WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Oct 15, 2023 – 06:11 AM EDT

PDB ID	:	7SNO
Title	:	Structure of Bacple_01701(H214N), a 6-O-galactose porphyran sulfatase
Authors	:	Ulaganathan, T.; Cygler, M.
Deposited on	:	2021-10-28
Resolution	:	2.10 Å(reported)

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

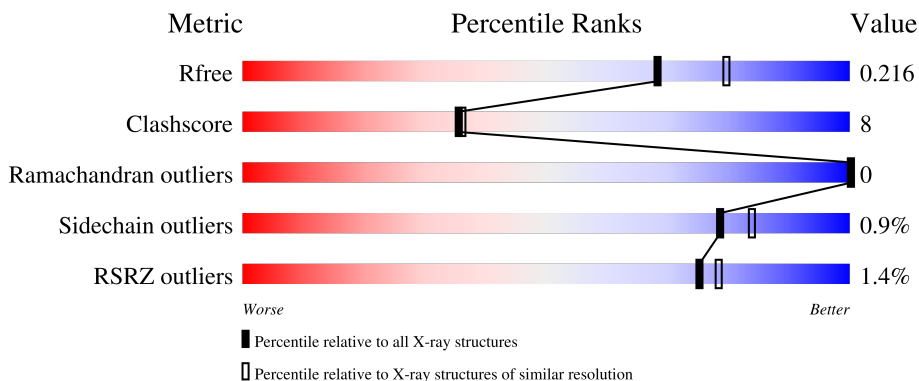|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance ⓘ

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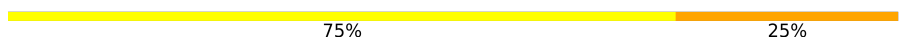*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.10 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		0.216
Clashscore		8
Ramachandran outliers		0
Sidechain outliers		0.9%
RSRZ outliers		1.4%

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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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	A	536	81% 13% 6%
2	B	4	75% 25%

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	PG4	A	606	-	-	X	-

2 Entry composition ⓘ

There are 8 unique types of molecules in this entry. The entry contains 4584 atoms, of which 116 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 Arylsulfatase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	506	Total 4133	C 2629	N 695	O 779	S 30	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	220	ASN	HIS	engineered mutation	UNP B5CYA4

- Molecule 2 is an oligosaccharide called 6-O-sulfo-alpha-L-galactopyranose-(1-3)-beta-D-gala ctopyranose-(1-4)-6-O-sulfo-alpha-L-galactopyranose-(1-3)-beta-D-galactopyranose.
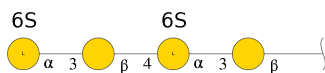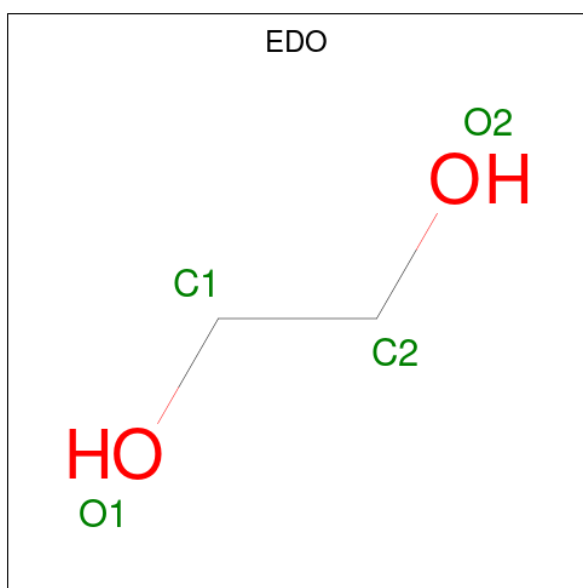

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	4	Total 93	C 24	H 40	O 27	S 2	0	0	0

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

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

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).

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

Continued on next page...

Continued from previous page...

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

- Molecule 5 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: $C_8H_{18}O_5$).
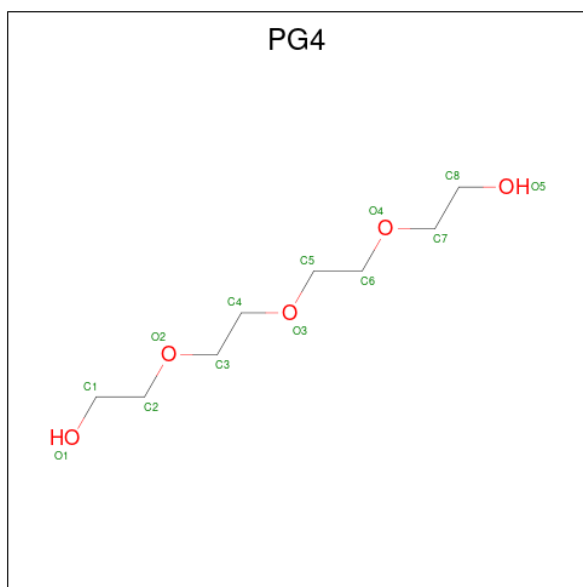
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total 31	C 8	H 18	O 5	0	0

- Molecule 6 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: $C_6H_{14}O_4$).
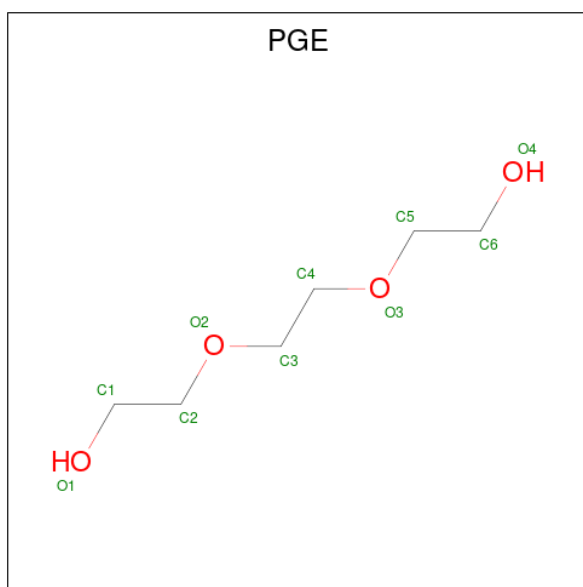
PGE

O4
OH

C5
C6

C4
O
O3

O2
O

C3

C1
C2

HO
O1

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	H	O		
6	A	1	24	6	14	4	0	0

- Molecule 7 is SODIUM ION (three-letter code: NA) (formula: Na).

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Na		
7	A	1	1	1	0	0

- Molecule 8 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
8	A	228	228	228	0	0

WORLDWIDE
wwPDB
PROTEIN DATA BANK

3 Residue-property plots ⓘ

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: Arylsulfatase
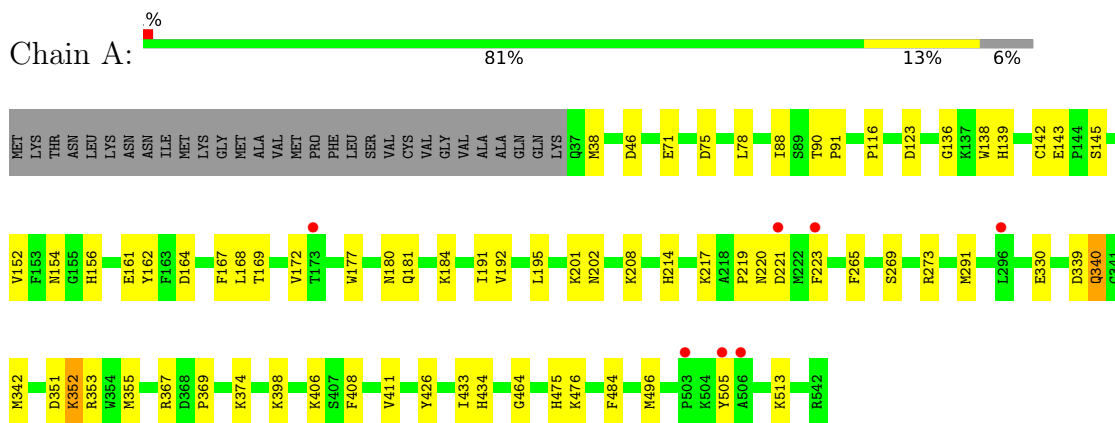
• Molecule 2: 6-O-sulfo-alpha-L-galactopyranose-(1-3)-beta-D-galactopyranose-(1-4)-6-O-sulfo-alpha-L-galactopyranose-(1-3)-beta-D-galactopyranose
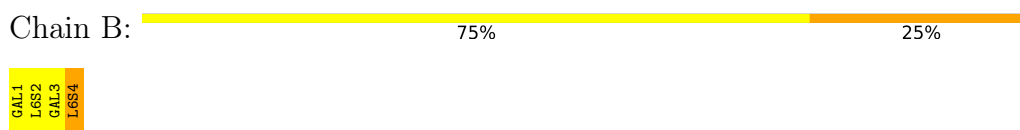
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 62	Depositor
Cell constants a, b, c, α, β, γ	154.72Å 154.72Å 45.97Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	44.66 – 2.10 44.66 – 2.10	Depositor EDS
% Data completeness (in resolution range)	99.9 (44.66-2.10) 99.9 (44.66-2.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.57 (at 2.10Å)	Xtriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.181 , 0.216 0.182 , 0.216	Depositor DCC
R_{free} test set	1853 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	34.5	Xtriage
Anisotropy	0.707	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 55.5	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.026 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	4584	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 3.43% 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: NA, PEG, PG4, L6S, PGE, GAL, EDO

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	0.47	0/4246	0.63	0/5745

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 ⓘ

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	4133	0	3914	67	0
2	B	53	40	39	1	0
3	A	14	20	20	2	0
4	A	16	24	24	1	0
5	A	13	18	18	8	0
6	A	10	14	14	0	0
7	A	1	0	0	0	0
8	A	228	0	0	11	0
All	All	4468	116	4029	67	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 8.

The worst 5 of 67 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:220:ASN:OD1	1:A:221:ASP:N	1.90	1.03
1:A:184:LYS:HZ1	5:A:606:PG4:H71	1.43	0.84
1:A:184:LYS:HZ3	5:A:606:PG4:H41	1.55	0.72
1:A:71:GLU:HG3	8:A:906:HOH:O	1.91	0.70
1:A:355:MET:HB2	1:A:426:TYR:CD2	2.26	0.70

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	A	504/536 (94%)	482 (96%)	22 (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	A	444/481 (92%)	440 (99%)	4 (1%)	78	84

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

Mol	Chain	Res	Type
1	A	88	ILE
1	A	156	HIS
1	A	340	GLN
1	A	352	LYS

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 ⓘ

4 monosaccharides 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	#$|Z| > 2$
2	GAL	B	1	2	12,12,12	0.68	0	17,17,17	2.07	5 (29%)
2	L6S	B	2	2	15,15,16	0.71	1 (6%)	20,22,24	1.18	1 (5%)
2	GAL	B	3	2	11,11,12	0.63	0	15,15,17	2.40	6 (40%)
2	L6S	B	4	7,2	15,15,16	0.61	0	20,22,24	1.67	3 (15%)

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
2	GAL	B	1	2	-	1/2/22/22	0/1/1/1
2	L6S	B	2	2	-	3/6/23/26	0/1/1/1
2	GAL	B	3	2	-	1/2/19/22	0/1/1/1
2	L6S	B	4	7,2	-	0/6/23/26	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	2	L6S	O3S-S1	2.28	1.64	1.50

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	3	GAL	C1-C2-C3	5.32	116.20	109.67
2	B	1	GAL	O3-C3-C2	-4.41	100.16	110.35
2	B	4	L6S	O6-C6-C5	4.11	115.29	107.62
2	B	1	GAL	C3-C4-C5	-4.06	103.00	110.24
2	B	3	GAL	O5-C5-C6	-3.79	101.26	107.20

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	2	L6S	C6-O6-S1-O1S
2	B	1	GAL	O5-C5-C6-O6
2	B	3	GAL	O5-C5-C6-O6
2	B	2	L6S	C6-O6-S1-O2S
2	B	2	L6S	C6-O6-S1-O3S

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	4	L6S	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 oligosaccharide.

Oligosaccharide Chain B
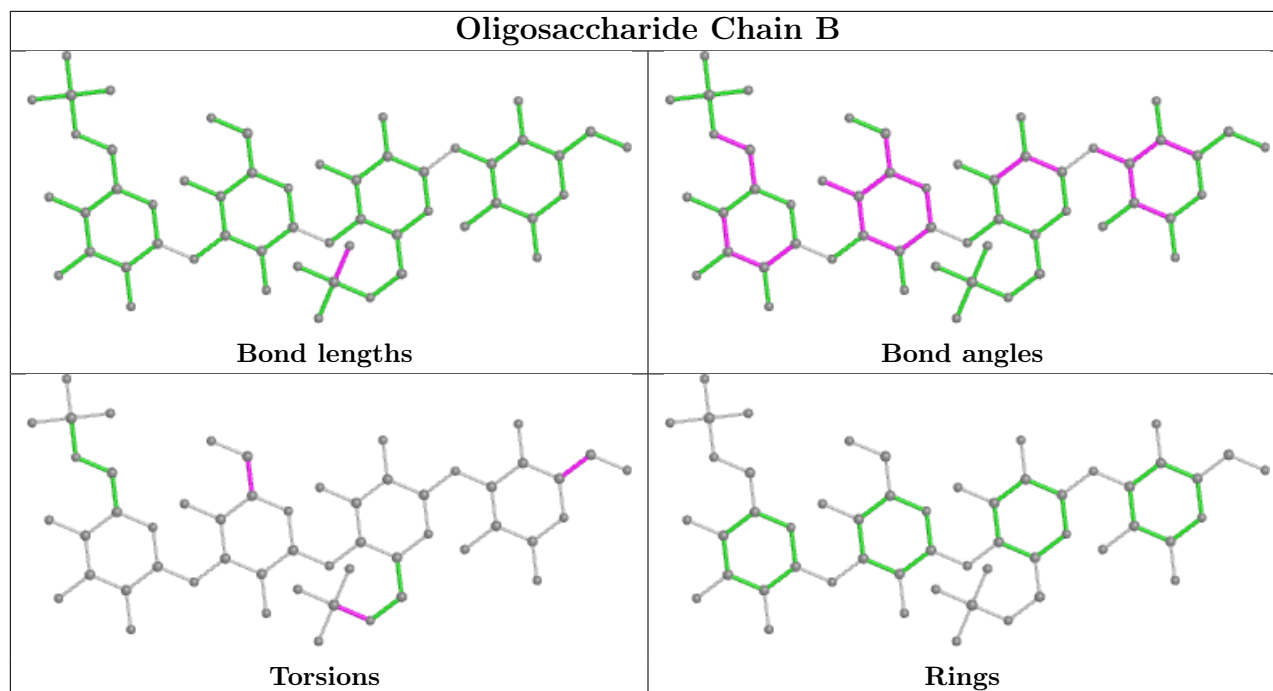
5.6 Ligand geometry ⓘ

Of 9 ligands modelled in this entry, 1 is monoatomic - leaving 8 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
4	EDO	A	605	-	3,3,3	0.38	0	2,2,2	0.44	0
4	EDO	A	603	-	3,3,3	0.46	0	2,2,2	0.48	0
4	EDO	A	604	-	3,3,3	0.43	0	2,2,2	0.35	0
5	PG4	A	606	-	12,12,12	0.25	0	11,11,11	0.54	0
4	EDO	A	608	-	3,3,3	0.62	0	2,2,2	0.50	0
6	PGE	A	607	-	9,9,9	0.39	0	8,8,8	0.27	0
3	PEG	A	601	-	6,6,6	0.15	0	5,5,5	0.13	0
3	PEG	A	602	-	6,6,6	0.12	0	5,5,5	0.12	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	EDO	A	605	-	-	0/1/1/1	-
4	EDO	A	603	-	-	0/1/1/1	-
4	EDO	A	604	-	-	1/1/1/1	-
5	PG4	A	606	-	-	5/10/10/10	-
4	EDO	A	608	-	-	1/1/1/1	-
6	PGE	A	607	-	-	4/7/7/7	-
3	PEG	A	601	-	-	2/4/4/4	-
3	PEG	A	602	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 15 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	602	PEG	O1-C1-C2-O2
5	A	606	PG4	O1-C1-C2-O2
3	A	601	PEG	O1-C1-C2-O2
3	A	602	PEG	O2-C3-C4-O4
5	A	606	PG4	O4-C7-C8-O5

There are no ring outliers.

3 monomers are involved in 11 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	603	EDO	1	0
5	A	606	PG4	8	0
3	A	602	PEG	2	0

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, 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	506/536 (94%)	-0.17	7 (1%)	75	78	33, 41, 59, 99	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	221	ASP	3.4
1	A	173	THR	3.3
1	A	223	PHE	2.8
1	A	505	TYR	2.6
1	A	503	PRO	2.5

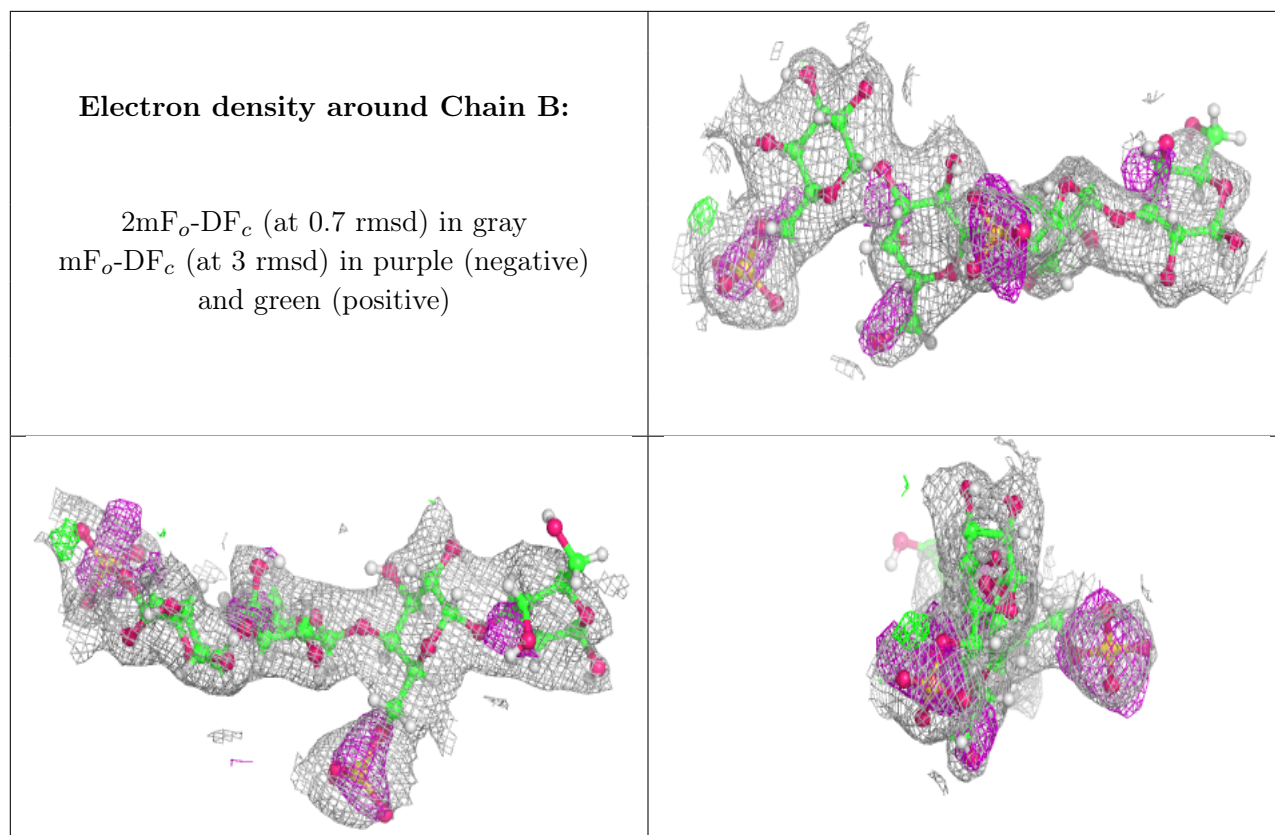## 6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

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	GAL	B	1	12/12	0.70	0.29	80,99,120,127	0
2	GAL	B	3	11/12	0.89	0.14	43,50,60,68	0
2	L6S	B	2	15/16	0.90	0.13	58,75,92,96	0
2	L6S	B	4	15/16	0.97	0.12	35,42,53,54	0

The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around Chain B:

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

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(Å^2)	Q<0.9
6	PGE	A	607	10/10	0.52	0.32	55,76,85,86	0
4	EDO	A	608	4/4	0.69	0.11	52,63,67,69	0
5	PG4	A	606	13/13	0.80	0.21	53,65,78,80	0
3	PEG	A	601	7/7	0.84	0.15	46,55,63,63	0
4	EDO	A	605	4/4	0.87	0.28	45,55,57,61	0
4	EDO	A	604	4/4	0.88	0.22	61,74,81,82	0
7	NA	A	609	1/1	0.88	0.09	39,39,39,39	0
3	PEG	A	602	7/7	0.90	0.31	53,64,74,76	0
4	EDO	A	603	4/4	0.90	0.20	54,65,71,73	0

6.5 Other polymers (i)

There are no such residues in this entry.