



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 21, 2020 – 11:40 pm BST

PDB ID : 5X6D  
Title : Crystal structure of PrfA-DNA binary complex  
Authors : Wang, Y.; Feng, H.; Zhu, Y.L.; Gao, P.  
Deposited on : 2017-02-21  
Resolution : 2.94 Å(reported)

This is a wwPDB X-ray Structure 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

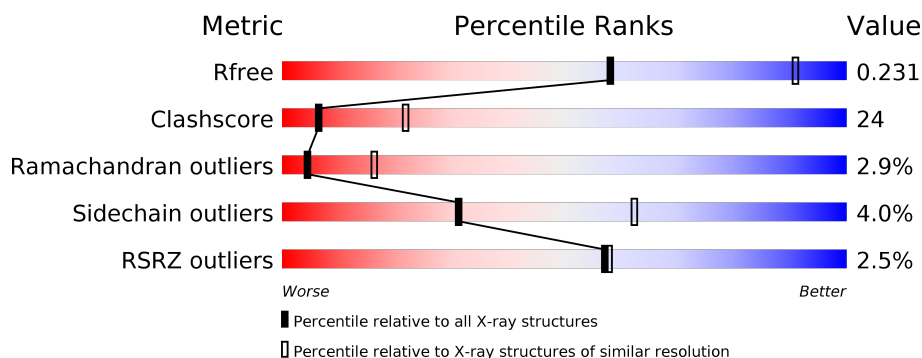
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.94 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))
$R_{free}$	130704	2969 (2.98-2.90)
Clashscore	141614	3218 (2.98-2.90)
Ramachandran outliers	138981	3122 (2.98-2.90)
Sidechain outliers	138945	3124 (2.98-2.90)
RSRZ outliers	127900	2902 (2.98-2.90)

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 respectively. 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  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	237	<div> <div>0%</div> <div> <div></div> <div>63%</div> <div>31%</div> <div>5%</div> </div> </div>
1	B	237	<div> <div>2%</div> <div> <div></div> <div>68%</div> <div>27%</div> <div>•</div> </div> </div>
1	G	237	<div> <div>3%</div> <div> <div></div> <div>56%</div> <div>34%</div> <div>6%</div> <div>••</div> </div> </div>
1	H	237	<div> <div>3%</div> <div> <div></div> <div>65%</div> <div>32%</div> <div>•</div> </div> </div>
1	M	237	<div> <div>3%</div> <div> <div></div> <div>58%</div> <div>34%</div> <div>5%</div> <div>••</div> </div> </div>
1	N	237	<div> <div>2%</div> <div> <div></div> <div>61%</div> <div>32%</div> <div>5%</div> <div>•</div> </div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	C	29	<div><div></div><div>10%</div><div>52%</div><div>48%</div></div>
2	K	29	<div><div></div><div>3%</div><div>38%</div><div>62%</div></div>
2	P	29	<div><div></div><div>28%</div><div>72%</div></div>
3	D	29	<div><div></div><div>7%</div><div>31%</div><div>66%</div><div></div><div>.</div></div>
3	L	29	<div><div></div><div>34%</div><div>62%</div><div></div><div>.</div></div>
3	O	29	<div><div></div><div>7%</div><div>34%</div><div>66%</div></div>



## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 15015 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Listeriolysin positive regulatory factor A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	236	Total	C	N	O	S	0	0	0
			1920	1253	301	359	7			
1	B	236	Total	C	N	O	S	0	0	0
			1920	1253	301	359	7			
1	G	235	Total	C	N	O	S	0	0	0
			1912	1249	299	357	7			
1	H	236	Total	C	N	O	S	0	0	0
			1920	1253	301	359	7			
1	M	235	Total	C	N	O	S	0	0	0
			1912	1249	299	357	7			
1	N	236	Total	C	N	O	S	0	0	0
			1920	1253	301	359	7			

- Molecule 2 is a DNA chain called DNA (29-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	29	Total	C	N	O	P	0	0	0
			596	286	110	172	28			
2	K	29	Total	C	N	O	P	0	0	0
			593	286	110	170	27			
2	P	29	Total	C	N	O	P	0	0	0
			596	286	110	172	28			

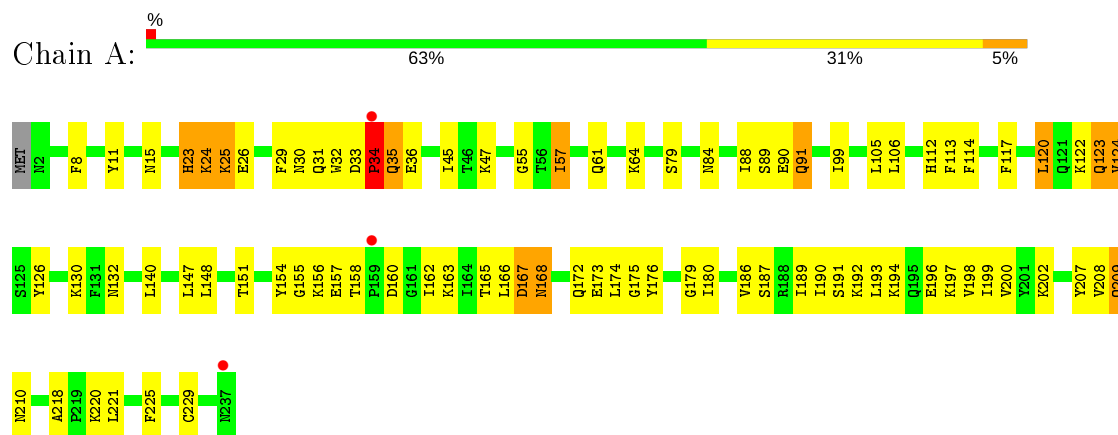
- Molecule 3 is a DNA chain called DNA (28-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	28	Total	C	N	O	P	0	0	0
			571	274	101	168	28			
3	L	28	Total	C	N	O	P	0	0	0
			568	274	101	166	27			
3	O	29	Total	C	N	O	P	0	0	0
			587	283	104	172	28			

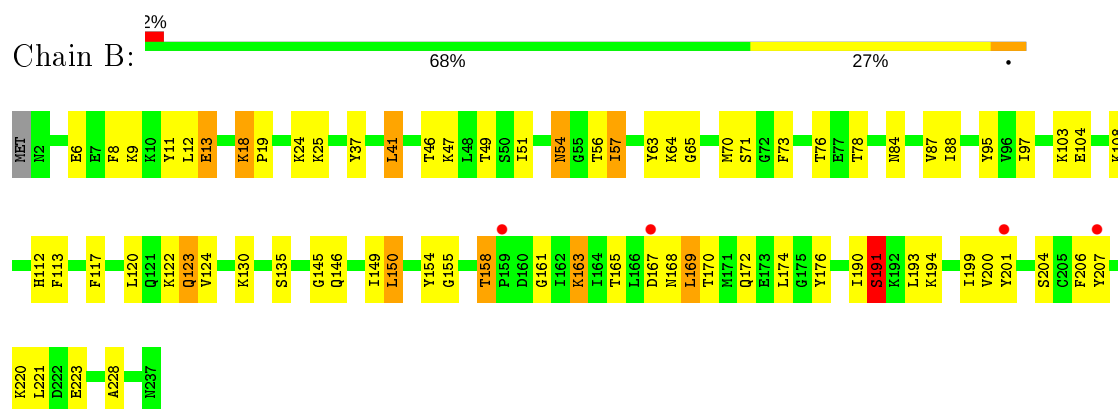
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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.

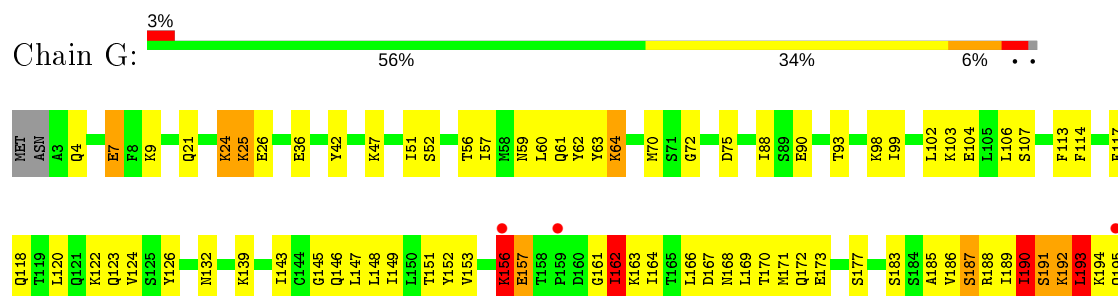
- Molecule 1: Listeriolysin positive regulatory factor A



- Molecule 1: Listeriolysin positive regulatory factor A

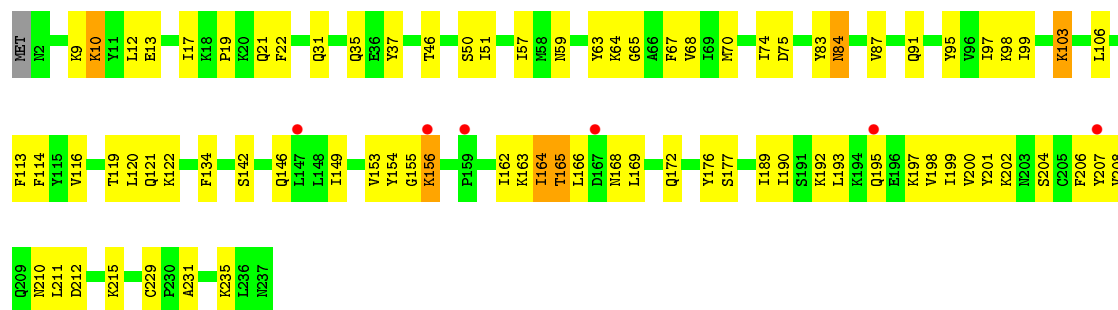


- Molecule 1: Listeriolysin positive regulatory factor A

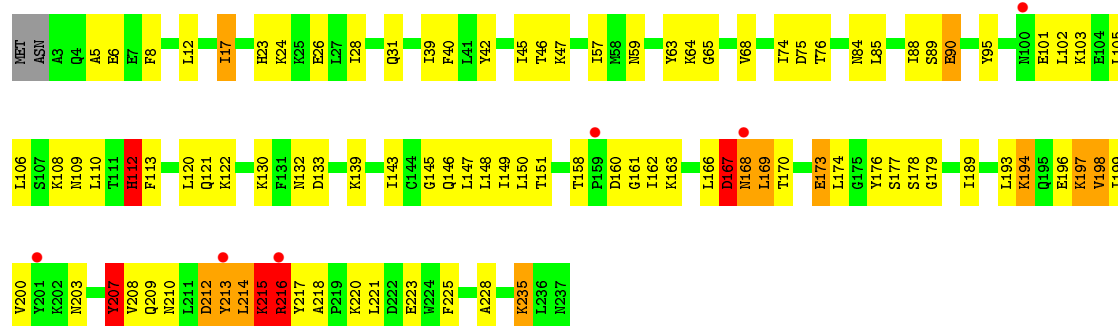




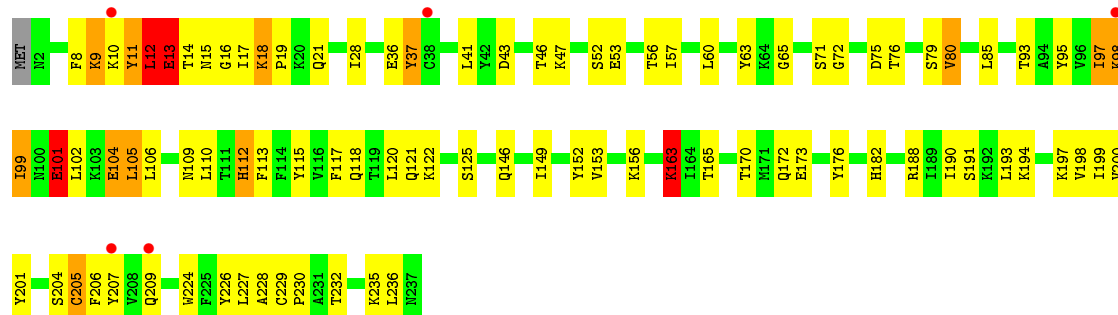
- Molecule 1: Listeriolysin positive regulatory factor A



- Molecule 1: Listeriolysin positive regulatory factor A

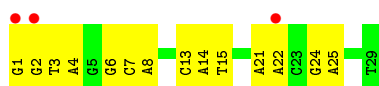


- Molecule 1: Listeriolysin positive regulatory factor A

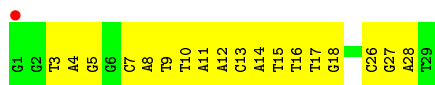


- Molecule 2: DNA (29-MER)

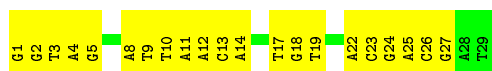




- Molecule 2: DNA (29-MER)



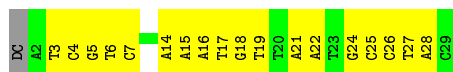
- Molecule 2: DNA (29-MER)



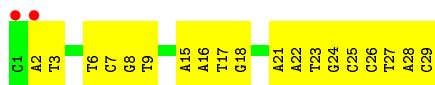
- Molecule 3: DNA (28-MER)



- Molecule 3: DNA (28-MER)



- Molecule 3: DNA (28-MER)



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	64.82Å 96.44Å 371.91Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.83 – 2.94 19.83 – 2.94	Depositor EDS
% Data completeness (in resolution range)	98.5 (19.83-2.94) 90.7 (19.83-2.94)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.98 (at 2.93Å)	Xtriage
Refinement program	PHENIX 1.8.4_1496	Depositor
R, $R_{free}$	0.213 , 0.266 0.216 , 0.231	Depositor DCC
$R_{free}$ test set	2473 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	80.0	Xtriage
Anisotropy	0.597	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 76.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	15015	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	117.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.04% of the height of the origin peak. No significant pseudotranslation is detected.*

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

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.



## 5 Model quality

### 5.1 Standard geometry

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	0.34	0/1964	0.63	1/2652 (0.0%)
1	B	0.33	0/1964	0.55	1/2652 (0.0%)
1	G	0.52	0/1956	0.86	7/2641 (0.3%)
1	H	0.35	0/1964	0.58	2/2652 (0.1%)
1	M	0.54	2/1956 (0.1%)	0.77	8/2641 (0.3%)
1	N	0.51	1/1964 (0.1%)	0.84	8/2652 (0.3%)
2	C	0.55	0/669	0.94	0/1032
2	K	0.57	0/665	0.94	0/1024
2	P	0.60	0/669	0.99	0/1032
3	D	0.55	0/639	0.98	0/983
3	L	0.59	0/635	1.01	0/975
3	O	0.58	0/657	0.99	0/1011
All	All	0.48	3/15702 (0.0%)	0.79	27/21947 (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	2
1	B	0	2
1	G	0	5
1	M	0	3
1	N	0	7
All	All	0	19

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	M	207	TYR	CD2-CE2	-9.09	1.25	1.39
1	N	101	GLU	CB-CG	-6.77	1.39	1.52
1	M	213	TYR	CD2-CE2	5.95	1.48	1.39

The worst 5 of 27 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	N	12	LEU	CA-CB-CG	13.85	147.16	115.30
1	N	163	LYS	CD-CE-NZ	10.31	135.41	111.70
1	G	207	TYR	CA-CB-CG	9.35	131.16	113.40
1	M	207	TYR	CB-CG-CD2	-9.14	115.51	121.00
1	G	193	LEU	CA-CB-CG	8.41	134.64	115.30

There are no chirality outliers.

5 of 19 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	167	ASP	Peptide
1	A	34	PRO	Peptide
1	B	18	LYS	Peptide
1	B	191	SER	Peptide
1	G	156	LYS	Peptide

## 5.2 Too-close contacts [i](#)

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 hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1920	0	1920	83	0
1	B	1920	0	1920	59	0
1	G	1912	0	1914	142	0
1	H	1920	0	1920	82	0
1	M	1912	0	1914	111	0
1	N	1920	0	1920	119	0
2	C	596	0	330	12	0
2	K	593	0	332	22	0
2	P	596	0	330	26	0
3	D	571	0	318	16	0
3	L	568	0	320	18	0
3	O	587	0	330	17	0
All	All	15015	0	13468	664	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 24.

The worst 5 of 664 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:G:198:VAL:H	1:G:199:ILE:HG13	1.16	1.08
1:G:106:LEU:HB3	1:G:113:PHE:HE1	1.22	1.02
1:H:106:LEU:HB3	1:H:113:PHE:HD1	1.24	1.01
1:G:106:LEU:HB3	1:G:113:PHE:CE1	1.97	1.00
1:G:197:LYS:HA	1:G:199:ILE:HD11	1.44	0.99

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

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	234/237 (99%)	213 (91%)	14 (6%)	7 (3%)	4	15
1	B	234/237 (99%)	216 (92%)	15 (6%)	3 (1%)	12	35
1	G	233/237 (98%)	191 (82%)	26 (11%)	16 (7%)	1	2
1	H	234/237 (99%)	218 (93%)	12 (5%)	4 (2%)	9	29
1	M	233/237 (98%)	212 (91%)	13 (6%)	8 (3%)	3	13
1	N	234/237 (99%)	200 (86%)	32 (14%)	2 (1%)	17	46
All	All	1402/1422 (99%)	1250 (89%)	112 (8%)	40 (3%)	4	16

5 of 40 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	91	GLN
1	A	168	ASN
1	B	167	ASP
1	B	169	LEU
1	G	99	ILE

### 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	211/212 (100%)	202 (96%)	9 (4%)	29	60
1	B	211/212 (100%)	204 (97%)	7 (3%)	38	69
1	G	210/212 (99%)	195 (93%)	15 (7%)	14	38
1	H	211/212 (100%)	207 (98%)	4 (2%)	57	81
1	M	210/212 (99%)	202 (96%)	8 (4%)	33	64
1	N	211/212 (100%)	203 (96%)	8 (4%)	33	64
All	All	1264/1272 (99%)	1213 (96%)	51 (4%)	31	62

5 of 51 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	193	LEU
1	G	208	VAL
1	N	112	HIS
1	G	199	ILE
1	G	210	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	59	ASN
1	G	61	GLN
1	H	121	GLN
1	B	123	GLN
1	G	31	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

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	236/237 (99%)	-0.05	3 (1%) 77 78	62, 92, 141, 215	0
1	B	236/237 (99%)	0.03	4 (1%) 70 71	65, 92, 135, 164	0
1	G	235/237 (99%)	0.14	8 (3%) 45 43	79, 111, 167, 197	0
1	H	236/237 (99%)	0.05	6 (2%) 57 58	75, 107, 152, 170	0
1	M	235/237 (99%)	0.12	6 (2%) 56 56	73, 104, 158, 216	0
1	N	236/237 (99%)	0.18	5 (2%) 63 64	74, 110, 168, 210	0
2	C	29/29 (100%)	0.81	3 (10%) 6 6	106, 148, 181, 191	0
2	K	29/29 (100%)	0.55	1 (3%) 45 43	115, 150, 189, 197	0
2	P	29/29 (100%)	0.19	0 100 100	107, 150, 195, 215	0
3	D	28/29 (96%)	0.50	2 (7%) 16 14	104, 145, 185, 200	0
3	L	28/29 (96%)	0.48	0 100 100	125, 153, 186, 195	0
3	O	29/29 (100%)	0.47	2 (6%) 16 15	114, 143, 195, 199	0
All	All	1586/1596 (99%)	0.12	40 (2%) 57 58	62, 106, 166, 216	0

The worst 5 of 40 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	1	DG	4.9
3	O	1	DC	4.0
1	N	10	LYS	4.0
1	G	207	TYR	4.0
1	G	200	VAL	3.7

### 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 [i](#)

There are no carbohydrates in this entry.

### 6.4 Ligands [i](#)

There are no ligands in this entry.

### 6.5 Other polymers [i](#)

There are no such residues in this entry.