



wwPDB X-ray Structure Validation Summary Report

Jun 16, 2024 – 11:34 PM EDT

PDB ID : 5MGA
Title : Structure of the Cpf1 endonuclease R-loop complex after DNA cleavage
Authors : Montoya, G.; Stella, S.
Deposited on : 2016-11-21
Resolution : 3.00 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

A user guide is available at

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

with specific help available everywhere you see the  symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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 : 2.37.1
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) : 2.37.1

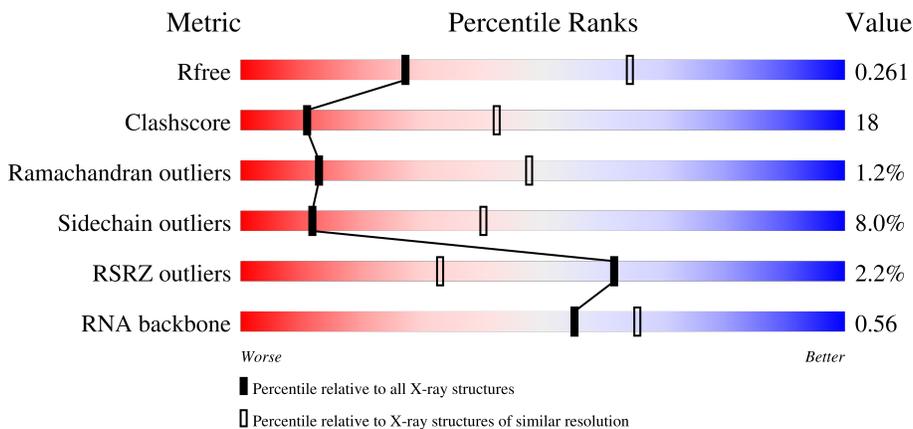
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.00 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	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)
RNA backbone	3102	1173 (3.30-2.70)

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 of the lower bar indicate the fraction of residues that contain outliers for ≥ 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	1323	
2	B	40	
3	C	26	
4	D	12	

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 11351 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-associated endonuclease Cpf1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1182	9595	6163	1588	1824	20	0	0	0

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	246	ASP	ILE	conflict	UNP A0Q7Q2
A	467	LEU	LYS	conflict	UNP A0Q7Q2
A	842	ILE	THR	conflict	UNP A0Q7Q2
A	1301	GLY	-	expression tag	UNP A0Q7Q2
A	1302	SER	-	expression tag	UNP A0Q7Q2
A	1303	GLU	-	expression tag	UNP A0Q7Q2
A	1304	PHE	-	expression tag	UNP A0Q7Q2
A	1305	GLU	-	expression tag	UNP A0Q7Q2
A	1306	LEU	-	expression tag	UNP A0Q7Q2
A	1307	GLU	-	expression tag	UNP A0Q7Q2
A	1308	ASN	-	expression tag	UNP A0Q7Q2
A	1309	LEU	-	expression tag	UNP A0Q7Q2
A	1310	TYR	-	expression tag	UNP A0Q7Q2
A	1311	PHE	-	expression tag	UNP A0Q7Q2
A	1312	GLN	-	expression tag	UNP A0Q7Q2
A	1313	GLY	-	expression tag	UNP A0Q7Q2
A	1314	GLU	-	expression tag	UNP A0Q7Q2
A	1315	LEU	-	expression tag	UNP A0Q7Q2
A	1316	ARG	-	expression tag	UNP A0Q7Q2
A	1317	ARG	-	expression tag	UNP A0Q7Q2
A	1318	GLN	-	expression tag	UNP A0Q7Q2
A	1319	ALA	-	expression tag	UNP A0Q7Q2
A	1320	SER	-	expression tag	UNP A0Q7Q2
A	1321	ALA	-	expression tag	UNP A0Q7Q2
A	1322	LEU	-	expression tag	UNP A0Q7Q2
A	1323	GLU	-	expression tag	UNP A0Q7Q2

- Molecule 2 is a RNA chain called RNA (40-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	40	847	381	148	279	39	0	0	0

- Molecule 3 is a DNA chain called DNA (26-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	26	529	254	91	158	26	0	0	0

- Molecule 4 is a DNA chain called DNA (5'-D(P*CP*GP*TP*TP*AP*GP*AP*GP*AP*AP*GP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	D	12	248	119	49	69	11	0	0	0

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	2	Total	Mg	0	0
			2	2		

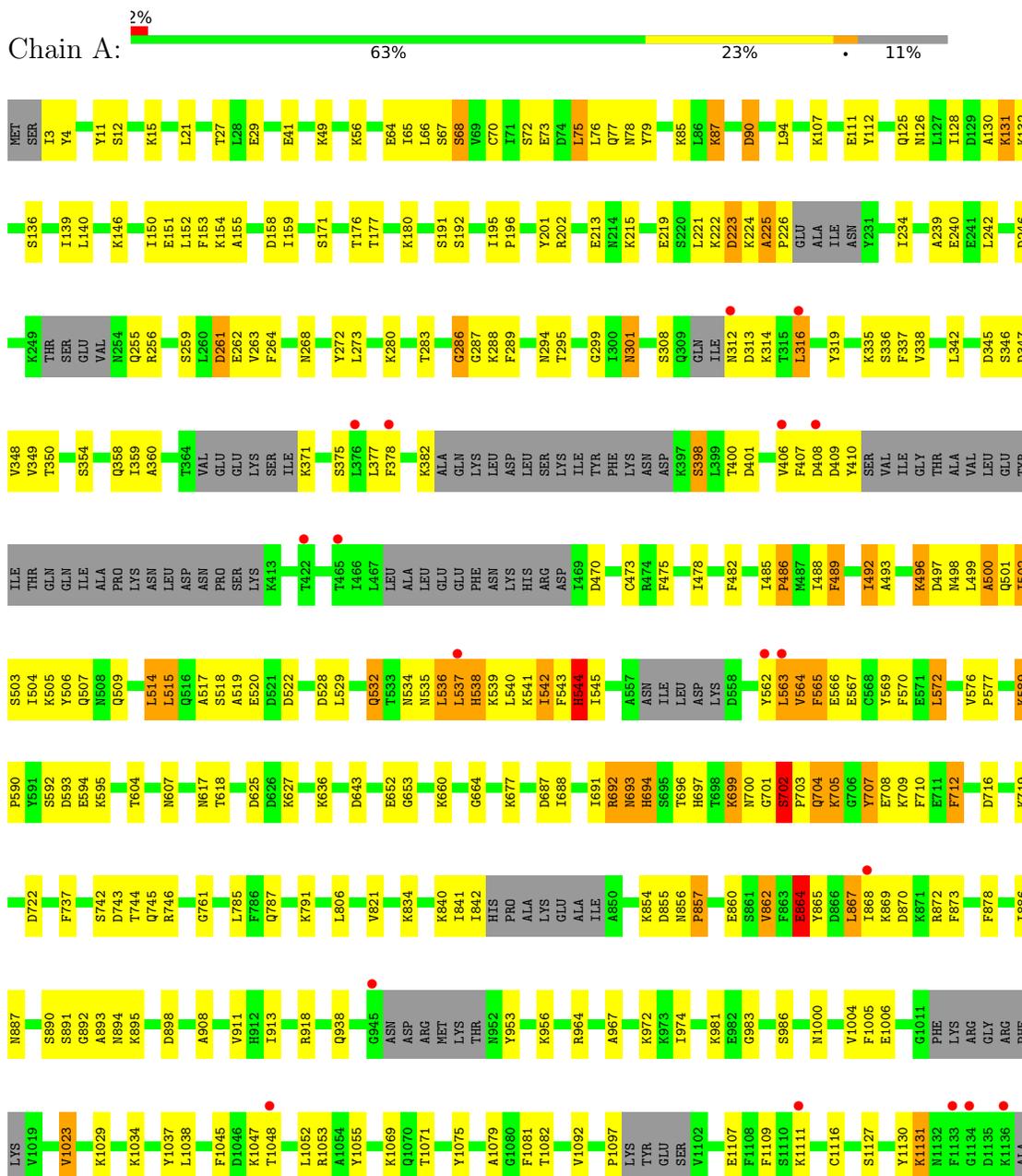
- Molecule 6 is water.

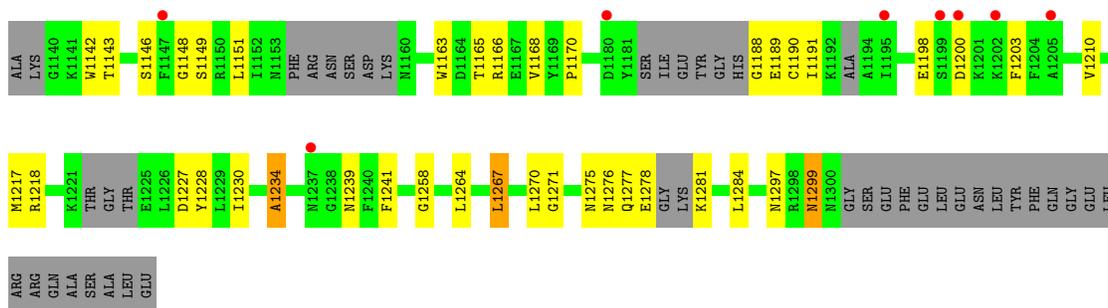
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	114	Total	O	0	0
			114	114		
6	B	9	Total	O	0	0
			9	9		
6	C	3	Total	O	0	0
			3	3		
6	D	4	Total	O	0	0
			4	4		

3 Residue-property plots

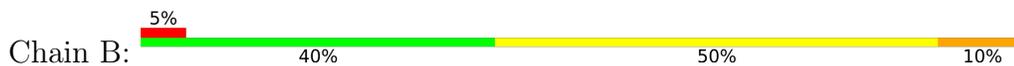
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: CRISPR-associated endonuclease Cpf1





- Molecule 2: RNA (40-MER)



- Molecule 3: DNA (26-MER)



- Molecule 4: DNA (5'-D(P*CP*GP*TP*TP*AP*GP*AP*GP*AP*AP*GP*T)-3')



4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	85.22Å 137.65Å 320.51Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.58 – 3.00 39.58 – 3.00	Depositor EDS
% Data completeness (in resolution range)	98.2 (39.58-3.00) 98.3 (39.58-3.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.54 (at 3.01Å)	Xtrriage
Refinement program	REFMAC 5.8.0155	Depositor
R, R_{free}	0.242 , 0.265 0.241 , 0.261	Depositor DCC
R_{free} test set	1843 reflections (4.91%)	wwPDB-VP
Wilson B-factor (Å ²)	69.0	Xtrriage
Anisotropy	0.554	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.25 , 59.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.43$, $\langle L^2 \rangle = 0.26$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	11351	wwPDB-VP
Average B, all atoms (Å ²)	97.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.09% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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 i

5.1 Standard geometry i

Bond lengths and bond angles in the following residue types are not validated in this section: 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	0.73	2/9763 (0.0%)	0.85	12/13094 (0.1%)
2	B	0.55	1/947 (0.1%)	0.82	4/1473 (0.3%)
3	C	0.75	3/591 (0.5%)	0.85	2/909 (0.2%)
4	D	0.72	0/279	0.88	0/430
All	All	0.72	6/11580 (0.1%)	0.85	18/15906 (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	1	5

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	544	HIS	CA-CB	-9.22	1.33	1.53
2	B	1	A	O3'-P	-5.54	1.54	1.61
1	A	41	GLU	C-O	5.40	1.33	1.23
3	C	3	DA	O3'-P	-5.39	1.54	1.61
3	C	-1	DT	O3'-P	-5.32	1.54	1.61

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	225	ALA	C-N-CD	-19.80	77.03	120.60
1	A	702	SER	C-N-CD	-8.85	101.14	120.60
1	A	694	HIS	N-CA-C	8.52	134.00	111.00
2	B	5	G	O5'-P-OP1	7.90	120.18	110.70

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	5	G	O5'-P-OP2	-6.54	99.82	105.70

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	694	HIS	CA

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	1234	ALA	Peptide
1	A	500	ALA	Peptide
1	A	543	PHE	Peptide
1	A	544	HIS	Peptide
1	A	75	LEU	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	9595	0	9345	360	0
2	B	847	0	428	17	0
3	C	529	0	296	13	0
4	D	248	0	137	18	0
5	B	2	0	0	0	0
6	A	114	0	0	1	2
6	B	9	0	0	0	0
6	C	3	0	0	0	0
6	D	4	0	0	0	0
All	All	11351	0	10206	388	2

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 18.

The worst 5 of 388 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:A:541:LYS:HD2	1:A:544:HIS:CB	1.28	1.64
1:A:694:HIS:N	1:A:705:LYS:HE3	1.35	1.40
1:A:541:LYS:CD	1:A:544:HIS:CB	2.00	1.39
1:A:566:GLU:HB3	1:A:570:PHE:CD1	1.71	1.25
1:A:700:ASN:OD1	1:A:709:LYS:CE	1.91	1.18

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
6:A:1481:HOH:O	6:A:1481:HOH:O[4_555]	1.84	0.36
6:A:1455:HOH:O	6:A:1455:HOH:O[3_554]	2.00	0.20

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	1136/1323 (86%)	1028 (90%)	94 (8%)	14 (1%)	13 48

5 of 14 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	150	ILE
1	A	408	ASP
1	A	693	ASN
1	A	864	GLU
1	A	1148	GLY

5.3.2 Protein sidechains [i](#)

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	1026/1203 (85%)	944 (92%)	82 (8%)	12 40

5 of 82 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	712	PHE
1	A	1109	PHE
1	A	806	LEU
1	A	867	LEU
1	A	1146	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	804	HIS
1	A	853	ASN
1	A	1094	GLN
1	A	495	ASN
1	A	534	ASN

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	39/40 (97%)	6 (15%)	1 (2%)

5 of 6 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	B	-9	G
2	B	-8	U
2	B	-7	U
2	B	-6	G
2	B	17	G

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	B	-8	U

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 monosaccharides 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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	6

The worst 5 of 6 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	426:ALA	C	459:LYS	N	18.99
1	A	550:GLU	C	555:ASP	N	7.52
1	A	422:THR	C	424:GLU	N	6.96
1	A	556:LYS	C	557:ALA	N	4.69
1	A	548:SER	C	550:GLU	N	3.41

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, 95th 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 ²)	Q<0.9
1	A	1182/1323 (89%)	-0.19	26 (2%) 62 33	38, 91, 183, 234	0
2	B	40/40 (100%)	-0.21	2 (5%) 28 10	39, 79, 163, 202	0
3	C	26/26 (100%)	-0.35	0 100 100	40, 62, 172, 181	0
4	D	12/12 (100%)	-0.27	0 100 100	37, 46, 201, 204	0
All	All	1260/1401 (89%)	-0.20	28 (2%) 62 33	37, 90, 184, 234	0

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	20	C	4.6
1	A	562	TYR	4.2
1	A	1205	ALA	4.2
1	A	1180	ASP	4.1
1	A	376	LEU	3.9

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 monosaccharides 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. The B-factors column lists the minimum, median, 95th 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	B-factors(\AA^2)	Q<0.9
5	MG	B	102	1/1	0.65	0.21	89,89,89,89	0
5	MG	B	101	1/1	0.73	0.32	68,68,68,68	0

6.5 Other polymers [i](#)

There are no such residues in this entry.