



# wwPDB X-ray Structure Validation Summary Report

Feb 28, 2014 – 05:51 AM GMT

PDB ID : 2GP1  
Title : Bacteriophage HK97 Prohead II crystal structure  
Authors : Gertsman, I.; Gan, L.; Johnson, J.E.  
Deposited on : 2006-04-15  
Resolution : 5.20 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

---

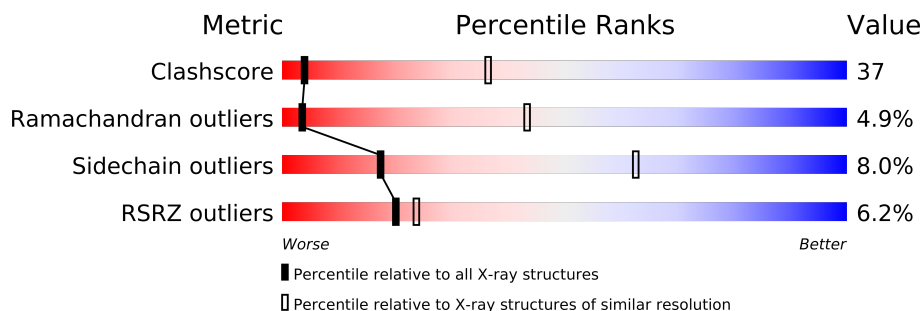
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 5.20 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.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	79885	1336 (6.80-3.50)
Ramachandran outliers	78287	1250 (6.80-3.50)
Sidechain outliers	78261	1228 (6.80-3.50)
RSRZ outliers	66119	1059 (6.80-3.50)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	282	
1	B	282	
1	C	282	
1	D	282	
1	E	282	
1	F	282	
1	G	282	

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 13797 atoms, of which 0 are hydrogen and 0 are deuterium.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Major capsid protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	254	Total	C	N	O	S	0	0	0
			1971	1233	345	383	10			
1	B	254	Total	C	N	O	S	0	0	0
			1971	1233	345	383	10			
1	C	254	Total	C	N	O	S	0	0	0
			1971	1233	345	383	10			
1	D	254	Total	C	N	O	S	0	0	0
			1971	1233	345	383	10			
1	E	254	Total	C	N	O	S	0	0	0
			1971	1233	345	383	10			
1	F	254	Total	C	N	O	S	0	0	0
			1971	1233	345	383	10			
1	G	254	Total	C	N	O	S	0	0	0
			1971	1233	345	383	10			

There are 7 discrepancies between the modelled and reference sequences:

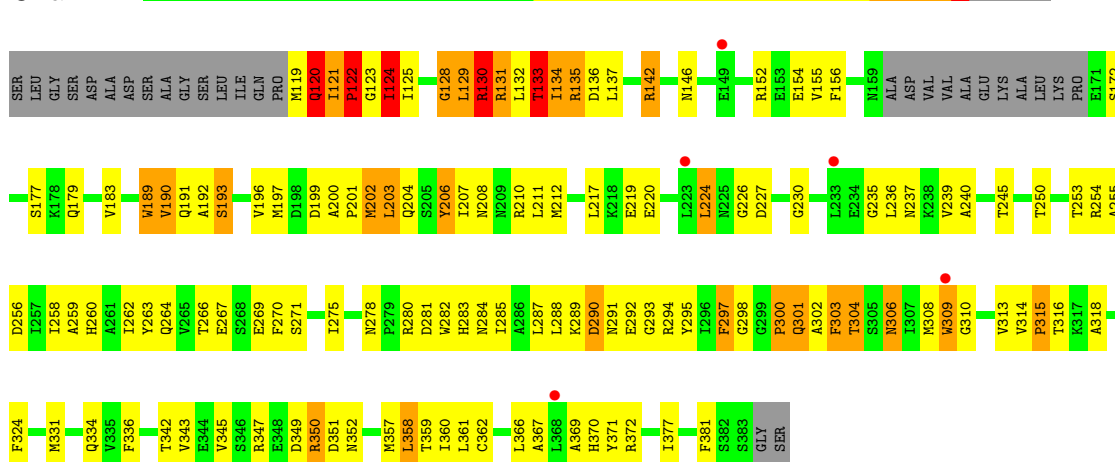
Chain	Residue	Modelled	Actual	Comment	Reference
A	336	PHE	TRP	ENGINEERED	UNP P49861
B	336	PHE	TRP	ENGINEERED	UNP P49861
C	336	PHE	TRP	ENGINEERED	UNP P49861
D	336	PHE	TRP	ENGINEERED	UNP P49861
E	336	PHE	TRP	ENGINEERED	UNP P49861
F	336	PHE	TRP	ENGINEERED	UNP P49861
G	336	PHE	TRP	ENGINEERED	UNP P49861

### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

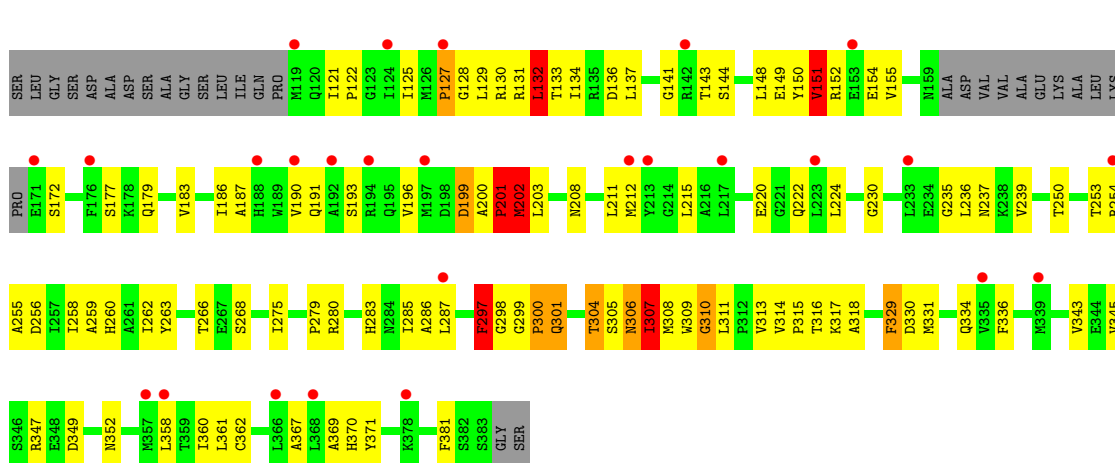
- Molecule 1: Major capsid protein

Chain A:



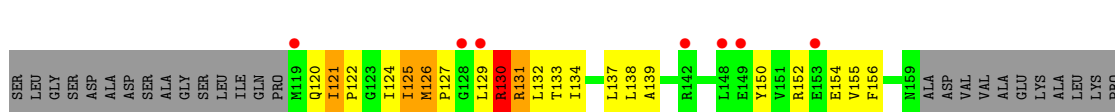
- Molecule 1: Major capsid protein

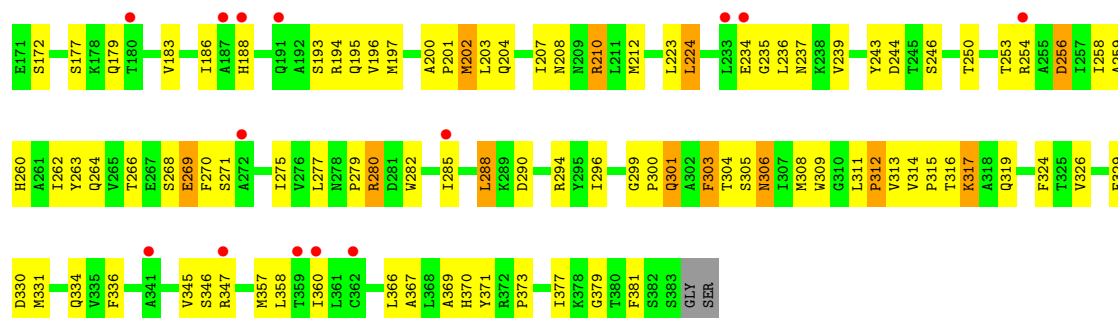
Chain B:



- Molecule 1: Major capsid protein

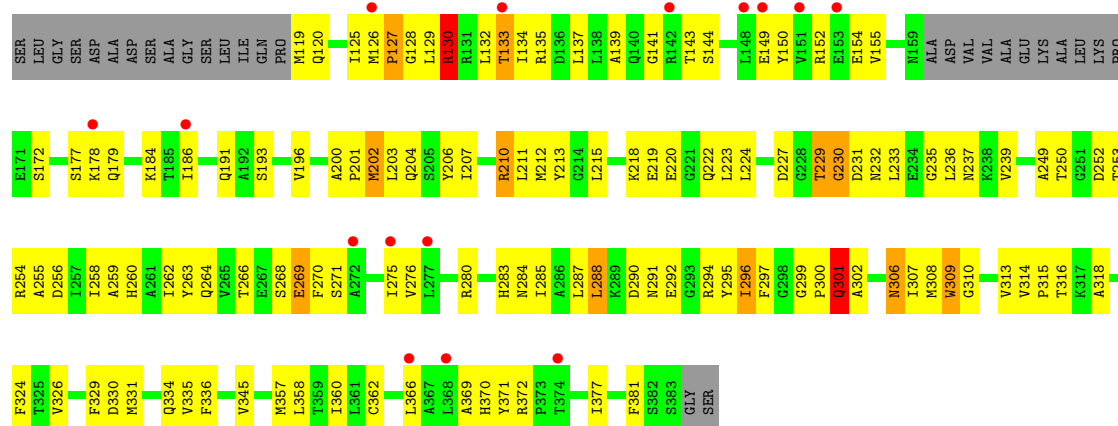
Chain C:





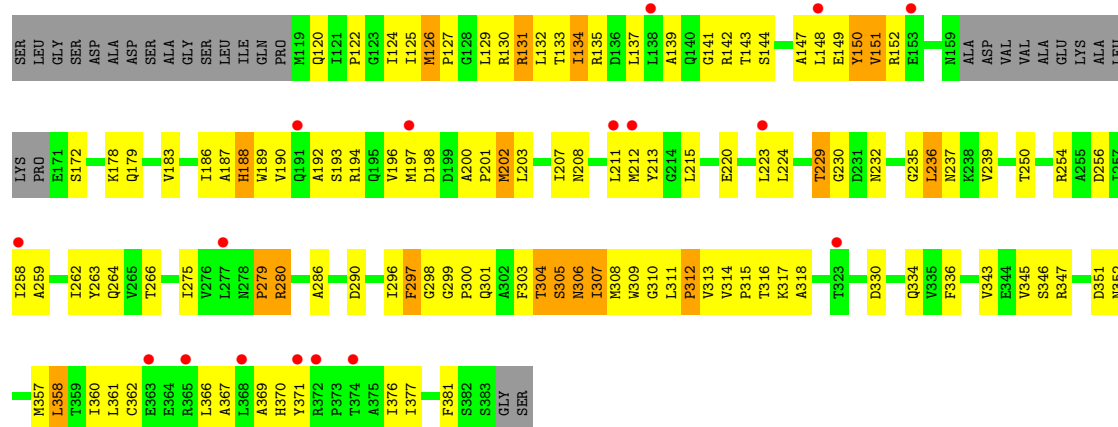
• Molecule 1: Major capsid protein

Chain D:



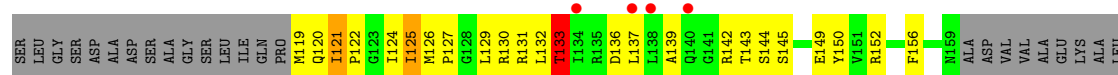
• Molecule 1: Major capsid protein

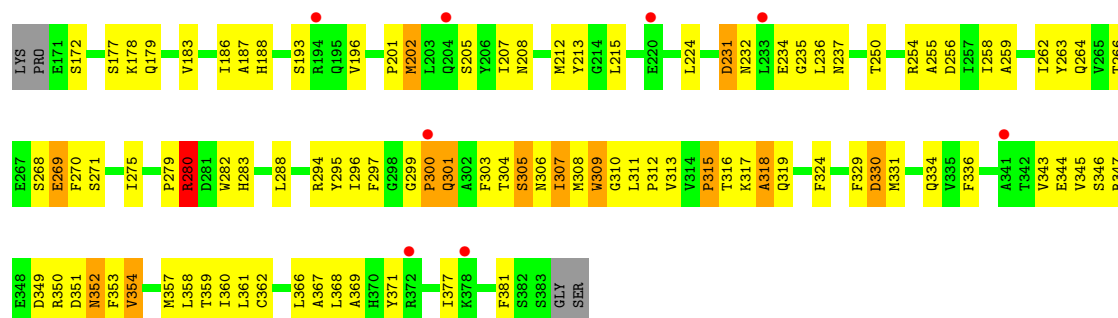
Chain E:



• Molecule 1: Major capsid protein

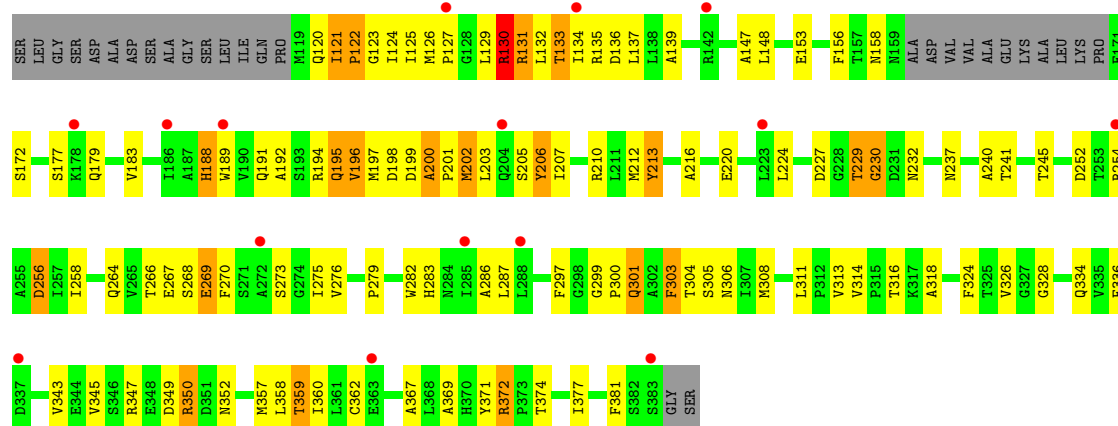
Chain F:





• Molecule 1: Major capsid protein

Chain G:



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	706.94Å 706.94Å 706.94Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 5.20 29.98 – 5.20	Depositor EDS
% Data completeness (in resolution range)	(Not available) (30.00-5.20) 84.2 (29.98-5.20)	Depositor EDS
$R_{merge}$	0.18	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.45 (at 5.34Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.410 , (Not available) 0.361 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	124.3	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 157.6	EDS
Estimated twinning fraction	0.076 for l,-k,h	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.37$ , $\langle L^2 \rangle = 0.20$	Xtriage
Outliers	0 of 373644 reflections	Xtriage
$F_o, F_c$ correlation	0.43	EDS
Total number of atoms	13797	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	111.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.16% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 5 Model quality i

### 5.1 Standard geometry i

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# $ Z  > 5$	RMSZ	# $ Z  > 5$
1	A	0.55	4/2004 (0.2%)	0.79	6/2715 (0.2%)
1	B	0.35	0/2004	0.64	1/2715 (0.0%)
1	C	0.35	0/2004	0.62	0/2715
1	D	0.36	0/2004	0.63	0/2715
1	E	0.36	0/2004	0.61	0/2715
1	F	0.61	2/2004 (0.1%)	0.68	3/2715 (0.1%)
1	G	0.36	0/2004	0.65	0/2715
All	All	0.43	6/14028 (0.0%)	0.66	10/19005 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
1	G	0	1
All	All	0	2

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	231	ASP	C-N	-21.67	0.84	1.34
1	A	133	THR	N-CA	-7.60	1.31	1.46
1	A	120	GLN	C-N	7.04	1.50	1.34
1	F	280	ARG	N-CA	6.11	1.58	1.46
1	A	128	GLY	N-CA	-5.95	1.37	1.46

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	122	PRO	CA-N-CD	-9.37	98.38	111.50
1	A	132	LEU	C-N-CA	-7.34	103.36	121.70

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	123	GLY	C-N-CA	7.03	139.26	121.70
1	F	318	ALA	N-CA-C	-5.91	95.04	111.00
1	A	133	THR	CA-C-N	-5.89	104.25	117.20

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	206	TYR	Sidechain
1	G	213	TYR	Sidechain

## 5.2 Close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1971	0	1937	219	0
1	B	1971	0	1938	138	0
1	C	1971	0	1938	176	0
1	D	1971	0	1937	185	0
1	E	1971	0	1938	202	0
1	F	1971	0	1936	192	0
1	G	1971	0	1938	124	0
All	All	13797	0	13562	1022	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 37.

The worst 5 of 1022 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:280:ARG:HH22	1:B:262:ILE:CD1	1.21	1.53
1:A:280:ARG:NH2	1:B:262:ILE:HD13	1.37	1.40
1:E:279:PRO:CB	1:F:266:THR:HG21	1.51	1.39
1:A:280:ARG:NH2	1:B:262:ILE:CD1	1.88	1.36
1:C:279:PRO:CB	1:D:266:THR:HG21	1.55	1.35

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone ⓘ

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution. The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	250/282 (89%)	198 (79%)	37 (15%)	15 (6%)	2	37
1	B	250/282 (89%)	201 (80%)	32 (13%)	17 (7%)	2	33
1	C	250/282 (89%)	212 (85%)	30 (12%)	8 (3%)	6	56
1	D	250/282 (89%)	205 (82%)	33 (13%)	12 (5%)	4	44
1	E	250/282 (89%)	206 (82%)	34 (14%)	10 (4%)	5	49
1	F	250/282 (89%)	201 (80%)	39 (16%)	10 (4%)	5	49
1	G	250/282 (89%)	200 (80%)	36 (14%)	14 (6%)	3	40
All	All	1750/1974 (89%)	1423 (81%)	241 (14%)	86 (5%)	3	43

5 of 86 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	122	PRO
1	A	124	ILE
1	A	135	ARG
1	A	193	SER
1	A	350	ARG

### 5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution. The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	211/231 (91%)	188 (89%)	23 (11%)	9	46
1	B	211/231 (91%)	198 (94%)	13 (6%)	26	73

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	C	211/231 (91%)	191 (90%)	20 (10%)	12	53
1	D	211/231 (91%)	200 (95%)	11 (5%)	32	78
1	E	211/231 (91%)	195 (92%)	16 (8%)	19	65
1	F	211/231 (91%)	195 (92%)	16 (8%)	19	65
1	G	211/231 (91%)	192 (91%)	19 (9%)	14	56
All	All	1477/1617 (91%)	1359 (92%)	118 (8%)	17	62

5 of 118 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	306	ASN
1	D	309	TRP
1	G	205	SER
1	C	317	LYS
1	D	202	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 44 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	204	GLN
1	D	334	GLN
1	G	232	ASN
1	D	208	ASN
1	D	283	HIS

### 5.3.3 RNA ⓘ

There are no RNA chains in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

There are no ligands in this entry.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues

There are no chain breaks in this entry.

## 6 Fit of model and data ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	254/282 (90%)	0.90	5 (1%) 62 50	47, 95, 95, 95	0
1	B	254/282 (90%)	1.09	26 (10%) 7 14	54, 114, 114, 114	0
1	C	254/282 (90%)	1.11	21 (8%) 11 17	57, 122, 122, 122	0
1	D	254/282 (90%)	1.04	15 (5%) 22 24	50, 103, 103, 103	0
1	E	254/282 (90%)	1.05	17 (6%) 17 22	54, 114, 114, 114	0
1	F	254/282 (90%)	1.03	12 (4%) 30 30	51, 104, 104, 104	0
1	G	254/282 (90%)	1.00	15 (5%) 22 24	61, 132, 132, 132	0
All	All	1778/1974 (90%)	1.03	111 (6%) 20 24	47, 114, 132, 132	0

The worst 5 of 111 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	119	MET	4.4
1	C	362	CYS	4.0
1	F	134	ILE	3.9
1	C	119	MET	3.4
1	C	347	ARG	3.3

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 6.4 Ligands ⓘ

There are no ligands in this entry.

## 6.5 Other polymers ⓘ

There are no such residues in this entry.