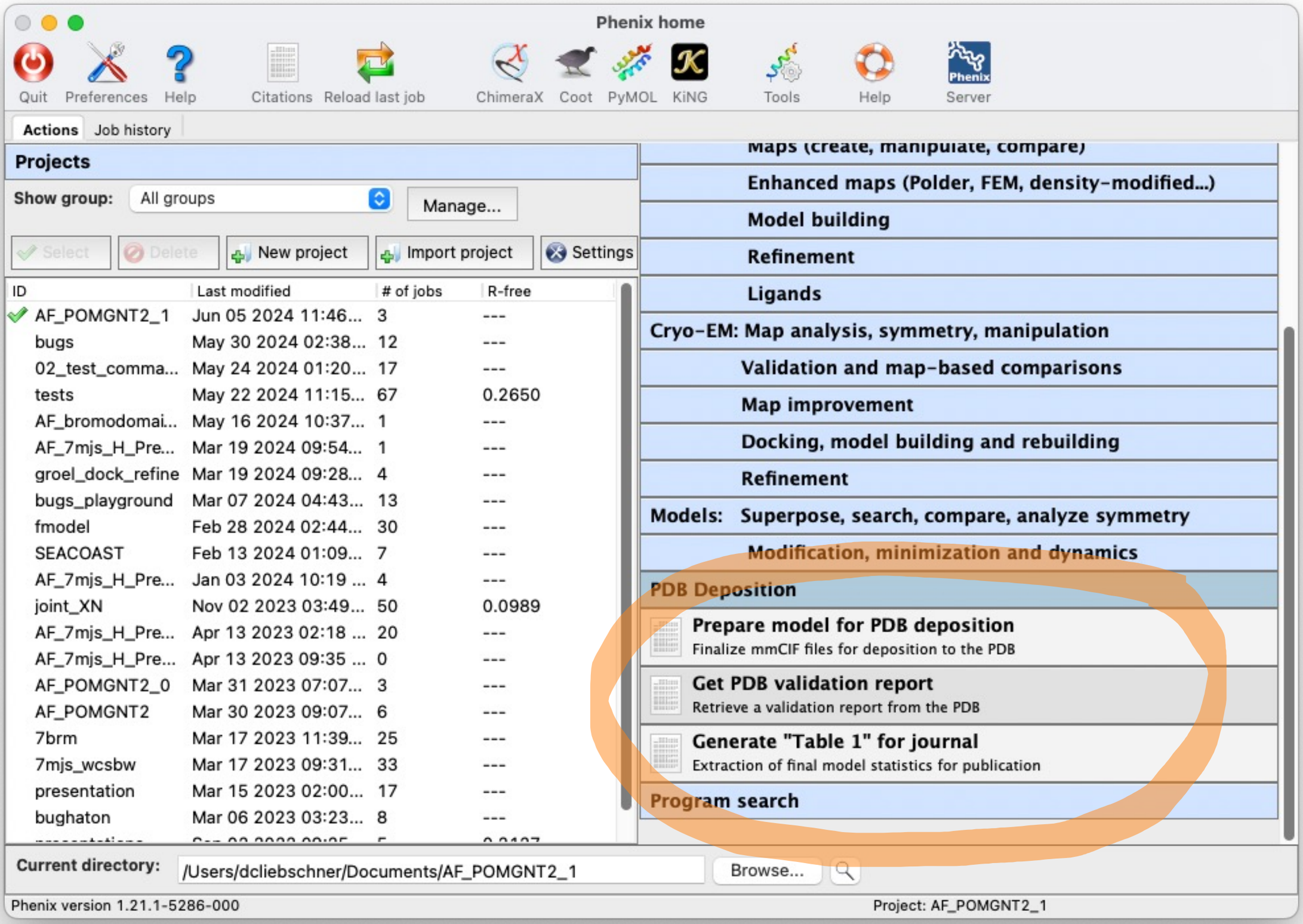
Phenix Tools for PDB deposition

Oleg Sobolev
Lawrence Berkeley Laboratory

Steps in crystallography



PDB deposition



mmCIF

mmCIF format is mandatory for deposition as of 2019

Acta Cryst D STRUCTURAL BIOLOGY

ISSN 2059-7983

Announcing mandatory submission of PDBx/mmCIF format files for crystallographic depositions to the Protein Data Bank (PDB)

Paul D. Adams,[a,b] Pavel V. Afonine,[a] Kumaran Baskaran,[c] Helen M. Berman,[d] John Berrisford,[e] Gerard Bricogne,[f] David G. Brown,[g] Stephen K. Burley,[d,h,i]* Minyu Chen,[j] Zukang Feng,[d] Claus Flensburg,[f] Aleksandras Gutmanas,[e] Jeffrey C. Hoch,[k]* Yasuyo Ikegawa,[j] Yumiko Kengaku,[j] Eugene Krissinel,[l] Genji Kurisu,[j]* Yuhe Liang,[d] Dorothee Liebschner,[a] Lora Mak,[e] John L. Markley,[c]* Nigel W. Moriarty,[a] Garib N. Murshudov,[m] Martin Noble,[n] Ezra Peisach,[d] Irina Persikova,[d] Billy K. Poon,[a] Oleg V. Sobolev,[a] Eldon L. Ulrich,[c] Sameer Velankar,[e]* Clemens Vonrhein,[f] John Westbrook,[d] Marcin Wojdyr,[f,l] Masashi Yokochi[j] and Jasmine Y. Young[d]

Received 21 February 2019
Accepted 3 April 2019

Edited by R. J. Read, University of Cambridge, England

mmCIF facts

- Contains a lot more information than PDB

- Not intended to be human editable
  - You can read it but it is (much) harder than PDB

- Phenix tools generally produce output in mmCIF format

- Avoid editing by hand
  - Easy to make hard-to-recover mistakes

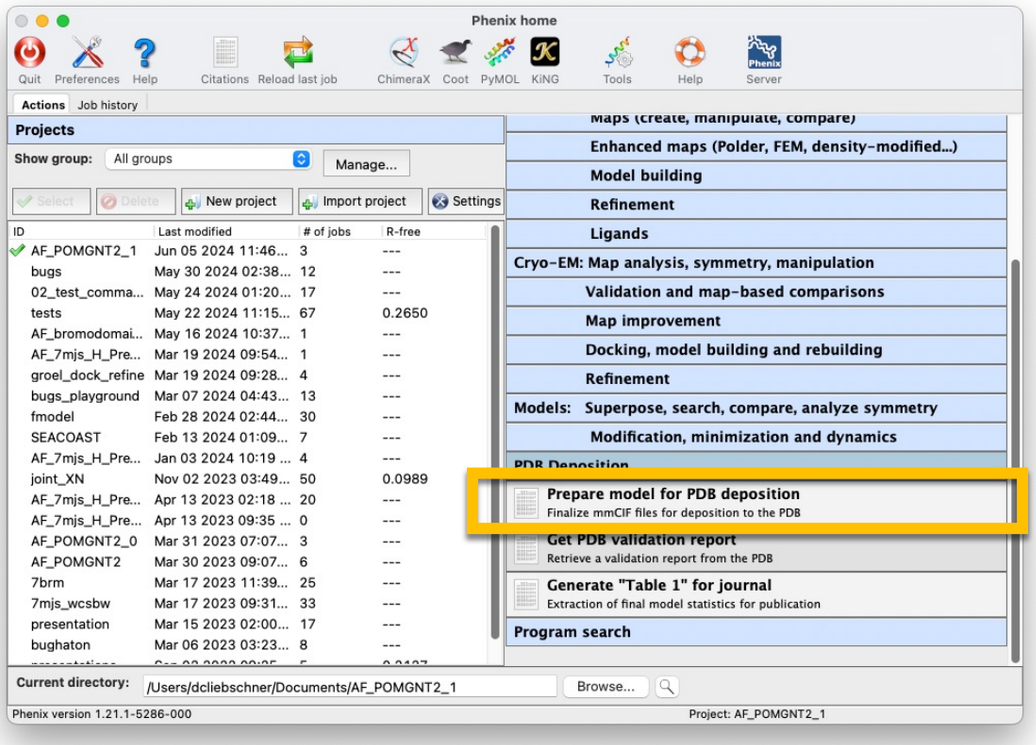
CIF file confusion

- CIF is a file format

- CIF file can contain:
  - Ligand information
  - Atomic model
  - Reflection data
  - Any mixture of three above

PDB deposition

You can get the model file in mmCIF format from phenix.refine.

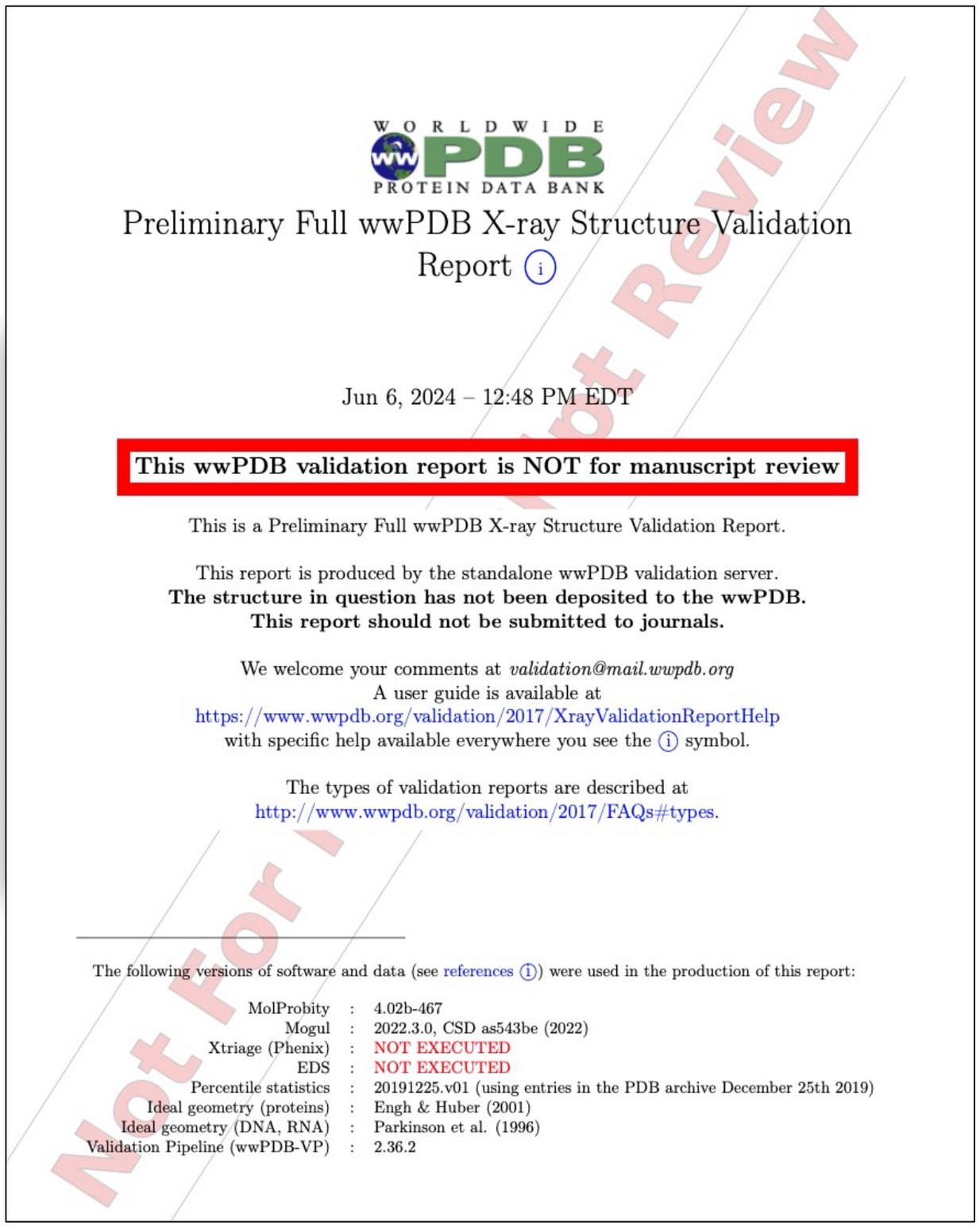


Prepare PDB deposition

Add sequence information to the mmCIF output from phenix.refine to make the file suitable for deposition into the PDB.

Minimum inputs: the model from phenix.refine and a sequence file.

Get PDB validation report



WORLDWIDE PDB PROTEIN DATA BANK

Preliminary Full wwPDB X-ray Structure Validation Report (i)

Jun 6, 2024 – 12:48 PM EDT

This wwPDB validation report is NOT for manuscript review

This is a Preliminary Full wwPDB X-ray Structure Validation Report.

This report is produced by the standalone wwPDB validation server.
The structure in question has not been deposited to the wwPDB.
This report should not be submitted to journals.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

You get the model file in mmCIF format from refinement by default.

Get PDB validation report

Get a validation report from PDB OneDep validation.
If validation completed successfully, you'll get a validation report in pdf and xml format.

Table 1

Table 1. Data collection and refinement statistics.

	1aba
Wavelength	
Resolution range	37.69 - 1.447 (1.46 - 1.45)
Space group	P 21 21 21
Unit cell	30.2 47.8 61.3 90 90 90
Total reflections	
Unique reflections	15730 (452)
Multiplicity	
Completeness (%)	95.66 (84.80)
Mean I/sigma(I)	
Wilson B-factor	14.51
R-merge	
R-meas	
R-pim	
CC1/2	
Reflections used in refinement	15730 (452)
Reflections used for R-free	0 (0)
R-work	0.2075 (0.2842)
R-free	0.2075 (0.2842)
Number of non-hydrogen atoms	880
macromolecules	728
ligands	12
solvent	140
Protein residues	87
RMS(bonds)	0.031
RMS(angles)	2.49

You can get the model file in mmCIF format from phenix.refine.

Generate ”Table 1” for journal

Generate the standard table of crystallographic statistics required by most scientific journals.
If you provide unmerged data, the table will include merging statistics.

PDB deposition dos and don’ts

- Do not change the content of files from refinement for any reason:
  - Add/remove atoms (hydrogens, water)
  - Edit labels, header information

- Run Comprehensive validation (Phenix GUI) to address all outstanding issues before deposition

- Don’t panic if validation statistics reported by Phenix does not match PDB validation report
  - If that happens and presents a problem – start conversation with PDB stuff and involve Phenix developers

- Once all is deposited and up on the web – check everything: mistakes at PDB end happen

Please take the survey

The Phenix Project

Lawrence Berkeley Laboratory

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

University of Cambridge

Randy Read, Airlie McCoy, Alisia Fadini

Los Alamos National Laboratory
New Mexico Consortium

Tom Terwilliger, Li-Wei Hung

UTHealth

Matt Baker

Duke University

Jane Richardson, Christopher Williams, Vincent Chen, Michael Prisant

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