wwPDB X-ray Structure Validation Summary Report (i)

Jul 18, 2022 – 04:17 pm BST

PDB ID	:	7P4C
Title	:	Crystal Structure of Agd31B, alpha-transglucosylase in Glycoside Hydrolase Family 31, in complex with noncovalent Cyclophellitol Sulfamidate probe KK131
Authors	:	Wu, L.; Davies, G.J.
Deposited on	:	2021-07-11
Resolution	:	1.86 Å(reported)

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

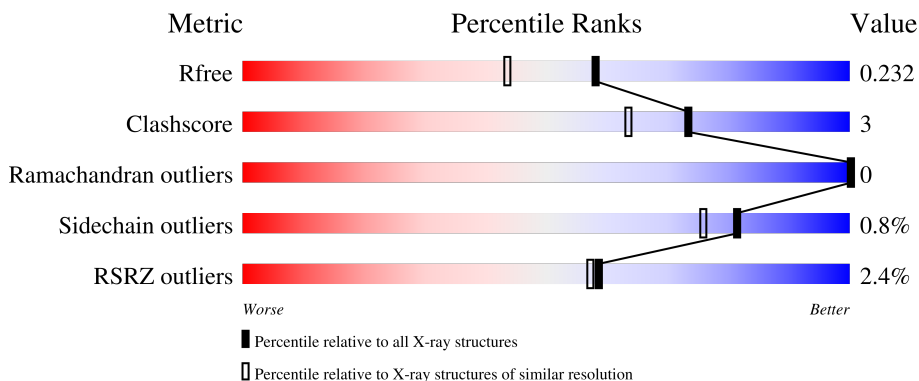|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.29
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.29

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 1.86 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	Percentile Ranks	Value
Rfree		0.232
Clashscore		3
Ramachandran outliers		0
Sidechain outliers		0.8%
RSRZ outliers		2.4%

Worse   *Better*

▊ Percentile relative to all X-ray structures

▊ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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 <=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	AAA	836	2% 84% 9% 7%

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 12946 atoms, of which 6157 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 Oligosaccharide 4-alpha-D-glucosyltransferase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	AAA	780	12395	4008	6120	1075	1171	21	172	5	0

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	24	MET	-	initiating methionine	UNP B3PEE6
AAA	817	LYS	-	expression tag	UNP B3PEE6
AAA	818	GLY	-	expression tag	UNP B3PEE6
AAA	819	GLY	-	expression tag	UNP B3PEE6
AAA	820	ARG	-	expression tag	UNP B3PEE6
AAA	821	ALA	-	expression tag	UNP B3PEE6
AAA	822	ASP	-	expression tag	UNP B3PEE6
AAA	823	PRO	-	expression tag	UNP B3PEE6
AAA	824	ALA	-	expression tag	UNP B3PEE6
AAA	825	PHE	-	expression tag	UNP B3PEE6
AAA	826	LEU	-	expression tag	UNP B3PEE6
AAA	827	TYR	-	expression tag	UNP B3PEE6
AAA	828	LYS	-	expression tag	UNP B3PEE6
AAA	829	VAL	-	expression tag	UNP B3PEE6
AAA	830	VAL	-	expression tag	UNP B3PEE6
AAA	831	ILE	-	expression tag	UNP B3PEE6
AAA	832	ASN	-	expression tag	UNP B3PEE6
AAA	833	SER	-	expression tag	UNP B3PEE6
AAA	834	LYS	-	expression tag	UNP B3PEE6
AAA	835	LEU	-	expression tag	UNP B3PEE6
AAA	836	GLU	-	expression tag	UNP B3PEE6
AAA	837	GLY	-	expression tag	UNP B3PEE6
AAA	838	LYS	-	expression tag	UNP B3PEE6
AAA	839	PRO	-	expression tag	UNP B3PEE6
AAA	840	ILE	-	expression tag	UNP B3PEE6
AAA	841	PRO	-	expression tag	UNP B3PEE6
AAA	842	ASN	-	expression tag	UNP B3PEE6

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	843	PRO	-	expression tag	UNP B3PEE6
AAA	844	LEU	-	expression tag	UNP B3PEE6
AAA	845	LEU	-	expression tag	UNP B3PEE6
AAA	846	GLY	-	expression tag	UNP B3PEE6
AAA	847	LEU	-	expression tag	UNP B3PEE6
AAA	848	ASP	-	expression tag	UNP B3PEE6
AAA	849	SER	-	expression tag	UNP B3PEE6
AAA	850	THR	-	expression tag	UNP B3PEE6
AAA	851	ARG	-	expression tag	UNP B3PEE6
AAA	852	THR	-	expression tag	UNP B3PEE6
AAA	853	GLY	-	expression tag	UNP B3PEE6
AAA	854	HIS	-	expression tag	UNP B3PEE6
AAA	855	HIS	-	expression tag	UNP B3PEE6
AAA	856	HIS	-	expression tag	UNP B3PEE6
AAA	857	HIS	-	expression tag	UNP B3PEE6
AAA	858	HIS	-	expression tag	UNP B3PEE6
AAA	859	HIS	-	expression tag	UNP B3PEE6

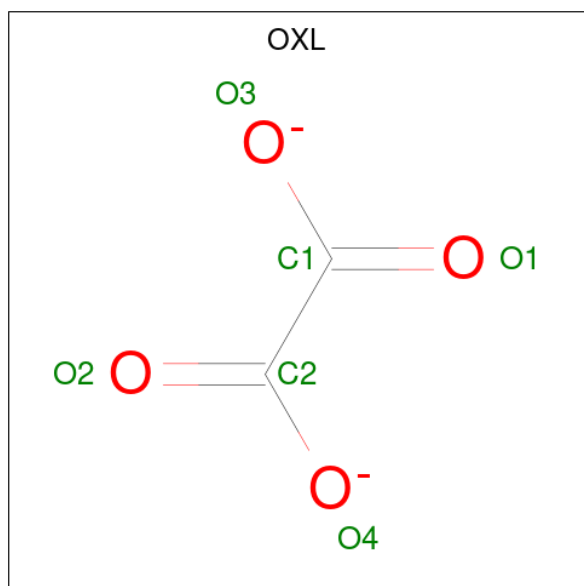- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	AAA	1	Total	C	H	O	1	0
			10	2	6	2		
2	AAA	1	Total	C	H	O	1	0
			10	2	6	2		
2	AAA	1	Total	C	H	O	1	0
			10	2	6	2		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	AAA	1	Total 10	C 2	H 6	O 2	1	0

- Molecule 3 is OXALATE ION (three-letter code: OXL) (formula: C_2O_4).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	AAA	1	Total 6	C 2	O 4	0	0

- Molecule 4 is (3aR,4S,5S,6R,7R,7aS)-7-(hydroxymethyl)-2,2-bis(oxidanylidene)-3a,4,5,6,7, 7a-hexahydro-3H-benzo[d][1,2,3]oxathiazole-4,5,6-triol (three-letter code: 5OV) (formula: $C_7H_{13}NO_7S$) (labeled as "Ligand of Interest" by depositor).
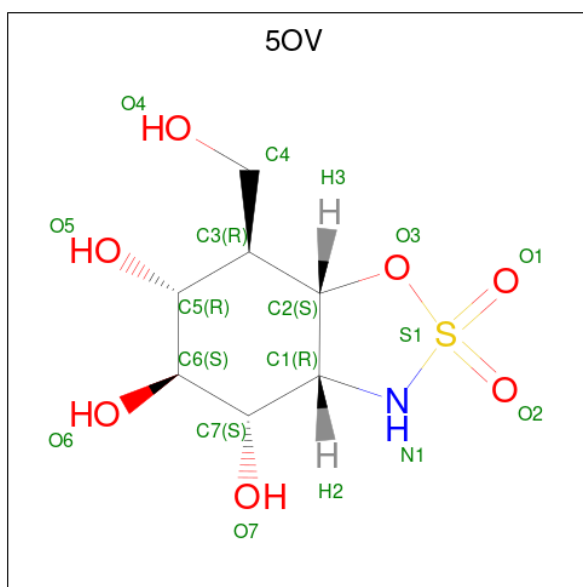
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
4	AAA	1	Total 29	C 7	H 13	N 1	O 7	S 1	4	0

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O_4S).

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

Continued on next page...

Continued from previous page...

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

• Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	AAA	431	Total 431	O 431	0	0

3 Residue-property plots (i)

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: Oligosaccharide 4-alpha-D-glucosyltransferase
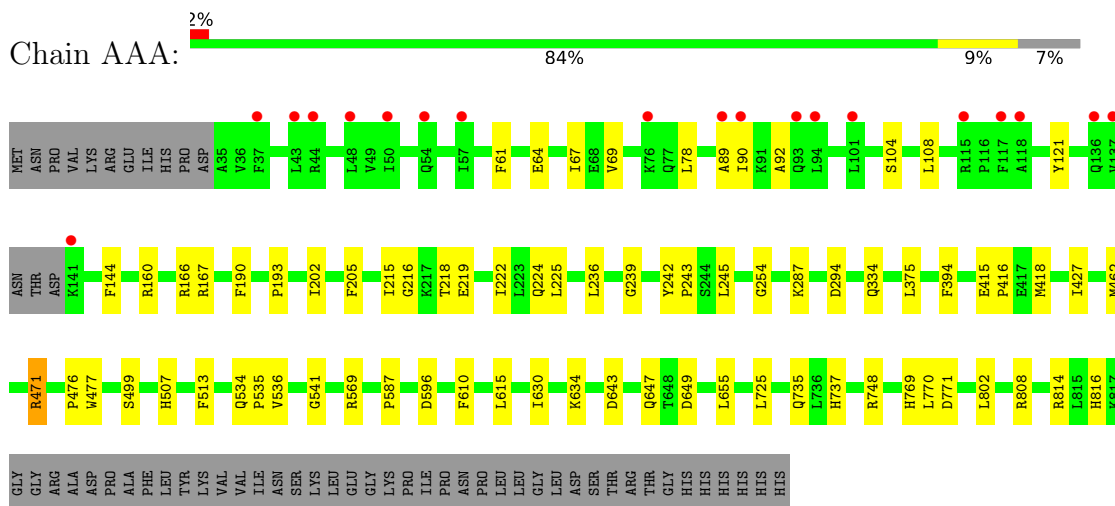
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 6 2 2	Depositor
Cell constants a, b, c, α, β, γ	197.00Å 197.00Å 102.16Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	102.31 – 1.86 102.16 – 1.86	Depositor EDS
% Data completeness (in resolution range)	100.0 (102.31-1.86) 100.0 (102.16-1.86)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.43 (at 1.86Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.192 , 0.229 0.198 , 0.232	Depositor DCC
R_{free} test set	4827 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å^2)	34.7	Xtriage
Anisotropy	0.476	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	12946	wwPDB-VP
Average B, all atoms (Å^2)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.35% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

5   Model quality ⓘ

5.1   Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: OXL, SO4, EDO, 5OV

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	AAA	0.75	0/6442	0.87	1/8734 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	AAA	471	ARG	NE-CZ-NH1	6.39	123.49	120.30

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	AAA	6275	6120	6093	41	0
2	AAA	16	24	24	0	0
3	AAA	6	0	0	0	0
4	AAA	16	13	0	0	0
5	AAA	45	0	0	1	0
6	AAA	431	0	0	6	0
All	All	6789	6157	6117	41	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 3.

The worst 5 of 41 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:AAA:569:ARG:HD2	6:AAA:1214:HOH:O	1.93	0.69
1:AAA:596:ASP:OD2	6:AAA:1001:HOH:O	2.16	0.61
1:AAA:239:GLY:HA2	5:AAA:914:SO4:O3	2.08	0.54
1:AAA:287:LYS:NZ	6:AAA:1010:HOH:O	2.41	0.53
1:AAA:205:PHE:HB2	1:AAA:225:LEU:HD13	1.91	0.52

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	AAA	781/836 (93%)	751 (96%)	30 (4%)	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	
1	AAA	662/706 (94%)	657 (99%)	5 (1%)	81	76

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AAA	190	PHE
1	AAA	294	ASP
1	AAA	394	PHE
1	AAA	477	TRP
1	AAA	748	ARG

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

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

15 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	AAA	914	-	4,4,4	0.34	0	6,6,6	0.12	0
5	SO4	AAA	912	-	4,4,4	0.36	0	6,6,6	0.07	0
4	5OV	AAA	905	-	14,17,17	0.64	0	18,27,27	1.22	1 (5%)
3	OXL	AAA	903	-	5,5,5	1.96	3 (60%)	6,6,6	1.20	1 (16%)
2	EDO	AAA	906	-	3,3,3	0.37	0	2,2,2	0.39	0
5	SO4	AAA	911	-	4,4,4	0.35	0	6,6,6	0.10	0

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#|Z| > 2$	Counts	RMSZ	$\#|Z| > 2$
5	SO4	AAA	915	-	4,4,4	0.34	0	6,6,6	0.06	0
5	SO4	AAA	909	-	4,4,4	0.44	0	6,6,6	0.18	0
2	EDO	AAA	904	-	3,3,3	0.12	0	2,2,2	0.03	0
2	EDO	AAA	902	-	3,3,3	0.29	0	2,2,2	0.32	0
5	SO4	AAA	908	-	4,4,4	0.42	0	6,6,6	0.16	0
5	SO4	AAA	913	-	4,4,4	0.33	0	6,6,6	0.14	0
2	EDO	AAA	901	-	3,3,3	0.14	0	2,2,2	0.40	0
5	SO4	AAA	907	-	4,4,4	0.38	0	6,6,6	0.22	0
5	SO4	AAA	910	-	4,4,4	0.33	0	6,6,6	0.13	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
4	5OV	AAA	905	-	-	0/2/36/36	0/2/2/2
3	OXL	AAA	903	-	-	0/4/4/4	-
2	EDO	AAA	906	-	-	1/1/1/1	-
2	EDO	AAA	904	-	-	0/1/1/1	-
2	EDO	AAA	902	-	-	1/1/1/1	-
2	EDO	AAA	901	-	-	1/1/1/1	-

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	AAA	903	OXL	O3-C1	-3.13	1.21	1.30
3	AAA	903	OXL	O4-C2	-2.20	1.24	1.30
3	AAA	903	OXL	C2-C1	2.09	1.60	1.54

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	AAA	905	5OV	O1-S1-N1	-3.92	106.13	111.57
3	AAA	903	OXL	O4-C2-C1	2.39	120.27	113.16

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	AAA	901	EDO	O1-C1-C2-O2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
2	AAA	902	EDO	O1-C1-C2-O2
2	AAA	906	EDO	O1-C1-C2-O2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	AAA	914	SO4	1	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand 5OV AAA 905

Bond lengths

Bond angles

Torsions

Rings
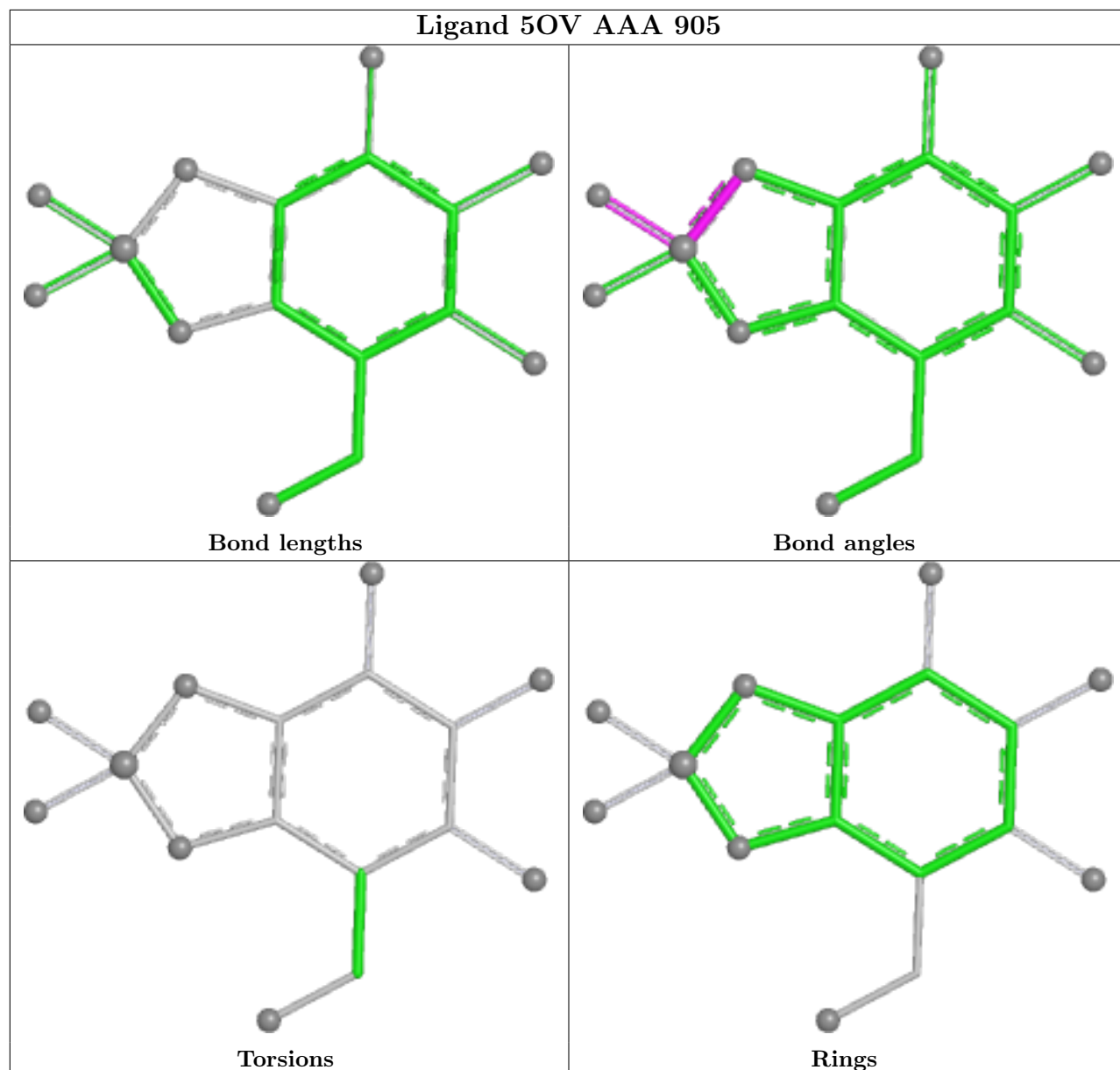
5.7 Other polymers (i)

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, 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	AAA	780/836 (93%)	0.04	19 (2%)	59	57	25, 39, 74, 112	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	54	GLN	4.4
1	AAA	76	LYS	3.5
1	AAA	93	GLN	3.2
1	AAA	50	ILE	3.0
1	AAA	44	ARG	3.0

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 ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

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.
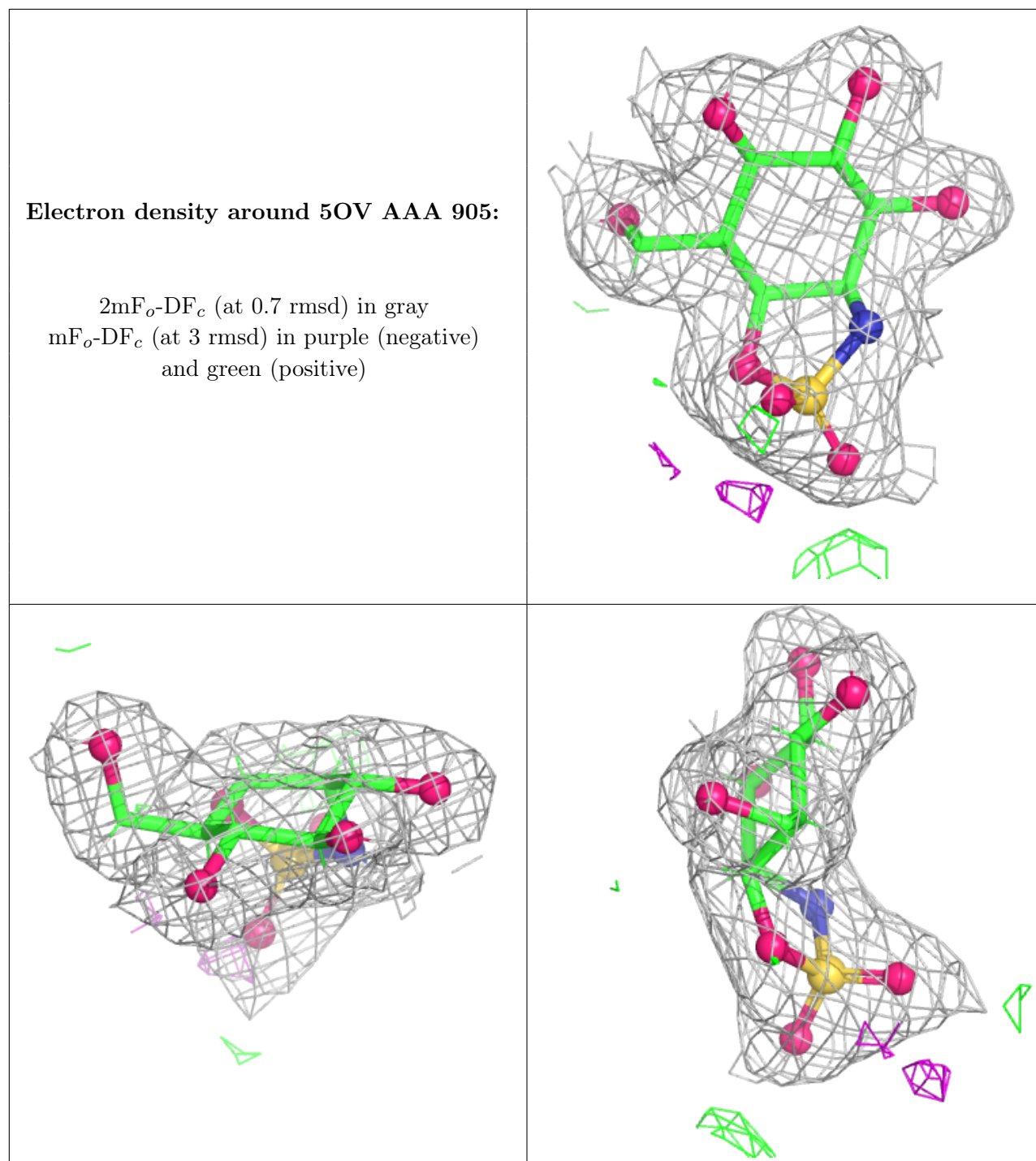
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	EDO	AAA	902	4/4	0.77	0.13	59,63,67,67	1
5	SO4	AAA	914	5/5	0.82	0.21	102,121,132,141	0

Continued on next page...

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	EDO	AAA	906	4/4	0.83	0.12	59,64,64,64	1
5	SO4	AAA	911	5/5	0.84	0.20	57,57,69,72	5
5	SO4	AAA	912	5/5	0.87	0.29	58,62,68,69	5
5	SO4	AAA	915	5/5	0.87	0.31	111,112,132,135	0
2	EDO	AAA	901	4/4	0.88	0.12	56,57,65,65	1
2	EDO	AAA	904	4/4	0.89	0.20	59,59,62,62	1
3	OXL	AAA	903	6/6	0.89	0.16	55,60,63,64	0
5	SO4	AAA	909	5/5	0.90	0.18	56,57,62,68	5
5	SO4	AAA	913	5/5	0.93	0.13	56,68,72,74	5
5	SO4	AAA	910	5/5	0.95	0.12	78,80,90,99	0
5	SO4	AAA	908	5/5	0.97	0.08	65,66,70,84	0
4	5OV	AAA	905	16/16	0.98	0.09	26,29,33,39	4
5	SO4	AAA	907	5/5	0.99	0.13	44,47,49,59	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around 5OV AAA 905:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)

6.5   Other polymers ⓘ

There are no such residues in this entry.