



wwPDB X-ray Structure Validation Summary Report ⓘ

Apr 14, 2025 – 04:53 PM JST

PDB ID : 8KAK / pdb_00008kak
Title : Crystal structure of SpyCas9 in complex with sgRNA and 18nt target DNA
Authors : Chen, Y.; Chen, J.; Liu, L.
Deposited on : 2023-08-03
Resolution : 3.60 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
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	2.0rc1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.006 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.42

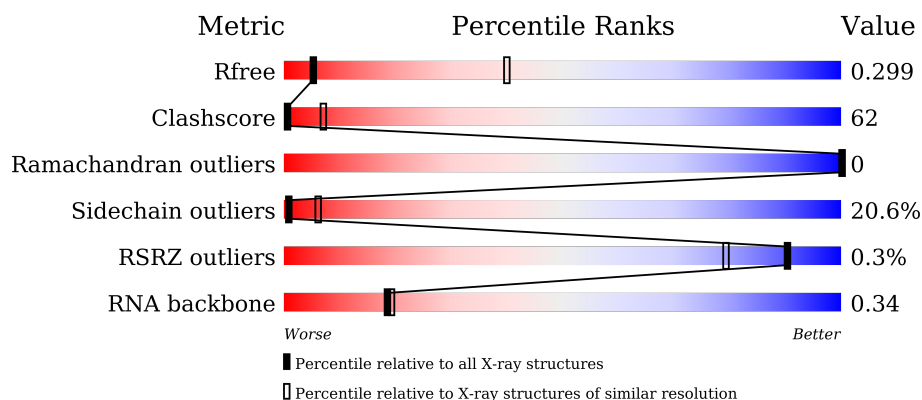
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.60 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.



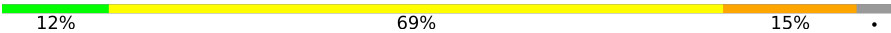
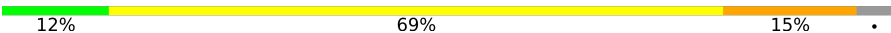


Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	1563 (3.70-3.50)
Clashscore	180529	1665 (3.70-3.50)
Ramachandran outliers	177936	1641 (3.70-3.50)
Sidechain outliers	177891	1640 (3.70-3.50)
RSRZ outliers	164620	1562 (3.70-3.50)
RNA backbone	3690	1108 (4.20-3.00)

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	98	<div> <div>12%</div> <div>47%</div> <div>32%</div> <div>5%</div> <div>.</div> </div>
1	E	98	<div> <div>12%</div> <div>45%</div> <div>34%</div> <div>5%</div> <div>.</div> </div>
2	B	1368	<div> <div>29%</div> <div>53%</div> <div>14%</div> <div>..</div> </div>
2	F	1368	<div> <div>%</div> <div>30%</div> <div>52%</div> <div>14%</div> <div>..</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	C	26	
3	G	26	
4	D	11	
4	H	11	

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	SO4	B	1401	-	-	X	-

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 27148 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 RNA (98-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	94	Total	C	N	O	P	0	0	0
			2009	899	362	654	94			
1	E	94	Total	C	N	O	P	0	0	0
			2009	899	362	654	94			

- Molecule 2 is a protein called CRISPR-associated endonuclease Cas9/Csn1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	1326	Total	C	N	O	S	0	0	0
			10827	6897	1879	2029	22			
2	F	1326	Total	C	N	O	S	0	0	0
			10827	6897	1879	2029	22			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	10	ALA	ASP	engineered mutation	UNP Q99ZW2
B	840	ALA	HIS	engineered mutation	UNP Q99ZW2
F	10	ALA	ASP	engineered mutation	UNP Q99ZW2
F	840	ALA	HIS	engineered mutation	UNP Q99ZW2

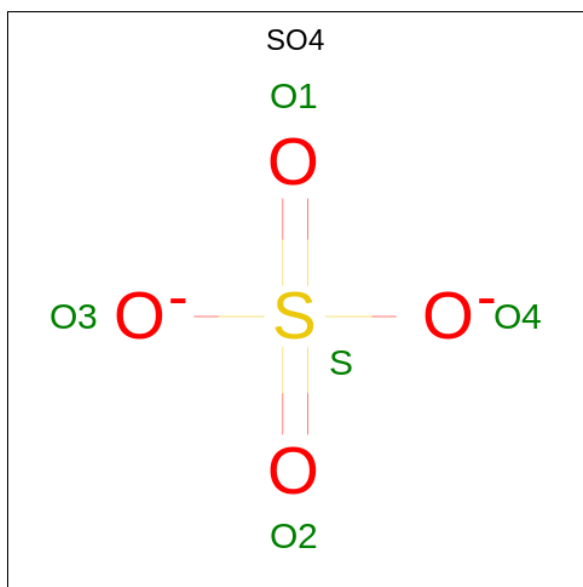
- Molecule 3 is a DNA chain called DNA (26-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	25	Total	C	N	O	P	0	0	0
			505	243	93	145	24			
3	G	25	Total	C	N	O	P	0	0	0
			505	243	93	145	24			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	11	Total	C	N	O	P	0	0	0
			225	110	37	68	10			
4	H	11	Total	C	N	O	P	0	0	0
			225	110	37	68	10			

- Molecule 5 is SULFATE ION (CCD ID: SO4) (formula: O₄S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	B	1	Total	O	S	0	0
			5	4	1		
5	F	1	Total	O	S	0	0
			5	4	1		

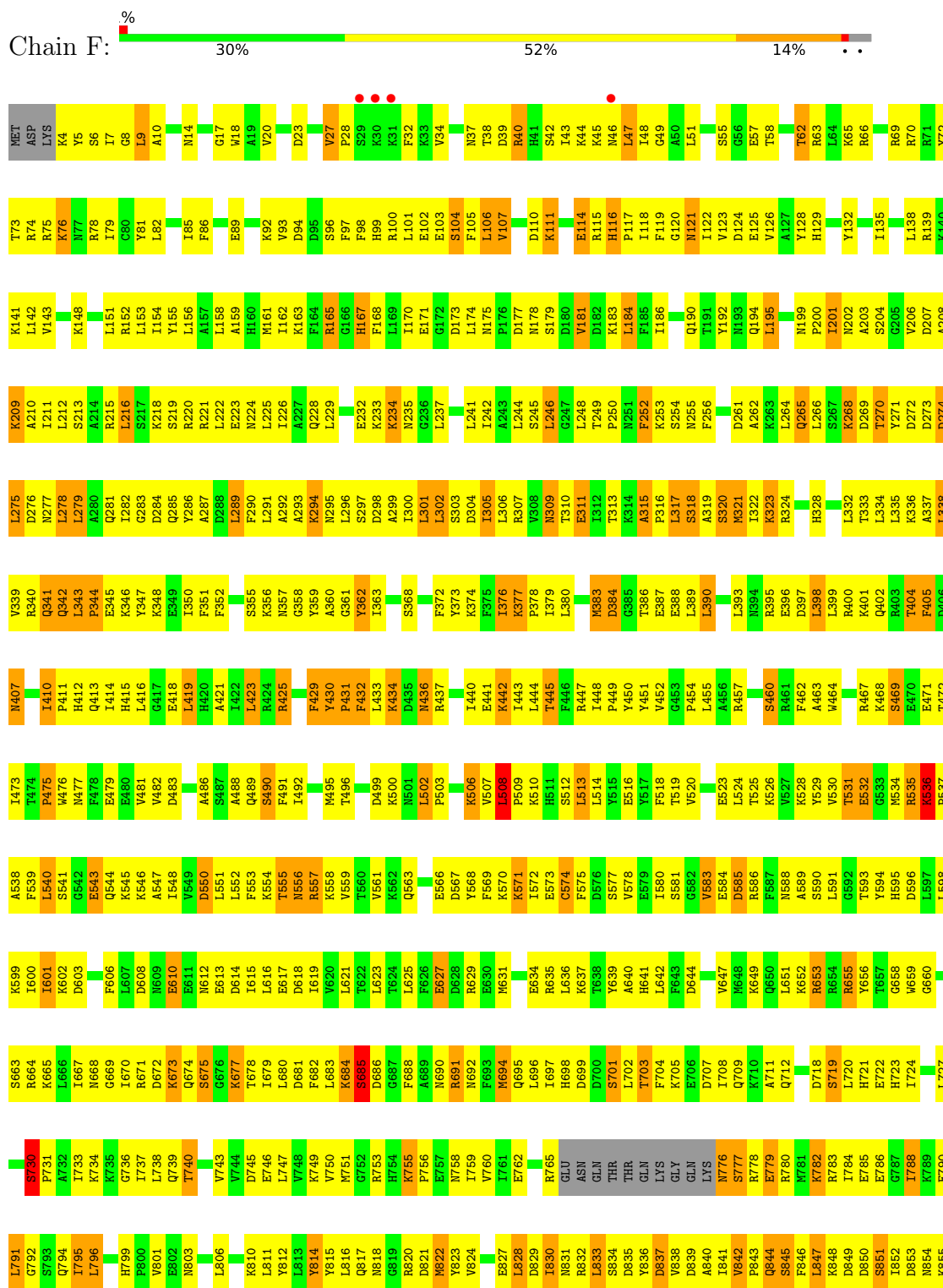
- Molecule 6 is water.

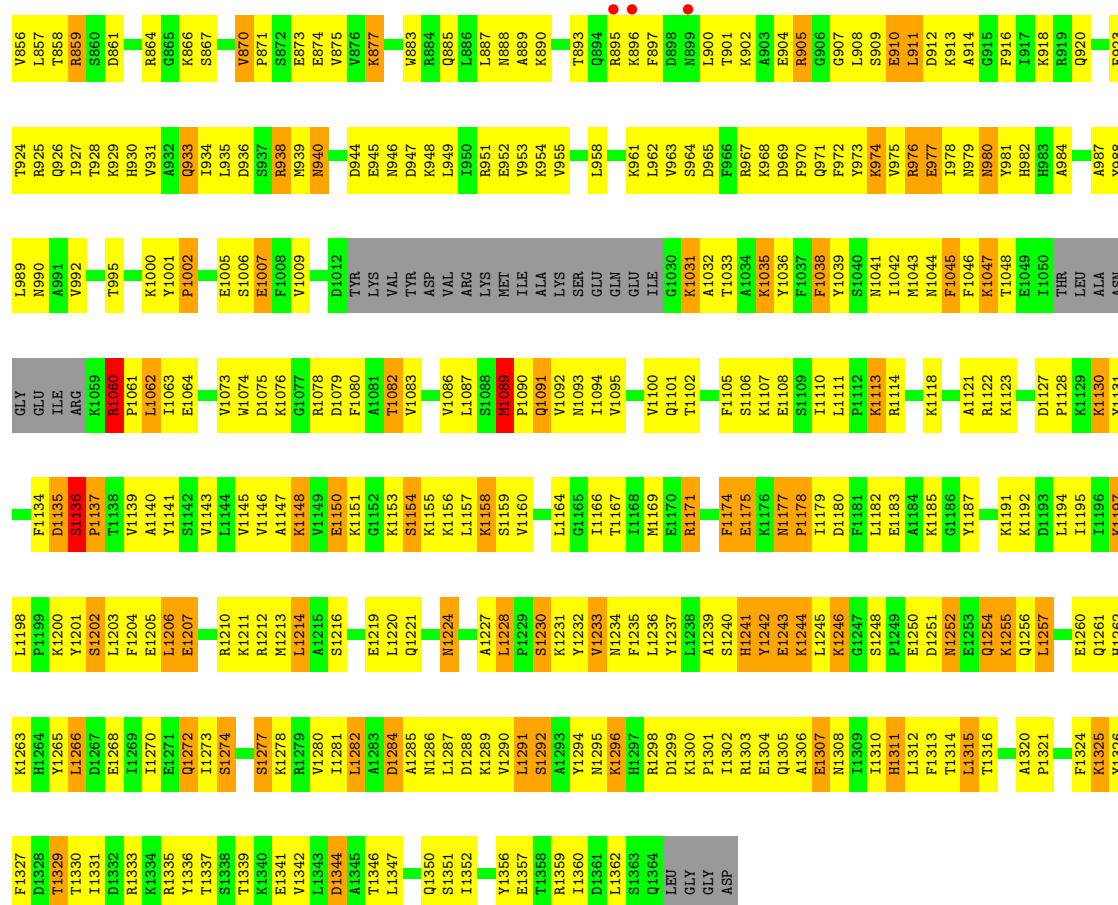
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	1	Total	O	0	0
			1	1		
6	B	1	Total	O	0	0
			1	1		
6	F	4	Total	O	0	0
			4	4		

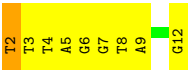
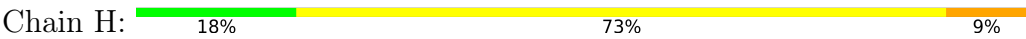




● Molecule 2: CRISPR-associated endonuclease Cas9/Csn1







4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	70.74Å 104.97Å 184.31Å 75.82° 78.67° 72.63°	Depositor
Resolution (Å)	24.97 – 3.60 24.97 – 3.60	Depositor EDS
% Data completeness (in resolution range)	73.7 (24.97-3.60) 73.9 (24.97-3.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.46 (at 3.57Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.265 , 0.297 0.264 , 0.299	Depositor DCC
R_{free} test set	2067 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å ²)	30.0	Xtriage
Anisotropy	0.942	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27 , 19.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.42$, $\langle L^2 \rangle = 0.24$	Xtriage
Estimated twinning fraction	0.107 for h,h-k,h-l	Xtriage
F_o, F_c correlation	0.89	EDS
Total number of atoms	27148	wwPDB-VP
Average B, all atoms (Å ²)	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.28% 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: SO4

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	0.87	1/2249 (0.0%)	1.58	39/3503 (1.1%)
1	E	0.87	1/2249 (0.0%)	1.58	41/3503 (1.2%)
2	B	0.61	11/11019 (0.1%)	0.74	31/14807 (0.2%)
2	F	0.61	10/11019 (0.1%)	0.74	32/14807 (0.2%)
3	C	1.25	2/566 (0.4%)	1.24	6/870 (0.7%)
3	G	1.25	2/566 (0.4%)	1.24	6/870 (0.7%)
4	D	1.40	2/251 (0.8%)	1.28	0/387
4	H	1.40	2/251 (0.8%)	1.27	0/387
All	All	0.72	31/28170 (0.1%)	0.98	155/39134 (0.4%)

The worst 5 of 31 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	3	DT	N1-C2	5.79	1.42	1.38
4	H	3	DT	N1-C2	5.78	1.42	1.38
2	F	27	VAL	CA-C	5.60	1.67	1.52
2	B	27	VAL	CA-C	5.57	1.67	1.52
3	G	8	DT	C3'-O3'	5.53	1.51	1.44

The worst 5 of 155 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	F	685	SER	N-CA-C	12.54	144.85	111.00
2	F	684	LYS	N-CA-C	-9.21	86.15	111.00
2	B	684	LYS	N-CA-C	-9.20	86.15	111.00
1	E	15	U	N3-C4-C5	8.90	119.94	114.60
1	A	15	U	N3-C4-C5	8.88	119.93	114.60

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	2009	0	1009	109	0
1	E	2009	0	1009	111	0
2	B	10827	0	10973	1508	0
2	F	10827	0	10972	1554	0
3	C	505	0	283	31	0
3	G	505	0	283	31	0
4	D	225	0	129	18	0
4	H	225	0	129	19	0
5	B	5	0	0	3	0
5	F	5	0	0	1	0
6	A	1	0	0	0	0
6	B	1	0	0	0	0
6	F	4	0	0	0	0
All	All	27148	0	24787	3189	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 62.

The worst 5 of 3189 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)
2:F:181:VAL:HG11	2:F:300:ILE:CD1	1.14	1.57
2:B:181:VAL:HG11	2:B:300:ILE:CD1	1.14	1.57
2:F:342:GLN:HG3	2:F:383:MET:CE	1.09	1.54
2:F:1270:ILE:HD12	2:F:1294:TYR:CE2	1.37	1.54
2:B:181:VAL:CG1	2:B:300:ILE:CD1	1.83	1.53

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	
2	B	1318/1368 (96%)	1279 (97%)	39 (3%)	0	100	100
2	F	1318/1368 (96%)	1281 (97%)	37 (3%)	0	100	100
All	All	2636/2736 (96%)	2560 (97%)	76 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	
2	B	1186/1225 (97%)	942 (79%)	244 (21%)	1	6
2	F	1186/1225 (97%)	942 (79%)	244 (21%)	1	6
All	All	2372/2450 (97%)	1884 (79%)	488 (21%)	1	6

5 of 488 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	1284	ASP
2	F	1150	GLU
2	F	302	LEU
2	F	1113	LYS
2	F	1277	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 47 such sidechains are listed below:

Mol	Chain	Res	Type
2	F	295	ASN
2	F	650	GLN
2	F	341	GLN
2	F	412	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	F	776	ASN

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	93/98 (94%)	31 (33%)	4 (4%)
1	E	93/98 (94%)	31 (33%)	4 (4%)
All	All	186/196 (94%)	62 (33%)	8 (4%)

5 of 62 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	9	U
1	A	11	C
1	A	20	G
1	A	24	U
1	A	27	G

5 of 8 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	E	68	A
1	E	42	A
1	E	8	G
1	A	68	A
1	E	27	G

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 oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

2 ligands are modelled in this entry.

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	$\# Z > 2$
5	SO4	F	1401	-	4,4,4	0.98	0	6,6,6	1.66	1 (16%)
5	SO4	B	1401	-	4,4,4	0.98	0	6,6,6	1.66	1 (16%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
5	B	1401	SO4	O4-S-O3	3.83	125.39	109.06
5	F	1401	SO4	O4-S-O3	3.83	125.39	109.06

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	F	1401	SO4	1	0
5	B	1401	SO4	3	0

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, 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	94/98 (95%)	-1.68	0	100	100	7, 25, 69, 93	0
1	E	94/98 (95%)	-1.65	0	100	100	11, 39, 87, 112	0
2	B	1326/1368 (96%)	-1.27	1 (0%)	92	88	5, 26, 65, 117	0
2	F	1326/1368 (96%)	-1.21	7 (0%)	87	72	11, 40, 74, 131	0
3	C	25/26 (96%)	-1.28	0	100	100	9, 17, 64, 79	0
3	G	25/26 (96%)	-0.94	0	100	100	19, 29, 78, 84	0
4	D	11/11 (100%)	-1.42	0	100	100	18, 32, 95, 99	0
4	H	11/11 (100%)	-1.07	0	100	100	27, 56, 102, 112	0
All	All	2912/3006 (96%)	-1.26	8 (0%)	90	79	5, 33, 74, 131	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	543	GLU	3.9
2	F	29	SER	3.9
2	F	31	LYS	3.1
2	F	896	LYS	2.6
2	F	46	ASN	2.5

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	SO4	F	1401	5/5	0.99	0.03	30,30,30,30	0
5	SO4	B	1401	5/5	1.00	0.03	30,30,30,30	0

6.5 Other polymers [i](#)

There are no such residues in this entry.