



# wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 17, 2023 – 02:54 PM EST

PDB ID : 4U31  
Title : Sco GlgEI-V279S in Complex with maltose-C-phosphonate  
Authors : Ronning, D.R.; Lindenberger, J.J.  
Deposited on : 2014-07-18  
Resolution : 1.85 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](mailto: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) : 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

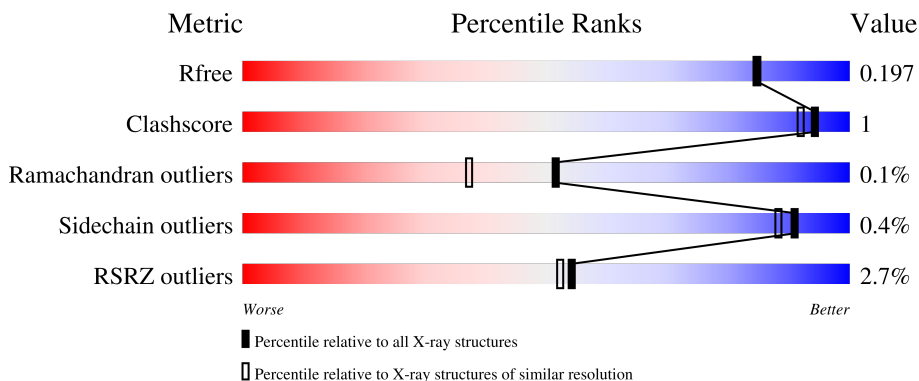
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.85 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	4003 (1.86-1.82)
Clashscore	141614	4233 (1.86-1.82)
Ramachandran outliers	138981	4185 (1.86-1.82)
Sidechain outliers	138945	4186 (1.86-1.82)
RSRZ outliers	127900	3957 (1.86-1.82)

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  $\geq 3$ , 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	683	 3% 91% 5%
1	B	683	 2% 92% 5%
2	C	2	 50% 50%
2	D	2	 100%

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
3	CIT	A	702	-	X	-	-
3	CIT	B	702	-	X	-	-

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 11906 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 Alpha-1,4-glucan:maltose-1-phosphate maltosyltransferase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	649	5207	3281	959	957	10	1	6	0
1	B	649	5152	3252	939	949	12	0	2	0

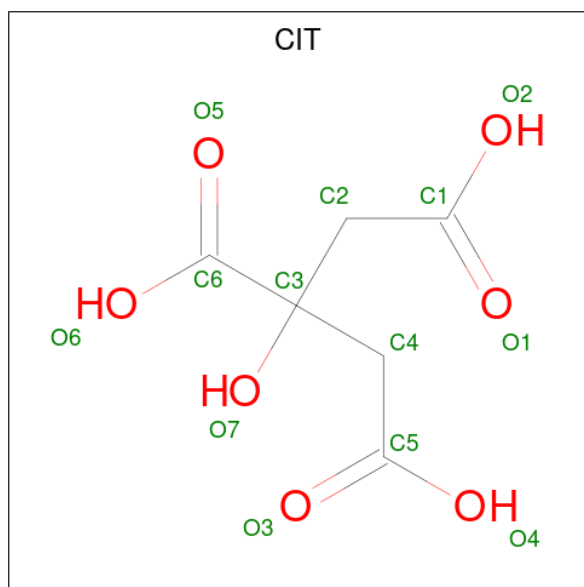
There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	279	SER	VAL	engineered mutation	UNP Q9L1K2
A	676	ALA	-	expression tag	UNP Q9L1K2
A	677	LEU	-	expression tag	UNP Q9L1K2
A	678	HIS	-	expression tag	UNP Q9L1K2
A	679	HIS	-	expression tag	UNP Q9L1K2
A	680	HIS	-	expression tag	UNP Q9L1K2
A	681	HIS	-	expression tag	UNP Q9L1K2
A	682	HIS	-	expression tag	UNP Q9L1K2
A	683	HIS	-	expression tag	UNP Q9L1K2
B	279	SER	VAL	engineered mutation	UNP Q9L1K2
B	676	ALA	-	expression tag	UNP Q9L1K2
B	677	LEU	-	expression tag	UNP Q9L1K2
B	678	HIS	-	expression tag	UNP Q9L1K2
B	679	HIS	-	expression tag	UNP Q9L1K2
B	680	HIS	-	expression tag	UNP Q9L1K2
B	681	HIS	-	expression tag	UNP Q9L1K2
B	682	HIS	-	expression tag	UNP Q9L1K2
B	683	HIS	-	expression tag	UNP Q9L1K2

- Molecule 2 is an oligosaccharide called alpha-D-glucopyranose-(1-4)-(1S)-1,5-anhydro-1-(phosphonomethyl)-D-glucitol.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	2	Total	C	O	P	0	0	0
			27	13	13	1			
2	D	2	Total	C	O	P	0	0	0
			27	13	13	1			

- Molecule 3 is CITRIC ACID (three-letter code: CIT) (formula:  $C_6H_8O_7$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			13	6	7		
3	B	1	Total	C	O	0	0
			13	6	7		

- 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 C O 4 2 2	0	0
4	A	1	Total C O 4 2 2	0	0
4	A	1	Total C O 4 2 2	0	0
4	A	1	Total C O 4 2 2	0	0
4	B	1	Total C O 4 2 2	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	710	Total O 710 710	0	0
5	B	737	Total O 737 737	0	0



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	113.56Å 113.56Å 315.14Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	42.69 – 1.85 42.68 – 1.85	Depositor EDS
% Data completeness (in resolution range)	100.0 (42.69-1.85) 97.5 (42.68-1.85)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.98 (at 1.86Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, $R_{free}$	0.164 , 0.194 0.167 , 0.197	Depositor DCC
$R_{free}$ test set	2000 reflections (1.14%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	20.9	Xtrriage
Anisotropy	0.162	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 41.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	11906	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	24.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.31% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GPM, EDO, GLC, CIT

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	0.35	0/5359	0.51	0/7324
1	B	0.36	0/5304	0.52	0/7251
All	All	0.36	0/10663	0.51	0/14575

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	5207	0	5050	13	0
1	B	5152	0	4992	11	0
2	C	27	0	22	1	0
2	D	27	0	22	0	0
3	A	13	0	2	2	0
3	B	13	0	7	2	0
4	A	16	0	24	0	0
4	B	4	0	6	0	0
5	A	710	0	0	1	2
5	B	737	0	0	3	2
All	All	11906	0	10125	27	2

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 1.

The worst 5 of 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)
3:A:702:CIT:O1	3:A:702:CIT:C1	1.66	1.42
3:B:702:CIT:O2	3:B:702:CIT:C1	1.71	1.37
1:A:25[B]:ASP:OD2	1:A:27[B]:ARG:NH1	2.16	0.79
1:A:69:ARG:NH2	1:A:96:GLU:OE2	2.16	0.78
3:A:702:CIT:O1	3:A:702:CIT:C2	2.35	0.74

All (2) 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)
5:A:892:HOH:O	5:B:1299:HOH:O[6_446]	2.12	0.08
5:A:892:HOH:O	5:B:863:HOH:O[6_446]	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	A	654/683 (96%)	640 (98%)	13 (2%)	1 (0%)	47 33
1	B	649/683 (95%)	636 (98%)	13 (2%)	0	100 100
All	All	1303/1366 (95%)	1276 (98%)	26 (2%)	1 (0%)	51 37

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	127	ASP

### 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	A	544/565 (96%)	542 (100%)	2 (0%)	91	88
1	B	539/565 (95%)	537 (100%)	2 (0%)	91	88
All	All	1083/1130 (96%)	1079 (100%)	4 (0%)	91	88

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

Mol	Chain	Res	Type
1	A	107	TRP
1	A	251	TYR
1	B	251	TYR
1	B	511	TYR

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](#)

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	GPM	C	1	2	15,16,16	1.62	2 (13%)	21,24,24	1.50	5 (23%)
2	GLC	C	2	2	11,11,12	1.31	1 (9%)	15,15,17	0.82	1 (6%)
2	GPM	D	1	2	15,16,16	1.46	2 (13%)	21,24,24	1.45	3 (14%)
2	GLC	D	2	2	11,11,12	1.45	1 (9%)	15,15,17	0.89	1 (6%)

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	GPM	C	1	2	-	0/7/27/27	0/1/1/1
2	GLC	C	2	2	-	0/2/19/22	0/1/1/1
2	GPM	D	1	2	-	0/7/27/27	0/1/1/1
2	GLC	D	2	2	-	0/2/19/22	0/1/1/1

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	1	GPM	P-C7	4.80	1.83	1.78
2	D	1	GPM	P-C7	3.92	1.82	1.78
2	D	2	GLC	C2-C3	-3.54	1.47	1.52
2	C	2	GLC	C2-C3	-3.43	1.47	1.52
2	D	1	GPM	C3-C2	-3.29	1.43	1.52

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	1	GPM	O5-C5-C4	3.97	116.90	109.69
2	C	1	GPM	O5-C5-C4	3.63	116.28	109.69
2	C	1	GPM	O4-C4-C5	-2.87	102.18	109.30
2	D	1	GPM	O4-C4-C5	-2.55	102.97	109.30
2	D	2	GLC	C1-C2-C3	2.44	112.67	109.67

There are no chirality outliers.

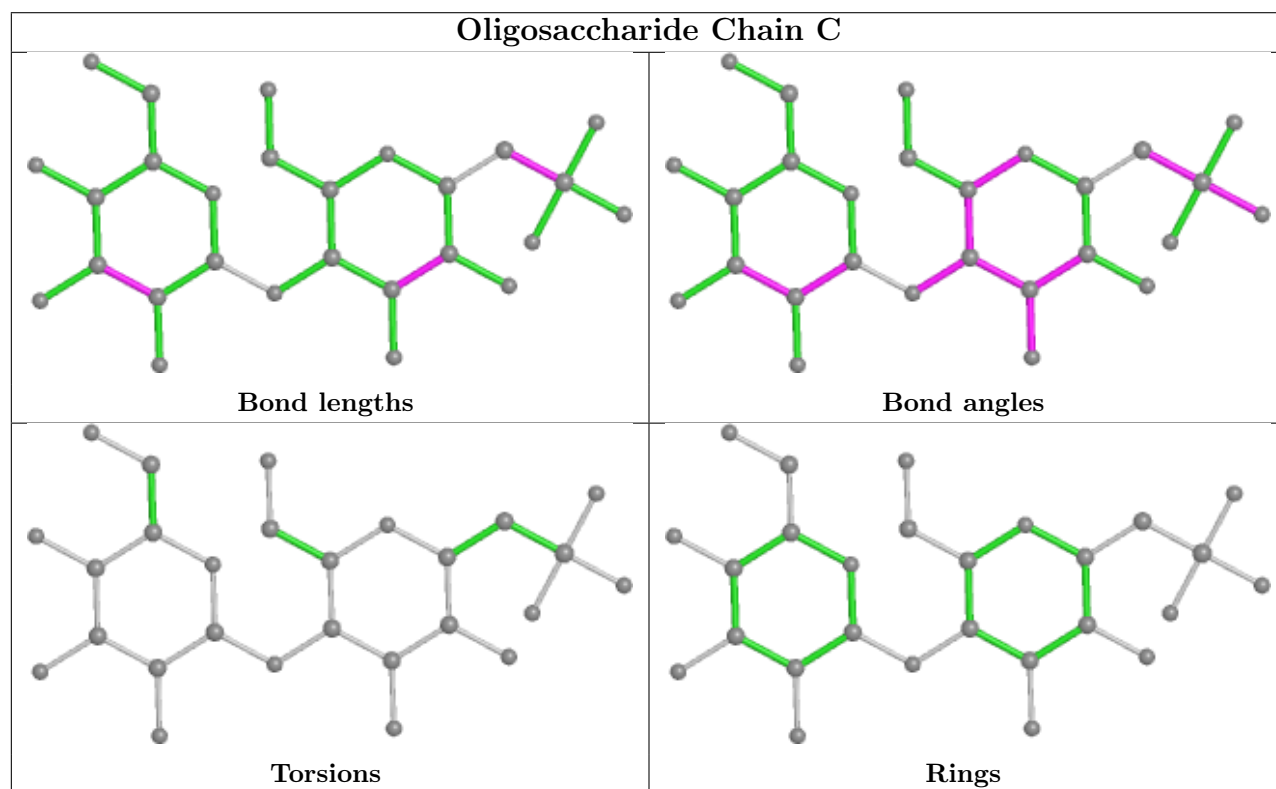
There are no torsion outliers.

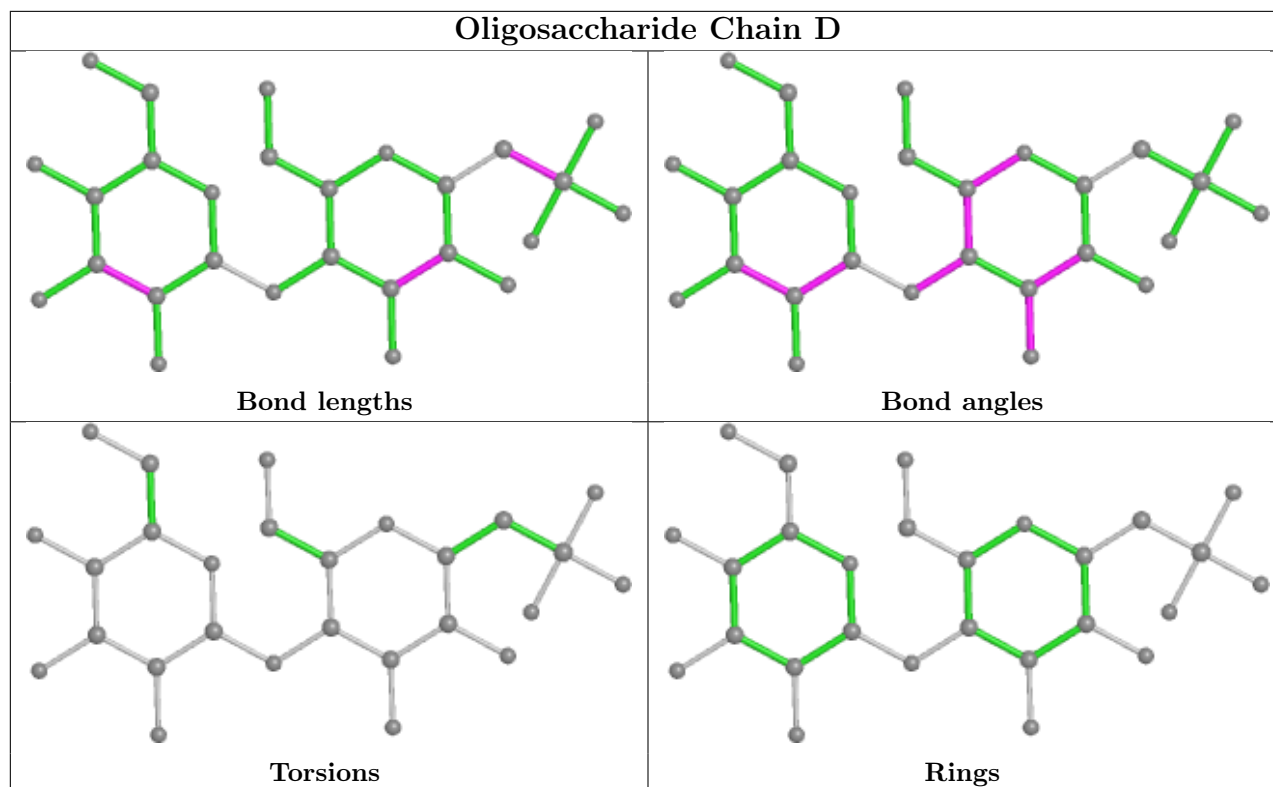
There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	C	1	GPM	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.





## 5.6 Ligand geometry [i](#)

7 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	EDO	B	703	-	3,3,3	0.52	0	2,2,2	0.16	0
3	CIT	A	702	3	12,12,12	5.10	6 (50%)	17,17,17	6.75	10 (58%)
4	EDO	A	704	-	3,3,3	0.49	0	2,2,2	0.52	0
4	EDO	A	705	-	3,3,3	0.47	0	2,2,2	0.40	0
4	EDO	A	703	-	3,3,3	0.55	0	2,2,2	0.33	0
4	EDO	A	706	-	3,3,3	0.55	0	2,2,2	0.34	0
3	CIT	B	702	3	12,12,12	6.05	9 (75%)	17,17,17	6.00	11 (64%)

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	B	703	-	-	0/1/1/1	-
3	CIT	A	702	3	-	3/16/16/16	-
4	EDO	A	704	-	-	0/1/1/1	-
4	EDO	A	705	-	-	0/1/1/1	-
4	EDO	A	703	-	-	0/1/1/1	-
4	EDO	A	706	-	-	0/1/1/1	-
3	CIT	B	702	3	-	7/16/16/16	-

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	702	CIT	O1-C1	13.38	1.66	1.22
3	B	702	CIT	O2-C1	11.98	1.71	1.30
3	B	702	CIT	C4-C3	11.00	1.67	1.53
3	A	702	CIT	C4-C5	-7.95	1.26	1.50
3	B	702	CIT	O4-C5	7.47	1.55	1.30

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	702	CIT	C3-C4-C5	13.92	147.53	113.81
3	A	702	CIT	O2-C1-O1	13.06	155.86	123.30
3	B	702	CIT	O7-C3-C4	-12.28	80.65	109.40
3	A	702	CIT	C3-C4-C5	11.67	142.07	113.81
3	A	702	CIT	C3-C2-C1	11.46	141.56	113.81

There are no chirality outliers.

5 of 10 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	702	CIT	C2-C3-C4-C5
3	B	702	CIT	C2-C3-C6-O5
3	B	702	CIT	C2-C3-C6-O6
3	A	702	CIT	O7-C3-C4-C5
3	B	702	CIT	O1-C1-C2-C3

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	702	CIT	2	0
3	B	702	CIT	2	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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	649/683 (95%)	-0.12	23 (3%) 44 40	13, 20, 44, 66	1 (0%)
1	B	649/683 (95%)	-0.17	12 (1%) 68 67	13, 21, 39, 65	5 (0%)
All	All	1298/1366 (95%)	-0.15	35 (2%) 54 52	13, 21, 42, 66	6 (0%)

The worst 5 of 35 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	126	LEU	7.5
1	B	625	HIS	4.1
1	A	625	HIS	3.8
1	B	226	HIS	3.6
1	A	146	VAL	3.6

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

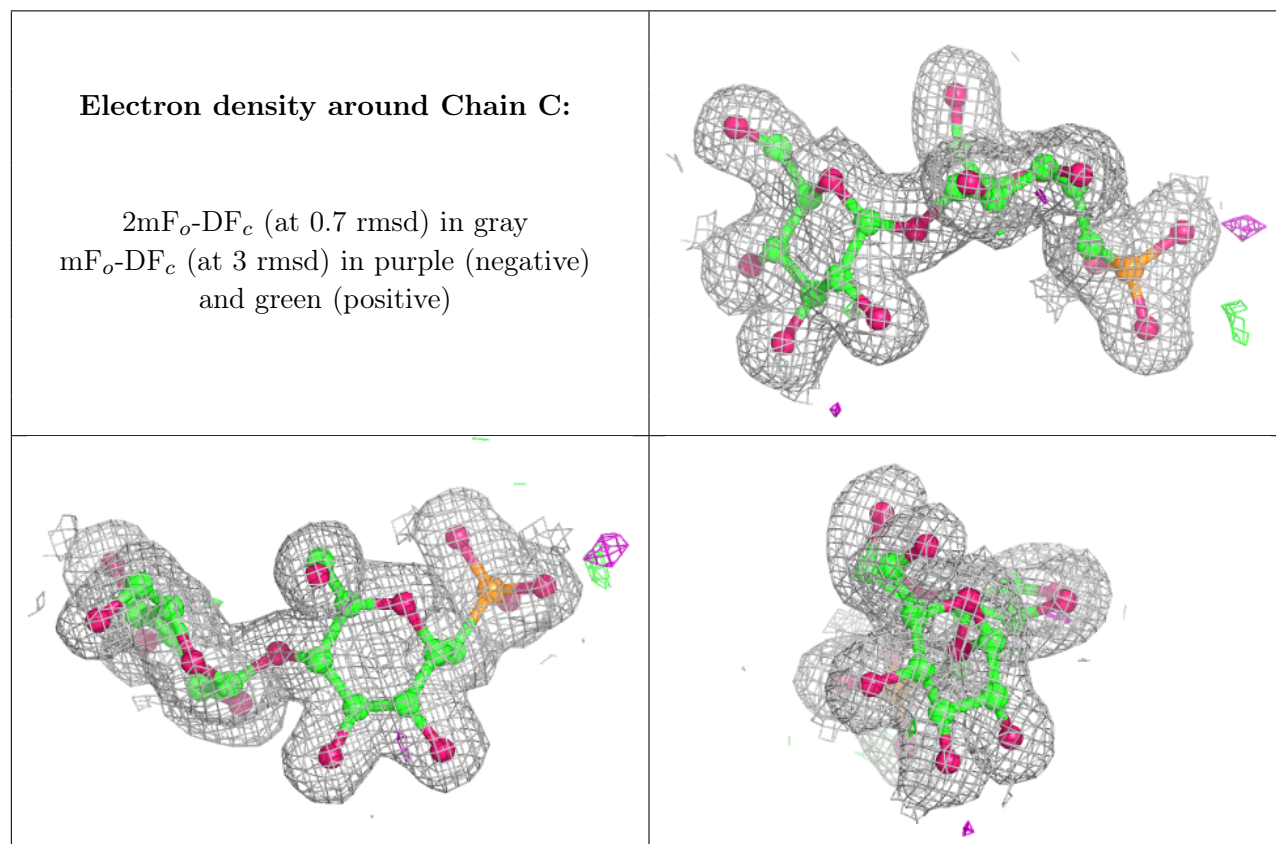
There are no non-standard protein/DNA/RNA residues in this entry.

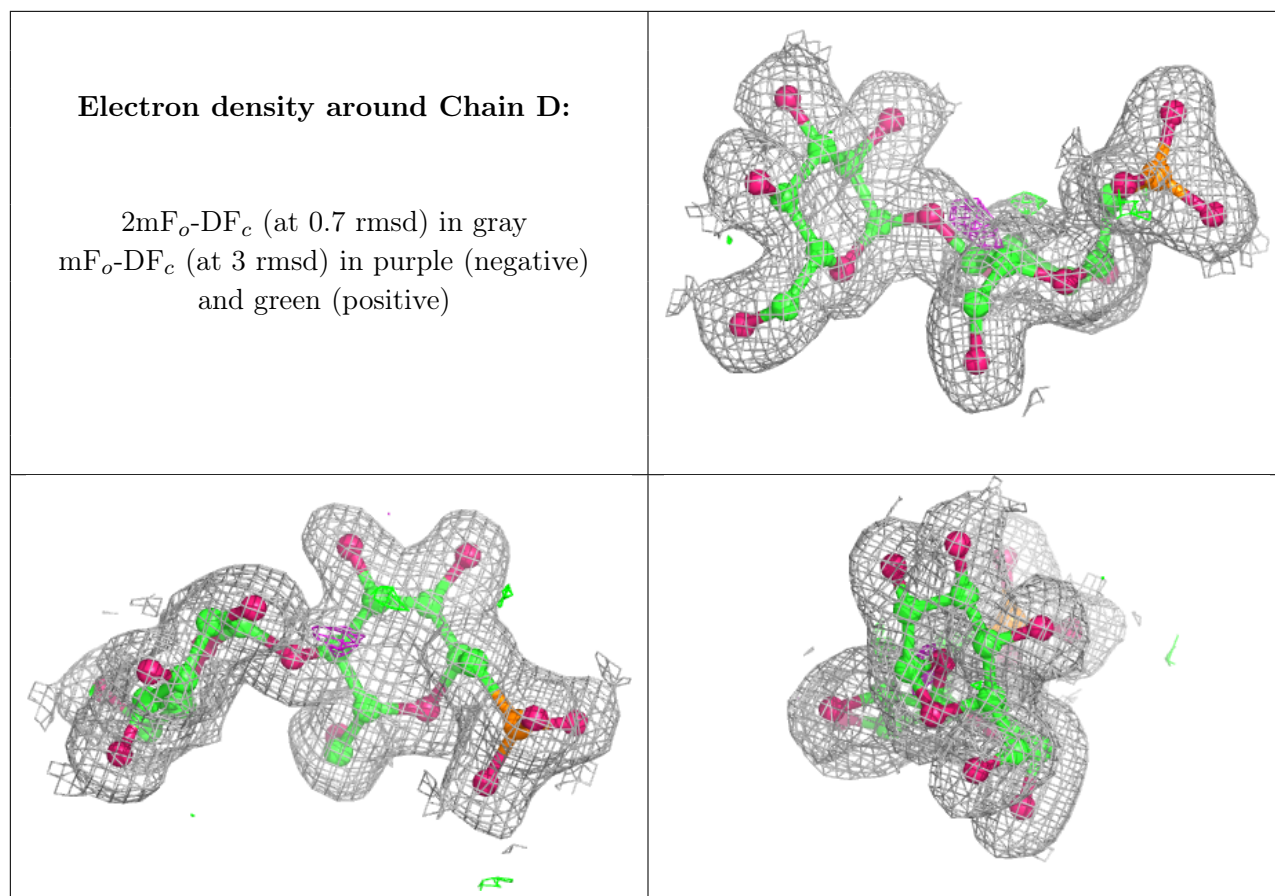
### 6.3 Carbohydrates [i](#)

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
2	GLC	C	2	11/12	0.97	0.07	16,17,18,19	0
2	GPM	D	1	16/16	0.97	0.10	12,17,22,22	0
2	GLC	D	2	11/12	0.97	0.07	14,16,18,19	0
2	GPM	C	1	16/16	0.98	0.09	14,17,22,22	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.





## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
4	EDO	A	704	4/4	0.71	0.20	31,34,34,39	4
4	EDO	A	706	4/4	0.84	0.21	25,26,28,30	4
4	EDO	B	703	4/4	0.85	0.19	25,29,33,36	0
3	CIT	B	702	13/13	0.87	0.15	28,33,37,38	13
3	CIT	A	702	13/13	0.88	0.14	28,33,36,38	13
4	EDO	A	703	4/4	0.91	0.16	22,23,24,27	4
4	EDO	A	705	4/4	0.93	0.20	32,45,46,48	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.