# Model Refinement and Validation

Damian Ekiert and Gira Bhabha NYU School of Medicine March 6, 2020

Goal of model refinement: To create a set of coordinates that 1) explains the data as best we can, but 2) also conforms with what we know about proteins in general

## Model Refinement vs Protein Folding Funnel

## $\Delta G_{unfolding}$

Energy

Image: Thomas Splettstoesser



## Refinement target describes differences between model and data



Image (adapted): Thomas Splettstoesser

**Different model parameters** (e.g., conformations)



### Refinement Target = (Model vs Data) Simplest:

## Map Model — (calculated from model)

(or simulated diffraction data "structure factors")

**Compare and quantify the differences** 

## Map (from experiment)

"Real-space" Refinement

7

(or experimental "structure factors" or "fourier coefficients")

"Reciprocal-space" Refinement



## At atomic resolution, position of individual atoms is well-defined



## But at "near-atomic" resolution, the position of residues and side chains is not always clear









## Refinement using only data



## Refinement using <u>only data</u>

## Harnessing prior knowledge of protein structure to bridge the gap



Refinement Target = (Model vs Data) + w<sub>1</sub> (Model vs PriorKnowledge)





## Stereochemistry



## Stereochemistry О 124° $120^{\circ}$ C Η 116° 0 1.51 Amino N + terminus Η Η R

**Torsion/dihedral Angles** 

# Constraints backbone conformations as well as side chain rotameric states

![](_page_11_Figure_3.jpeg)

## **Secondary Structure and Hydrogen Bonds**

![](_page_12_Picture_2.jpeg)

## **Distance restraints between H-bonding atoms** Torsion angle restraints to maintain appropriate backbone conformation

Image: http://book.bionumbers.org/what-is-the-energy-of-a-hydrogen-bond/

## "Non-crystallographic symmetry" (NCS), reference model

![](_page_13_Picture_1.jpeg)

Image: O'Dell, et al. Angewandte Chemie (2016).

## **Restrain to be similar Constrain to be identical**

## **Especially helpful at** lower resolution with non-symmetrized maps

Chains can be restrained to be similar to other chains in structure, or a "reference", higher resolution structure (or the starting model)

![](_page_13_Figure_7.jpeg)

## **B factor / ADP restraints**

### Higher B factor = **fatter ribbon**, warmer color

![](_page_14_Picture_2.jpeg)

### Image: Harry Jubb, https://github.com/arose/ngl/issues/291

# B factors are not randomly distributed

B factor of a particular residue is a good predictor of the residue just before and after

Therefore, we can retrain B factors such that connected atoms/residues must have similar B factors

![](_page_14_Picture_7.jpeg)

### Atom 1

Atom 2

**Optimal** center-to-center distance ~ sum of VDW radii

If atoms get too close together, need a force to push them apart

## Steric repulsion

![](_page_15_Figure_5.jpeg)

![](_page_15_Figure_6.jpeg)

![](_page_15_Picture_7.jpeg)

## **Force fields**

- to reduce atomic overlap, but also tends to push things apart
- Historically, X-ray refinement has relied "repulsive" forces (steric clashes) More complex force fields commonly used elsewhere:
  - NMR structure calculations
  - Molecular dynamics
- Growing interest in using more complex force fields, including "attractive" terms that may help stabilize model at lower resolutions
  - Rosetta energy

  - Implementation of Amber force field in latest version of Phenix • Especially helpful at lower resolutions!

## More complete refinement target includes many terms

![](_page_17_Picture_1.jpeg)

Refinement Target = (Model vs Data) +  $w_1$ (Model vs Stereo) +  $w_2$ (Model vs ForceField) +  $w_3$ (Model vs NCS) + ...

![](_page_18_Figure_2.jpeg)

\*Please don't be this person **Always give credit to software developers!** 

	Refinement Target	GL
PHENIX (phenix.real_space_refine)	Real-space	Ye
CCP-EM (REFMAC5)	Reciprocal- space	Ye
Rosetta	Real-space	Nc

# Some Highlights

![](_page_19_Figure_3.jpeg)

## Refinement moves model towards local minimum

![](_page_20_Figure_1.jpeg)

Image (adapted): Thomas Splettstoesser

**Different model parameters** (e.g., conformations)

![](_page_20_Figure_4.jpeg)

## **Optimization protocols: Rigid body refinement**

![](_page_21_Picture_1.jpeg)

Image: Woetzel, et al. J Struct Biol. 2011.

## **Optimization protocols: Gradient driven minimization**

![](_page_22_Picture_1.jpeg)

Slide adapted from: Pavel Afonine, LBNL (Phenix)

## Refinement "radius of convergence"

![](_page_23_Picture_1.jpeg)

## **Optimization protocols: Simulated annealing**

![](_page_24_Picture_1.jpeg)

Video: Darrell Hurt, <u>https://www.youtube.com/watch?v=59llFNEt8F4</u>

## **Optimization protocols: Simulated annealing**

![](_page_25_Picture_1.jpeg)

Slide adapted from: Pavel Afonine, LBNL (Phenix)

## **Optimization protocols: Torsion-angle Grid Search**

![](_page_26_Figure_1.jpeg)

![](_page_26_Figure_2.jpeg)

![](_page_26_Figure_3.jpeg)

m

р

Disallowed t

### Disallowed

![](_page_26_Figure_6.jpeg)

 Can allow for larger shifts in model than simple minimization

![](_page_26_Figure_8.jpeg)

## **Optimization protocols: Torsion-angle Grid Search**

![](_page_27_Picture_1.jpeg)

Slide adapted from: Pavel Afonine, LBNL (Phenix)

## **Optimization protocols: Morphing**

![](_page_28_Picture_1.jpeg)

Slide adapted from: Tom Terwilliger, Los Alamos (Phenix)

## **Optimization protocols: Morphing**

![](_page_29_Picture_1.jpeg)

Slide adapted from: Tom Terwilliger, Los Alamos (Phenix)

	Real-space r	refinement (Project: tmp)	
🕺 🛞 🍋 🖾			
Preferences Help Run Abort Save Ask	for help		
Input/Output Definement Settings DealSpaceDefine 4			d b
input/output Rennement Settings RealSpaceRenne_4			4 1
Job title : run1			
Input			
File path	Format	Data type	
Users/dcekiert/Synch/PDBs/YebT_20180804/	PDB	Input model	
/Users/dcekiert/Synch/PDBs/YebT_20180804	CCP4 map	Map file	
Add file Remove file Modify file data type			
Resolution : 3.0	Map coefficient	s label : 🗘	
- Ignore symmetry conflicts			
Output			
File name prefix : run1			
✓ Write initial geo file 🛛 Write final geo file	Write all states	s 🔽 Write final model in mmCIF 🛛 🗹 Run validation	
•			
1 ioh(s) running		Project: tmp	
		Project. tilp	

🔴 😑 🔵 Real-space refinement (Project: tmp)	
🕺 💮 🔂 🛄 🛛 🚯	
Preferences Help Run Abort Save Ask for help	
Input/Output Refinement Settings RealSpaceRefine_4	⊲ ⊳
Strategy	
Run : morphing simulated_annealing dp	
Max iterations : 100 Macro cycles : 5	
Target bonds rmsd :    0.01    Target angles rmsd :    1.0	
Select Atoms Vse secondary structure restraints Vse NCS	
Strategy Options	
Morphing : once	
Simulated annealing : once Options	
Reference model restraints Options	
Other Options	
Scattering table : electron 🗘 Weight : Resolution factor :	0.25
Nproc: 1 Random seed : 0	
🗹 Rotamer restraints 🛛 🗹 Ramachandran restraints 🔂 Ncs constraints	
🗹 Refine ncs operators 🛛 🗹 Show per residue	
Model interpretation Automatic linking All parameters	
O 1 job(s) running	Project: tmp

Input/Output Refinement Settings   RealSpaceRefine_4     Run status     Status     Running on Damians-MacBook.local (PID 858)	4 Þ 4 Þ
Preferences Help   Run Abort   Status     Running on Damians-MacBook.local (PID 858)	4 Þ 4 Þ
Preferences     Help     Run     Abort     Save     Ask for help       Input/Output     Refinement     Settings     RealSpaceRefine_4       Run status     Status       Status       Running on Damians-MacBook.local (PID 858)	4 Þ 4 Þ
Input/Output Refinement Settings RealSpaceRefine_4 Run status Status Running on Damians-MacBook.local (PID 858)	4 Þ 4 Þ
Running on Damians-MacBook.local (PID 858)	
Status Running on Damians-MacBook.local (PID 858)	
Running on Damians-MacBook.local (PID 858)	
Log output	
CC=0.5261 (Dest to keep CC=0.5292), moved from start (max/mean)= 0.468 0.231	
refining group 2	
CC=0.5202 (best to keep $CC=0.5202$ ), moved from start (max/mean)= 0.000 0.000	
CC=0.5096 (best to keep $CC=0.5202$ ), moved from start (max/mean)= 0.762 0.346	
CC=0.5089 (best to keep CC=0.5202), moved from start (max/mean)= 0.816 0.374 CC=0.5189 (best to keep CC=0.5202), moved from start (max/mean)= 0.211 0.136 refining group 3	
CC=0.5560 (best to keep $CC=0.5560$ ), moved from start (max/mean)= 0.000 0.000	
CC=0.5393 (best to keep CC=0.5560), moved from start (max/mean)= 0.950 0.446	
CC=0.5415 (best to keep CC=0.5560), moved from start (max/mean)= 0.759 0.430	
CC=0.5452 (best to keep CC=0.5560), moved from start (max/mean)= 0.441 0.366 refining group 4	
CC=0.5418 (best to keep CC=0.5418), moved from start (max/mean)= 0.000 0.000	
CC=0.5196 (best to keep CC=0.5418), moved from start (max/mean)= 0.936 0.489	
CC=0.5216 (best to keep CC=0.5418), moved from start (max/mean)= 0.806 0.472	
CC=0.5255 (best to keep CC=0.5418), moved from start (max/mean)= 0.500 0.426 refining group 5	
CC=0.5200 (best to keep CC=0.5200), moved from start (max/mean)= 0.000 0.000	
CC=0.5173 (best to keep CC=0.5200), moved from start (max/mean)= 0.529 0.208	
CC=0.5165 (best to keep CC=0.5200), moved from start (max/mean)= 0.626 0.232	
CC=0.5200 (best to keep CC=0.5200), moved from start (max/mean)= 0.041 0.018	
model-to-map fit, CC_mask: 0.8077	
moved from start: 0.3344	

<pre>CC=0.5261 (best to keep CC=0.5292), moved from start (ma refining group 2 CC=0.5202 (best to keep CC=0.5202), moved from start (ma CC=0.5096 (best to keep CC=0.5202), moved from start (ma CC=0.5189 (best to keep CC=0.5202), moved from start (ma refining group 3 CC=0.5560 (best to keep CC=0.5560), moved from start (ma CC=0.5393 (best to keep CC=0.5560), moved from start (ma CC=0.5415 (best to keep CC=0.5560), moved from start (ma CC=0.5415 (best to keep CC=0.5560), moved from start (ma CC=0.5452 (best to keep CC=0.5560), moved from start (ma refining group 4 CC=0.5418 (best to keep CC=0.5560), moved from start (ma CC=0.5196 (best to keep CC=0.5418), moved from start (ma CC=0.5216 (best to keep CC=0.5418), moved from start (ma CC=0.5255 (best to keep CC=0.5418), moved from start (ma CC=0.5255 (best to keep CC=0.5418), moved from start (ma CC=0.5200 (best to keep CC=0.5200), moved from start (ma CC=0.5173 (best to keep CC=0.5200), moved from start (ma CC=0.5165 (best to keep CC=0.5200), moved from start (ma CC=0.5165 (best to keep CC=0.5200), moved from start (ma CC=0.5200 (best to keep CC=0.5200), moved from start</pre>										
<pre>refining group 2 CC=0.5202 (best to keep CC=0.5202), moved from start (ma CC=0.5096 (best to keep CC=0.5202), moved from start (ma CC=0.5089 (best to keep CC=0.5202), moved from start (ma CC=0.5189 (best to keep CC=0.5202), moved from start (ma refining group 3 CC=0.5560 (best to keep CC=0.5560), moved from start (ma CC=0.5393 (best to keep CC=0.5560), moved from start (ma CC=0.5415 (best to keep CC=0.5560), moved from start (ma CC=0.5415 (best to keep CC=0.5560), moved from start (ma CC=0.5452 (best to keep CC=0.5560), moved from start (ma refining group 4 CC=0.5196 (best to keep CC=0.5418), moved from start (ma CC=0.5196 (best to keep CC=0.5418), moved from start (ma CC=0.5216 (best to keep CC=0.5418), moved from start (ma CC=0.5255 (best to keep CC=0.5418), moved from start (ma CC=0.5200 (best to keep CC=0.5200), moved from start (ma CC=0.5173 (best to keep CC=0.5200), moved from start (ma CC=0.5165 (best to keep CC=0.5200), moved from start (ma CC=0.5200 (best to keep CC=0.5200), moved from start (ma CC=0.5165 (best to keep CC=0.5200), moved from start (ma CC=0.5200 (best to keep CC=0.5200), moved from start</pre>	ſ	CC=0.5201	(Dest	το	кеер	CC=U.5292),	movea	ILOW	start	(ma
<pre>CC=0.5202 (best to keep CC=0.5202), moved from start (ma CC=0.5096 (best to keep CC=0.5202), moved from start (ma CC=0.5189 (best to keep CC=0.5202), moved from start (ma refining group 3 CC=0.5560 (best to keep CC=0.5560), moved from start (ma CC=0.5393 (best to keep CC=0.5560), moved from start (ma CC=0.5415 (best to keep CC=0.5560), moved from start (ma CC=0.5452 (best to keep CC=0.5560), moved from start (ma refining group 4 CC=0.5418 (best to keep CC=0.5560), moved from start (ma cC=0.5196 (best to keep CC=0.5418), moved from start (ma cC=0.5216 (best to keep CC=0.5418), moved from start (ma cC=0.5255 (best to keep CC=0.5418), moved from start (ma cC=0.5255 (best to keep CC=0.5418), moved from start (ma cC=0.5216 (best to keep CC=0.5418), moved from start (ma cC=0.5255 (best to keep CC=0.5200), moved from start (ma cC=0.5173 (best to keep CC=0.5200), moved from start (ma cC=0.5165 (best to keep CC=0.5200), moved from start (ma cC=0.5200 (b</pre>		refining	g group	p 2						
<pre>CC=0.5096 (best to keep CC=0.5202), moved from start (ma CC=0.5089 (best to keep CC=0.5202), moved from start (ma refining group 3 CC=0.5560 (best to keep CC=0.5560), moved from start (ma CC=0.5393 (best to keep CC=0.5560), moved from start (ma CC=0.5415 (best to keep CC=0.5560), moved from start (ma CC=0.5452 (best to keep CC=0.5560), moved from start (ma refining group 4 CC=0.5418 (best to keep CC=0.5560), moved from start (ma CC=0.5196 (best to keep CC=0.5418), moved from start (ma CC=0.5216 (best to keep CC=0.5418), moved from start (ma CC=0.5255 (best to keep CC=0.5418), moved from start (ma refining group 5 CC=0.5200 (best to keep CC=0.5200), moved from start (ma CC=0.5165 (best to keep CC=0.5200), moved from start (ma model-to-map fit, CC_mask: 0.8077 moved from start: 0.3344</pre>		CC=0.5202	(best	to	keep	CC=0.5202),	moved	from	start	(ma
CC=0.5089 (best to keep CC=0.5202), moved from start (ma CC=0.5189 (best to keep CC=0.5202), moved from start (ma refining group 3 CC=0.5560 (best to keep CC=0.5560), moved from start (ma CC=0.5393 (best to keep CC=0.5560), moved from start (ma CC=0.5415 (best to keep CC=0.5560), moved from start (ma CC=0.5452 (best to keep CC=0.5560), moved from start (ma refining group 4 CC=0.5418 (best to keep CC=0.5418), moved from start (ma CC=0.5216 (best to keep CC=0.5418), moved from start (ma CC=0.5216 (best to keep CC=0.5418), moved from start (ma CC=0.5255 (best to keep CC=0.5418), moved from start (ma refining group 5 CC=0.5200 (best to keep CC=0.5200), moved from start (ma CC=0.5165 (best to keep CC=0.5200), moved from start (ma CC=0.5165 (best to keep CC=0.5200), moved from start (ma CC=0.5200 (best to keep CC=0.5200), moved from start (ma model-to-map fit, CC_mask: 0.8077 moved from start: 0.3344		CC=0.5096	(best	to	keep	CC=0.5202),	moved	from	start	(ma
<pre>CC=0.5189 (best to keep CC=0.5202), moved from start (ma refining group 3 CC=0.5560 (best to keep CC=0.5560), moved from start (ma CC=0.5393 (best to keep CC=0.5560), moved from start (ma CC=0.5415 (best to keep CC=0.5560), moved from start (ma refining group 4 CC=0.5452 (best to keep CC=0.5560), moved from start (ma refining group 4 CC=0.5418 (best to keep CC=0.5418), moved from start (ma CC=0.5196 (best to keep CC=0.5418), moved from start (ma CC=0.5216 (best to keep CC=0.5418), moved from start (ma CC=0.5255 (best to keep CC=0.5418), moved from start (ma refining group 5 CC=0.5200 (best to keep CC=0.5200), moved from start (ma CC=0.5173 (best to keep CC=0.5200), moved from start (ma CC=0.5165 (best to keep CC=0.5200), moved from start (ma CC=0.5200 (bes CC=0.5200), moved fr</pre>		CC=0.5089	(best	to	keep	CC=0.5202),	moved	from	start	(ma
<pre>refining group 3 CC=0.5560 (best to keep CC=0.5560), moved from start (ma CC=0.5393 (best to keep CC=0.5560), moved from start (ma CC=0.5415 (best to keep CC=0.5560), moved from start (ma CC=0.5452 (best to keep CC=0.5560), moved from start (ma refining group 4 CC=0.5418 (best to keep CC=0.5418), moved from start (ma CC=0.5196 (best to keep CC=0.5418), moved from start (ma CC=0.5216 (best to keep CC=0.5418), moved from start (ma CC=0.5255 (best to keep CC=0.5418), moved from start (ma refining group 5 CC=0.5200 (best to keep CC=0.5200), moved from start (ma CC=0.5173 (best to keep CC=0.5200), moved from start (ma CC=0.5165 (best to keep CC=0.5200), moved from start (ma CC=0.5200 (best to keep CC=0.5200), moved from start</pre>		CC=0.5189	(best	to	keep	CC=0.5202),	moved	from	start	(ma
<pre>CC=0.5560 (best to keep CC=0.5560), moved from start (ma CC=0.5393 (best to keep CC=0.5560), moved from start (ma CC=0.5415 (best to keep CC=0.5560), moved from start (ma CC=0.5452 (best to keep CC=0.5560), moved from start (ma refining group 4 CC=0.5418 (best to keep CC=0.5418), moved from start (ma CC=0.5196 (best to keep CC=0.5418), moved from start (ma CC=0.5216 (best to keep CC=0.5418), moved from start (ma CC=0.5255 (best to keep CC=0.5418), moved from start (ma refining group 5 CC=0.5200 (best to keep CC=0.5200), moved from start (ma CC=0.5173 (best to keep CC=0.5200), moved from start (ma CC=0.5165 (best to keep CC=0.5200), moved from start (ma CC=0.5200 (b</pre>		refining	g group	<b>3</b>						
<pre>CC=0.5393 (best to keep CC=0.5560), moved from start (ma CC=0.5415 (best to keep CC=0.5560), moved from start (ma refining group 4 CC=0.5452 (best to keep CC=0.5560), moved from start (ma CC=0.5418 (best to keep CC=0.5418), moved from start (ma CC=0.5196 (best to keep CC=0.5418), moved from start (ma CC=0.5216 (best to keep CC=0.5418), moved from start (ma CC=0.5255 (best to keep CC=0.5418), moved from start (ma refining group 5 CC=0.5200 (best to keep CC=0.5200), moved from start (ma CC=0.5173 (best to keep CC=0.5200), moved from start (ma CC=0.5165 (best to keep CC=0.5200), moved from start (ma CC=0.5200 (b</pre>		CC=0.5560	(best	to	keep	CC=0.5560),	moved	from	start	(ma
<pre>CC=0.5415 (best to keep CC=0.5560), moved from start (ma CC=0.5452 (best to keep CC=0.5560), moved from start (ma refining group 4 CC=0.5418 (best to keep CC=0.5418), moved from start (ma CC=0.5196 (best to keep CC=0.5418), moved from start (ma CC=0.5216 (best to keep CC=0.5418), moved from start (ma CC=0.5255 (best to keep CC=0.5418), moved from start (ma refining group 5 CC=0.5200 (best to keep CC=0.5200), moved from start (ma CC=0.5173 (best to keep CC=0.5200), moved from start (ma CC=0.5165 (best to keep CC=0.5200), moved from start (ma CC=0.5200 (b</pre>		CC=0.5393	(best	to	keep	CC=0.5560),	moved	from	start	(ma
<pre>CC=0.5452 (best to keep CC=0.5560), moved from start (matrix refining group 4 CC=0.5418 (best to keep CC=0.5418), moved from start (matrix CC=0.5196 (best to keep CC=0.5418), moved from start (matrix CC=0.5216 (best to keep CC=0.5418), moved from start (matrix refining group 5 CC=0.5255 (best to keep CC=0.5200), moved from start (matrix CC=0.5173 (best to keep CC=0.5200), moved from start (matrix CC=0.5165 (best to keep CC=0.5200), moved from start (matrix CC=0.5200 (best to keep CC=0.5200), moved from start (matrix CC=0.5165 (best to keep CC=0.5200), moved from start (matrix CC=0.5200 (best CC=0.5200), moved from start (matrix CC</pre>		CC=0.5415	(best	to	keep	CC=0.5560),	moved	from	start	(ma
<pre>refining group 4 CC=0.5418 (best to keep CC=0.5418), moved from start (ma CC=0.5196 (best to keep CC=0.5418), moved from start (ma CC=0.5216 (best to keep CC=0.5418), moved from start (ma CC=0.5255 (best to keep CC=0.5418), moved from start (ma refining group 5 CC=0.5200 (best to keep CC=0.5200), moved from start (ma CC=0.5173 (best to keep CC=0.5200), moved from start (ma CC=0.5165 (best to keep CC=0.5200), moved from start (ma CC=0.5200 (best to keep CC=0.5200), moved from start (ma model-to-map fit, CC_mask: 0.8077 moved from start: 0.3344</pre>		CC=0.5452	(best	to	keep	CC=0.5560),	moved	from	start	(ma
<pre>CC=0.5418 (best to keep CC=0.5418), moved from start (ma CC=0.5196 (best to keep CC=0.5418), moved from start (ma CC=0.5216 (best to keep CC=0.5418), moved from start (ma CC=0.5255 (best to keep CC=0.5418), moved from start (ma refining group 5 CC=0.5200 (best to keep CC=0.5200), moved from start (ma CC=0.5173 (best to keep CC=0.5200), moved from start (ma CC=0.5165 (best to keep CC=0.5200), moved from start (ma CC=0.5200 (best to keep CC=0.5200), moved from start (ma model-to-map fit, CC_mask: 0.8077 moved from start: 0.3344</pre>		refining	g group	<b>4</b>						
<pre>CC=0.5196 (best to keep CC=0.5418), moved from start (ma CC=0.5216 (best to keep CC=0.5418), moved from start (ma CC=0.5255 (best to keep CC=0.5418), moved from start (ma refining group 5 CC=0.5200 (best to keep CC=0.5200), moved from start (ma CC=0.5173 (best to keep CC=0.5200), moved from start (ma CC=0.5165 (best to keep CC=0.5200), moved from start (ma CC=0.5200 (best to keep CC=0.5200), moved from start (ma model-to-map fit, CC_mask: 0.8077 moved from start: 0.3344</pre>		CC=0.5418	(best	to	keep	CC=0.5418),	moved	from	start	(ma
<pre>CC=0.5216 (best to keep CC=0.5418), moved from start (ma CC=0.5255 (best to keep CC=0.5418), moved from start (ma refining group 5 CC=0.5200 (best to keep CC=0.5200), moved from start (ma CC=0.5173 (best to keep CC=0.5200), moved from start (ma CC=0.5165 (best to keep CC=0.5200), moved from start (ma CC=0.5200 (best to keep CC=0.5200), moved from start (ma model-to-map fit, CC_mask: 0.8077 moved from start: 0.3344</pre>		CC=0.5196	(best	to	keep	CC=0.5418),	moved	from	start	(ma
<pre>CC=0.5255 (best to keep CC=0.5418), moved from start (ma refining group 5 CC=0.5200 (best to keep CC=0.5200), moved from start (ma CC=0.5173 (best to keep CC=0.5200), moved from start (ma CC=0.5165 (best to keep CC=0.5200), moved from start (ma CC=0.5200 (best to keep CC=0.5200), moved from start (ma model-to-map fit, CC_mask: 0.8077 moved from start: 0.3344</pre>		CC=0.5216	(best	to	keep	CC=0.5418),	moved	from	start	(ma
<pre>refining group 5 CC=0.5200 (best to keep CC=0.5200), moved from start (ma CC=0.5173 (best to keep CC=0.5200), moved from start (ma CC=0.5165 (best to keep CC=0.5200), moved from start (ma CC=0.5200 (best to keep CC=0.5200), moved from start (ma model-to-map fit, CC_mask: 0.8077 moved from start: 0.3344</pre>		CC=0.5255	(best	to	keep	CC=0.5418),	moved	from	start	(ma
<pre>CC=0.5200 (best to keep CC=0.5200), moved from start (ma CC=0.5173 (best to keep CC=0.5200), moved from start (ma CC=0.5165 (best to keep CC=0.5200), moved from start (ma CC=0.5200 (best to keep CC=0.5200), moved from start (ma model-to-map fit, CC_mask: 0.8077 moved from start: 0.3344</pre>		refining	g group	5 5						
CC=0.5173 (best to keep CC=0.5200), moved from start (ma CC=0.5165 (best to keep CC=0.5200), moved from start (ma CC=0.5200 (best to keep CC=0.5200), moved from start (ma model-to-map fit, CC_mask: 0.8077 moved from start: 0.3344		CC=0.5200	(best	to	keep	CC=0.5200),	moved	from	start	(ma
CC=0.5165 (best to keep CC=0.5200), moved from start (ma CC=0.5200 (best to keep CC=0.5200), moved from start (ma model-to-map fit, CC_mask: 0.8077 moved from start: 0.3344		CC=0.5173	(best	to	keep	CC=0.5200),	moved	from	start	(ma
CC=0.5200 (best to keep CC=0.5200), moved from start (ma model-to-map fit, CC_mask: 0.8077 moved from start: 0.3344		CC=0.5165	(best	to	keep	CC=0.5200),	moved	from	start	(ma
<pre>model-to-map fit, CC_mask: 0.8077 moved from start: 0.3344</pre>		CC=0.5200	(best	to	keep	CC=0.5200),	moved	from	start	(ma
moved from start: 0.3344		model-to	-map 1	Eit	, CC r	nask: 0.8077				10 N
		moved fr	om sta	art	19	0.3344				
		20 * Ad 24 16	1954 - 1760 - 179							

![](_page_32_Picture_5.jpeg)

1 job(s) running

• •	Real-space refinement (Project: tmp)	
🛛 🖉 🔊 🔊 🖓 🖬		
Preferences Help Run Abort Save	Ask for help	
Input/Output Refinement Settings RealSpace	Refine_4	
Run status Results Validation		
Files		
Output Directory /Users/deakiert/Decum	ants /tmp / PoolSpace Pofing 4	
Output Directory. /Osers/ucekient/Docume	ents/timp/kealspacekenne_4	
Filename	Directory	
long_ring2to4hexa-r09-coot-161.pdb	/Users/dcekiert/Synch/PDBs/YebT_20180804/ring24_long	
ring24_long_m24C6130_cl2_locfilt.mrc	/Users/dcekiert/Synch/PDBs/Yeb1_20180804/ring24_long	
run1_real_space_refined.log	/Users/dcekiert/Documents/tmp/RealSpaceRefine_4	
run1_real_space_refined.clf	/Users/dcekiert/Documents/tmp/RealSpaceRefine_4	
run1_real_space_renned.pdb	/Users/dcekiert/Documents/tmp/RealSpaceRefine_4	
run1_initial.geo	/Users/dcekiert/Documents/tmp/RealSpaceRenne_4	
Open in Coot		

O Idle

Project: tmp

## What does the output look like?

Summary MolProbity Model vs. I	Data Data					٩	Þ
Composition (#)		Box					
Chains	6	Lengths (Å)		113.97, 103.49, 120.52			
Atoms	15942 (Hydrogens: 0)	Angles (°)		90.00, 90.00, 90.00			
Residues	Protein: 2106 Nucleotide: 0	Supplied Resolution (Å	<b>(</b> )	3.0			
Water	0	Resolution Estimates (	Å)	Masked	Unmasked		
Ligands	0	d FSC (half maps;	0.143)				
Bonds (RMSD)		d 99 (full/half1/h	nalf2)	3.4//	3.4//		
Length (Å) (# > 4σ)	0.002 (0)	d model		3.3	3.4		
Angles (°) (# > 4σ)	0.573 (0)	d FSC model (0/0	.143/0.5)	3.0/3.1/3.4	3.1/3.2/3.4		
MolProbity score	2.26	Map min/max/mean		-0.29/0.53/0.00			1
Clash score	7.67						
Ramachandran plot (%)		Model vs. Data					
Outliers	0.00	CC (mask)	0.78				
Allowed	4.30	CC (box)	0.69				
Favored	95.70	CC (peaks)	0.64				
Rotamer outliers (%)	5.12	CC (volume)	0.78				
Cβ outliers (%)	0.00	Mean CC for ligands					
Peptide plane (%)							
Cis proline/general	0.0/0.0						
Twisted proline/general	0.0/0.0						
CaBLAM outliers (%)	2.02						
ADP (B-factors)							
lso/Aniso (#)	15942/0						
min/max/mean							
Protein	60.42/162.58/90.50						
Nucleotide							
Ligand							
Water							
Occupancy							
Mean	1.00						
occ = 1 (%)	100.00						
0 < occ < 1 (%)	0.00						2
occ > 1 (%)	0.00						

😑 Idle

# Goal of model validation:

# 1) To assess refinement strategies and progress 3) To assess overall and local model quality/reliability

- 2) To identify problem areas requiring manual intervention

"Self-assessment": We want to create the most accurate and reliable model we can, and validation stats clue us in to regions of the mode that may have issues

## Model validation metrics - By Problem type

- Overall Quality Indicators
  - Model/Map CC
  - RMS deviations
  - Unmodeled densities
  - Molprobity score and clash score
- Backbone issues
  - Ramachandran plot
  - Cis peptide bonds
- Side chain issues:
  - Rotamer outliers (EM ringer?)
  - Cbeta deviations
- B factor / ADP outliers

## Model validation metrics - By source of problem

- Model building problem
  - Unmodeled densities
  - Cbeta deviations
  - Ramachandran plot
  - Cis peptide bonds
  - Model/Map CC
- Refinement problem
  - RMS deviations
  - B factor / ADP outliers
- Either/both?
  - Molprobity score and clash score
  - Rotamer outliers (EM ringer?)

## Model/Map CC

### Table 3 Summary of map correlation coefficients used in this work.

Metric	Region of the map used in calculation
CChor	Whole map
CCmask	Jiang & Brünger (1994) mask with a fixed radius
CC <sub>volume</sub>	Mask of points with the highest values in the mod
CC <sub>peaks</sub>	Mask of points with the highest values in the mode target maps
CC <sub>vr_mask</sub>	Same as $CC_{mask}$ but atomic radii are variable and resolution, atom type and ADP

Afonine, et al. "New tools for the analysis and validation of cryo-EM maps and atomic models" Acta Cryst. 2018

	Purpose
	Similarity of maps
	Fit of the atomic centers
del map	Fit of the molecular envelope defined by the model
el and in the	Fit of the strongest peaks in the model and target r
function of	Fit of the atomic images in the given map

![](_page_38_Picture_6.jpeg)

## Root mean square (RMS) deviations

**Typical RMS Bonds for protein structure:** 0.005 - 0.015 Å

![](_page_39_Figure_3.jpeg)

Image: Janez Stepisnik

## **Covalent bond lengths and angles exhibit known, narrow distributions**

### **Typical RMS Angles for protein structure:** 0.5 - 1.5 Å

![](_page_39_Figure_7.jpeg)

## **Root mean square (RMS) deviations**

**Typical RMS Bonds for protein structure:** 0.005 - 0.015 Å

REMARK	3					
REMARK	3	DEVIATIONS	FROM	IDEAL	VALUES	•
REMARK	3			RMSD	MAX	COUNT
REMARK	3	BOND	: 0	.010	0.098	16212
REMARK	3	ANGLE	: 1	.328	19.912	21996
REMARK	3	CHIRALITY	: 0	.074	0.294	2586
REMARK	3	PLANARITY	: 0	.009	0.061	2856
REMARK	3	DIHEDRAL	: 9	.311 1	79.980	9852
REMARK	3	MIN NONBOR	NDED	DISTAN	CE: 2.	433
REMARK	3					

### **Covalent bond lengths and angles exhibit known, narrow distributions**

**Typical RMS Angles for protein structure:** 0.5 - 1.5 Å

# **Unmodeled densities**

	X Coot 0.8.9.2-pre EL (revision count 7766)
<u>File Edit Calculate Draw Measures</u>	<u>Validate</u> HID About Ligand E <u>x</u> tensions
🖹 🔍 Reset View 📃 Display Manager	<ul> <li>Ramachandran Plot</li> <li>Kleywegt Plot</li> <li>Incorrect Chiral Volumes</li> <li>Unmodelled blobs</li> <li>Difference Man Deeks</li> </ul>
	<ul> <li>Check/Delete Waters</li> <li>Geometry analysis</li> <li>Peptide omega analysis</li> <li>Temp. fact. variance analysis</li> </ul>
	<ul> <li>Average Temp. fact. analysis</li> <li>GLN and ASN B-factor Outliers</li> <li>Rotamer analysis</li> <li>Density fit analysis</li> <li>Probe clashes</li> <li>NCS Differences</li> </ul>
	Highly coordinated waters Pukka Puckers? Alignment vs PIR
(mol. por 3) CA (1/A/160 ASP occ. 1.00 h	of: 83.75 ele: C pos: (166.20.151.54.102.21)

5)	Unmodelled blobs of density:
ons	There are unexplained blobs of density (too big to be waters):
	Blob 1
	Blob 2
	Blob 3
	Blob 4
	Blob 5
	Blob 6
KARAKA 7	Blob 7
	Blob 8
	Blob 9
	Blob 10
9	Blob 11
	Blob 12
	Blob 13
	Blob 14
	Blob 15
	Blob 16
- HSBA	Blob 17
	Blob 18
	Blob 19
	Blob 20
	Blob 21
	<u>۱</u>
31)	Dismiss

## **Unmodeled densities**

000	X Coot	0.8.9.2-pre EL (revision co	ount 7766)		000	X Unmodelled blobs of density:		
<u>F</u> ile <u>E</u> dit <u>C</u> alculate <u>D</u> raw	<u>M</u> easures <u>V</u> alidate	HID About Ligand B	E <u>x</u> tensions			There are unexplained blobs of		
📙 🔍 Reset View 📃 Displa	ay Manager 🍛 🗞 🗍	Sphere Refine +				density (too big to be waters):		
	CALDY-	<b>ISKAS</b>				Blob 1		-
Ey		HS LACE	X			Blob 2		
	🔀 Fin	d Unmodelled Blobs of der	sity			Blob 3		
• 4 /Lisers/dcekier	rt/Synch/PDBs/VehT	Select Map: 20180804/ripg24_lopg/	ring24 long m24C6T	30 cl2 locfilt mrc		Blob 4		
		20100004/111924_10119/1				Blob 5		
O 0 /Licerc /deal/io	rt/Ourseh/DDDsA/ahT (	Select Protein Model:	IIClass boys r10 see	t 227 now2 ndh		Blob 6		
O 1 /Users/dcekier	rt/Synch/PDBs/Yeb1_/	exameric w-grafted-12	5loop ABCDEF.pdb	L-227_new2.pdb		Blob 7		
O 2 /Users/dcekier	rt/xtal/bex2628_Pr130	7/refinement/phenix1/b	ex2628_phenix1_001	1.pdb		Blob 8		
● 3 /Users/dcekier	rt/Synch/PDBs/YebT_:	20180804/ring24_long/l	long_ring2to4hexa-r0	9-coot-161.pdb		Blob 9		
		Find Blobs above:				Blob 10		
	r.m.s.a.					Blob 11		
			Find Blobs	😹 <u>C</u> ancel		Blob 12		8
	KT/XV		ZZA XA KSAMIA	1.		Blob 13		
				A A		Blob 14		
KA II				4		Blob 15		
		- Tend	X H A T	J		Blob 16		
			W/AKK	A		Blob 17		
						Blob 18		
			A			Blob 19		
	AR					Blob 20		
						Blob 21		•
					4		Þ	
	and 1 00 bf. 00 75 -	a. C pac. /100 00 151	F 4 100 01)			C	Dismiss	

![](_page_43_Picture_0.jpeg)

## **Unmodeled densities**

Unmodelled blobs of density: There are unexplained blobs of density (too big to be waters): Blob 1 Blob 2 Blob 3 Blob 4 Blob 5 Blob 6 Blob 7 Blob 8 Blob 9 Blob 10 Blob 11 Blob 12 \* Blob 13 Blob 14 Blob 15 Blob 16 Blob 17 Blob 18 Blob 19 Blob 20 Blob 21 4 Dismiss

# Molprobity score and clash score

### **Summary statistics**

All-Atom	Clashscore, all atoms:	3.82		96 <sup>th</sup> percentile <sup>*</sup> (N=1784, all resolutions)		
Contacts	Clashscore is the number of serious steric of	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.				
	Poor rotamers	12	0.68%	Goal: <0.3%		
	Favored rotamers	1674	95.22%	Goal: >98%		
	Ramachandran outliers	0	0.00%	Goal: <0.05%		
Protein	Ramachandran favored	2070	98.85%	Goal: >98%		
Geometry	MolProbity score <sup>^</sup>	1.17		99 <sup>th</sup> percentile <sup>*</sup> (N=27675, 0Å - 99Å)		
	Cβ deviations >0.25Å	6	0.32%	Goal: 0		
	Bad bonds:	0 / 16212	0.00%	Goal: 0%		
	Bad angles:	24 / 21996	0.11%	Goal: <0.1%		
Peptide Omegas	Cis Prolines:	0/114	0.00%	Expected: $\leq 1$ per chain, or $\leq 5\%$		
Low-resolution Criteria	CaBLAM outliers	42	2.0%	Goal: <1.0%		
	CA Geometry outliers	24	1.15%	Goal: <0.5%		
Additional validations	Pseudochiral naming errors	6				
	Waters with clashes	0/0	0.00%	See UnDowser table for details		

In the two column results, the left column gives the raw count, right column gives the percentage. \* 100<sup>th</sup> percentile is the best among structures of comparable resolution; 0<sup>th</sup> percentile is the worst. For clashscore the comparative set of structures was selected in 2004, for MolProbity score in 2006. ^ MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.

### <u>http://molprobity.biochem.duke.edu/</u> Williams et al. Protein Science (2018).

100
_

## Ramachandran plot

000		X Coot	0.8.9.	2-pre EL	(revision	count 776
<u>F</u> ile <u>E</u> dit <u>C</u> alculate <u>D</u> raw	<u>M</u> easures	<u>V</u> alidate	HID	About	Ligand	E <u>x</u> tensi
Enc Lat Galoarato Later Reset View Displa	y Manager	<ul> <li>Rama</li> <li>Kleyw</li> <li>Incorr</li> <li>Incorr</li> <li>Unmo</li> <li>Differe</li> <li>Check</li> <li>Geom</li> <li>Geom</li> <li>Feptic</li> <li>Averation</li> <li>Averation</li> <li>Rotant</li> <li>Rotant</li> <li>Probe</li> <li>NCS D</li> <li>Highly</li> <li>Pukkat</li> </ul>	chance egt P ect C delle ence (Dele ence de orr fact, ge Te nd As ner ar ty fit a clash Differe coor Puck	dran Plo lot hiral Volu d blobs. Map Pea ete Wate analysis nega ana variance mp. fact SN B-fac nalysis analysis nes ences dinated kers?	t umes aks aks alysis e analysis tor Outlie tor Outlie	s s ers
		Angrin				

Successfully read coordinates file /Lisers/deakiert/Synch/DDRs/VehT 20180804/ring24\_long/long\_ring2to/heya\_r

![](_page_45_Figure_3.jpeg)

## Ramachandran plot

![](_page_46_Figure_1.jpeg)

# Ramachandran plot

![](_page_47_Figure_1.jpeg)

### Allowed Regions (non-Pro/Gly)

![](_page_47_Figure_4.jpeg)

![](_page_48_Picture_1.jpeg)

### Figure 2.20 Biochemistry, Seventh Edition © 2012 W. H. Freeman and Company

![](_page_48_Picture_3.jpeg)

![](_page_49_Picture_1.jpeg)

## Trans

### Figure 2.21 Biochemistry, Seventh Edition © 2012 W. H. Freeman and Company

![](_page_49_Picture_4.jpeg)

![](_page_50_Picture_1.jpeg)

Coot highlights all cis and non-planar peptide bonds, and color codes them to make potential problems easy to ID Green = cis-Proline (probably OK); Yellow = non-planar peptide bond (check!); Red = non-proline cis peptide bond (check!)

![](_page_50_Picture_3.jpeg)

![](_page_51_Picture_1.jpeg)

Coot highlights all cis and non-planar peptide bonds, and color codes them to make potential problems easy to ID Green = cis-Proline (probably OK); Yellow = non-planar peptide bond (check!); Red = non-proline cis peptide bond (check!)

![](_page_51_Picture_3.jpeg)

![](_page_52_Picture_1.jpeg)

https://www.phenix-online.org/documentation/tutorials/molprobity.html

## Rotamer outliers: Coot

Elle       Edit       Calculate       Draw       Measures       Validate       HID       About       Ligand       Extensions         Image: Calculate       Display Manager       Image: Calculate       Ramachandran Plot       Image: Calculate       Image: Calcula	Elle Edit Calculate Draw Measures Validate HID About Ligand Extensions	000	🔀 Coot 0.8.9.2-pre EL (revision count 7766)
Reset View       Display Manager         Kleywegt Plot       Incorrect Chiral Volumes         Difference Map Peaks       Difference Map Peaks         Check/Delete Waters       Emp. fact. variance analysis         Femp. fact. variance analysis       Emp. fact. analysis         Rotamer analysis       Oring2to4hexa-r09-coot-1         Probe clashes       NCS Differences         NCS Differences       Highly coordinated waters         Publika Puckkars?       Alignment vs PIR	Reset View Display Manager     Kleywegt Plot     Uncorrect Chiral Volumes     Difference Map Peaks     Check/Delete Waters     Mediate analysis     Temp. fact. variance analysis     Average Temp. fact. analysis     Pobe clashes   NCS Differences   Highly coordinated waters   Pickerv?     Alignment vs PIR     Alignment vs PIR	<u>F</u> ile <u>E</u> dit <u>C</u> alculate	<u>⊃</u> raw <u>M</u> easures <u>V</u> alidate HID About Ligand E <u>x</u> tensions
Image: Solution of the second sec	Kleywegt Plot Unmodelled blobs Difference Map Peaks Check/Delete Waters Hendied omega analysis Temp. fact. variance analysis Kotamer analysis Pobe clashes NCS Differences Highly coordinated waters NCS Differences Highly coordinated waters Alignment vs PIR	📄 🔍 Reset View 目	Display Manager 🚺 Ramachandran Plot 🔹 🕨
		Reset View	Display Manager   Display Manager Ramachandran Plot   Kleywegt Plot   Incorrect Chiral Volumes   Unmodelled blobs   Difference Map Peaks   Check/Delete Waters   Check/Delete Waters   Accometry analysis   Peptide omega analysis   Temp, fact. variance analysis   Average Temp. fact. analysis   Average Temp. fact. analysis   Chosh NB-Factor Outliers   Rotamer analysis   Pobe clashes   NCS Differences   Highly coordinated waters   Pukka Puckers?   Alignment vs PIR

Successfully read coordinates file /Users/dcekiert/xtal/bex2669\_Pr1307/refinement\_20191126/phenix2/bex2669\_phenix2\_001....

![](_page_53_Figure_3.jpeg)

Green/Short - Happy rotamer Lilac(?): Missing side chain atoms

# Rotamer outliers: EM ringer

![](_page_54_Picture_1.jpeg)

Fraser, et al. Nature 2009. Barad, et al. Nature Methods 2015.

![](_page_54_Figure_3.jpeg)

![](_page_54_Picture_4.jpeg)

## **Cbeta deviations**

![](_page_55_Picture_1.jpeg)

### Which Leu rotamer is correct???

![](_page_55_Picture_3.jpeg)

## **Cbeta deviations**

![](_page_56_Picture_1.jpeg)

![](_page_56_Figure_2.jpeg)

![](_page_56_Picture_3.jpeg)

## **Cbeta deviations**

![](_page_57_Picture_1.jpeg)

## Which Leu rotamer is correct???

## **Correct answer:** YELLOW

![](_page_57_Picture_4.jpeg)

Summary MolProbity Model vs. D	ata Data					4
Composition (#)		Box				
Chains	6	Lengths (Å)		113.97, 103.49, 120.52		
Atoms	15942 (Hydrogens: 0)	Angles (°)		90.00, 90.00, 90.00		
Residues	Protein: 2106 Nucleotide: 0	Supplied Resolution (Å	)	3.0		
Water	0	Resolution Estimates (/	Å)	Masked	Unmasked	
Ligands	0	d FSC (half maps;	0.143)			
Bonds (RMSD)		d 99 (full/half1/h	alf2)	3.4//	3.4//	
Length (Å) (# > 4σ)	0.002 (0)	d model		3.3	3.4	
Angles (°) ( $\# > 4\sigma$ )	0.573 (0)	d FSC model (0/0	143/0.5)	3.0/3.1/3.4	3.1/3.2/3.4	
MolProbity score	2.26	Map min/max/mean		-0.29/0.53/0.00		
Clash score	7.67					
Ramachandran plot (%)		Model vs. Data				
Outliers	0.00	CC (mask)	0.78			
Allowed	4.30	CC (box)	0.69			
Favored	95.70	CC (peaks)	0.64			
Rotamer outliers (%)	5.12	CC (volume)	0.78			
Cβ outliers (%)	0.00	Mean CC for ligands	1 <del>00</del>			
Peptide plane (%)						
Cis proline/general	0.0/0.0					
Twisted proline/general	0.0/0.0					
CaBLAM outliers (%)	2.02					
ADP (B-factors)						
lso/Aniso (#)	15942/0					
min/max/mean						
Protein	60.42/162.58/90.50					
Nucleotide						
Ligand						
Water						
Occupancy		•				
Mean	1.00					
occ = 1 (%)	100.00					
0 < occ < 1 (%)	0.00					
occ > 1 (%)	0.00					

😑 Idle

## **ADP outliers**

![](_page_59_Picture_0.jpeg)

### Full wwPDB/EMDataBank EM Map/Model Validation Report (i)

Dec 2, 2019 – 05:08 PM EST

PDB ID	:	6Q04
EMDB ID:		EMD-20542
Title	•	MERS-CoV S a
Authors	:	Park, Y.J.; Wa
		B.J.; DiMaio, H
		tious Disease (S
Deposited on	•	2019-08-01
Resolution	:	2.50 Å(reported
ed on PDB ID	:	6BN3

Based

This is a Full wwPDB/EMDataBank EM Map/Model Validation Report for a publicly released PDB/EMDB entry.

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

MolProbity:4.02b-467Mogul:1.8.0 (224370), CSD as540be (2019)Percentile statistics:20171227.v01 (using entries in the PDB archive December 27th 2017)Ideal geometry (proteins):Engh & Huber (2001)Ideal geometry (DNA, RNA):Parkinson et. al. (1996)

Validation Pipeline (wwPDB-VP) : 2.4

structure in complex with 5-N-acetyl neuraminic acid alls, A.C.; Wang, Z.; Sauer, M.; Li, W.; Tortorici, M.A.; Bosch, F.D.; Veesler, D.; Seattle Structural Genomics Center for Infec-(SSGCID)

d)

### Overall quality at a glance (i) 1

The following experimental techniques were used to determine the structure: ELECTRON MICROSCOPY

The reported resolution of this entry is 2.50 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

![](_page_60_Figure_5.jpeg)

Metric	$\begin{array}{c} \textbf{Whole archive} \\ \textbf{(\#Entries)} \end{array}$	${f EM} {f structures} \ (\#{f Entries})$
Clashscore	136327	1886
Ramachandran outliers	132723	1663
Sidechain outliers	132532	1531

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5%

Mol	Chain	Length	
1	A	1359	
1	В	1359	
1	С	1359	

![](_page_60_Figure_12.jpeg)

## More Resources

- http://molprobity.biochem.duke.edu/
- https://www.phenix-online.org/documentation/index.html
- https://www.ccpem.ac.uk/
- https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
- Afonine, et al. "New tools for the analysis and validation of cryo-EM maps and atomic models" Acta Cryst. 2018
- Wang, et al. "Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta" eLife 2016
- Nicholls, et al. "Current approaches for the fitting and refinement of atomic models into cryo-EM maps using CCP-EM" Acta Cryst. 2018