Full wwPDB X-ray Structure Validation Report ⓘ

Jun 20, 2022 – 01:20 pm BST

PDB ID	:	7OZA
Title	:	Sulfated host glycan recognition by carbohydrate sulfatases of the human gut microbiota (BT3796_S1_16)
Authors	:	Cartmell, A.
Deposited on	:	2021-06-27
Resolution	:	1.50 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 following versions of software and data (see references ⓘ) were used in the production of this report:

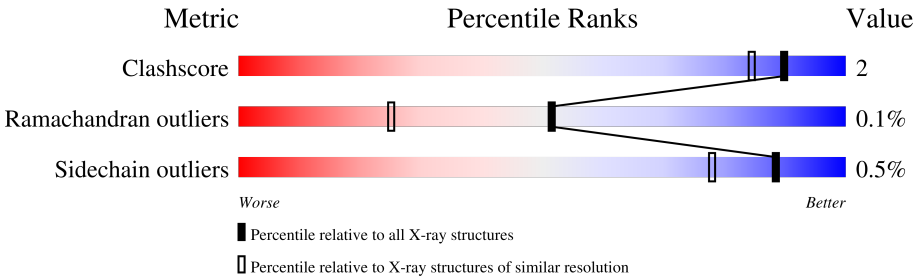
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	FAILED
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.28.1

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 1.50 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))
Clashscore	141614	3144 (1.50-1.50)
Ramachandran outliers	138981	3066 (1.50-1.50)
Sidechain outliers	138945	3064 (1.50-1.50)

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%

Note EDS failed to run properly.

Mol	Chain	Length	Quality of chain
1	AAA	522	92% · 7%
1	CCC	522	89% · 7%

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 8630 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 Putative secreted sulfatase ydeN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	487	Total	C	N	O	S	0	6	0
			3853	2451	646	735	21			
1	CCC	488	Total	C	N	O	S	0	12	0
			3923	2492	662	747	22			

There are 46 discrepancies between the modelled and reference sequences:

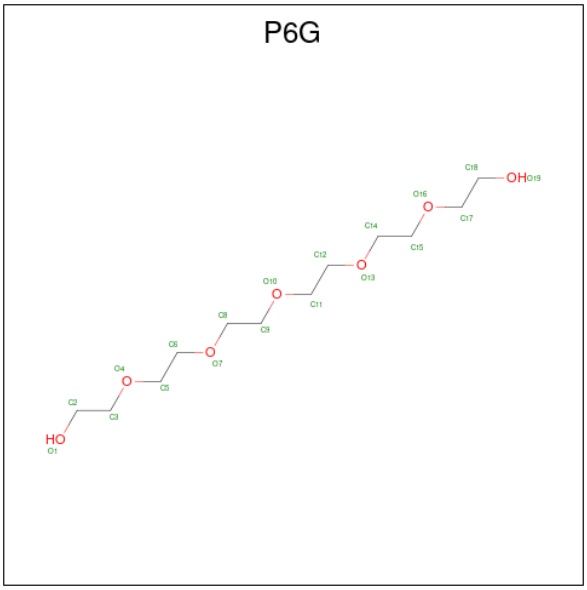
Chain	Residue	Modelled	Actual	Comment	Reference
AAA	-3	MET	-	initiating methionine	UNP Q8A171
AAA	-2	GLY	-	expression tag	UNP Q8A171
AAA	-1	SER	-	expression tag	UNP Q8A171
AAA	0	SER	-	expression tag	UNP Q8A171
AAA	1	HIS	-	expression tag	UNP Q8A171
AAA	2	HIS	-	expression tag	UNP Q8A171
AAA	3	HIS	-	expression tag	UNP Q8A171
AAA	4	HIS	-	expression tag	UNP Q8A171
AAA	5	HIS	-	expression tag	UNP Q8A171
AAA	6	HIS	-	expression tag	UNP Q8A171
AAA	7	SER	-	expression tag	UNP Q8A171
AAA	8	SER	-	expression tag	UNP Q8A171
AAA	9	GLY	-	expression tag	UNP Q8A171
AAA	10	LEU	-	expression tag	UNP Q8A171
AAA	11	VAL	-	expression tag	UNP Q8A171
AAA	12	PRO	-	expression tag	UNP Q8A171
AAA	13	ARG	-	expression tag	UNP Q8A171
AAA	14	GLY	-	expression tag	UNP Q8A171
AAA	15	SER	-	expression tag	UNP Q8A171
AAA	16	HIS	-	expression tag	UNP Q8A171
AAA	17	MET	-	expression tag	UNP Q8A171
AAA	18	ALA	-	expression tag	UNP Q8A171
AAA	19	SER	-	expression tag	UNP Q8A171
CCC	-3	MET	-	initiating methionine	UNP Q8A171
CCC	-2	GLY	-	expression tag	UNP Q8A171

Continued on next page...

Continued from previous page...

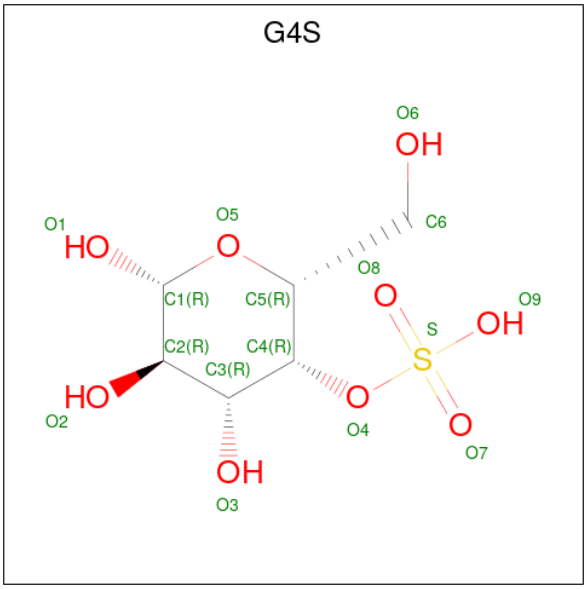
Chain	Residue	Modelled	Actual	Comment	Reference
CCC	-1	SER	-	expression tag	UNP Q8A171
CCC	0	SER	-	expression tag	UNP Q8A171
CCC	1	HIS	-	expression tag	UNP Q8A171
CCC	2	HIS	-	expression tag	UNP Q8A171
CCC	3	HIS	-	expression tag	UNP Q8A171
CCC	4	HIS	-	expression tag	UNP Q8A171
CCC	5	HIS	-	expression tag	UNP Q8A171
CCC	6	HIS	-	expression tag	UNP Q8A171
CCC	7	SER	-	expression tag	UNP Q8A171
CCC	8	SER	-	expression tag	UNP Q8A171
CCC	9	GLY	-	expression tag	UNP Q8A171
CCC	10	LEU	-	expression tag	UNP Q8A171
CCC	11	VAL	-	expression tag	UNP Q8A171
CCC	12	PRO	-	expression tag	UNP Q8A171
CCC	13	ARG	-	expression tag	UNP Q8A171
CCC	14	GLY	-	expression tag	UNP Q8A171
CCC	15	SER	-	expression tag	UNP Q8A171
CCC	16	HIS	-	expression tag	UNP Q8A171
CCC	17	MET	-	expression tag	UNP Q8A171
CCC	18	ALA	-	expression tag	UNP Q8A171
CCC	19	SER	-	expression tag	UNP Q8A171

- Molecule 2 is HEXAETHYLENE GLYCOL (three-letter code: P6G) (formula: $C_{12}H_{26}O_7$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	AAA	1	Total	C	O	0	0
			13	8	5		

- Molecule 3 is 4-O-sulfo-beta-D-galactopyranose (three-letter code: G4S) (formula: $C_6H_{12}O_9S$) (labeled as "Ligand of Interest" by depositor).

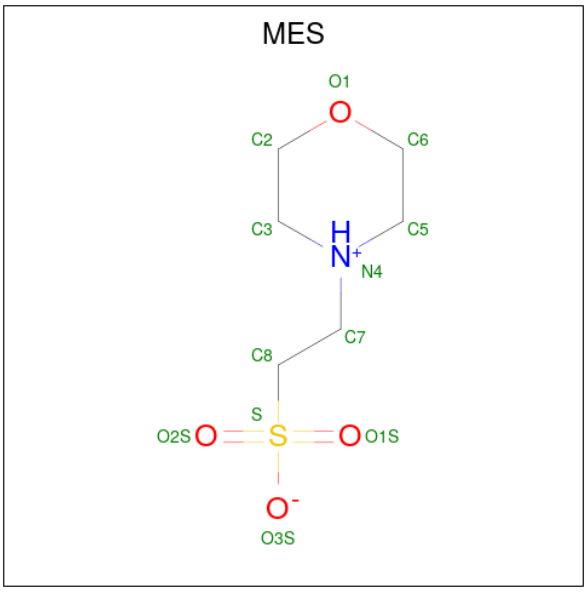


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	AAA	1	Total C O S 32 12 18 2	0	1
3	CCC	1	Total C O S 16 6 9 1	0	0

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

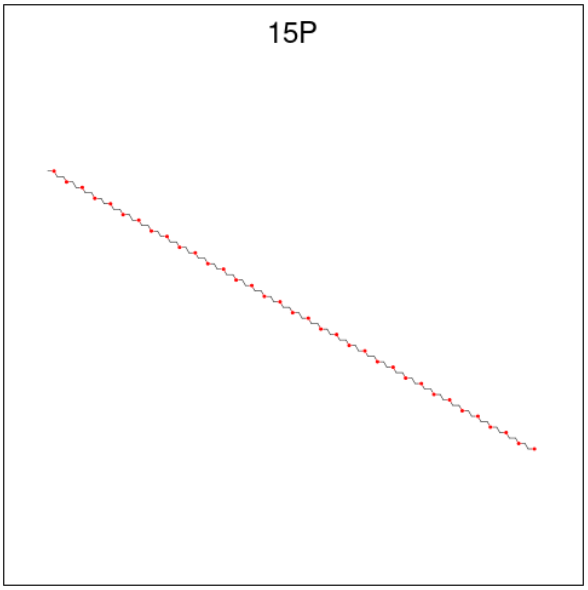
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	AAA	1	Total Ca 1 1	0	0
4	CCC	1	Total Ca 1 1	0	0

- Molecule 5 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: $C_6H_{13}NO_4S$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	CCC	1	Total	C	N	O	S	0	0
			12	6	1	4	1		
5	CCC	1	Total	C	N	O	S	0	0
			12	6	1	4	1		

- Molecule 6 is POLYETHYLENE GLYCOL (N=34) (three-letter code: 15P) (formula: $C_{69}H_{140}O_{35}$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	CCC	1	Total	C	O	0	0
			11	7	4		

- Molecule 7 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	CCC	1	Total Cl 1 1	0	0

- Molecule 8 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	AAA	355	Total O 355 355	0	0
8	CCC	400	Total O 400 400	0	0

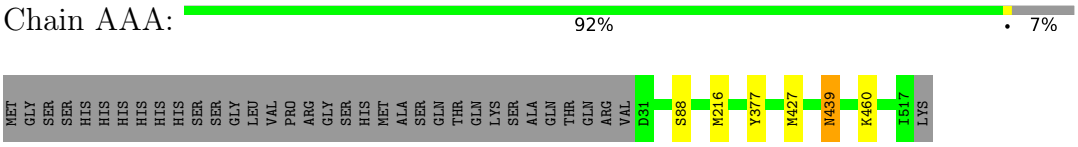
3 Residue-property plots (i)

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. 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. 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.

Note EDS failed to run properly.

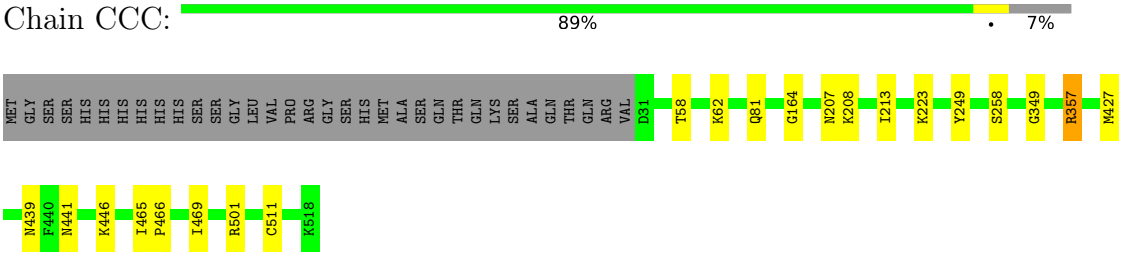
- Molecule 1: Putative secreted sulfatase ydeN

Chain AAA:



- Molecule 1: Putative secreted sulfatase ydeN

Chain CCC:



4 Data and refinement statistics (i)

EDS failed to run properly - this section is therefore incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	73.18Å 88.95Å 87.91Å 90.00° 112.05° 90.00°	Depositor
Resolution (Å)	81.61 – 1.50	Depositor
% Data completeness (in resolution range)	99.5 (81.61-1.50)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.42 (at 1.45Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.141 , 0.176	Depositor
Wilson B-factor (Å^2)	16.4	Xtriage
Anisotropy	0.344	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.018 for h,-k,-h-l	Xtriage
Total number of atoms	8630	wwPDB-VP
Average B, all atoms (Å^2)	19.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.36% 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: CA, G4S, P6G, MES, CL, 15P

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	0.64	0/3959	0.76	1/5383 (0.0%)
1	CCC	0.68	0/4029	0.79	2/5474 (0.0%)
All	All	0.66	0/7988	0.78	3/10857 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	CCC	357	ARG	NE-CZ-NH2	-6.80	116.90	120.30
1	AAA	216	MET	CG-SD-CE	-5.63	91.19	100.20
1	CCC	249	TYR	CB-CG-CD2	-5.20	117.88	121.00

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	3853	0	3655	4	0
1	CCC	3923	0	3732	20	0
2	AAA	13	0	16	2	0
3	AAA	32	0	22	1	0
3	CCC	16	0	11	0	0
4	AAA	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	CCC	1	0	0	0	0
5	CCC	24	0	26	0	0
6	CCC	11	0	13	11	0
7	CCC	1	0	0	0	0
8	AAA	355	0	0	2	0
8	CCC	400	0	0	1	1
All	All	8630	0	7475	27	1

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.

All (27) 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:CCC:427[A]:MET:HE3	1:CCC:439:ASN:H	1.57	0.69
1:AAA:427[A]:MET:HE2	8:AAA:945:HOH:O	1.95	0.65
1:CCC:81:GLN:HE22	6:CCC:602:15P:H16	1.68	0.59
1:CCC:357:ARG:HH22	6:CCC:602:15P:C2	2.16	0.58
1:CCC:427[A]:MET:O	1:CCC:427[A]:MET:HG2	2.05	0.57
1:AAA:88[B]:SER:OG	3:AAA:802[B]:G4S:O9	2.24	0.55
2:AAA:801:P6G:C6	8:AAA:910:HOH:O	2.55	0.53
1:CCC:427[A]:MET:HE3	1:CCC:439:ASN:N	2.25	0.51
1:CCC:427[A]:MET:CE	1:CCC:439:ASN:H	2.21	0.51
1:CCC:427[A]:MET:O	1:CCC:427[A]:MET:CG	2.60	0.50
1:CCC:357:ARG:HH22	6:CCC:602:15P:H12	1.76	0.50
1:CCC:357:ARG:HH12	6:CCC:602:15P:H12	1.79	0.48
1:CCC:466:PRO:HG3	6:CCC:602:15P:C5	2.43	0.48
1:CCC:466:PRO:HG3	6:CCC:602:15P:H15	1.96	0.48
6:CCC:602:15P:H23	6:CCC:602:15P:H25	1.55	0.45
1:CCC:58:THR:O	1:CCC:62[A]:LYS:HE3	2.17	0.44
1:CCC:446:LYS:HE3	8:CCC:798:HOH:O	2.16	0.44
1:AAA:427[A]:MET:HE1	1:AAA:439:ASN:H	1.81	0.44
1:AAA:460:LYS:HD2	2:AAA:801:P6G:H112	2.00	0.44
1:CCC:207:ASN:HB3	1:CCC:213[B]:ILE:HD11	2.01	0.43
1:CCC:469:ILE:HG13	6:CCC:602:15P:H21	1.99	0.43
1:CCC:357:ARG:NH2	6:CCC:602:15P:H12	2.35	0.42
1:CCC:465:ILE:O	6:CCC:602:15P:C1	2.68	0.42
1:CCC:349:GLY:O	1:CCC:427[A]:MET:HE1	2.20	0.41
1:CCC:501:ARG:HD2	1:CCC:511:CYS:O	2.21	0.41
6:CCC:602:15P:H13	6:CCC:602:15P:H11	1.58	0.41
1:CCC:164:GLY:HA3	1:CCC:258:SER:O	2.21	0.40

All (1) 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)
8:CCC:983:HOH:O	8:CCC:1044:HOH:O[2_655]	2.14	0.06

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	491/522 (94%)	479 (98%)	11 (2%)	1 (0%)	47 23
1	CCC	499/522 (96%)	487 (98%)	12 (2%)	0	100 100
All	All	990/1044 (95%)	966 (98%)	23 (2%)	1 (0%)	51 25

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	AAA	439	ASN

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	404/444 (91%)	403 (100%)	1 (0%)	93 86
1	CCC	412/444 (93%)	409 (99%)	3 (1%)	84 69
All	All	816/888 (92%)	812 (100%)	4 (0%)	88 78

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AAA	377	TYR
1	CCC	208	LYS
1	CCC	223	LYS
1	CCC	441	ASN

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)

Of 10 ligands modelled in this entry, 3 are monoatomic - leaving 7 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	$\#\|Z\| > 2$
3	G4S	CCC	604	4	16,16,16	0.63	0	19,24,24	0.79	0
5	MES	CCC	601	-	12,12,12	0.74	0	14,16,16	0.57	0
2	P6G	AAA	801	-	12,12,18	0.35	0	11,11,17	0.32	0
3	G4S	AAA	802[B]	4	16,16,16	0.44	0	19,24,24	0.60	0
3	G4S	AAA	802[A]	4	16,16,16	0.45	0	19,24,24	0.66	0
5	MES	CCC	603	-	12,12,12	0.72	0	14,16,16	0.50	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
6	15P	CCC	602	-	10,10,103	0.35	0	9,9,102	0.21	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
3	G4S	CCC	604	4	-	1/7/27/27	0/1/1/1
5	MES	CCC	601	-	-	0/6/14/14	0/1/1/1
2	P6G	AAA	801	-	-	6/10/10/16	-
3	G4S	AAA	802[B]	4	-	2/7/27/27	0/1/1/1
3	G4S	AAA	802[A]	4	-	1/7/27/27	0/1/1/1
5	MES	CCC	603	-	-	1/6/14/14	0/1/1/1
6	15P	CCC	602	-	-	7/8/8/101	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (18) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	CCC	603	MES	N4-C7-C8-S
6	CCC	602	15P	C3-C4-O2-C5
6	CCC	602	15P	O1-C3-C4-O2
3	AAA	802[B]	G4S	O5-C5-C6-O6
6	CCC	602	15P	C5-C6-O3-C7
6	CCC	602	15P	C1-C2-O1-C3
3	AAA	802[B]	G4S	C4-C5-C6-O6
2	AAA	801	P6G	O13-C14-C15-O16
6	CCC	602	15P	OXT-C1-C2-O1
2	AAA	801	P6G	O10-C11-C12-O13
6	CCC	602	15P	O2-C5-C6-O3
2	AAA	801	P6G	O4-C5-C6-O7
3	CCC	604	G4S	O5-C5-C6-O6
3	AAA	802[A]	G4S	O5-C5-C6-O6
2	AAA	801	P6G	O7-C8-C9-O10
2	AAA	801	P6G	C8-C9-O10-C11
6	CCC	602	15P	C6-C5-O2-C4
2	AAA	801	P6G	C12-C11-O10-C9

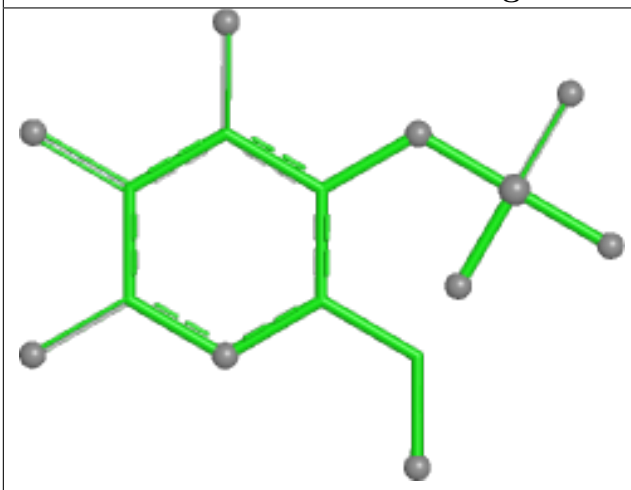
There are no ring outliers.

3 monomers are involved in 14 short contacts:

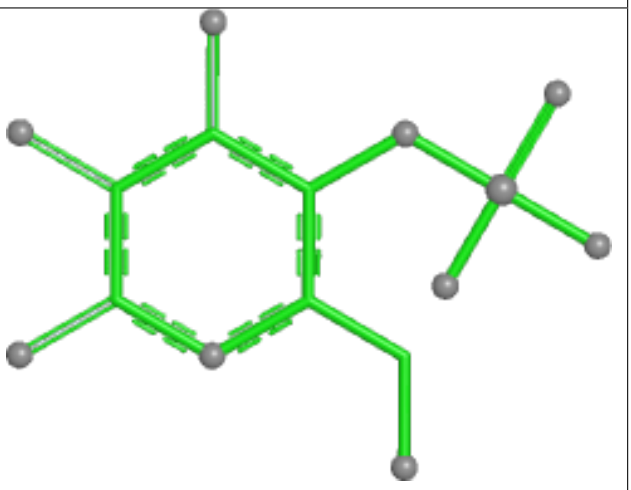
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	AAA	801	P6G	2	0
3	AAA	802[B]	G4S	1	0
6	CCC	602	15P	11	0

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.

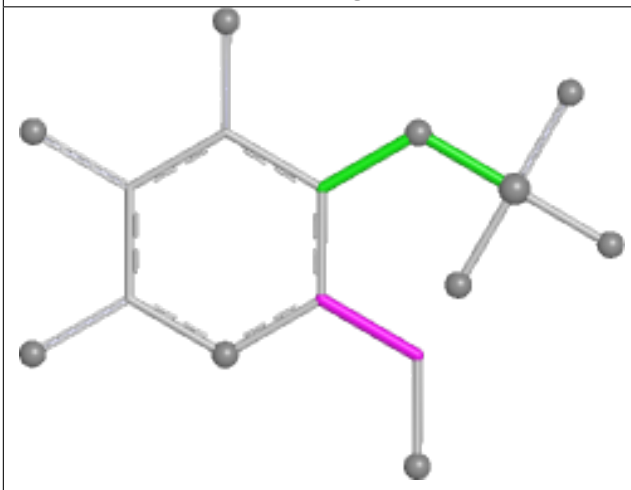
Ligand G4S CCC 604



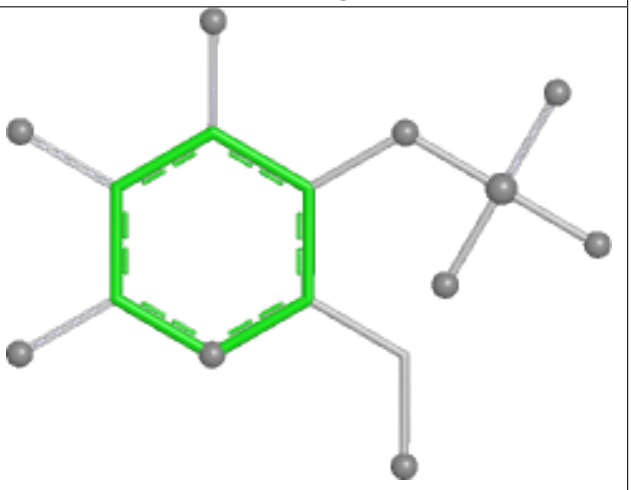
Bond lengths



Bond angles



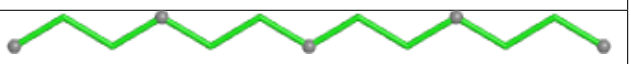
Torsions



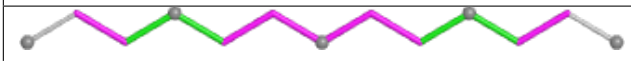
Rings

Ligand P6G AAA 801

Bond lengths



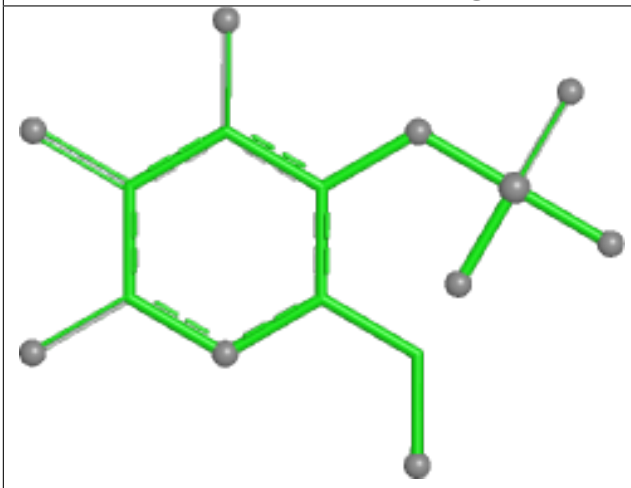
Bond angles



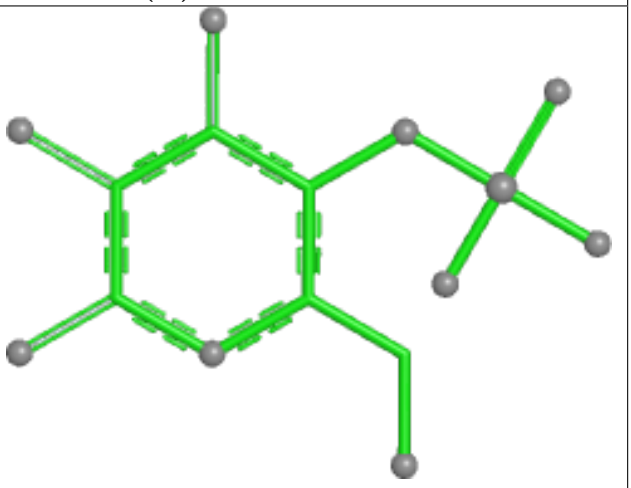
Torsions

Rings

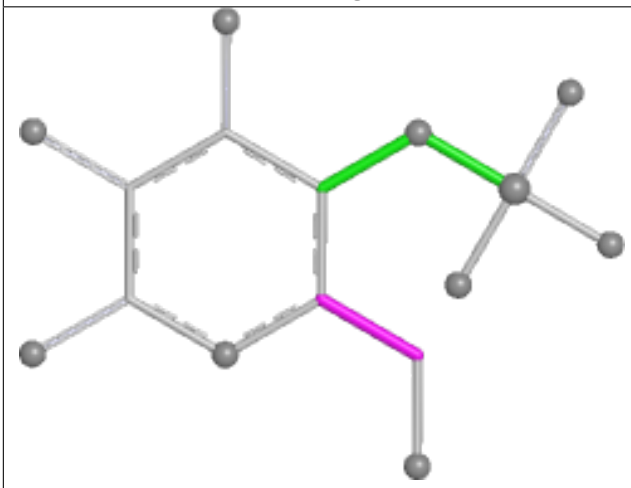
Ligand G4S AAA 802 (B)



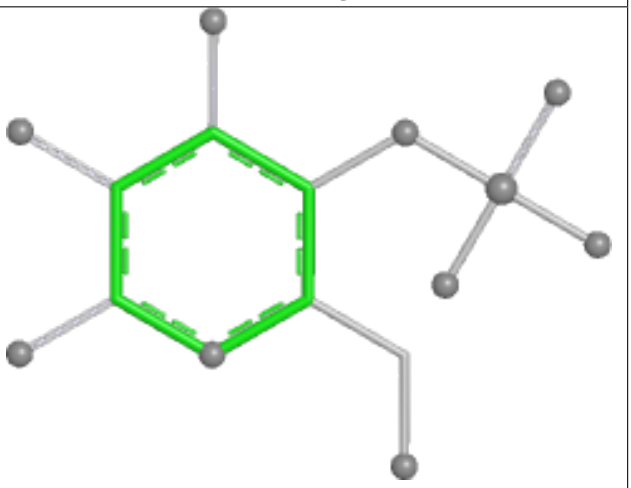
Bond lengths



Bond angles

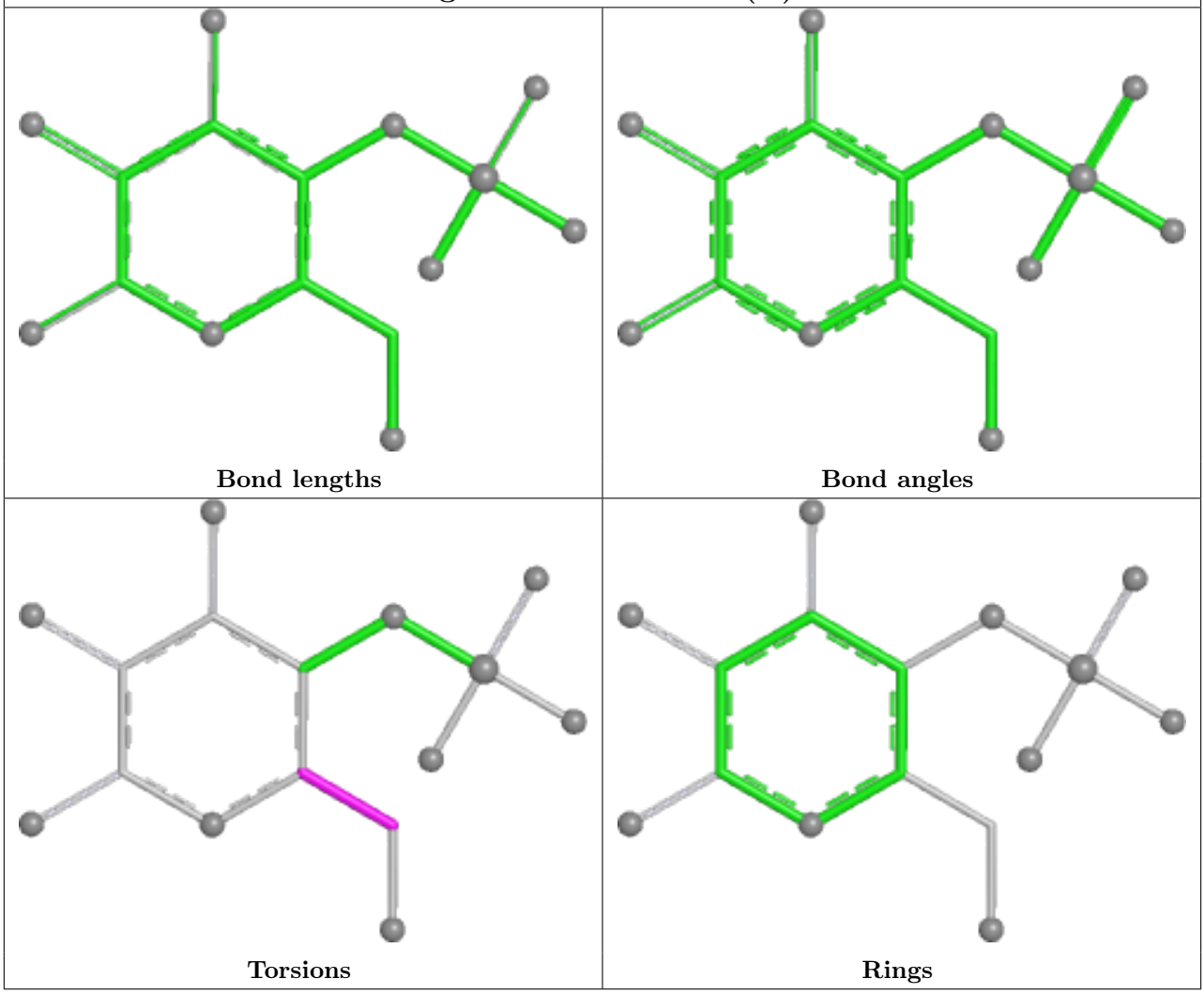


Torsions

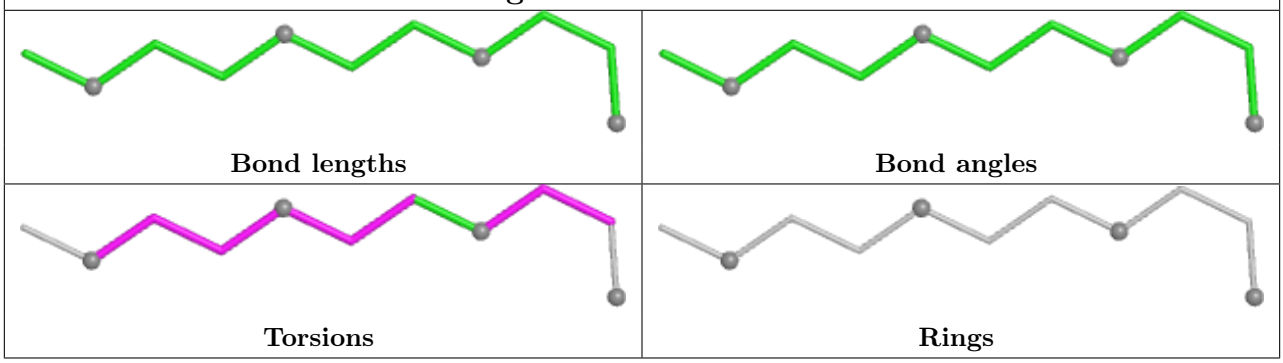


Rings

Ligand G4S AAA 802 (A)



Ligand 15P CCC 602



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

EDS failed to run properly - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

EDS failed to run properly - this section is therefore empty.

6.3 Carbohydrates

EDS failed to run properly - this section is therefore empty.

6.4 Ligands

EDS failed to run properly - this section is therefore empty.

6.5 Other polymers

EDS failed to run properly - this section is therefore empty.