wwPDB X-ray Structure Validation Summary Report (i)

Jan 15, 2024 – 04:59 pm GMT

PDB ID	:	6SXU
Title	:	GH51 a-l-arabinofuranosidase soaked with cyclic sulfate inhibitor
Authors	:	McGregor, N.G.S.; Davies, G.J.
Deposited on	:	2019-09-26
Resolution	:	1.40 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 (i) 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 (i)) were used in the production of this report:

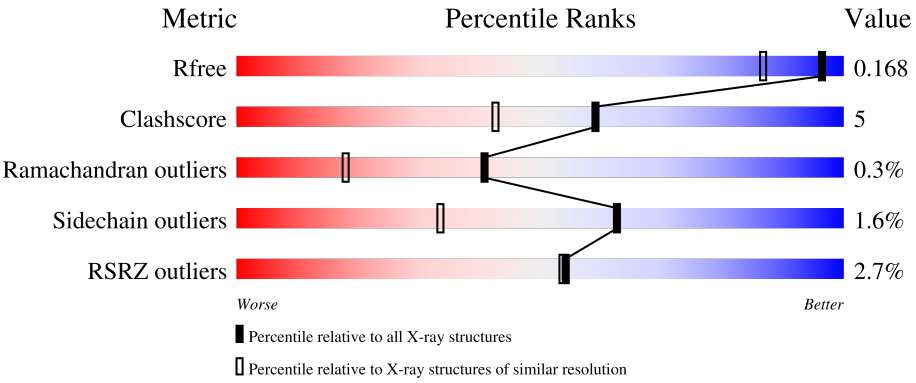
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.40 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	1714 (1.40-1.40)
Clashscore	141614	1812 (1.40-1.40)
Ramachandran outliers	138981	1763 (1.40-1.40)
Sidechain outliers	138945	1762 (1.40-1.40)
RSRZ outliers	127900	1674 (1.40-1.40)

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	502	3% 92% 7%
1	BBB	502	2% 91% 7%

2 Entry composition

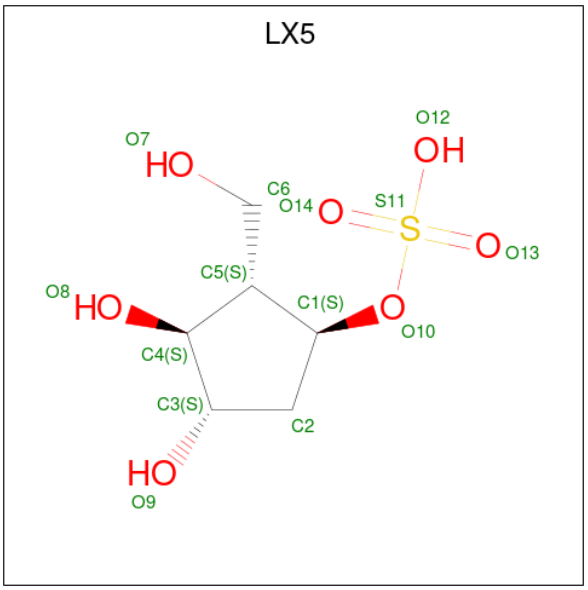
There are 6 unique types of molecules in this entry. The entry contains 16857 atoms, of which 7907 are hydrogens and 0 are deuteriums.

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 Intracellular exo-alpha-(1->5)-L-arabinofuranosidase.

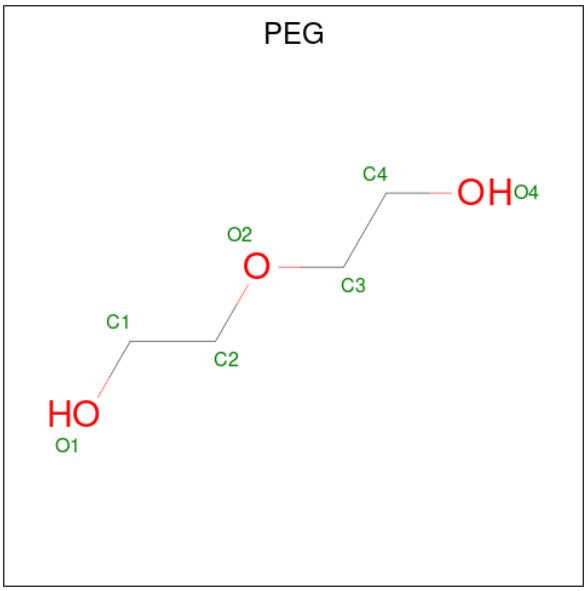
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	AAA	500	Total	C	H	N	O	S	222	13	0
			7873	2557	3901	664	727	24			
1	BBB	500	Total	C	H	N	O	S	224	16	0
			7904	2566	3916	667	731	24			

- Molecule 2 is [(1 {S},2 {S},3 {S},4 {S})-2-(hydroxymethyl)-3,4-bis(oxidanyl)cyclopentyl] hydrogen sulfate (three-letter code: LX5) (formula: $C_6H_{12}O_7S$) (labeled as "Ligand of Interest" by depositor).



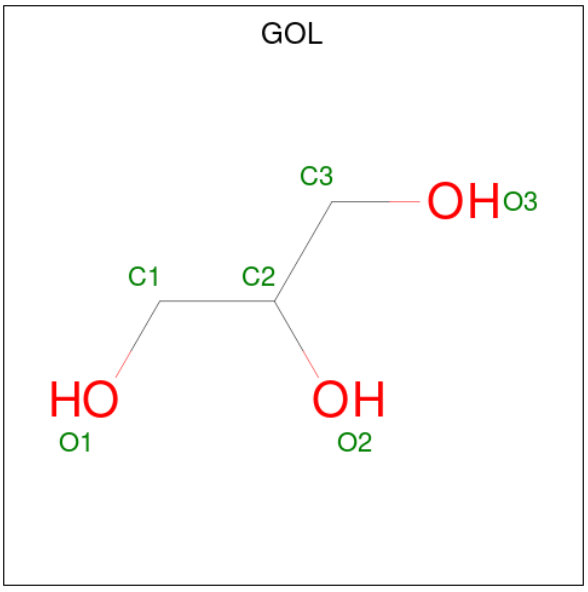
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	AAA	1	Total	C	H	O	S	3	0
			24	6	10	7	1		
2	BBB	1	Total	C	H	O	S	3	0
			24	6	10	7	1		

- Molecule 3 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	AAA	1	Total	C	H	O	1	0
			17	4	10	3		
3	AAA	1	Total	C	H	O	1	0
			17	4	10	3		
3	BBB	1	Total	C	H	O	1	0
			17	4	10	3		

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



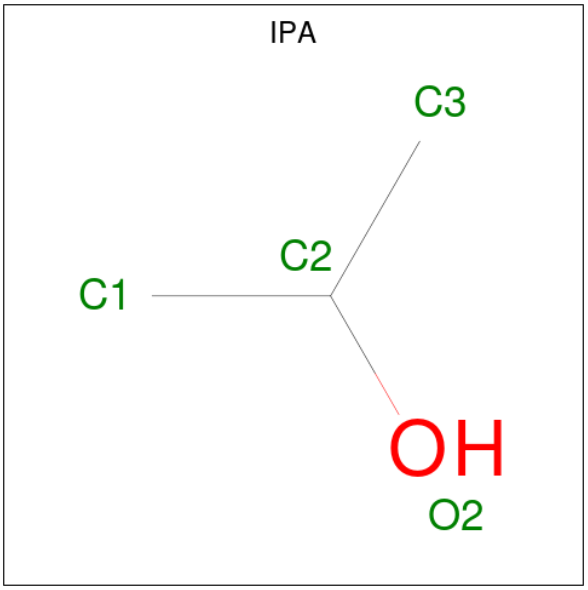
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	AAA	1	Total	C	H	O	2	0
			14	3	8	3		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	AAA	1	Total	C	H	O	2	0
			14	3	8	3		
4	BBB	1	Total	C	H	O	2	0
			14	3	8	3		
4	BBB	1	Total	C	H	O	2	0
			14	3	8	3		

- Molecule 5 is ISOPROPYL ALCOHOL (three-letter code: IPA) (formula: C_3H_8O).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	BBB	1	Total	C	H	O	0	0
			12	3	8	1		

- Molecule 6 is water.

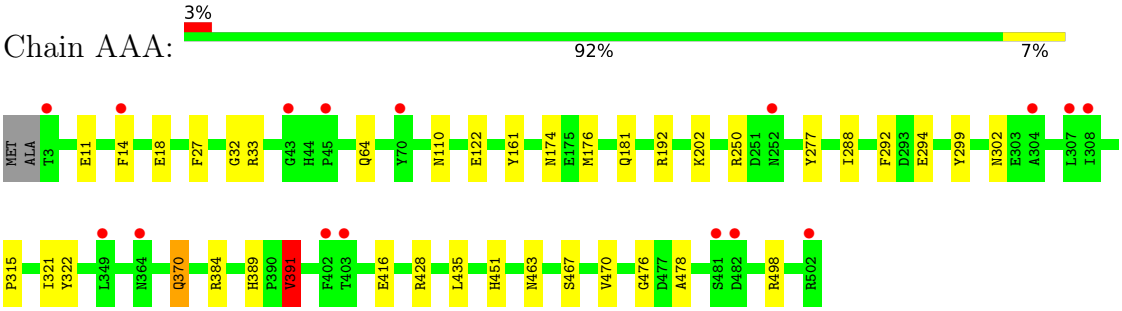
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	AAA	439	Total	O	0	1
			440	440		
6	BBB	472	Total	O	0	1
			473	473		

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.

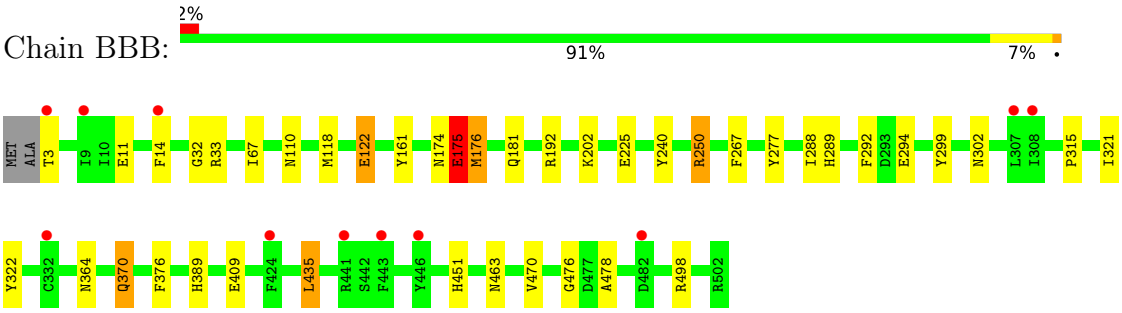
• Molecule 1: Intracellular exo-alpha-(1->5)-L-arabinofuranosidase

Chain AAA:



• Molecule 1: Intracellular exo-alpha-(1->5)-L-arabinofuranosidase

Chain BBB:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α, β, γ	178.47Å 178.47Å 100.41Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	29.75 – 1.40 29.75 – 1.40	Depositor EDS
% Data completeness (in resolution range)	99.8 (29.75-1.40) 99.8 (29.75-1.40)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.82 (at 1.40Å)	Xtriage
Refinement program	REFMAC 5.8.0253	Depositor
R, R_{free}	0.155 , 0.162 0.162 , 0.168	Depositor DCC
R_{free} test set	12421 reflections (5.28%)	wwPDB-VP
Wilson B-factor (Å^2)	14.5	Xtriage
Anisotropy	0.070	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 53.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.011 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.98	EDS
Total number of atoms	16857	wwPDB-VP
Average B, all atoms (Å^2)	18.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.92% 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.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: LX5, PEG, GOL, IPA

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	0.84	5/4114 (0.1%)	0.99	11/5601 (0.2%)
1	BBB	0.85	7/4139 (0.2%)	1.04	12/5636 (0.2%)
All	All	0.85	12/8253 (0.1%)	1.02	23/11237 (0.2%)

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	BBB	294	GLU	CD-OE1	8.73	1.35	1.25
1	AAA	294	GLU	CD-OE1	8.40	1.34	1.25
1	BBB	294	GLU	CB-CG	-7.87	1.37	1.52
1	AAA	294	GLU	CB-CG	-7.70	1.37	1.52
1	AAA	416	GLU	CD-OE2	-6.58	1.18	1.25

The worst 5 of 23 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	BBB	33[A]	ARG	NE-CZ-NH2	14.46	127.53	120.30
1	BBB	33[B]	ARG	NE-CZ-NH2	14.46	127.53	120.30
1	BBB	33[A]	ARG	NE-CZ-NH1	-12.97	113.82	120.30
1	BBB	33[B]	ARG	NE-CZ-NH1	-12.97	113.82	120.30
1	AAA	498	ARG	NE-CZ-NH2	-8.84	115.88	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	3972	3901	3823	41	0
1	BBB	3988	3916	3843	51	0
2	AAA	14	10	0	0	0
2	BBB	14	10	0	0	0
3	AAA	14	20	20	2	0
3	BBB	7	10	10	0	0
4	AAA	12	16	16	0	0
4	BBB	12	16	16	0	0
5	BBB	4	8	4	0	0
6	AAA	440	0	0	5	0
6	BBB	473	0	0	6	0
All	All	8950	7907	7732	74	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 5.

The worst 5 of 74 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:14[A]:PHE:CE2	1:BBB:14[A]:PHE:CE2	2.02	1.47
1:AAA:14[A]:PHE:CD2	1:BBB:14[A]:PHE:CZ	2.01	1.47
1:AAA:14[A]:PHE:HE2	1:BBB:14[A]:PHE:CE2	1.31	1.45
1:AAA:14[A]:PHE:CD2	1:BBB:14[A]:PHE:CE1	2.17	1.32
1:AAA:14[A]:PHE:HD2	1:BBB:14[A]:PHE:CZ	1.47	1.18

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
1	AAA	511/502 (102%)	490 (96%)	20 (4%)	1 (0%)	47 21
1	BBB	514/502 (102%)	493 (96%)	18 (4%)	3 (1%)	25 7
All	All	1025/1004 (102%)	983 (96%)	38 (4%)	4 (0%)	41 12

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	AAA	299	TYR
1	BBB	299	TYR
1	BBB	175[A]	GLU
1	BBB	175[B]	GLU

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	AAA	410/434 (94%)	403 (98%)	7 (2%)	60 31
1	BBB	414/434 (95%)	407 (98%)	7 (2%)	60 31
All	All	824/868 (95%)	810 (98%)	14 (2%)	62 31

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	BBB	176[A]	MET
1	BBB	176[B]	MET
1	BBB	435	LEU
1	BBB	292	PHE
1	BBB	370	GLN

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

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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)

10 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	$\#\|Z\| > 2$
4	GOL	BBB	604	-	5,5,5	0.39	0	5,5,5	0.55	0
3	PEG	AAA	602	-	6,6,6	0.18	0	5,5,5	0.72	0
2	LX5	BBB	601	1	14,14,14	1.99	4 (28%)	15,21,21	2.26	5 (33%)
3	PEG	AAA	605	-	6,6,6	0.37	0	5,5,5	0.21	0
4	GOL	BBB	603	-	5,5,5	0.51	0	5,5,5	0.61	0
3	PEG	BBB	602	-	6,6,6	0.57	0	5,5,5	0.89	0
5	IPA	BBB	605	-	3,3,3	2.26	1 (33%)	3,3,3	0.79	0
4	GOL	AAA	604	-	5,5,5	0.52	0	5,5,5	0.51	0
2	LX5	AAA	601	1	14,14,14	1.71	2 (14%)	15,21,21	2.05	6 (40%)
4	GOL	AAA	603	-	5,5,5	0.69	0	5,5,5	0.61	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	GOL	BBB	604	-	-	0/4/4/4	-
3	PEG	AAA	602	-	-	2/4/4/4	-
2	LX5	BBB	601	1	-	0/7/23/23	0/1/1/1
3	PEG	AAA	605	-	-	2/4/4/4	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	GOL	BBB	603	-	-	0/4/4/4	-
3	PEG	BBB	602	-	-	2/4/4/4	-
4	GOL	AAA	604	-	-	0/4/4/4	-
2	LX5	AAA	601	1	-	0/7/23/23	0/1/1/1
4	GOL	AAA	603	-	-	0/4/4/4	-

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	BBB	601	LX5	O10-C1	-4.43	1.40	1.47
2	AAA	601	LX5	O10-C1	-4.26	1.40	1.47
5	BBB	605	IPA	C1-C2	-3.90	1.22	1.48
2	BBB	601	LX5	C2-C1	-3.35	1.45	1.52
2	BBB	601	LX5	C3-C4	-2.98	1.48	1.53

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	AAA	601	LX5	C1-O10-S11	4.65	123.97	117.91
2	BBB	601	LX5	C1-O10-S11	4.62	123.94	117.91
2	BBB	601	LX5	O9-C3-C4	4.20	119.21	111.27
2	BBB	601	LX5	O14-S11-O13	3.05	124.47	112.22
2	BBB	601	LX5	O12-S11-O14	3.02	118.98	108.49

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	AAA	605	PEG	O2-C3-C4-O4
3	BBB	602	PEG	O2-C3-C4-O4
3	AAA	602	PEG	O1-C1-C2-O2
3	BBB	602	PEG	C4-C3-O2-C2
3	AAA	602	PEG	C1-C2-O2-C3

There are no ring outliers.

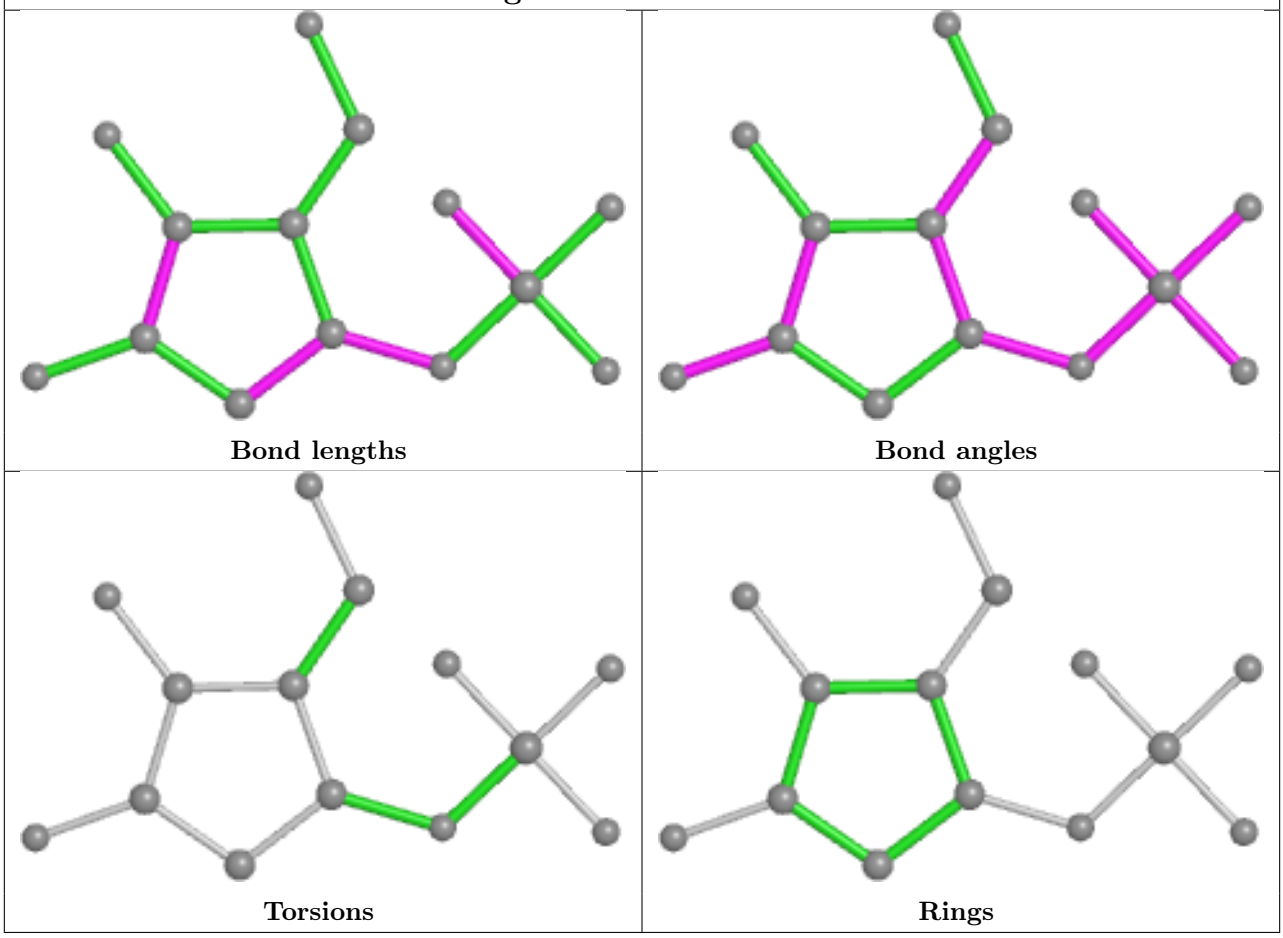
1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	AAA	605	PEG	2	0

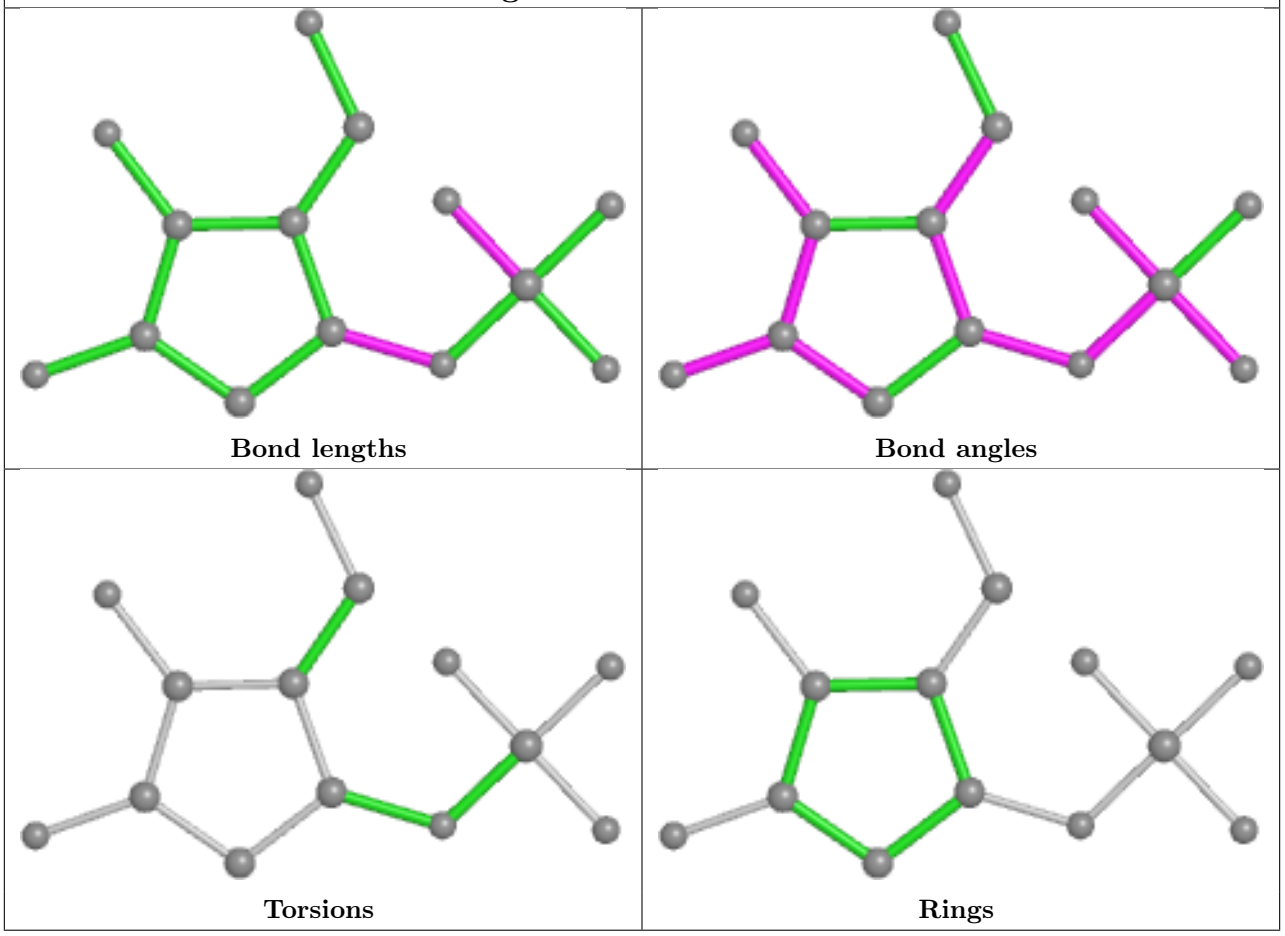
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths,

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 LX5 BBB 601



Ligand LX5 AAA 601



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, 95^{th} 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	500/502 (99%)	0.16	16 (3%)	47	46	10, 15, 32, 53	0
1	BBB	500/502 (99%)	0.15	11 (2%)	62	61	9, 15, 33, 52	0
All	All	1000/1004 (99%)	0.16	27 (2%)	54	54	9, 15, 32, 53	0

The worst 5 of 27 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	BBB	14[A]	PHE	5.0
1	AAA	14[A]	PHE	4.3
1	AAA	307	LEU	4.0
1	BBB	443	PHE	3.6
1	BBB	307	LEU	3.5

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

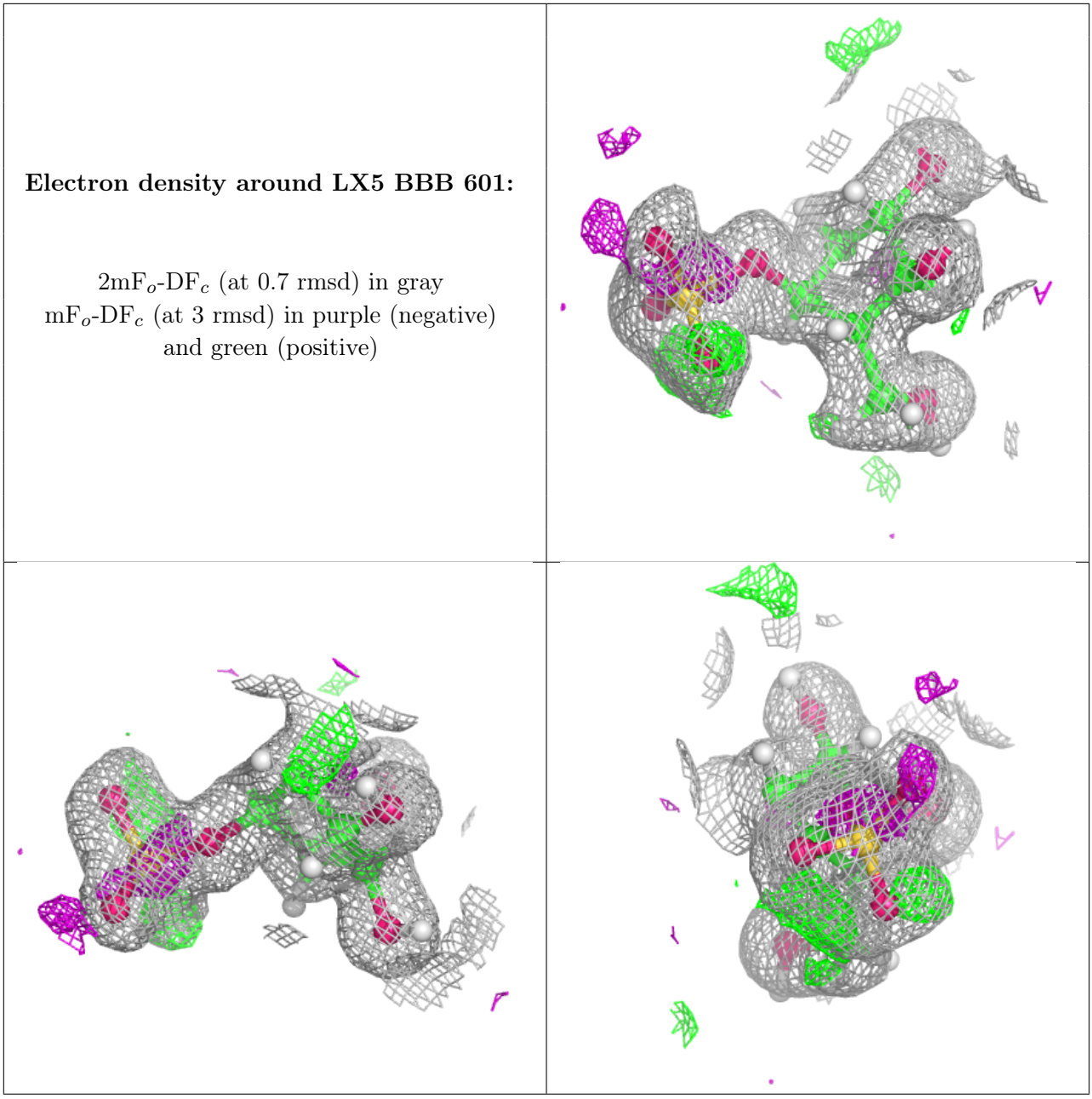
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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
3	PEG	AAA	605	7/7	0.68	0.16	48,64,78,78	1
5	IPA	BBB	605	4/4	0.80	0.29	21,29,39,39	0
3	PEG	BBB	602	7/7	0.81	0.23	20,22,37,37	17
3	PEG	AAA	602	7/7	0.85	0.23	20,22,38,38	17
4	GOL	AAA	603	6/6	0.86	0.18	23,25,32,32	2
4	GOL	BBB	604	6/6	0.88	0.13	25,27,35,35	2
4	GOL	BBB	603	6/6	0.89	0.17	23,24,31,31	2
4	GOL	AAA	604	6/6	0.90	0.12	26,28,34,34	2
2	LX5	BBB	601	14/14	0.94	0.10	10,14,26,32	3
2	LX5	AAA	601	14/14	0.95	0.09	11,16,25,29	3

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.

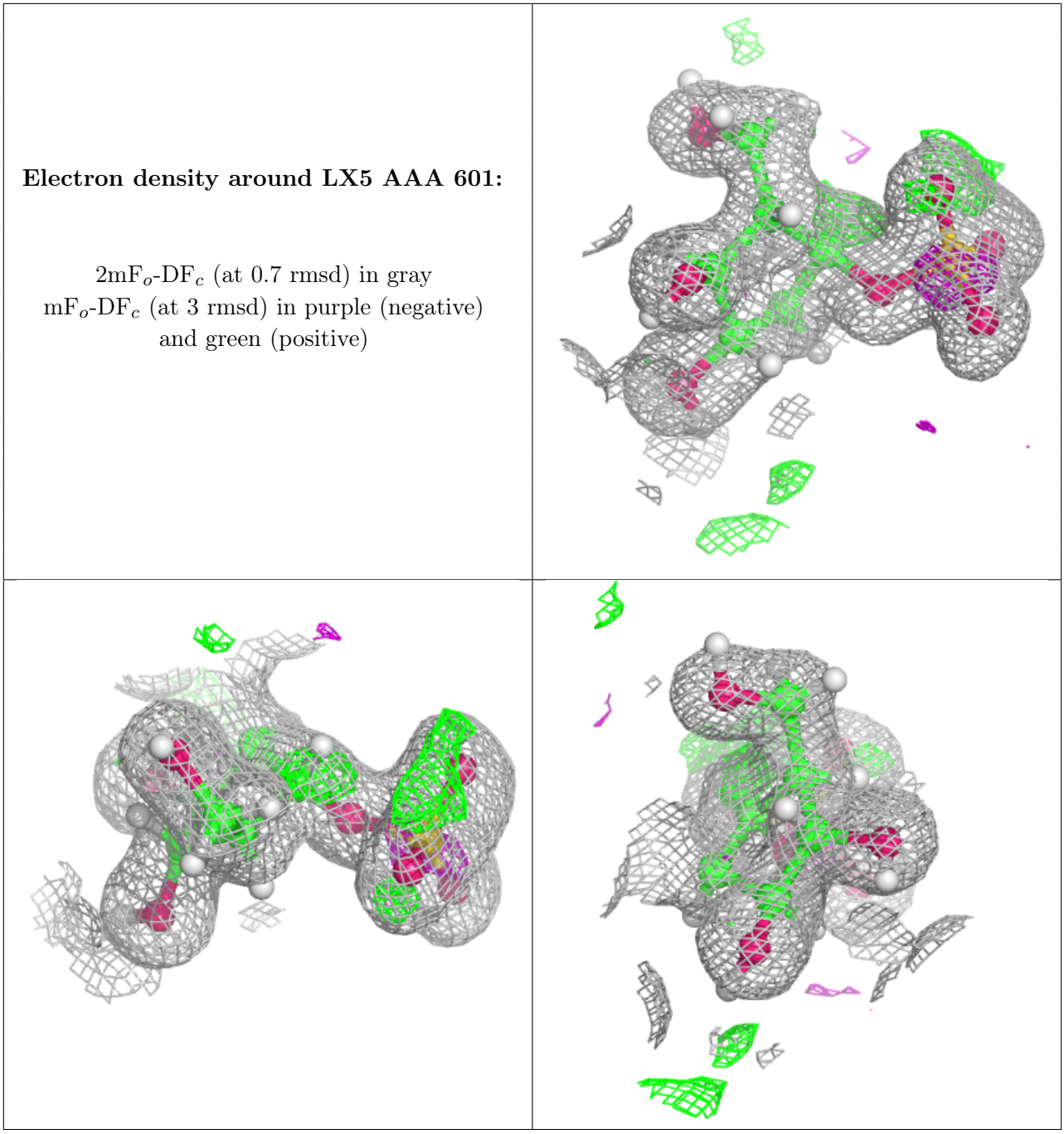
Electron density around LX5 BBB 601:

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



Electron density around LX5 AAA 601:

$2mF_o$-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.