

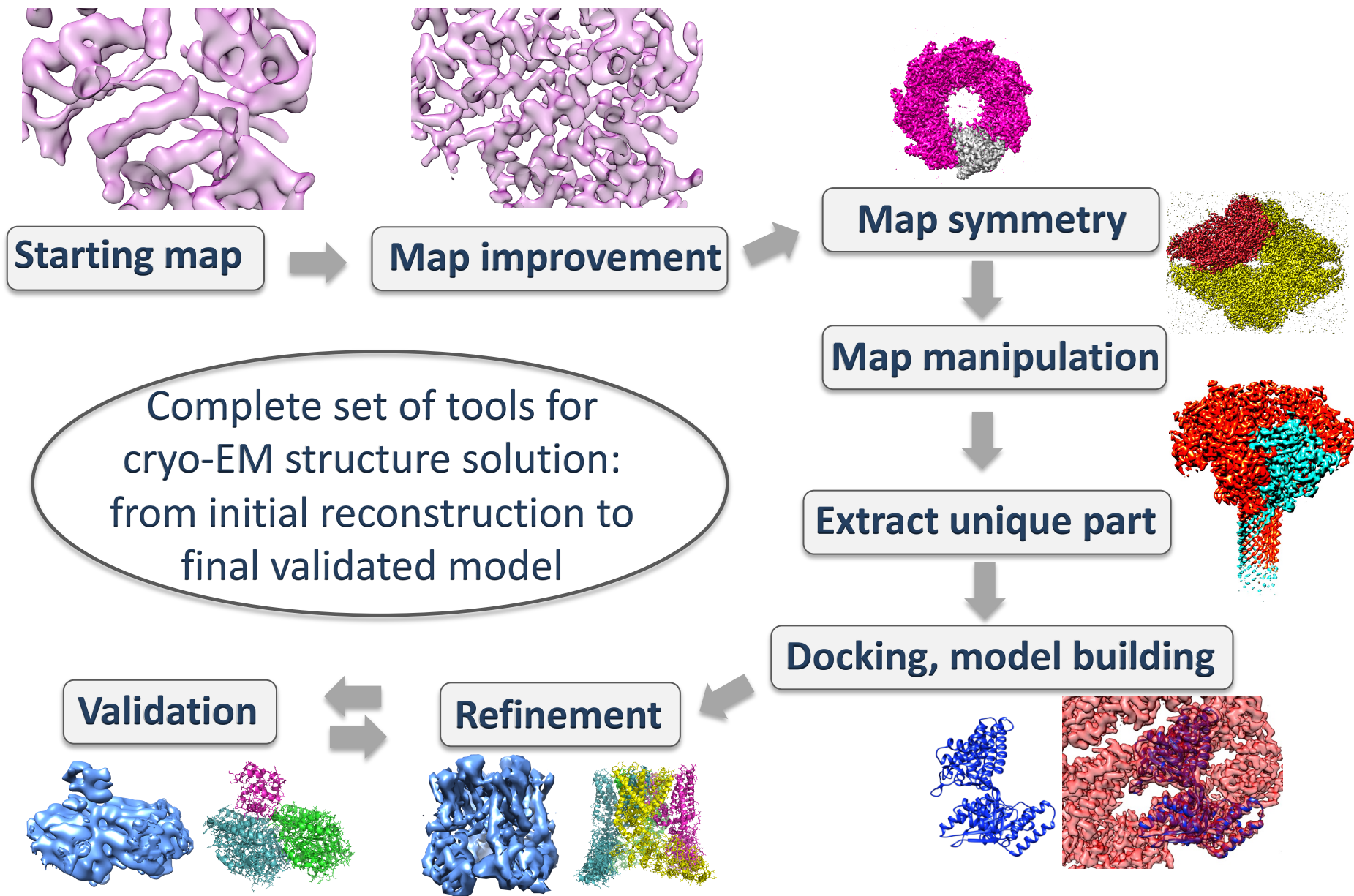
***Phenix* Tools for Cryo-EM: Validation**

Pavel Afonine

LBNL, Berkeley, California, USA

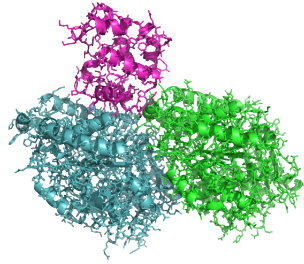
January 30, 2020

Cryo-EM tools in *Phenix*

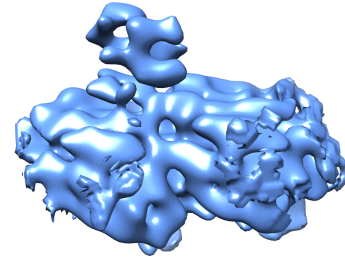


Validation

Model

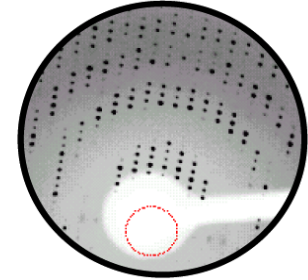


Data



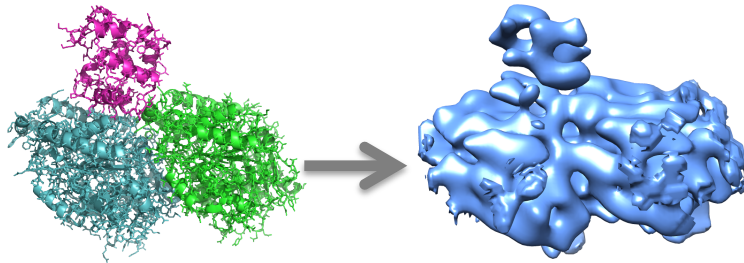
Cryo-EM

or



Diffraction

Model to data fit

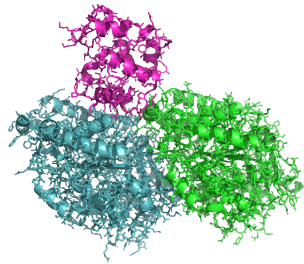


Validation = checking model, data and model-to-data fit are all make sense and obey to prior expectations

Validation tools: Crystallography vs Cryo-EM

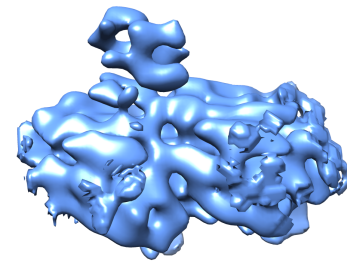
Exact same

Model



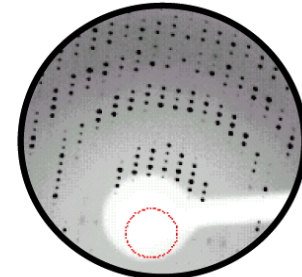
Different

Data



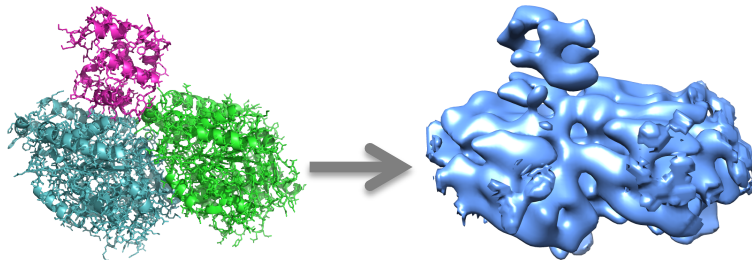
Cryo-EM

or



Diffraction

Model to data fit



Similar

Validation

- **Helps to save time**
- **Helps to produce better models**
- **Helps to set correct expectations**

Validation

Validation for crystallography (X-ray, neutron) and cryo-EM

The screenshot displays the PHENIX software interface. At the top, there is a menu bar with icons for Quit, Preferences, Help, Citations, Coot, PyMOL, KING, Other tools, and Ask for help. Below the menu bar, there are tabs for 'Actions' and 'Job history'. The main window is divided into two panes. The left pane, titled 'Projects', contains a table with columns for ID, Last modified, # of jobs, and R-free. The right pane, titled 'Favorites', contains a list of validation and analysis tools. A black arrow points from the top of the Favorites pane to the 'Validation' section.

Projects

Show group: All groups [v] Manage...

Select Delete New project Settings

ID	Last modified	# of jobs	R-free
✓ real-space-refin...	Mar 20 2019 11:08 ...	1	---
sbgrid	Mar 20 2019 09:55 ...	6	0.1627
tmp31	Mar 08 2019 10:43 ...	7	---
tmp30	Mar 05 2019 03:52 ...	4	---
tmp25	Feb 07 2019 06:32 ...	4	---
tmp24	Feb 05 2019 11:56 ...	2	0.3555
Lan	Jan 30 2019 10:28 ...	1	---
tmp22	Jan 29 2019 09:47 ...	4	---
paper4	Jan 25 2019 11:19 AM	2	0.2520
tmp11	Nov 21 2018 03:39 ...	1	---
tmp10	Nov 20 2018 08:02 ...	9	0.2213
tmp07	Nov 19 2018 10:24 ...	1	---
tmp09	Nov 19 2018 06:34 ...	3	---
tmp08	Nov 18 2018 10:31 ...	1	0.1900
1us0	Nov 18 2018 02:38 ...	2	0.1032
scott	Nov 17 2018 08:07 ...	1	---
KEuser	Oct 28 2018 09:35 ...	1	---
emma	Oct 08 2018 02:34 ...	1	---
cele	Oct 07 2018 08:08 ...	1	0.3513
5c11	Oct 04 2018 11:55 ...	3	0.2819
almu	Oct 03 2018 08:25 ...	2	---

Favorites

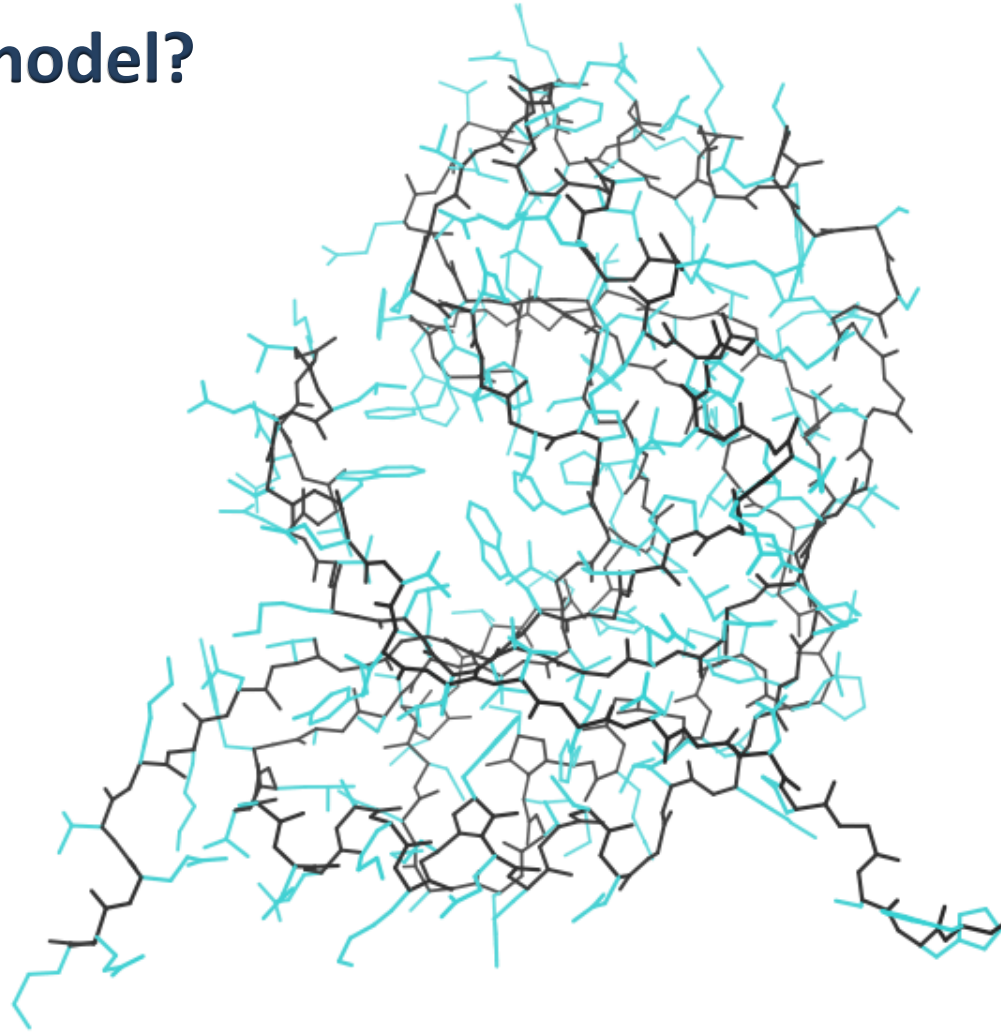
- Data analysis
- Experimental phasing
- Molecular replacement
- Model building
- Refinement
- Cryo-EM
- Validation**
 - Comprehensive validation (X-ray/Neutron)**
Model quality assessment, including real-space correlation and geometry inspection using MolProbity tools
 - Comprehensive validation (cryo-EM)**
Model quality assessment, including real-space correlation, for cryo-EM structures
 - Structure comparison**
Identify differences between multiple structures of the same protein, using multiple criteria
 - Calculate CC***
Comparison of unmerged data quality with refined model, as described in Karplus & Diederichs (2012)
 - EMRinger**
Model validation for de novo electron microscopy structures
- Ligands

Current directory: /Users/pafonine/Documents/real-space-refine-5ljv Browse... 🔍

PHENIX version dev-svn-000 Project: real-space-refine-5ljv

Model validation

Is it a good model?

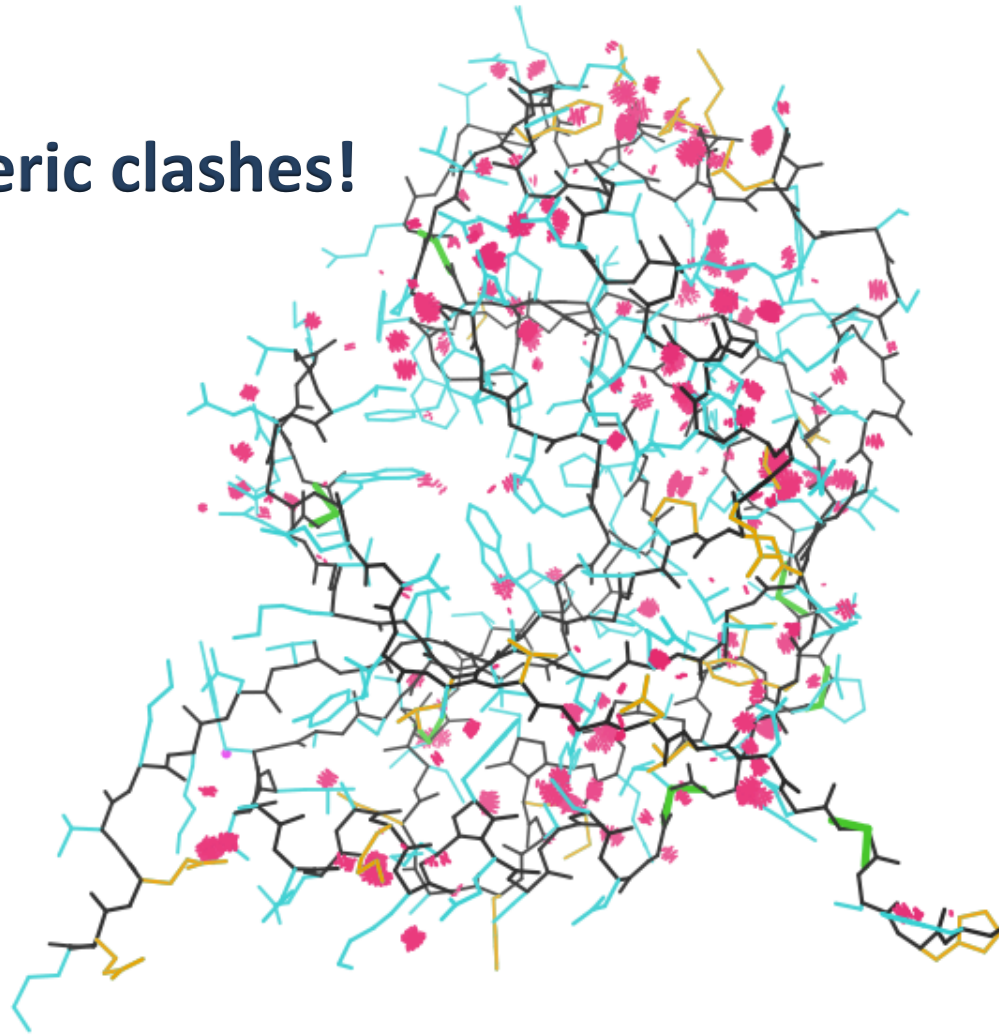


Cyclic Nucleotide Phosphodiesterase (2.4 Å)

Model validation

No.

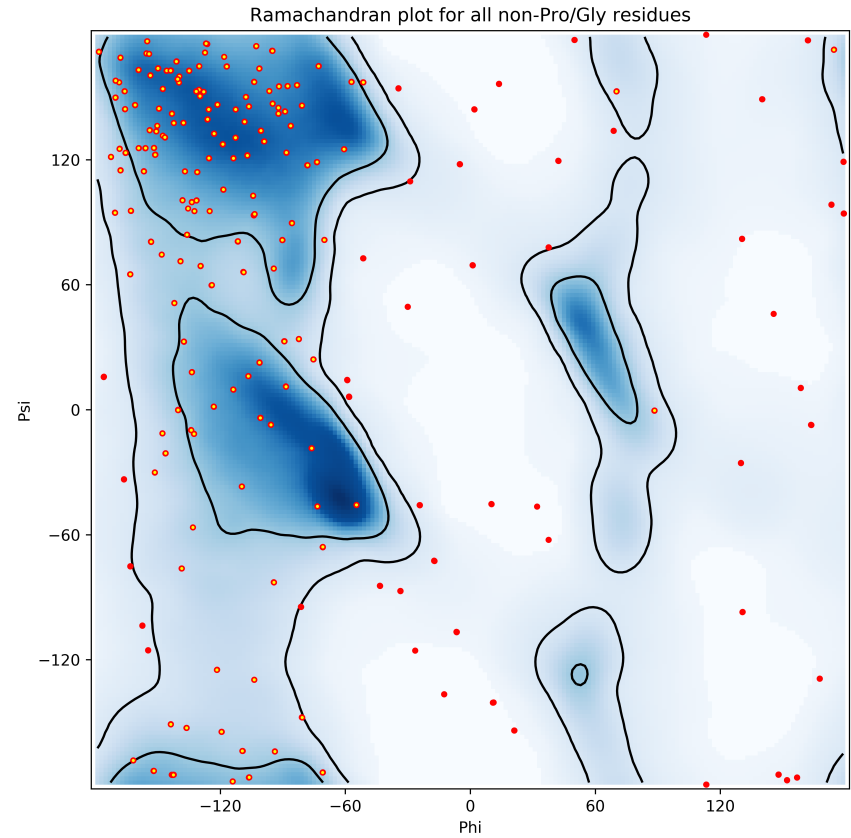
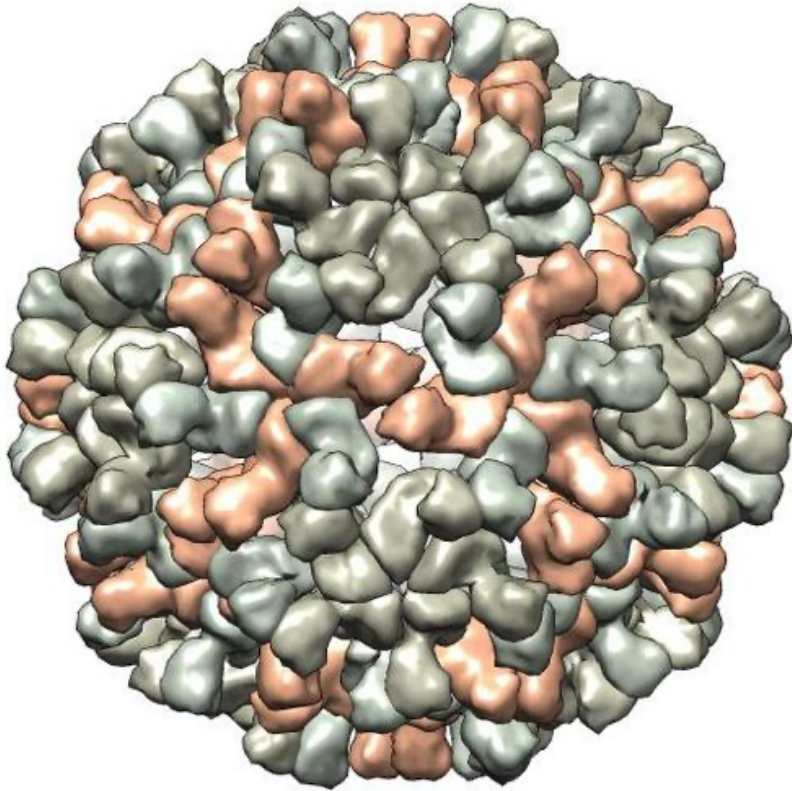
Too many steric clashes!



Cyclic Nucleotide Phosphodiesterase (2.4 Å)

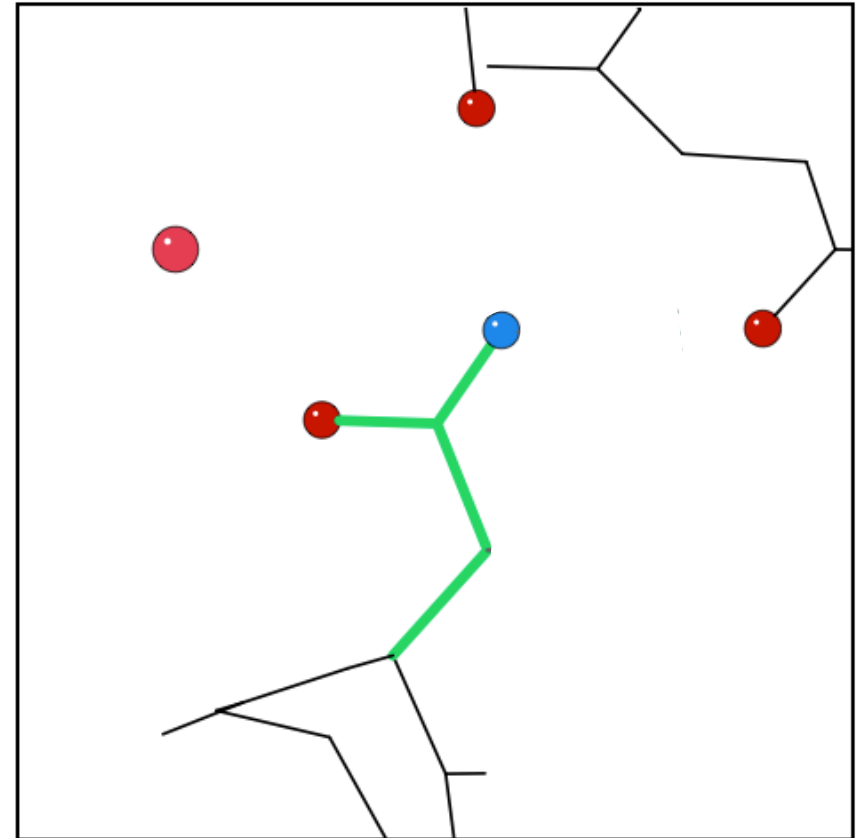
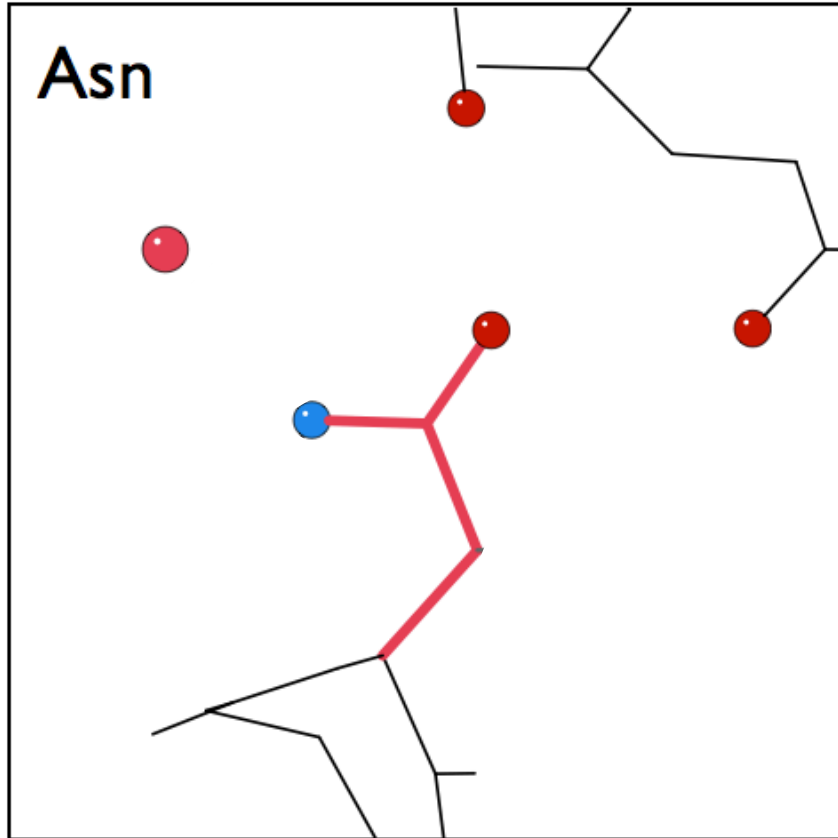
Model validation

3zx9

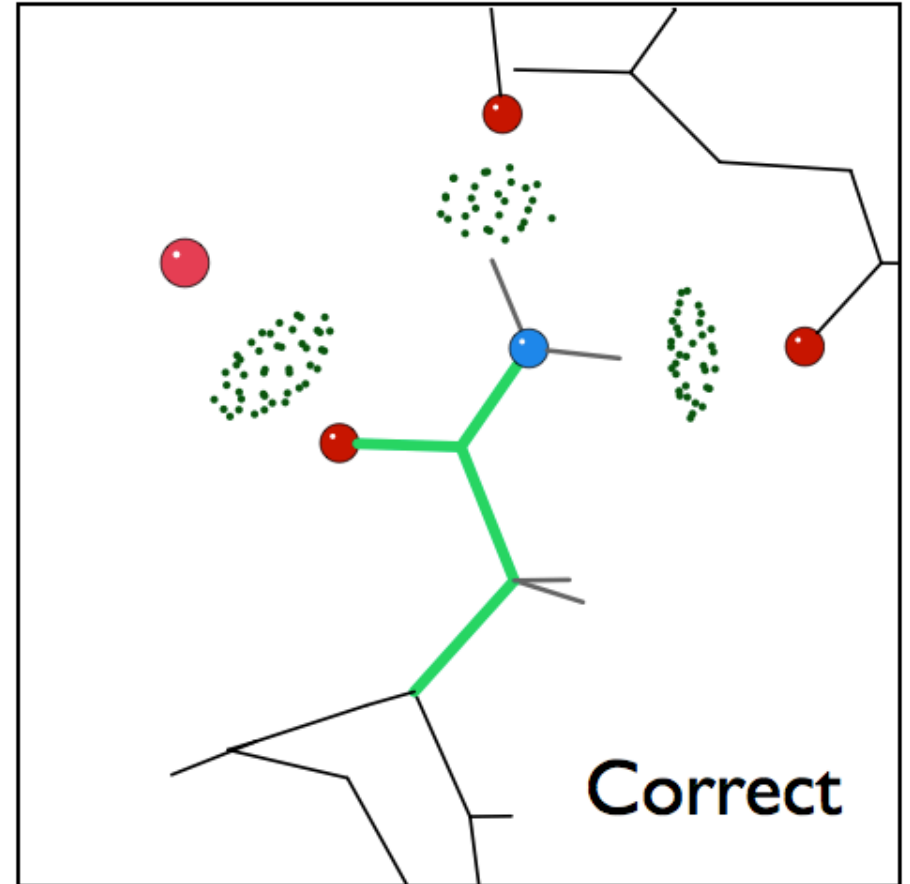
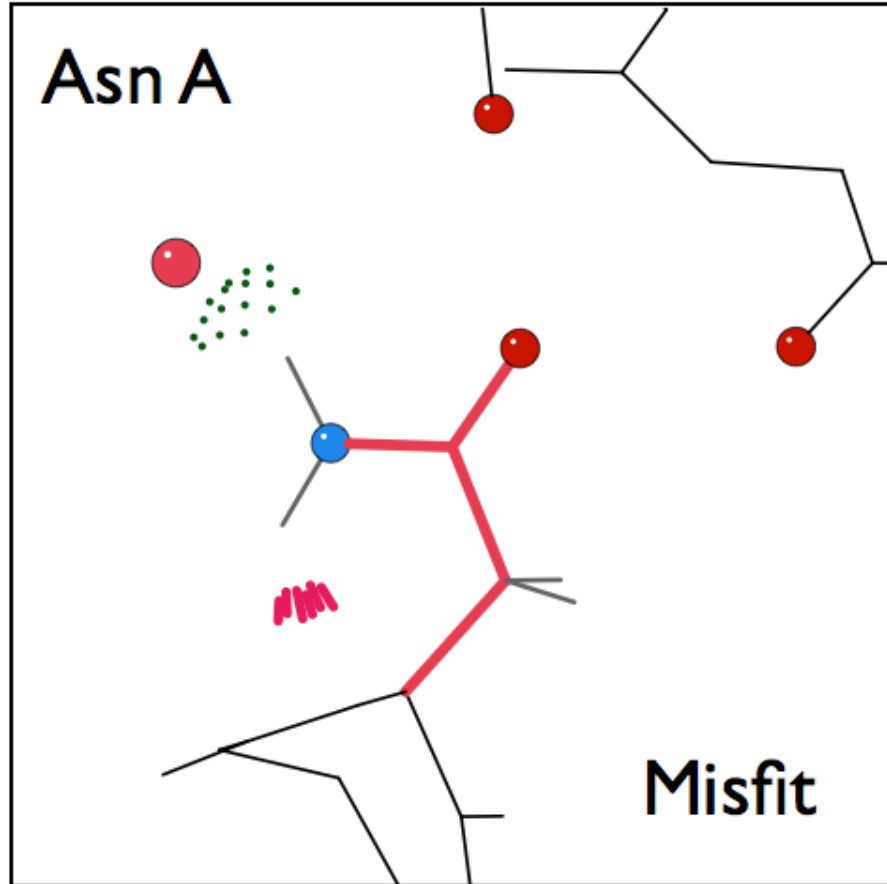


Very unlikely Ramachandran plot!

Clashes – N/Q/H flips



Clashes – N/Q/H flips



Validation – Table 1 (Crystallography)

• Data

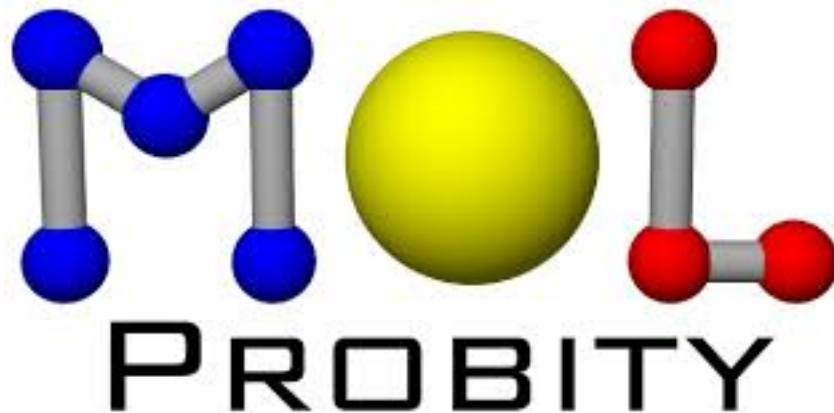
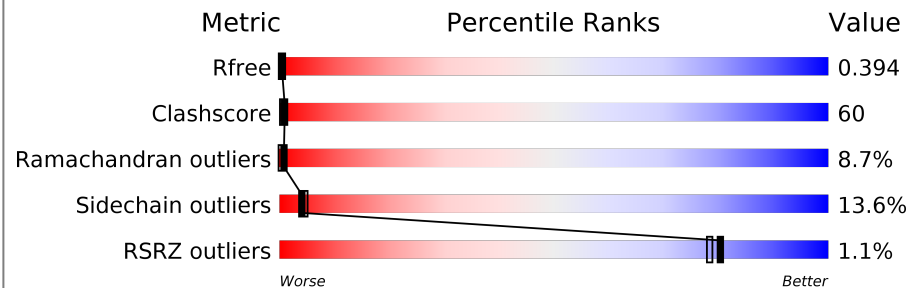
- Unit cell parameters & space group
- Data collection details (T, λ , instrument,...)
- Resolution & Completeness
- I/σ
- Redundancy
- Wilson B
- Various CC and R factors

• Model

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

• Model-to-Data fit

- R_{WORK} , R_{FREE}



Validation (cryo-EM)

Comprehensive validation (cryo-EM) (Project: real-space-refine-5ljv_0)

Preferences Help Run Abort Ask for help

Input/Output ValidationCryoEM_1

Run status Summary MolProbity Model vs. Data Data

Open in Coot Export Table 1

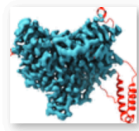
Model		Data		
Composition (#)		Box		
Chains	2	Lengths (Å)	50.92, 68.34, 83.08	
Atoms	2500 (Hydrogens: 0)	Angles (°)	90.00, 90.00, 90.00	
Residues	Protein: 325 Nucleotide: 0	Supplied Resolution (Å)	3.6	
Water	0	Resolution Estimates (Å)	Masked	Unmasked
Ligands	MG: 1 ADP: 1	d FSC (half maps; 0.143)	---	---
Bonds (RMSD)		d 99 (full/half1/half2)	3.7/---/---	3.1/---/---
Length (Å) (# > 4σ)	0.029 (146)	d model	3.7	3.7
Angles (°) (# > 4σ)	2.853 (122)	d FSC model (0/0.143/0.5)	3.4/3.5/3.6	3.4/3.6/3.9
MolProbity score	3.14	Man min/max/mean	-0.42/0.80/0.03	
Clash score	19.06	Model vs. Data		
Ramachandran plot (%)		CC (mask)	0.83	
Outliers	3.10	CC (box)	0.55	
Allowed	7.12	CC (peaks)	0.34	
Favored	89.78	CC (volume)	0.83	
Rotamer outliers (%)	11.57	Mean CC for ligands	0.86	
Cβ outliers (%)	3.68			
Peptide plane (%)				
Cis proline/general	5.6/0.0			
Twisted proline/general	11.1/0.7			
CaBLAM outliers (%)	2.18			
ADP (B-factors)				
Iso/Aniso (#)	2500/0			
min/max/mean				
Protein	30.26/493.42/109.69			
Nucleotide	---			
Ligand	57.57/99.69/75.15			
Water	---			
Occupancy				
Mean	1.00			
occ = 1 (%)	100.00			
0 < occ < 1 (%)	0.00			
occ > 1 (%)	0.00			

Idle Project: real-space-refine-5ljv_0

RESEARCH PAPERS

Acta Cryst. (2018). D74, 814-840
<https://doi.org/10.1107/S2059798318009324>

Cited by 71 OPEN



New tools for the analysis and validation of cryo-EM maps and atomic models

P. V. Afonine¹, B. P. Klaholz², N. W. Moriarty³, B. K. Poon⁴, O. V. Sobolev⁵, T. C. Terwilliger⁶, P. D. Adams⁷ and A. Urzhumtsev

Recent advances in the field of electron cryomicroscopy (cryo-EM) have resulted in a rapidly increasing number of atomic models of biomacromolecules that have been solved using this technique and deposited in the Protein Data Bank and the Electron Microscopy Data Bank. Similar to macromolecular crystallography, validation tools for these models and maps are required. While some of these validation tools may be borrowed from crystallography, new methods specifically designed for cryo-EM validation are required. Here, new computational methods and tools implemented in PHENIX are discussed, including *d*₉₉ to estimate resolution, *phenix.auto_sharpen* to improve maps and *phenix.mtriage* to analyze cryo-EM maps. It is suggested that cryo-EM half-maps and masks should be deposited to facilitate the evaluation and validation of cryo-EM-derived atomic models and maps. The application of these tools to deposited cryo-EM atomic models and maps is also presented.

Keywords: cryo-EM; atomic models; model quality; data quality; validation; resolution.

Read article Similar articles

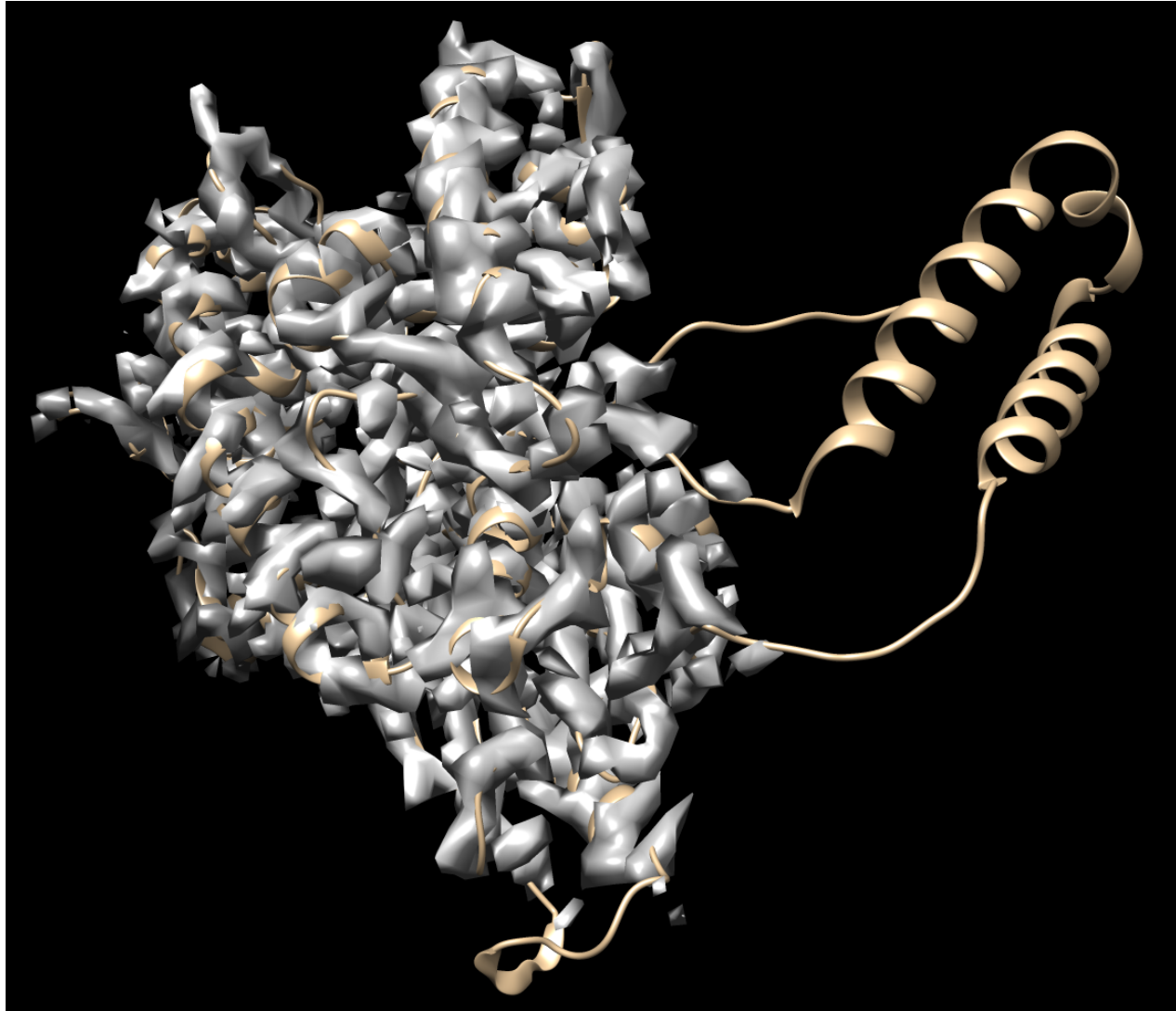
Validation: model-to-map fit

3a5x (emd_1641) | 4.0Å | CC \approx 0



Validation: model-to-map fit

3j9e (emd_6240) | 3.3Å | CC = 0.85 | Year: 2015



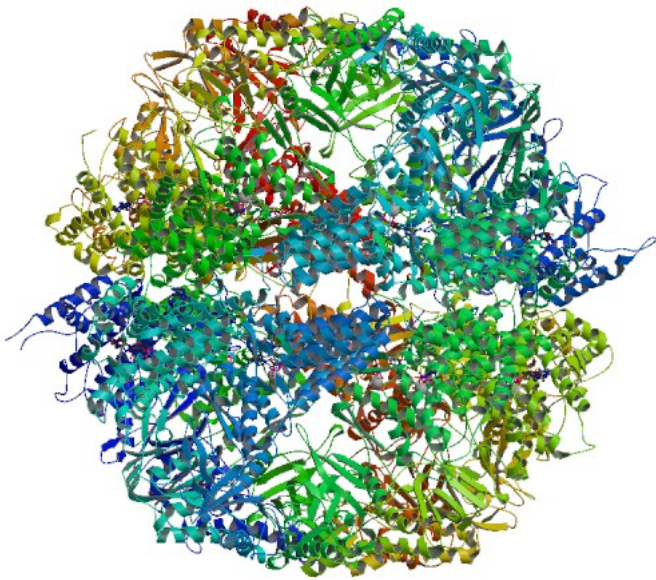
Latest trends, developments and... issues

Latest trends and developments

- Validation metrics progressively become refinement goals
 - `phenix.refine` and `phenix.real_space_refine` use:
 - Ramachandran plot restraints
 - C β deviation restraints
 - Secondary structure restraints
 - Restraints on χ angles of amino-acid side-chain rotamers
- As result, validation becomes less capable of catching problems

Latest trends and developments

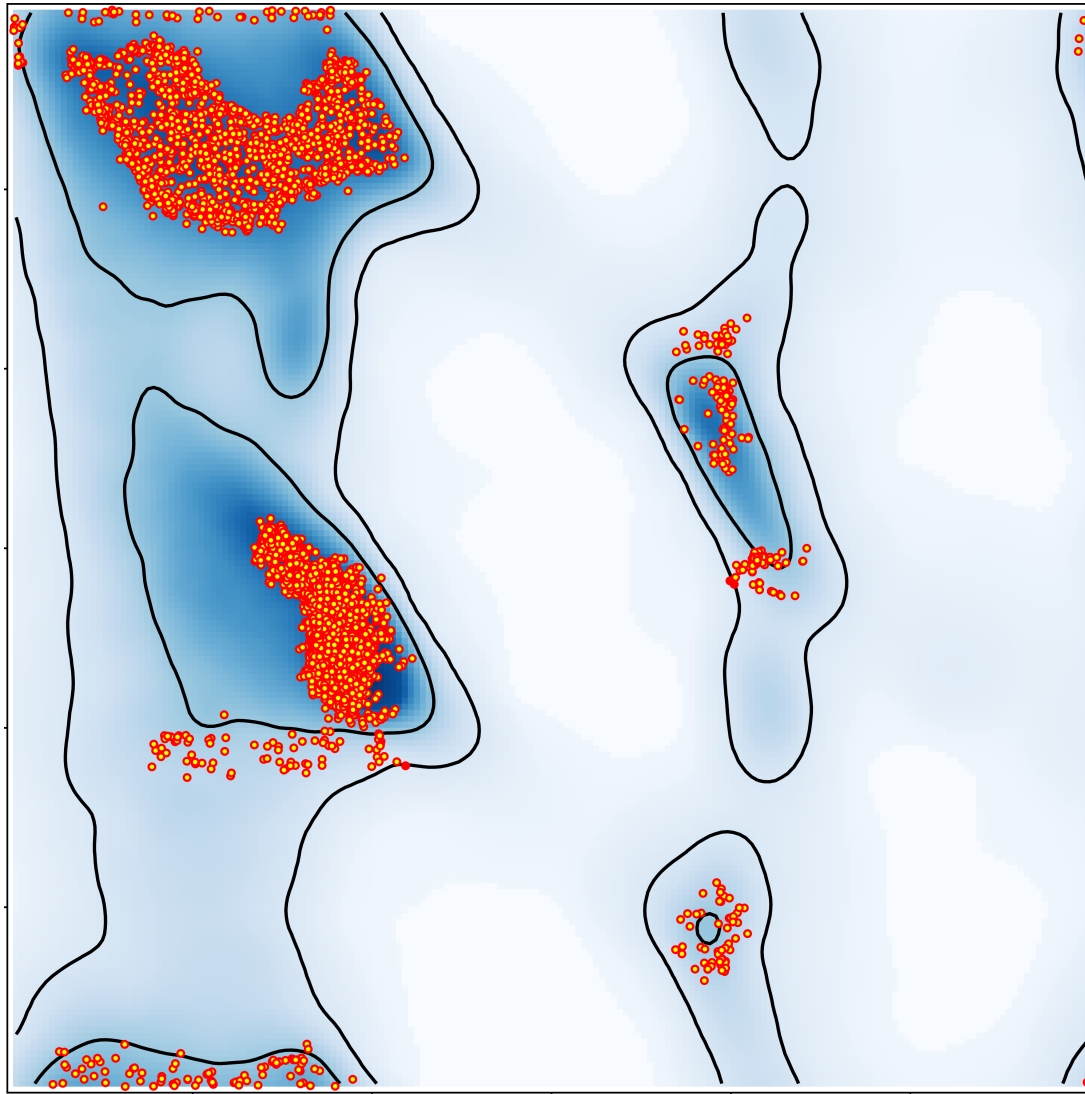
PNAS, 2019 116 (39) 19513-19522



Metric / PDB code		6KS6
Clashscore		7.7
Ramachandran (%)	avored	96.4
	outliers	0.2
Rotamer outliers (%)		0
C _β deviations		0
RMSD	Bond (Å)	0.001
	Angle (°)	0.396
Resolution (Å)		3.0

- Perfect statistics

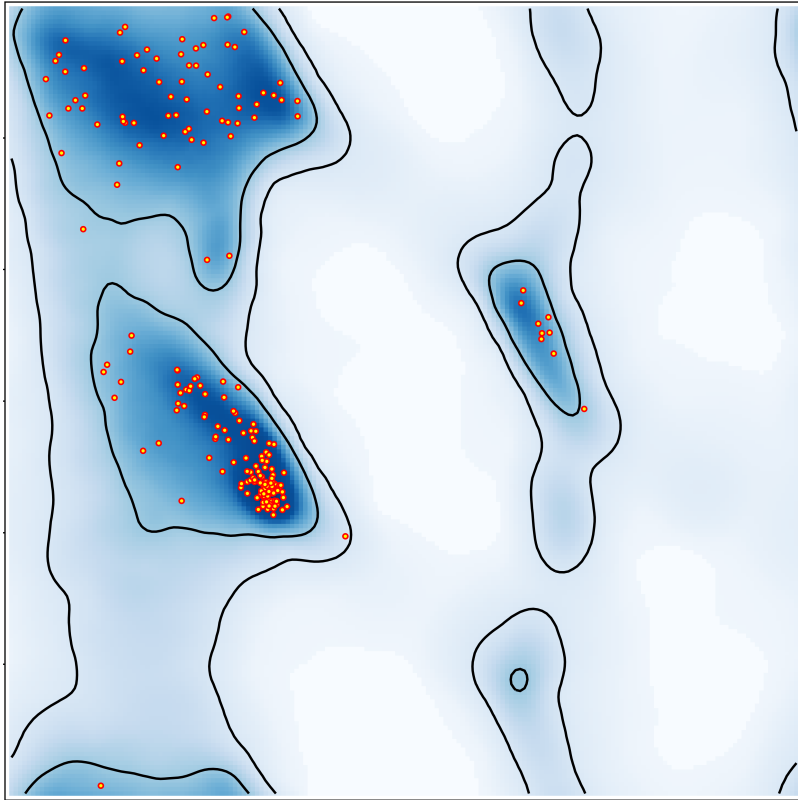
Latest trends and developments



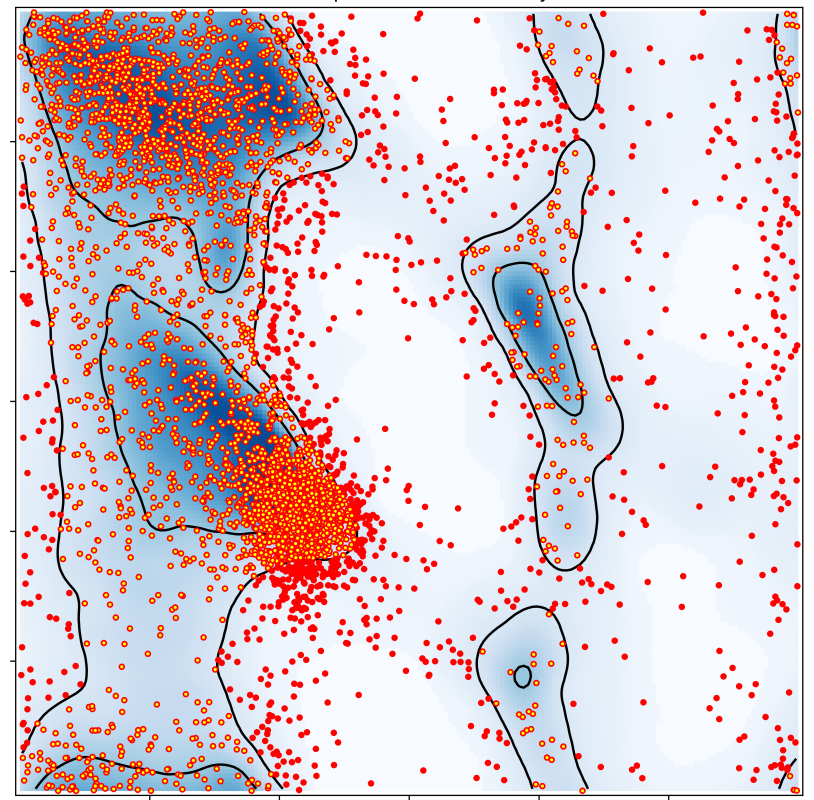
- Is it a good plot?

Example: good vs bad plots

1us0 | 0.66 Å



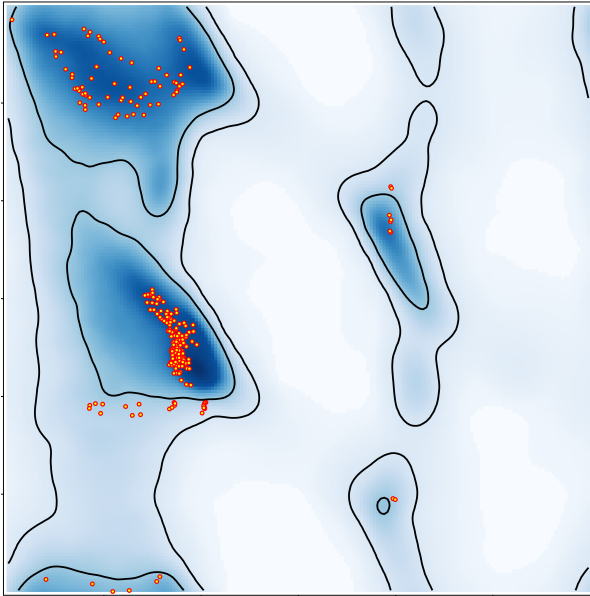
5a9z | 4.7 Å



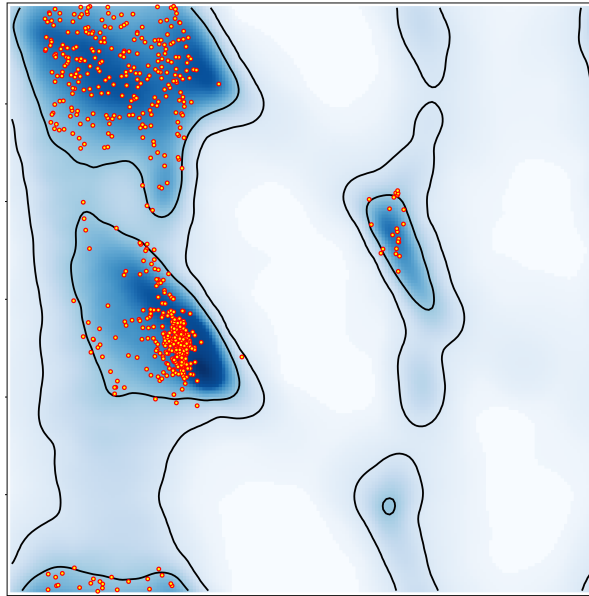
- A trained eye can distinguish good and bad plots

Example: poor plots

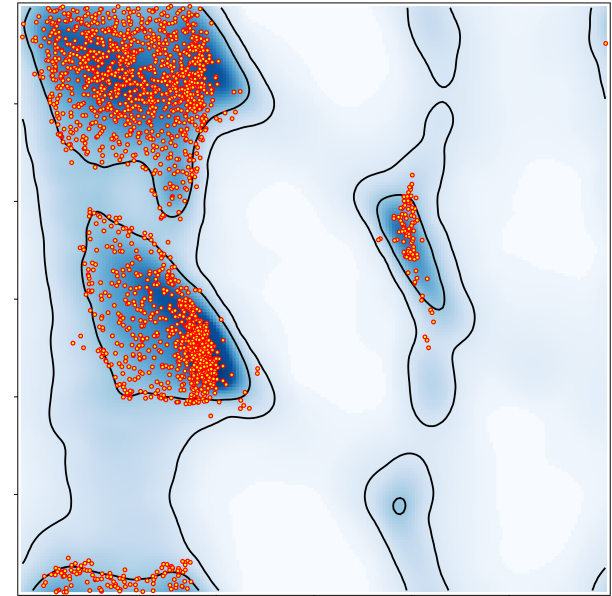
6bu9 | 6.8 Å



6dzv | 4.2 Å



6cs1 | 4.6 Å



- Overall Ramachandran plot counts (favored/outlier/allowed) will not flag this
- A trained eye required to appreciate the issue
- How to tell good vs bad without looking at the plot?

Ramachandran plot Z-score

CABIOS

Vol. 13 no. 4 1997
Pages 425-430

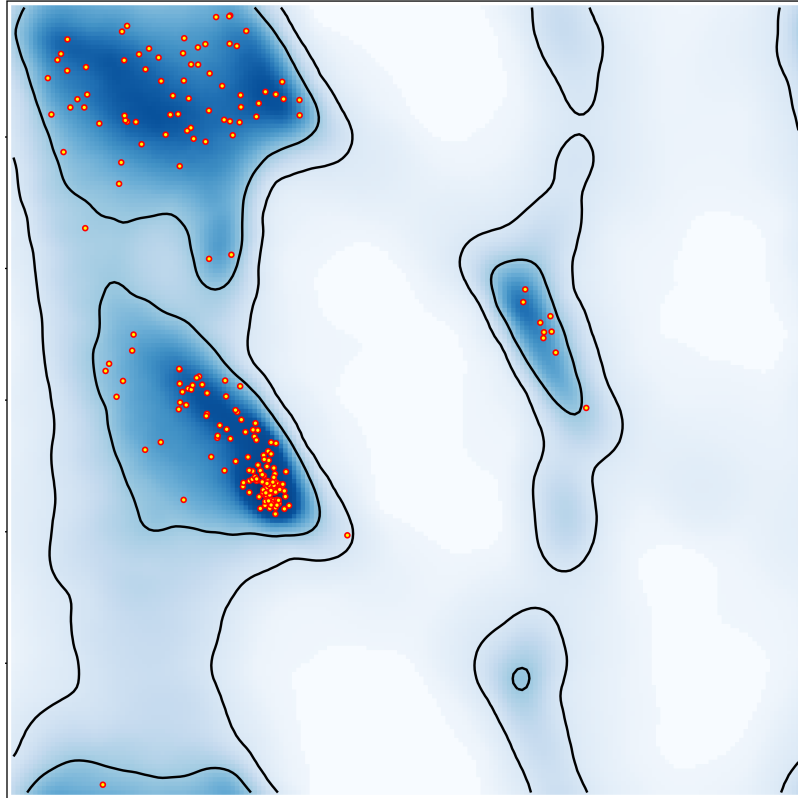
Objectively judging the quality of a protein structure from a Ramachandran plot

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

- Ramachandran Z-score is good at identifying odd-looking Ramachandran plots!
- Used in PDBREDO and WhatCheck. Implemented in *Phenix* (Oleg Sobolev)
 - One number, simple criteria:
 - $Z < -3$: Poor
 - $-3 < Z < -2$: Suspicious
 - $Z > -2$: Good

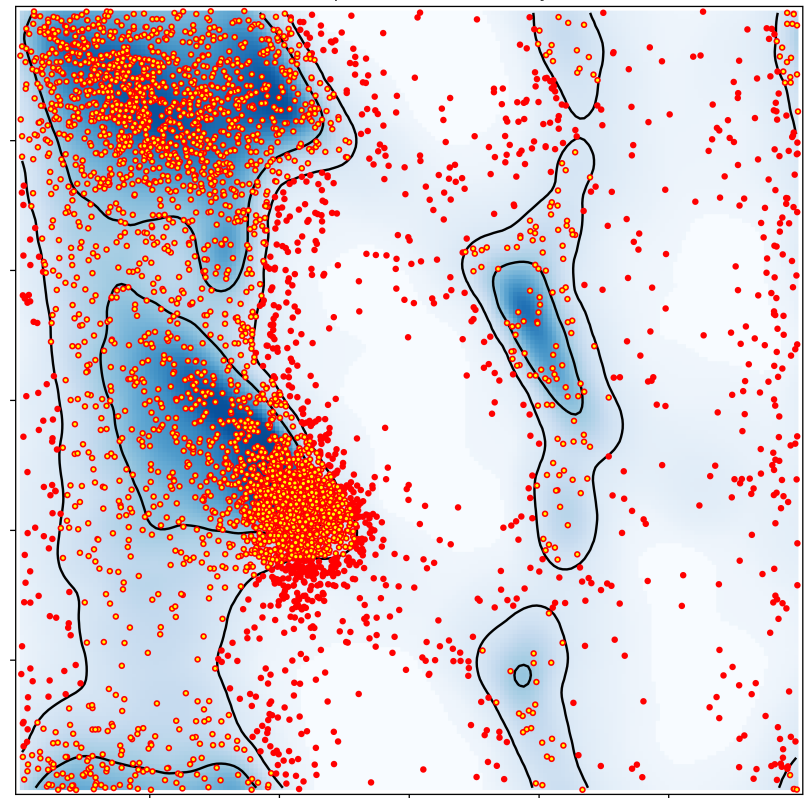
Example: good vs bad plots

1us0 | 0.66 Å



RamaZ = -0.5

5a9z | 4.7 Å

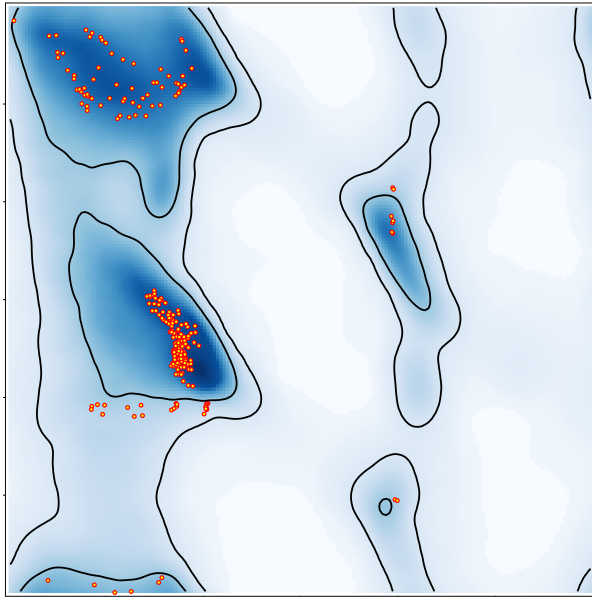


RamaZ = -7.7

- $Z < -3$: Poor
- $-3 < Z < -2$: Suspicious
- $Z > -2$: Good

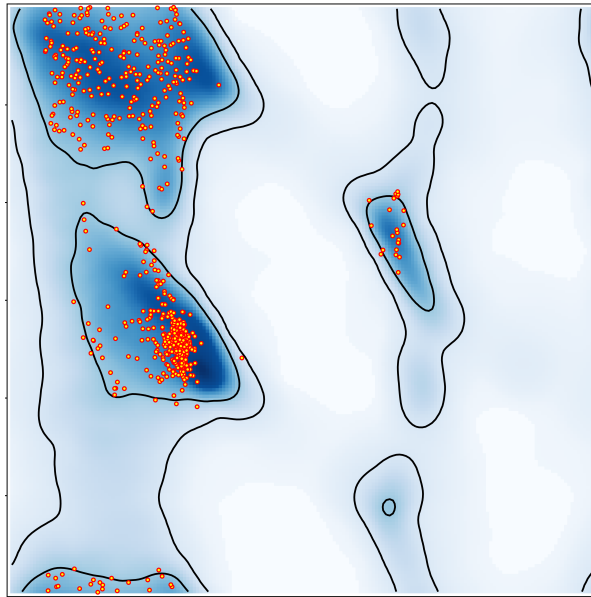
Example: odd plots

6bu9 | 6.8 Å



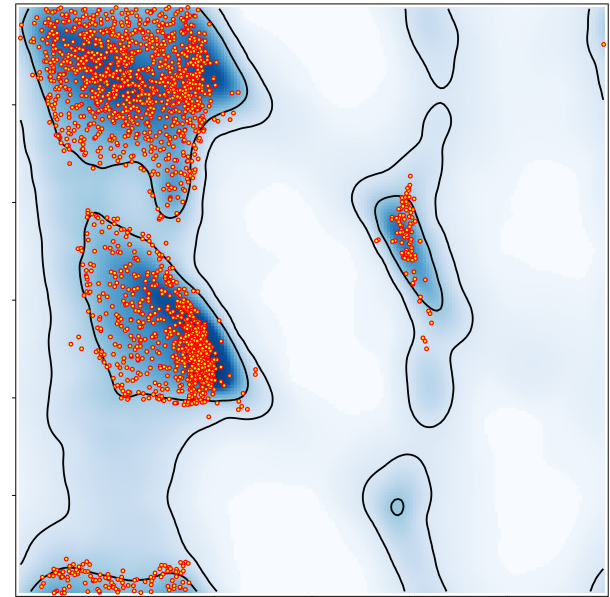
RamaZ = -5.0

6dzv | 4.2 Å



RamaZ = -4.1

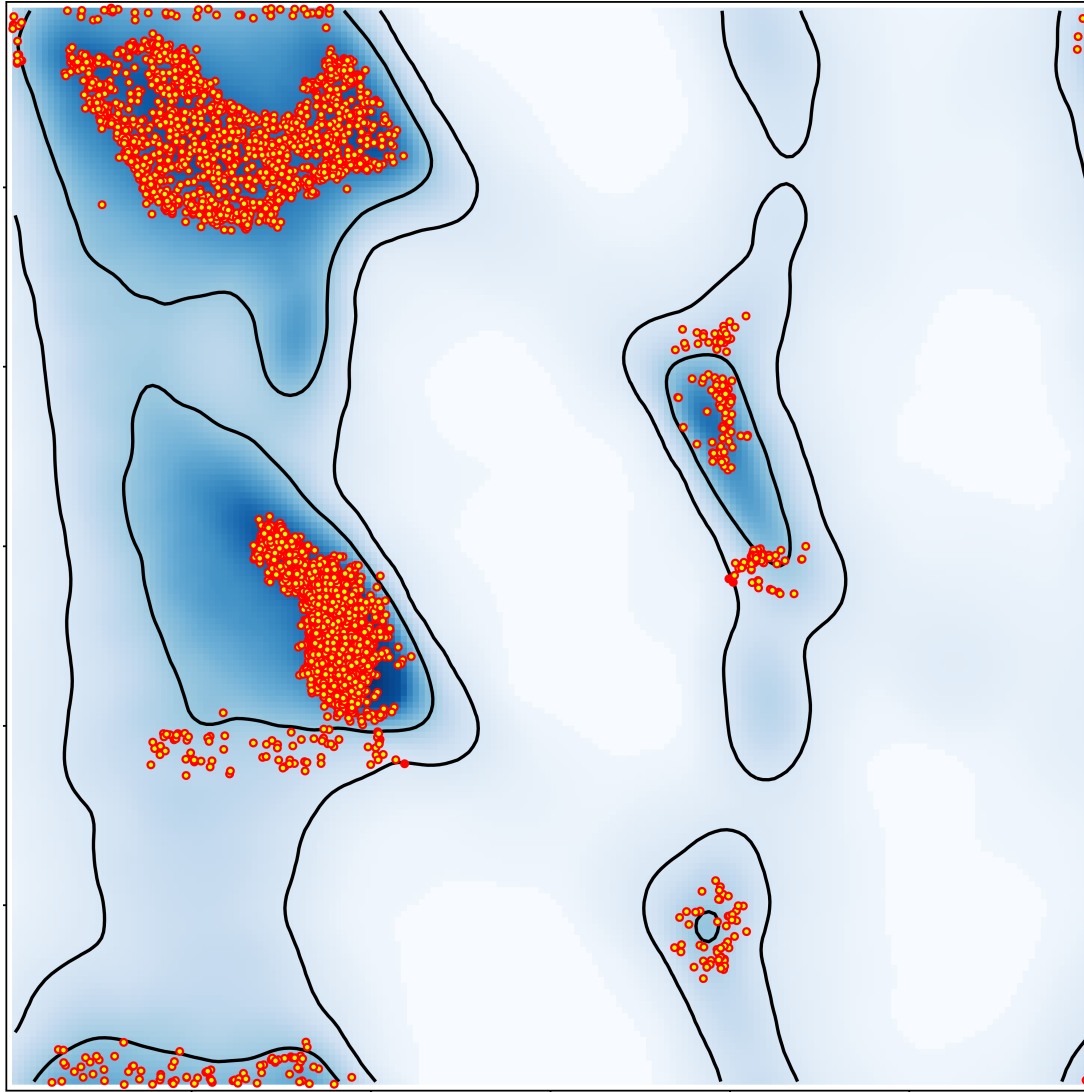
6cs1 | 4.6 Å



RamaZ = -4.2

- $Z < -3$: Poor
- $-3 < Z < -2$: Suspicious
- $Z > -2$: Good

6KS6



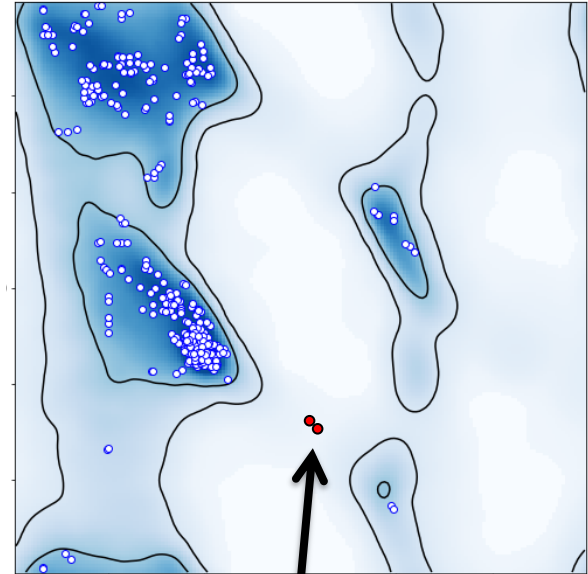
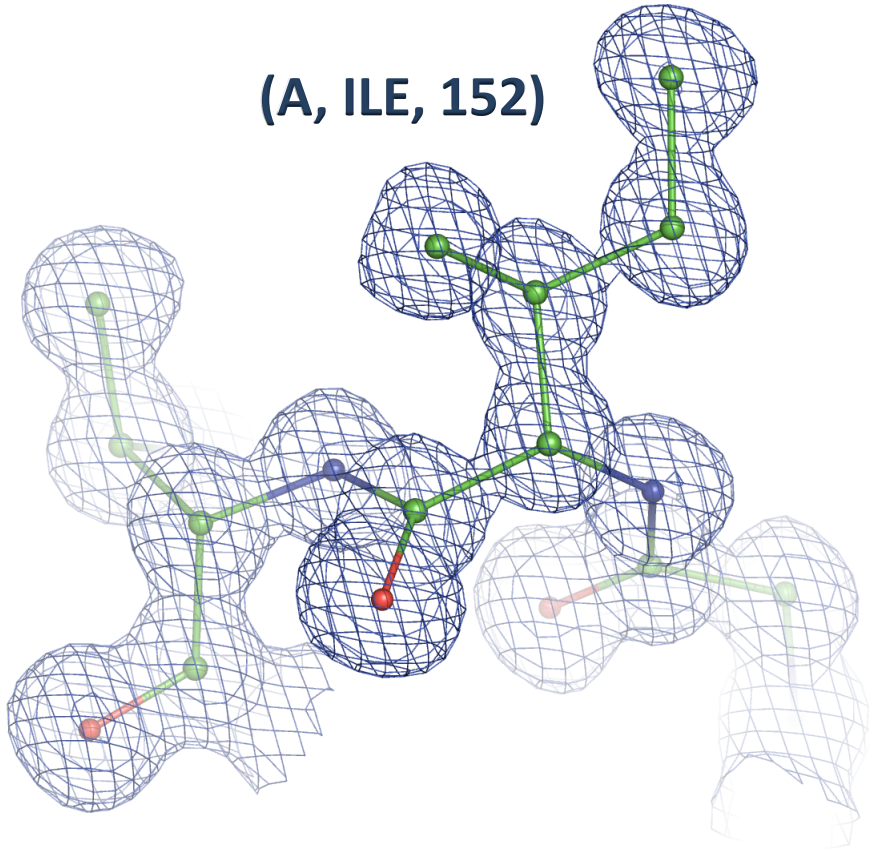
- $Z < -3$: Poor
- $-3 < Z < -2$: Suspicious
- $Z > -2$: Good

RamaZ = -3.3

Ramachandran plot

PDB code 3NOQ, 1 Å

(A, ILE, 152)



Outliers:

(A, ILE, 152), (B, ILE, 154)

Valid Ramachandran plot outliers: justified by the data (density map)

Ramachandran plot facts

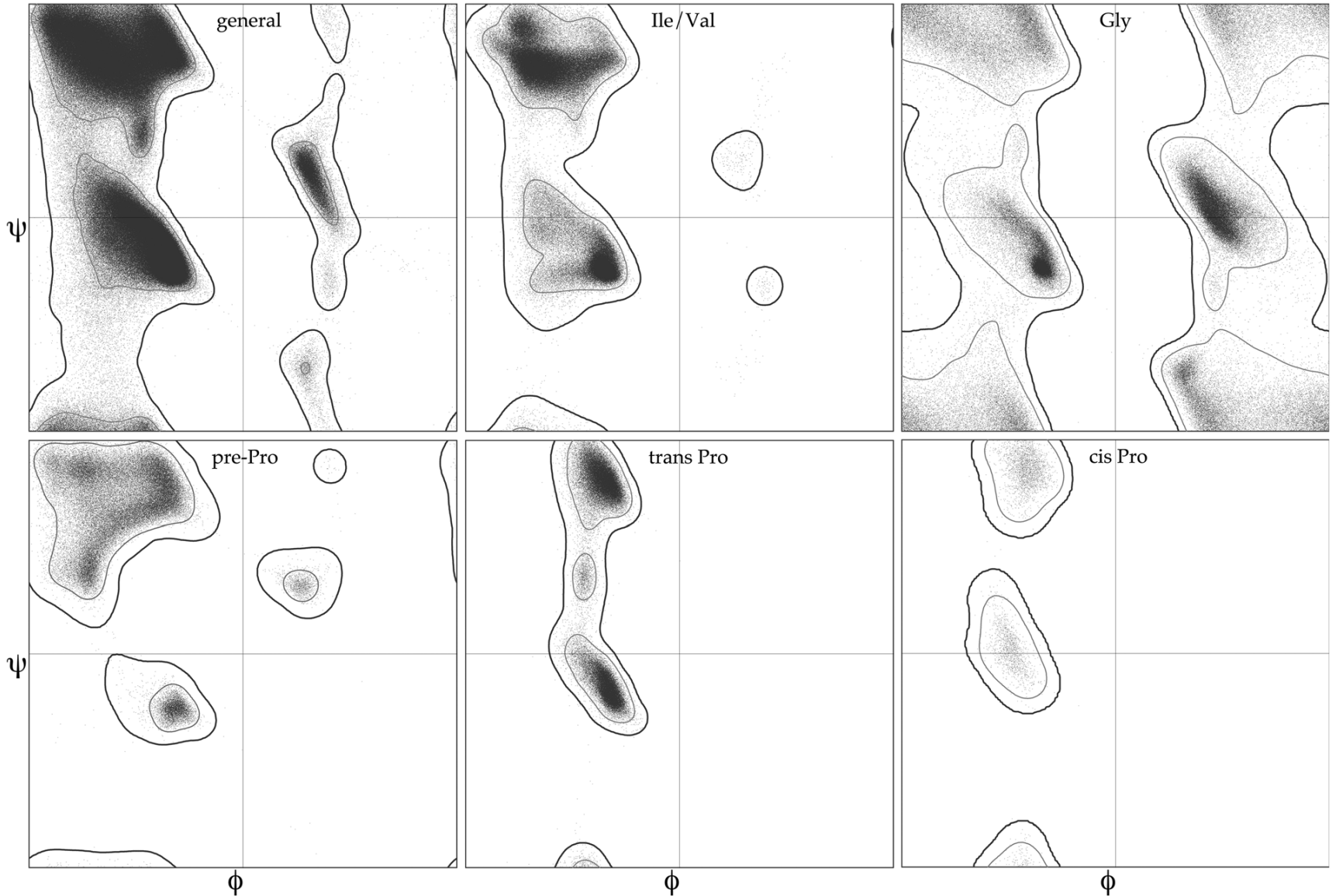


Image from Jane and David Richardson, Duke University

Ramachandran plot

Comprehensive validation (cryo-EM) (Project: real-space-refine-5ljl)

Preferences Help Run Abort Ask for help

Input/Output ValidationCryoEM_1

Run status Summary MolProbity Model vs. Data Data

Clashes CaBLAM C β Cis/Twisted Rotamers **Ramachandran** Geometry Restraints

Outlier list

Chain	Residue	Residue type	Score	Phi	Psi
A	GLY 77	Gly	0.01	59.0	-53.5
A	VAL 78	Ile/Val	0.06	-32.4	136.4
A	PRO 103	trans-Pro	0.00	-137.8	132.5
A	PRO 179	trans-Pro	0.00	-138.9	1.5
A	LYS 187	General	0.03	-57.5	98.4

Ramachandran graphs

Position type: All Residue name: * Save graph

Show data points: Any Color scheme: Blue

Ramachandran plot for all non-Pro/Gly residue

Ramachandran plot for Glycine

Ramachandran plot for cis-Proline

Ramachandran plot for trans-Proline

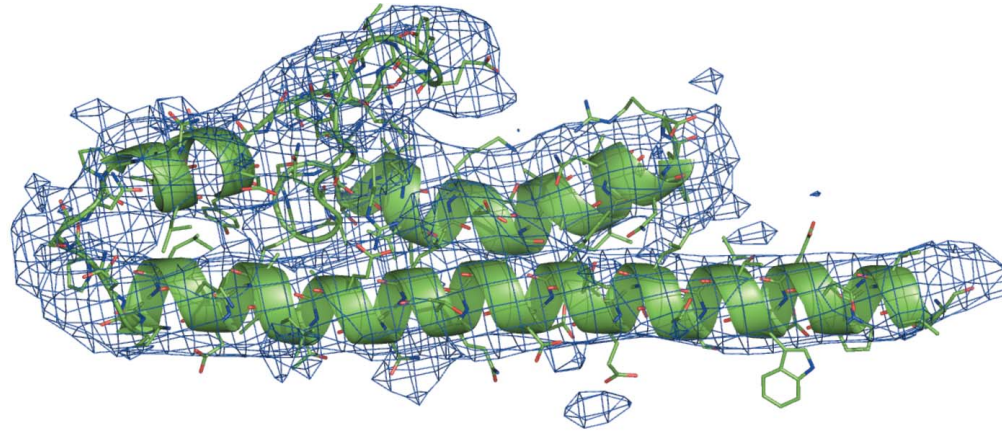
Ramachandran plot for pre-Proline residues

Ramachandran plot for Ile or Val

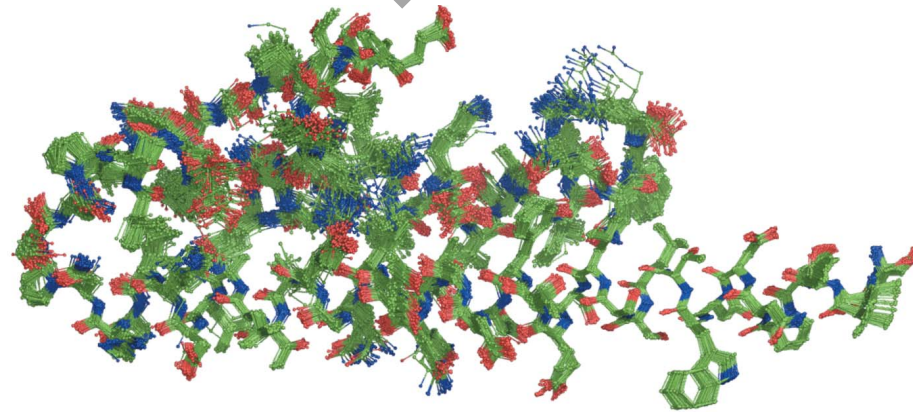
Idle Project: real-space-refine-5ljl

Multiple interpretation of low-res maps

- Low-resolution maps allow non-unique interpretation



phenix.mia



Model-map correlation coefficient (CC)

- **Definition**

- **With or w/o subtracting mean**

$$CC(\rho_1, \rho_2) = \left(\sum_{\mathbf{n}} (\rho_1(\mathbf{n}))^2 \right)^{-1/2} \left(\sum_{\mathbf{n}} (\rho_2(\mathbf{n}))^2 \right)^{-1/2} \left(\sum_{\mathbf{n}} \rho_1(\mathbf{n}) \rho_2(\mathbf{n}) \right)$$

$$CC(\rho_1, \rho_2) = \left(\sum_{\mathbf{n}} (\rho_1(\mathbf{n}) - \langle \rho_1 \rangle)^2 \right)^{-1/2} \left(\sum_{\mathbf{n}} (\rho_2(\mathbf{n}) - \langle \rho_2 \rangle)^2 \right)^{-1/2} \left(\sum_{\mathbf{n}} (\rho_1(\mathbf{n}) - \langle \rho_1 \rangle) (\rho_2(\mathbf{n}) - \langle \rho_2 \rangle) \right)$$

- **How model map is calculated**

- **Approximation (e.g. N-gaussian)**
- **Form-factors (electron vs crystallographic, eg. X-ray)**
- **Fourier map**
 - **Box or sphere of Fourier map coefficients**

- **Region in the map used to calculate CC**

- **Whole box**
- **Mask around atoms**
 - **Atom radius**