

Pavel Afonine

Phenix team

Lawrence Berkeley National Lab, California, USA

May 10th, 2023 MCCS, Madrid

Validation





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

Validation tools: Crystallography vs Cryo-EM





Validation: why to do?

- Can help to:
 - save (a lot of) time
 - produce better models
 - Set correct expectations
- Subjectivity:
 - lot's of manual steps that depend on skills, pressure and ethics
- Software isn't perfect
- Databases are not perfect

Lack of validation will be discovered (sooner or later)!

Validation: why to do?

(2019) Nature 570: 400-404 | PDB: 609j 3.9Å



Validation: why to do?



Validation tools in Phenix

PHENIX home					
🕑 🔏 ? 🛛	🛫 🚀 🥑	C 🤞 😳			
Quit Preferences Help Citat	ions Coot PyMOL Ki	G Other tools Ask for help			
Actions Job history					
Projects	Projects Favorites				
Show group: All groups	S Manage	Data analysis			
Select 🕖 Delete 🛃 New p	project 😵 Settings	Xtriage Analysis of data quality and crystal defects			
ID Last modified ✓ ChrisF Apr 13 2020	# of jobs R-free 09:42 28 0.194	4 Merging statistics Calculates a variety of statistics for unmerged intensities, including I/sigma, R-merge, R-meas, and			
real-space-refin Apr 03 2020 zzz1 Mar 21 2020	07:42 2 09:10 1	CC1/2.			
chris Mar 12 2020 dan Mar 11 2020	12:27 11 0.189 05:44 1	0 Analyze quality of maps in CCP4 format			
3j63 Mar 11 2020	02:28 1	Experimental phasing			
jason Mar 11 2020 rt6 Mar 11 2020	11:36 1 10:31 1 0.245	Molecular replacement			
mate Mar 10 2020	01:36 1	Pofinement			
emily Mar 09 2020	03:52 3	Kennement			
Mar 05 2020	08:25 3 0.192	3 Cryo-EM			
alex Feb 27 2020	11:33 6	Validation			
rt20201 Feb 18 2020	12:50 4 0.221 09:00 1 0.197	Comprehensive validation (X-rav/Neutron)			
real-space-refin Jan 30 2020	02:38 2	Model quality assessment, including real-space correlation and geometry inspection using MolProbity tools			
real-space-refin Jan 29 2020 ion_channel_den Jan 27 2020	10:56 1 07:36 3	Comprehensive validation (cryo-EM)			
10101 Jan 27 2020	12:38 2	Model quality assessment, including real-space correlation, for cryo-EM structures			
demos Jan 27 2020	10:57 3	Structure comparison			
ion_cnannei_den Jan 27 2020	10:03 2	by Structure comparison light differences between multiple structures of the same protein, using multiple criteria			
maicoim Jan 22 2020	04:28 3				
3NIR Dec 05 2019	10:2 1	Calculate CC*			
leighton Sep 02 2019	05:1 2	Comparison of unmerged data quality with refined model, as described in Karplus & Diederichs			
5pti Aug 27 2019	03:4 3	(2012)			
·		EMRinger Model validation for de novo electron microscopy structures			
	Ligands				
Current directory: /Users/pafoning	e/Desktop/all/people/ChrisF	Browse			
PHENIX version dev-svn-000 Project: ChrisF					

Model validation

• Table 1 items (not a complete list!)

- Content (macromolecule, ligands, NCS, ...)
- Bond/angle RMSDs / RMSZ
- Molprobity:
 - Clashscore
 - Ramachandran plot (favorite, outliers)
 - Rotamer outliers
 - C-beta deviations
- Incomplete residues
- Solvent content
- ADP (mean, Bonded <B_i-B_j>)

Model validation: amino-acid side-chain rotamers

- An outlier \neq wrong
 - However, each outlier has to be explained



Model validation: amino-acid side-chain rotamers

Thr O 3 from 1YHQ



Model validation: amino-acid side-chain rotamers

• Low-resolution maps

Side chain lacking density may be forced into main chain density and become a rotamer outlier

Phenix refinement (real- and reciprocal-space) use rotamer-specific restraints on torsion chi-angles

Model validation: clashes

- Half of the atoms in a protein molecule
- Make most interatomic contacts
 - Using H in refinement helps prevent or eliminate clashes



Model validation: clashes

- N/Q/H flips (asparagine/glutamine/histidine)
 - Based on clash analysis
 - Requires H present



Model validation: clashes

- N/Q/H flips
 - Based on clash analysis
 - Requires H present



• Different plots for different classes of residues





• A Ramachandran plot outlier \neq wrong





• All outliers need to be explained (supported by the data)

Model validation and refinement

- Validation metrics progressively become refinement goals
 - Ramachandran plot restraints
 - Cβ deviation restraints
 - Secondary structure restraints
 - Restraints on χ angles of amino-acid side-chain rotamers
- As result, validation becomes less capable of catching problems

Model validation

PNAS, 2019 116 (39) 19513-19522



Metric	6KS6		
Clashsc	7.7		
Rama. (%)		favored	96.4
		outliers	0.2
Rotame	0		
C_{β} devia	0		
	Bon	d (Å)	0.001
RMSD	Angle (°)		0.396
Resolution (Å)			3.0

Perfect statistics! All looks just great!



Odd Ramachandran plot. How we know this?



Model validation: Ramachandran plot Z-score



Vol. 13 no. 4 1997 Pages 425-430

Objectively judging the quality of a protein structure from a Ramachandran plot

Rob W.W.Hooft, Chris Sander and Gerrit Vriend

- Ramachandran Z-score is good at spotting odd plots
- Used in PDBREDO. Implemented in *Phenix* in all validation
- One number, simple criteria:
 - Poor: |Z| > 3 Suspicious: 2 < |Z| < 3 Good: |Z| < 2



Model validation: Ramachandran plot Z-score



Model validation: Ramachandran plot Z score

PNAS, 2019 116 (39) 19513-19522





Metric	6KS6			
Clashso	7.7			
Rama. (%)		favored	96.4	
		outliers	0.2	
Rotame	0			
C_{β} deviations			0	
	Bon	d (Å)	0.001	
RMSD	Ang	le (°)	0.396	
Resolut	3.0			
RamaZ = -3.3				

Poor: |Z| > 3 Suspicious: 2 < |Z| < 3 Good: |Z| < 2

Always think Local vs Global

- 2.5Å: R_{WORK}/R_{FREE} = 17.1/21.2% bonds = 0.01Å angles = 1.6°
 - R-factors are great, overall geometry is great, but...

Local vs Global

- 2.5Å: R_{WORK}/R_{FREE} = 17.1/21.2% bonds = 0.01Å angles = 1.6°
 - R-factors are great, overall geometry is great, but...

Histogram of	deviations from ideal values				les	
Bonds			Angles			
0.000 - 0.035:	2645		0.000	—	9.313:	4208
0.035 - 0.070:	19		9.313	—	18.626:	9
0.070 - 0.106:	13		18.626	—	27.939:	3
0.106 - 0.141:	5		27.939	—	37.252:	4
0.141 - 0.176:	3		37.252	—	46.565:	0
0.176 - 0.211:	0		46.565	—	55.878:	0
0.211 - 0.246:	0		55.878	—	65.191:	2
0.246 - 0.281:	0		65.191	—	74.504:	1
0.281 - 0.317:	2		74.504	-	83.817:	0
0.317 - 0.352:	18		83.817	-	93.130:	8

- Problem with a few atoms, while the rest is ok
 - Poor ligand geometry

Data completeness

• PDB code: 1NH2, resolution 1.9Å, showing E6-E8

2mFo-DFc , 1σ



Local vs Global

Completeness by resolution:

Overall completeness in d_{min} -inf: 0.95



Data incompleteness distorts maps

Validation – Sequence register errors

MASTER GFVDLTLHDQVSMEHPVKLLFGKCVEGMVEIVYTFLSSTLKSLE Chain A GFVDLTRHDQVSMEHPGKLLFGK--EGMVEIVYTF-----KSLE Chain B GFVDLTRHDQVSMEHPGKLLFGK--EGMVEIVYTFVSSTLKSLE Chain C GFVDLTRHDQVSMEHPGKLLFGKKVEGMVEIVYTFVSSTLKSLE Chain D GFVDLTRHDQVSMEHPGKLLFGKKVEGMVEIVYTFLSSTLKSLE *****

Map and model errors



- Errors in model parameters
- Fourier artifacts

Things to try:

- Refine occupancy, anisotropic ADP, anomalous f' & f'', define charge.
- Do many refinement cycles (so refinement converges)
- Consider alternative ion

Ligands and Polder map

PDB code: 1ABA, Resolution: 1.45 Å



Ligands and Polder map

PDB code: 1ABA, Resolution: 1.45 Å



Ligands and Polder map

PHENIX home					
(b) 🐰	2	à 🛛 🛫	` 🌌 🛣	<u>s</u>	
Quit Preferences	Help Citations Reload	last job Coot	PyMOL KING	Other tools Ask for help	
Actions Job histor	v				
Projects				Create Fo-Fo map from isomorphous datasets	
Show group: All groups Show group:		*	Superpose maps Superpose two PDB files and transform the associated map coefficients to the new		
Select 🖉 De	elete 🚽 📣 New project 🐼	Settings		orientation	
ID	Last modified #	of jobs R-free		Cut out density	
؇ 1aba-polder	Jun 27 2016 02:43 2		- Aut		
qun	Jun 22 2016 06:22 1		1 Se	Multi-crystal averaging	
secretase	Jun 18 2016 06:37 9		299	Density modification with multi-crystal averaging of maps	
haim	Jun 02 2016 02:12 2			Find difference map peaks and holes	
scottc	Jun 01 2016 02:18 1			Identify local maxima and minima in mFo-DFc map (and anomalous map if	
wtempel	Jun 01 2016 02:16 5			available) and flag waters with excess density	
polder_rev	May 31 2016 01:56 1			Create maximum entropy map	
rnase-s	May 28 2016 10:07 2	0.2573	8	Statistical map modification procedure to remove artifacts due to missing data	
kay	May 28 2016 09:26 2		Pur		
5j8p	May 27 2016 01:35 5	0.2998		Create feature-enhanced map	
chen	May 18 2016 08:11 1		8	Calculate a 2mFo-DFc map locally scaled and density-modified to enhance fine	
sukumar	May 18 2016 07:24 4	0.1077	700	details	
pka-validate	May 18 2016 02:07 1			Map Comparison	
zohar	May 17 2016 06:37 0		8	Tool for calculating equivalent sigma levels when visually comparing two maps	
tt_test	May 15 2016 04:17 38	3	223		
msf	May 09 2016 06:44 0			Polder Mans	
olga	Apr 29 2016 03:50 7		X	Tool for calculating Polder maps	
ramal	Apr 27 2016 10:45 5				
fTI S	Mar 21 2016 08.22 0		МО		
Current directory:	Current directory: /Users/pafonine/demo/1aba-polder Browse				
PHENIX version dev-s	PHENIX version dev-svn-000 Project: 1aba-polder				

Estimating and using uncertainty

100 identical refinement runs each one starting with slightly perturbed model



Refinement run

Support

• Feedback, questions, help

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

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