



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 04:22 am GMT

PDB ID : 3X1L  
Title : Crystal Structure of the CRISPR-Cas RNA Silencing Cmr Complex Bound to a Target Analog  
Authors : Osawa, T.; Numata, T.  
Deposited on : 2014-11-20  
Resolution : 2.10 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

<http://wwpdb.org/validation/2016/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) : 1.9-1692  
EDS : trunk28620  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : recalc28949

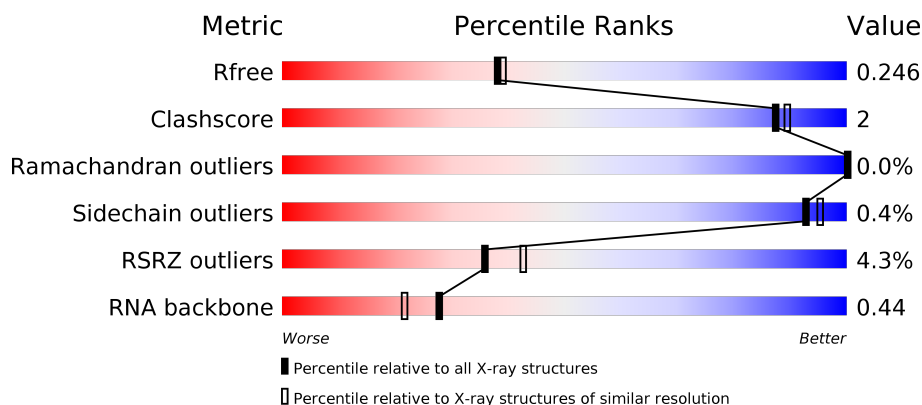
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.10 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))
$R_{free}$	100719	4243 (2.10-2.10)
Clashscore	112137	4788 (2.10-2.10)
Ramachandran outliers	110173	4740 (2.10-2.10)
Sidechain outliers	110143	4741 (2.10-2.10)
RSRZ outliers	101464	4275 (2.10-2.10)
RNA backbone	2435	1063 (2.70-1.50)

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. 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	677	<div> <div>4%</div> <div>79% 6% 15%</div> </div>
2	B	322	<div> <div>%</div> <div>94% 5%</div> </div>
3	C	357	<div> <div>85% 6% 10%</div> </div>
3	D	357	<div> <div>3%</div> <div>85% 7% 7%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	E	357	<div><div></div><div>9%</div><div>85%</div><div>8%</div><div>8%</div></div>
4	F	155	<div><div></div><div>5%</div><div>89%</div><div>5%</div><div>6%</div></div>
4	G	155	<div><div></div><div>6%</div><div>88%</div><div>10%</div><div></div></div>
5	H	349	<div><div></div><div>6%</div><div>81%</div><div></div><div>15%</div></div>
6	I	39	<div><div></div><div>56%</div><div>26%</div><div>18%</div></div>
7	J	31	<div><div></div><div>65%</div><div>6%</div><div>29%</div></div>

## 2 Entry composition

There are 10 unique types of molecules in this entry. The entry contains 19952 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 CRISPR system Cmr subunit Cmr2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	573	4470	2920	749	789	12	0	0	0

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	195	MET	-	EXPRESSION TAG	UNP Q8U1S6
A	196	GLY	-	EXPRESSION TAG	UNP Q8U1S6
A	197	SER	-	EXPRESSION TAG	UNP Q8U1S6
A	198	SER	-	EXPRESSION TAG	UNP Q8U1S6
A	199	HIS	-	EXPRESSION TAG	UNP Q8U1S6
A	200	HIS	-	EXPRESSION TAG	UNP Q8U1S6
A	201	HIS	-	EXPRESSION TAG	UNP Q8U1S6
A	202	HIS	-	EXPRESSION TAG	UNP Q8U1S6
A	203	HIS	-	EXPRESSION TAG	UNP Q8U1S6
A	204	HIS	-	EXPRESSION TAG	UNP Q8U1S6
A	205	SER	-	EXPRESSION TAG	UNP Q8U1S6
A	206	SER	-	EXPRESSION TAG	UNP Q8U1S6
A	207	GLY	-	EXPRESSION TAG	UNP Q8U1S6
A	208	LEU	-	EXPRESSION TAG	UNP Q8U1S6
A	209	VAL	-	EXPRESSION TAG	UNP Q8U1S6
A	210	PRO	-	EXPRESSION TAG	UNP Q8U1S6
A	211	ARG	-	EXPRESSION TAG	UNP Q8U1S6
A	212	GLY	-	EXPRESSION TAG	UNP Q8U1S6
A	213	SER	-	EXPRESSION TAG	UNP Q8U1S6
A	214	HIS	-	EXPRESSION TAG	UNP Q8U1S6
A	215	MET	-	EXPRESSION TAG	UNP Q8U1S6

- Molecule 2 is a protein called CRISPR system Cmr subunit Cmr3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	319	Total	C	N	O	S	0	0	0
			2458	1611	394	449	4			

- Molecule 3 is a protein called Cmr4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	322	Total	C	N	O	S	0	0	0
			2405	1549	411	444	1			
3	D	331	Total	C	N	O	S	0	0	0
			2428	1571	413	442	2			
3	E	330	Total	C	N	O	S	0	0	0
			2395	1550	410	434	1			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-1	MET	-	EXPRESSION TAG	UNP O28416
C	0	GLY	-	EXPRESSION TAG	UNP O28416
D	-1	MET	-	EXPRESSION TAG	UNP O28416
D	0	GLY	-	EXPRESSION TAG	UNP O28416
E	-1	MET	-	EXPRESSION TAG	UNP O28416
E	0	GLY	-	EXPRESSION TAG	UNP O28416

- Molecule 4 is a protein called CRISPR system Cmr subunit Cmr5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	F	146	Total	C	N	O	S	0	0	0
			1110	722	180	206	2			
4	G	152	Total	C	N	O	S	0	0	0
			1167	753	186	226	2			

- Molecule 5 is a protein called Cmr6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	H	297	Total	C	N	O	S	0	0	0
			2260	1469	363	425	3			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
H	344	HIS	-	EXPRESSION TAG	UNP O28418
H	345	HIS	-	EXPRESSION TAG	UNP O28418

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
H	346	HIS	-	EXPRESSION TAG	UNP O28418
H	347	HIS	-	EXPRESSION TAG	UNP O28418
H	348	HIS	-	EXPRESSION TAG	UNP O28418
H	349	HIS	-	EXPRESSION TAG	UNP O28418

- Molecule 6 is a RNA chain called RNA (32-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	I	32	Total	C	N	O	P	0	0	0
			679	304	117	227	31			

- Molecule 7 is a DNA chain called DNA (5'-D(\*TP\*GP\*CP\*TP\*CP\*TP\*CP\*AP\*GP\*CP\*CP\*GP\*CP\*AP\*AP\*GP\*GP\*AP\*CP\*CP\*GP\*CP\*AP\*TP\*AP\*CP\*TP\*AP\*CP\*AP\*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	J	22	Total	C	N	O	P	0	0	0
			448	212	88	126	22			

- Molecule 8 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	1	Total	Zn	0	0
			1	1		

- Molecule 9 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	B	1	Total	Mg	0	0
			1	1		

- Molecule 10 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
10	A	32	Total	O	0	0
			32	32		
10	B	23	Total	O	0	0
			23	23		
10	C	16	Total	O	0	0
			16	16		
10	D	14	Total	O	0	0
			14	14		

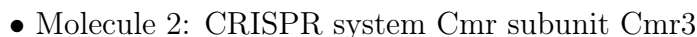
*Continued on next page...*

*Continued from previous page...*

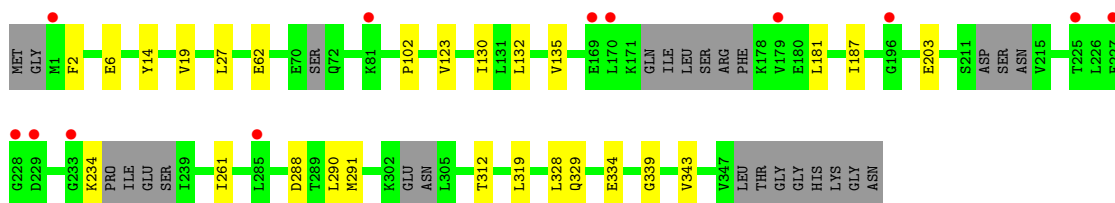
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
10	E	13	Total 13	O 13	0	0
10	G	5	Total 5	O 5	0	0
10	H	4	Total 4	O 4	0	0
10	I	23	Total 23	O 23	0	0



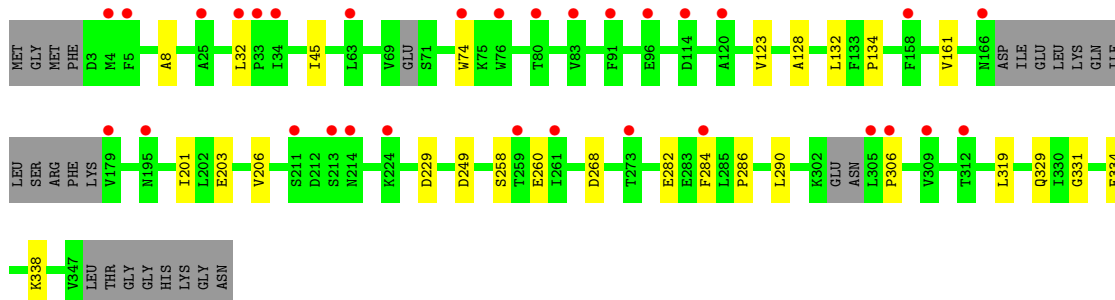
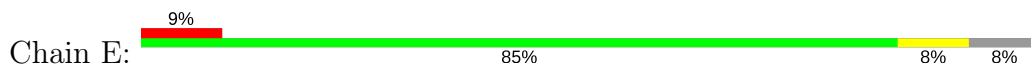
- Molecule 1: CRISPR system Cmr subunit Cmr2



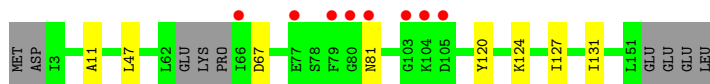
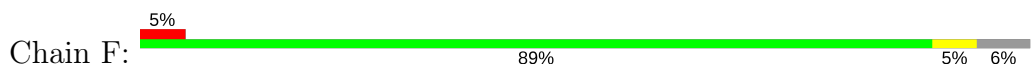




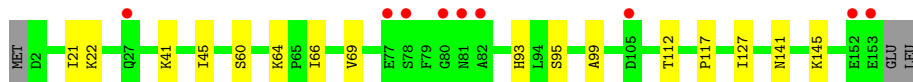
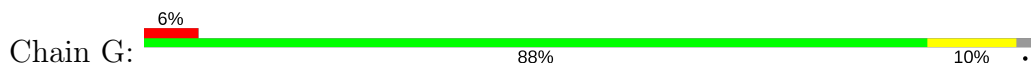
- Molecule 3: Cmr4



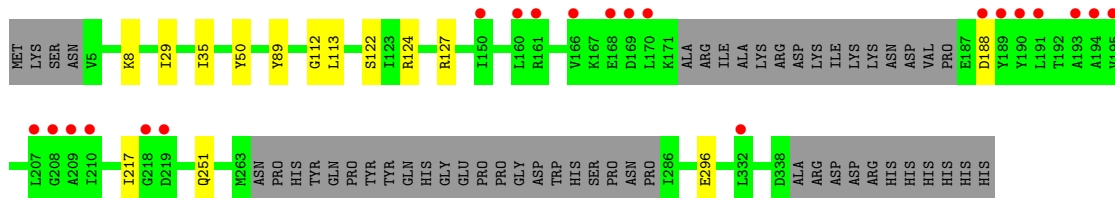
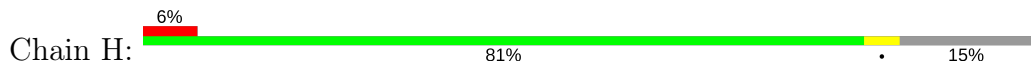
- Molecule 4: CRISPR system Cmr subunit Cmr5



- Molecule 4: CRISPR system Cmr subunit Cmr5



- Molecule 5: Cmr6

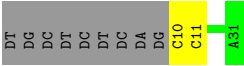


- Molecule 6: RNA (32-MER)





● Molecule 7: DNA (5'-D(\*TP\*GP\*CP\*TP\*CP\*TP\*CP\*AP\*GP\*CP\*CP\*GP\*CP\*AP\*AP\*GP\*GP\*AP\*CP\*CP\*GP\*CP\*AP\*TP\*AP\*CP\*TP\*AP\*CP\*AP\*A)-3')



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	75.49Å 76.22Å 139.19Å 90.32° 104.83° 118.58°	Depositor
Resolution (Å)	44.34 – 2.10 44.34 – 2.10	Depositor EDS
% Data completeness (in resolution range)	97.4 (44.34-2.10) 92.8 (44.34-2.10)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.56 (at 2.10Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.4_1496)	Depositor
R, $R_{free}$	0.208 , 0.246 0.208 , 0.246	Depositor DCC
$R_{free}$ test set	7441 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	32.8	Xtriage
Anisotropy	0.397	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 44.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.005 for k,-h-k,h+l 0.005 for -h-k,h,h+k+l 0.017 for h,-h-k,-h-l 0.017 for -h-k,k,-l 0.016 for k,h,-h-k-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	19952	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.90% 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

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, MG

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	0.23	0/4561	0.36	0/6184
2	B	0.24	0/2512	0.39	0/3401
3	C	0.23	0/2445	0.39	0/3320
3	D	0.24	0/2467	0.38	0/3348
3	E	0.24	0/2436	0.38	0/3315
4	F	0.23	0/1131	0.35	0/1527
4	G	0.23	0/1190	0.35	0/1609
5	H	0.24	0/2306	0.35	0/3124
6	I	0.24	0/758	0.84	1/1180 (0.1%)
7	J	0.52	0/503	0.84	0/772
All	All	0.25	0/20309	0.42	1/27780 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	I	15	U	O4'-C1'-N1	5.16	112.32	108.20

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-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 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	4470	0	4434	23	0
2	B	2458	0	2480	9	0
3	C	2405	0	2362	12	0
3	D	2428	0	2351	14	0
3	E	2395	0	2289	15	0
4	F	1110	0	1097	8	0
4	G	1167	0	1137	10	0
5	H	2260	0	2132	9	0
6	I	679	0	343	5	0
7	J	448	0	245	3	0
8	A	1	0	0	0	0
9	B	1	0	0	0	0
10	A	32	0	0	0	0
10	B	23	0	0	0	0
10	C	16	0	0	0	0
10	D	14	0	0	0	0
10	E	13	0	0	0	0
10	G	5	0	0	0	0
10	H	4	0	0	0	0
10	I	23	0	0	0	0
All	All	19952	0	18870	92	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 2.

The worst 5 of 92 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)
3:E:331:GLY:O	6:I:20:G:N2	2.27	0.68
3:C:201:ILE:HD13	4:F:131:ILE:HG12	1.77	0.66
5:H:29:ILE:HD13	5:H:35:ILE:HB	1.80	0.64
2:B:42:ILE:HD13	2:B:178:ILE:HD11	1.82	0.62
1:A:422:LYS:HB3	2:B:108:PRO:HG3	1.83	0.61

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	559/677 (83%)	551 (99%)	8 (1%)	0	100	100
2	B	315/322 (98%)	304 (96%)	11 (4%)	0	100	100
3	C	316/357 (88%)	309 (98%)	7 (2%)	0	100	100
3	D	319/357 (89%)	310 (97%)	9 (3%)	0	100	100
3	E	322/357 (90%)	313 (97%)	9 (3%)	0	100	100
4	F	142/155 (92%)	141 (99%)	1 (1%)	0	100	100
4	G	150/155 (97%)	143 (95%)	7 (5%)	0	100	100
5	H	291/349 (83%)	283 (97%)	7 (2%)	1 (0%)	44	44
All	All	2414/2729 (88%)	2354 (98%)	59 (2%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
5	H	217	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	457/604 (76%)	456 (100%)	1 (0%)	94	97
2	B	257/280 (92%)	256 (100%)	1 (0%)	93	95
3	C	243/309 (79%)	242 (100%)	1 (0%)	93	95
3	D	238/309 (77%)	237 (100%)	1 (0%)	93	95

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	E	228/309 (74%)	225 (99%)	3 (1%)	73	80
4	F	111/131 (85%)	111 (100%)	0	100	100
4	G	119/131 (91%)	119 (100%)	0	100	100
5	H	224/309 (72%)	223 (100%)	1 (0%)	93	95
All	All	1877/2382 (79%)	1869 (100%)	8 (0%)	93	95

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	2	PHE
5	H	89	TYR
3	E	268	ASP
3	C	39	HIS
3	E	229	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
6	I	31/39 (79%)	3 (9%)	0

All (3) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
6	I	2	U
6	I	14	G
6	I	27	G

There are no RNA pucker outliers to report.

## 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 [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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	573/677 (84%)	0.13	25 (4%) 35 41	21, 35, 58, 65	0
2	B	319/322 (99%)	0.06	2 (0%) 89 91	21, 33, 51, 64	0
3	C	322/357 (90%)	0.10	1 (0%) 93 94	21, 34, 52, 65	0
3	D	331/357 (92%)	0.31	12 (3%) 43 50	23, 37, 64, 81	0
3	E	330/357 (92%)	0.60	31 (9%) 9 12	25, 39, 64, 77	0
4	F	146/155 (94%)	0.20	8 (5%) 26 32	24, 37, 60, 72	0
4	G	152/155 (98%)	0.13	9 (5%) 23 29	24, 35, 52, 61	0
5	H	297/349 (85%)	0.29	21 (7%) 17 21	25, 39, 68, 77	0
6	I	32/39 (82%)	-0.62	0 100 100	25, 29, 61, 79	0
7	J	22/31 (70%)	-0.22	0 100 100	32, 46, 60, 62	0
All	All	2524/2799 (90%)	0.21	109 (4%) 36 43	21, 36, 61, 81	0

The worst 5 of 109 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	D	228	GLY	5.4
5	H	219	ASP	4.9
1	A	527	VAL	4.9
2	B	50	GLY	4.7
3	D	179	VAL	4.7

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

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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
8	ZN	A	901	1/1	1.00	0.12	0.07	30,30,30,30	0
9	MG	B	401	1/1	0.98	0.10	-	24,24,24,24	0

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

There are no such residues in this entry.