



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 26, 2017 – 08:18 PM EDT

PDB ID : 5B2S  
Title : Crystal structure of the Streptococcus pyogenes Cas9 EQR variant in complex with sgRNA and target DNA (TGAG PAM)  
Authors : Hirano, S.; Nishimasu, H.; Ishitani, R.; Nureki, O.  
Deposited on : unknown  
Resolution : 2.20 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  
Mogul : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20030345  
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) : rb-20030345

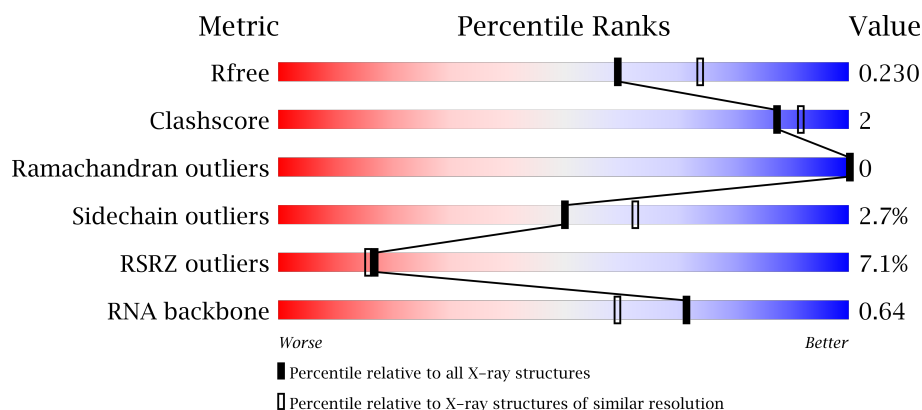
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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))
$R_{free}$	100719	4002 (2.20-2.20)
Clashscore	112137	4730 (2.20-2.20)
Ramachandran outliers	110173	4656 (2.20-2.20)
Sidechain outliers	110143	4657 (2.20-2.20)
RSRZ outliers	101464	4033 (2.20-2.20)
RNA backbone	2435	1007 (2.74-1.66)

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	81	<div> <div>2%</div> <div> <div></div> <div>79%</div> <div>20%</div> <div>.</div> </div> </div>
2	B	1372	<div> <div>7%</div> <div> <div></div> <div>88%</div> <div>8%</div> <div>.</div> </div> </div>
3	C	28	<div> <div></div> <div> <div></div> <div>79%</div> <div>21%</div> </div> </div>
4	D	8	<div> <div>13%</div> <div> <div></div> <div>100%</div> </div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	K	B	1408	-	-	-	X
7	EDO	B	1411	-	-	-	X
7	EDO	B	1412	-	-	-	X
8	ACT	A	107	-	-	-	X

## 2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 13535 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 RNA chain called Guide RNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	81	Total	C	N	O	P	0	0	0
			1739	778	319	561	81			

- Molecule 2 is a protein called CRISPR-associated endonuclease Cas9.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	1319	Total	C	N	O	S	0	1	0
			10494	6695	1820	1959	20			

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-3	GLY	-	expression tag	UNP Q99ZW2
B	-2	SER	-	expression tag	UNP Q99ZW2
B	-1	GLY	-	expression tag	UNP Q99ZW2
B	0	HIS	-	expression tag	UNP Q99ZW2
B	10	ALA	ASP	engineered mutation	UNP Q99ZW2
B	80	LEU	CYS	engineered mutation	UNP Q99ZW2
B	574	GLU	CYS	engineered mutation	UNP Q99ZW2
B	840	ALA	HIS	engineered mutation	UNP Q99ZW2
B	1135	GLU	ASP	engineered mutation	UNP Q99ZW2
B	1335	GLN	ARG	engineered mutation	UNP Q99ZW2
B	1337	ARG	THR	engineered mutation	UNP Q99ZW2

- Molecule 3 is a DNA chain called Target DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	28	Total	C	N	O	P	0	0	0
			560	269	100	164	27			

- Molecule 4 is a DNA chain called Non-target DNA, DNA (5'-D(\*TP\*GP\*AP\*GP\*AP\*TP\*TP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	8	Total	C	N	O	P	0	0	0
			165	80	31	47	7			

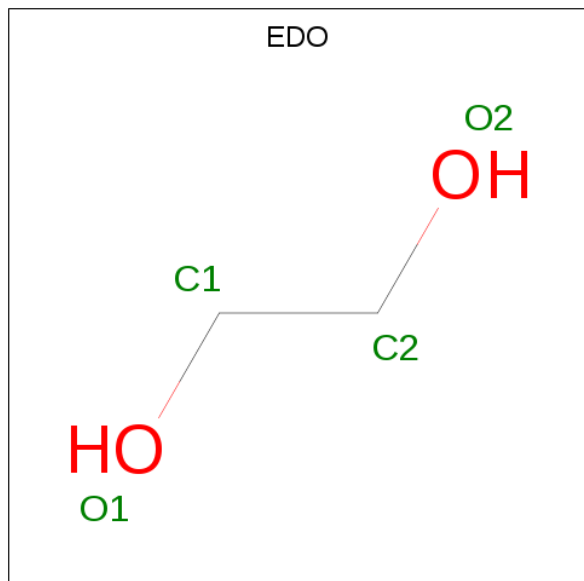
- Molecule 5 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	8	Total	K	0	0
			8	8		
5	A	3	Total	K	0	0
			3	3		

- Molecule 6 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	B	1	Total	Mg	0	0
			1	1		
6	A	2	Total	Mg	0	0
			2	2		

- Molecule 7 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



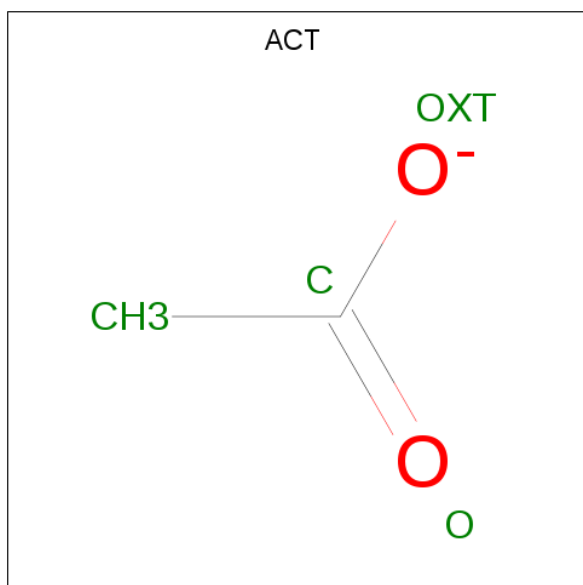
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	A	1	Total	C	O	0	0
			4	2	2		
7	B	1	Total	C	O	0	0
			4	2	2		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	B	1	Total	C	O	0	0
			4	2	2		
7	B	1	Total	C	O	0	0
			4	2	2		

- Molecule 8 is ACETATE ION (three-letter code: ACT) (formula:  $C_2H_3O_2$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
8	A	1	Total	C	O	0	0
			4	2	2		

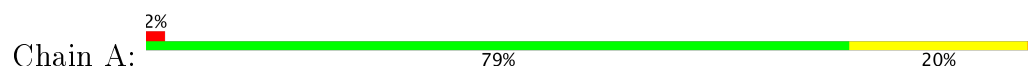
- Molecule 9 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	A	161	Total	O	0	0
			161	161		
9	B	343	Total	O	0	0
			343	343		
9	C	30	Total	O	0	0
			30	30		
9	D	9	Total	O	0	0
			9	9		

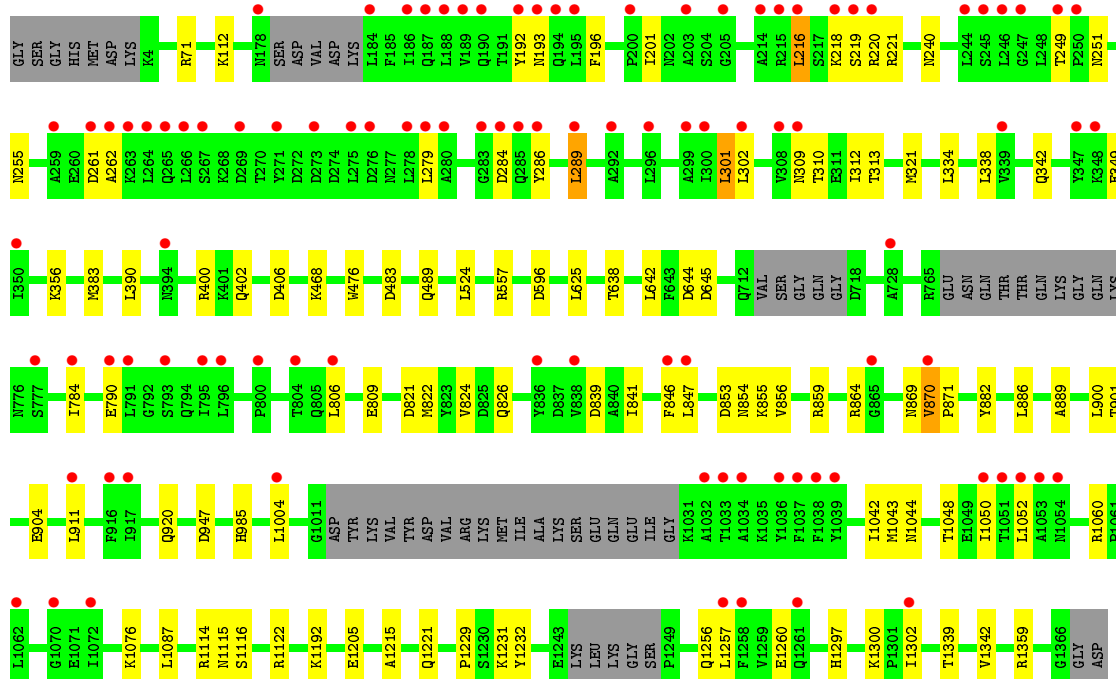
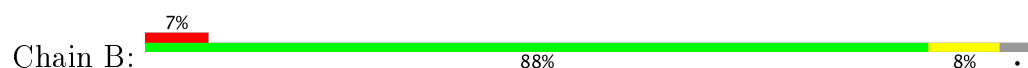
### 3 Residue-property plots [i](#)

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.

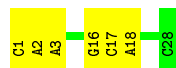
#### • Molecule 1: Guide RNA



#### • Molecule 2: CRISPR-associated endonuclease Cas9



#### • Molecule 3: Target DNA



#### • Molecule 4: Non-target DNA, DNA (5'-D(\*TP\*GP\*AP\*GP\*AP\*TP\*TP\*G)-3')

Chain D:  100%



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	178.02Å 67.82Å 187.57Å 90.00° 111.12° 90.00°	Depositor
Resolution (Å)	48.01 – 2.20 48.01 – 2.20	Depositor EDS
% Data completeness (in resolution range)	97.4 (48.01-2.20) 97.5 (48.01-2.20)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.81 (at 2.20Å)	Xtriage
Refinement program	PHENIX 1.10 _2155: ???	Depositor
R, $R_{free}$	0.207 , 0.231 0.205 , 0.230	Depositor DCC
$R_{free}$ test set	5161 reflections (4.98%)	DCC
Wilson B-factor (Å <sup>2</sup> )	40.4	Xtriage
Anisotropy	0.540	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 35.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	13535	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.31% 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 [i](#)

### 5.1 Standard geometry [i](#)

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

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	0.18	0/1949	0.71	0/3037
2	B	0.24	0/10682	0.40	0/14409
3	C	0.54	0/626	0.96	0/961
4	D	0.50	0/185	0.95	0/285
All	All	0.26	0/13442	0.52	0/18692

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 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	1739	0	870	6	0
2	B	10494	0	10384	49	0
3	C	560	0	316	5	0
4	D	165	0	93	0	0
5	A	3	0	0	0	0
5	B	8	0	0	0	0
6	A	2	0	0	0	0
6	B	1	0	0	0	0
7	A	4	0	6	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	B	12	0	18	0	0
8	A	4	0	3	0	0
9	A	161	0	0	0	0
9	B	343	0	0	0	0
9	C	30	0	0	0	0
9	D	9	0	0	0	0
All	All	13535	0	11690	58	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 58 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:18:G:N7	2:B:71:ARG:NH2	2.29	0.81
2:B:240:ASN:ND2	2:B:255:ASN:OD1	2.31	0.62
2:B:251:ASN:HD21	2:B:262:ALA:H	1.48	0.62
2:B:1004:LEU:HD11	2:B:1042:ILE:HD11	1.83	0.60
3:C:2:DA:H2'	3:C:3:DA:C8	2.36	0.60

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	
2	B	1308/1372 (95%)	1264 (97%)	44 (3%)	0	100	100

There are no Ramachandran outliers to report.

### 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
2	B	1106/1227 (90%)	1076 (97%)	30 (3%)	50 62

5 of 30 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	524	LEU
2	B	644	ASP
2	B	1192	LYS
2	B	642	LEU
2	B	645	ASP

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

Mol	Chain	Res	Type
2	B	726	ASN
2	B	926	GLN
2	B	1261	GLN
2	B	511	HIS
2	B	1115	ASN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	80/81 (98%)	8 (10%)	0

5 of 8 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	28	A
1	A	40	C
1	A	51	A
1	A	56	U
1	A	59	U

There are no RNA pucker outliers to report.

## 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](#)

Of 19 ligands modelled in this entry, 14 are monoatomic - leaving 5 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
7	EDO	A	106	-	3,3,3	0.45	0	2,2,2	0.40	0
8	ACT	A	107	-	1,3,3	1.16	0	0,3,3	0.00	-
7	EDO	B	1410	-	3,3,3	0.47	0	2,2,2	0.33	0
7	EDO	B	1411	-	3,3,3	0.46	0	2,2,2	0.35	0
7	EDO	B	1412	-	3,3,3	0.44	0	2,2,2	0.40	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
7	EDO	A	106	-	-	0/1/1/1	0/0/0/0
8	ACT	A	107	-	-	0/0/0/0	0/0/0/0
7	EDO	B	1410	-	-	0/1/1/1	0/0/0/0
7	EDO	B	1411	-	-	0/1/1/1	0/0/0/0
7	EDO	B	1412	-	-	0/1/1/1	0/0/0/0

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

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 [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	81/81 (100%)	0.14	2 (2%) 58 55	28, 42, 127, 182	0
2	B	1319/1372 (96%)	0.38	99 (7%) 15 14	26, 53, 106, 140	0
3	C	28/28 (100%)	-0.20	0 100 100	35, 47, 72, 87	0
4	D	8/8 (100%)	0.10	1 (12%) 4 4	37, 51, 97, 114	0
All	All	1436/1489 (96%)	0.35	102 (7%) 17 16	26, 53, 107, 182	0

The worst 5 of 102 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	244	LEU	7.9
2	B	247	GLY	7.4
2	B	216	LEU	7.4
2	B	189	VAL	6.6
2	B	286	TYR	6.2

### 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
5	K	B	1408	1/1	-0.11	0.44	15.43	160,160,160,160	0
8	ACT	A	107	4/4	0.94	0.24	6.39	50,50,50,51	0
7	EDO	B	1411	4/4	0.83	0.24	5.73	50,50,51,51	0
7	EDO	B	1412	4/4	0.95	0.18	4.00	39,42,45,47	0
7	EDO	B	1410	4/4	0.97	0.16	0.93	31,33,34,36	0
7	EDO	A	106	4/4	0.95	0.15	-0.73	30,32,34,35	0
5	K	B	1405	1/1	0.91	0.11	-0.94	81,81,81,81	0
5	K	B	1404	1/1	0.96	0.08	-2.13	52,52,52,52	0
5	K	B	1401	1/1	0.99	0.07	-2.56	43,43,43,43	0
5	K	B	1403	1/1	0.88	0.09	-2.95	74,74,74,74	0
5	K	A	102	1/1	0.99	0.10	-3.10	38,38,38,38	0
5	K	A	101	1/1	0.99	0.13	-3.36	28,28,28,28	0
6	MG	A	104	1/1	0.77	0.06	-7.74	43,43,43,43	0
5	K	B	1402	1/1	0.99	0.11	-	47,47,47,47	0
5	K	B	1406	1/1	0.71	0.13	-	93,93,93,93	0
6	MG	B	1409	1/1	0.92	0.13	-	41,41,41,41	0
5	K	B	1407	1/1	0.95	0.15	-	60,60,60,60	0
5	K	A	103	1/1	0.98	0.10	-	38,38,38,38	0
6	MG	A	105	1/1	0.84	0.26	-	49,49,49,49	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.