wwPDB X-ray Structure Validation Summary Report (i)

Nov 11, 2021 – 11:08 am GMT

PDB ID	:	7P39
Title	:	4,6-alpha-glucanotransferase GtfB from Limosilactobacillus reuteri NCC 2613 complexed with acarbose
Authors	:	Pijning, T.; te Poele, E.; Gangoiti, J.; Boerner, T.; Dijkhuizen, L.
Deposited on	:	2021-07-07
Resolution	:	2.90 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 following versions of software and data (see references (1)) were used in the production of this report:

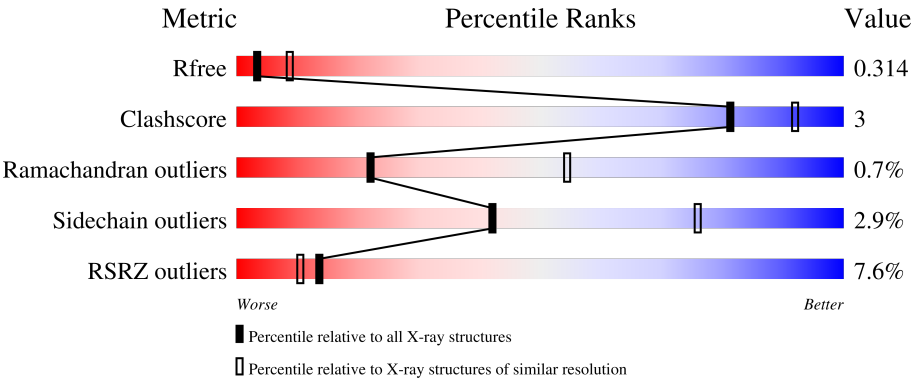
MolProbity	:	4.02b-467
Mogul	:	1.8.4 (270009), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.90 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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	881	8% 86% 8% • 6%
1	B	881	7% 87% 7% 6%
2	D	3	100%
2	G	3	100%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 13136 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 Dextransucrase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	832	Total	C	N	O	S	0	0	0
			6511	4054	1101	1334	22			
1	B	832	Total	C	N	O	S	0	0	0
			6511	4054	1101	1334	22			

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	397	MET	-	initiating methionine	UNP A0A1Z2RUH3
A	398	ALA	-	expression tag	UNP A0A1Z2RUH3
A	399	HIS	-	expression tag	UNP A0A1Z2RUH3
A	400	HIS	-	expression tag	UNP A0A1Z2RUH3
A	401	HIS	-	expression tag	UNP A0A1Z2RUH3
A	402	HIS	-	expression tag	UNP A0A1Z2RUH3
A	403	HIS	-	expression tag	UNP A0A1Z2RUH3
A	404	HIS	-	expression tag	UNP A0A1Z2RUH3
A	405	SER	-	expression tag	UNP A0A1Z2RUH3
A	406	ALA	-	expression tag	UNP A0A1Z2RUH3
A	407	ALA	-	expression tag	UNP A0A1Z2RUH3
A	408	LEU	-	expression tag	UNP A0A1Z2RUH3
A	409	GLU	-	expression tag	UNP A0A1Z2RUH3
A	410	VAL	-	expression tag	UNP A0A1Z2RUH3
A	411	LEU	-	expression tag	UNP A0A1Z2RUH3
A	412	PHE	-	expression tag	UNP A0A1Z2RUH3
A	413	GLN	-	expression tag	UNP A0A1Z2RUH3
A	414	GLY	-	expression tag	UNP A0A1Z2RUH3
A	415	PRO	-	expression tag	UNP A0A1Z2RUH3
A	416	GLY	-	expression tag	UNP A0A1Z2RUH3
B	397	MET	-	initiating methionine	UNP A0A1Z2RUH3
B	398	ALA	-	expression tag	UNP A0A1Z2RUH3
B	399	HIS	-	expression tag	UNP A0A1Z2RUH3
B	400	HIS	-	expression tag	UNP A0A1Z2RUH3
B	401	HIS	-	expression tag	UNP A0A1Z2RUH3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	402	HIS	-	expression tag	UNP A0A1Z2RUH3
B	403	HIS	-	expression tag	UNP A0A1Z2RUH3
B	404	HIS	-	expression tag	UNP A0A1Z2RUH3
B	405	SER	-	expression tag	UNP A0A1Z2RUH3
B	406	ALA	-	expression tag	UNP A0A1Z2RUH3
B	407	ALA	-	expression tag	UNP A0A1Z2RUH3
B	408	LEU	-	expression tag	UNP A0A1Z2RUH3
B	409	GLU	-	expression tag	UNP A0A1Z2RUH3
B	410	VAL	-	expression tag	UNP A0A1Z2RUH3
B	411	LEU	-	expression tag	UNP A0A1Z2RUH3
B	412	PHE	-	expression tag	UNP A0A1Z2RUH3
B	413	GLN	-	expression tag	UNP A0A1Z2RUH3
B	414	GLY	-	expression tag	UNP A0A1Z2RUH3
B	415	PRO	-	expression tag	UNP A0A1Z2RUH3
B	416	GLY	-	expression tag	UNP A0A1Z2RUH3

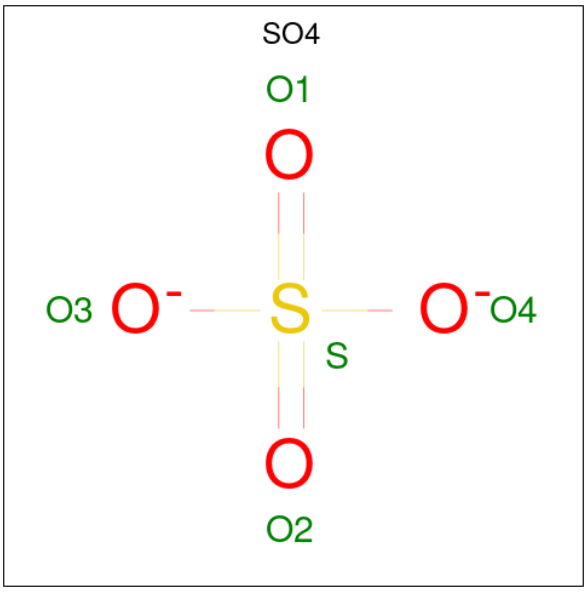
- Molecule 2 is an oligosaccharide called 4,6-dideoxy-4-{[(1S,4R,5S,6S)-4,5,6-trihydroxy-3-(hydroxymethyl)cyclohex-2-en-1-yl]amino}-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-beta-D-glucopyranose.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	D	3	Total	C	N	O	0	0	0
			44	25	1	18			
2	G	3	Total	C	N	O	0	0	0
			44	25	1	18			

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Ca	0	0
			1	1		
3	B	1	Total	Ca	0	0
			1	1		

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



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

- Molecule 5 is water.

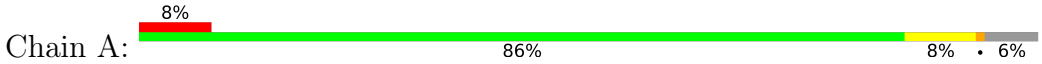
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	10	Total O 10 10	0	0
5	B	9	Total O 9 9	0	0

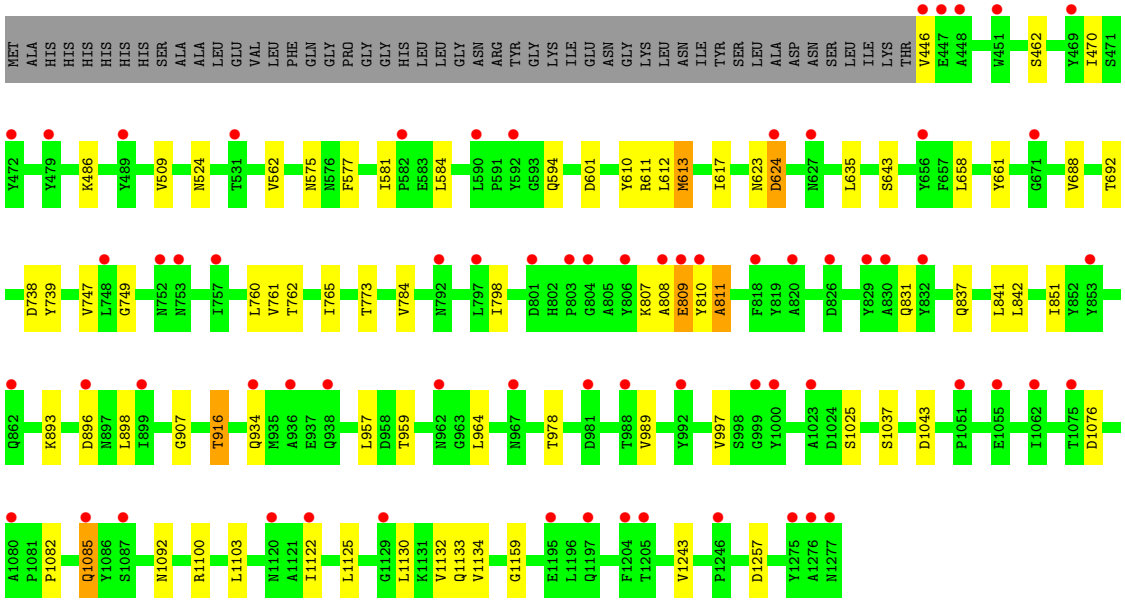
3 Residue-property plots (i)

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: Dextransucrase

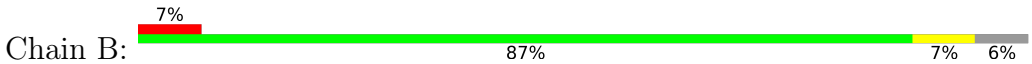
Chain A:

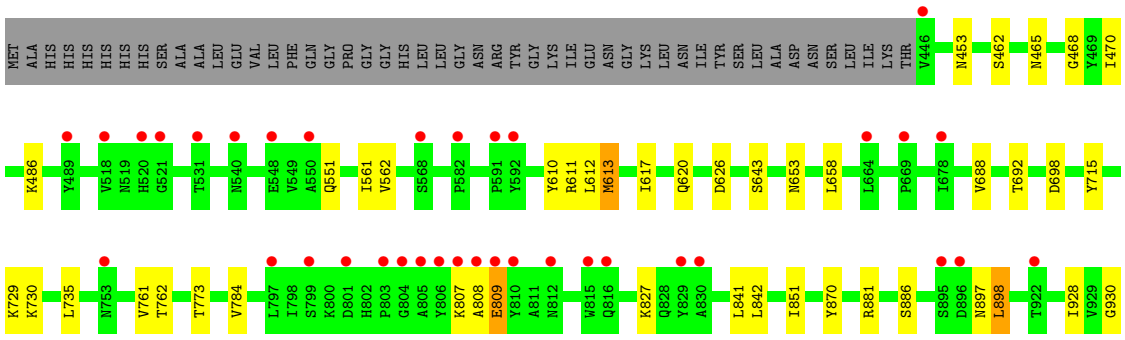


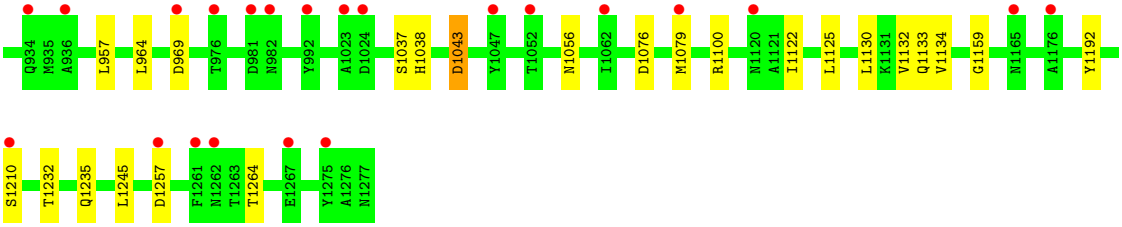


• Molecule 1: Dextransucrase

Chain B:





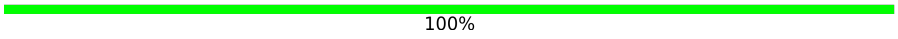


• Molecule 2: 4,6-dideoxy-4-{[(1S,4R,5S,6S)-4,5,6-trihydroxy-3-(hydroxymethyl)cyclohex-2-en-1-yl]amino}-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-beta-D-glucopyranose

Chain D:

• Molecule 2: 4,6-dideoxy-4-{[(1S,4R,5S,6S)-4,5,6-trihydroxy-3-(hydroxymethyl)cyclohex-2-en-1-yl]amino}-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-beta-D-glucopyranose

Chain G:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	107.22Å 133.81Å 147.87Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.29 – 2.90 47.16 – 2.90	Depositor EDS
% Data completeness (in resolution range)	97.8 (49.29-2.90) 97.8 (47.16-2.90)	Depositor EDS
R_{merge}	0.43	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.83 (at 2.91Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
R, R_{free}	0.293 , 0.315 0.295 , 0.314	Depositor DCC
R_{free} test set	2250 reflections (4.82%)	wwPDB-VP
Wilson B-factor (Å^2)	35.0	Xtriage
Anisotropy	0.619	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	0.85	EDS
Total number of atoms	13136	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 46.53 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.1296e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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: AC1, SO4, BGC, CA, GLC

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	0.37	0/6653	0.57	0/9063
1	B	0.37	0/6653	0.57	0/9063
All	All	0.37	0/13306	0.57	0/18126

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	6511	0	6129	40	0
1	B	6511	0	6129	26	0
2	D	44	0	30	0	0
2	G	44	0	30	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	5	0	0	0	0
5	A	10	0	0	0	0
5	B	9	0	0	0	0
All	All	13136	0	12318	66	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 66 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:1122:ILE:HG23	1:A:1132:VAL:HG11	1.61	0.83
1:B:1122:ILE:HG23	1:B:1132:VAL:HG11	1.61	0.82
1:A:577:PHE:CZ	1:A:581:ILE:HD11	2.15	0.82
1:B:1076:ASP:HB3	1:B:1133:GLN:HE22	1.52	0.74
1:A:1076:ASP:HB3	1:A:1133:GLN:HE22	1.51	0.73

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	A	830/881 (94%)	793 (96%)	30 (4%)	7 (1%)	19	51
1	B	830/881 (94%)	791 (95%)	34 (4%)	5 (1%)	25	58
All	All	1660/1762 (94%)	1584 (95%)	64 (4%)	12 (1%)	22	54

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	808	ALA
1	A	809	GLU
1	A	811	ALA
1	A	1243	VAL
1	B	808	ALA

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	702/741 (95%)	681 (97%)	21 (3%)	41	75
1	B	702/741 (95%)	682 (97%)	20 (3%)	43	76
All	All	1404/1482 (95%)	1363 (97%)	41 (3%)	42	76

5 of 41 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	698	ASP
1	B	1056	ASN
1	B	729	LYS
1	B	969	ASP
1	B	1210	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 34 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	973	ASN
1	B	1012	GLN
1	B	1133	GLN
1	A	780	ASN
1	A	753	ASN

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)

6 monosaccharides are modelled in this entry.

There are no bond length outliers.

There are no bond angle outliers.

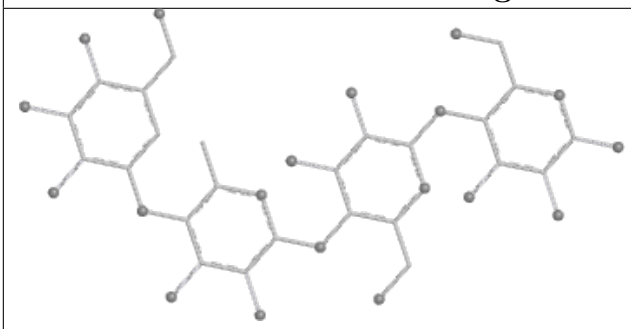
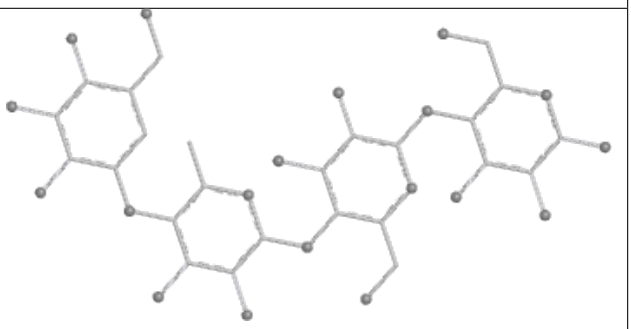
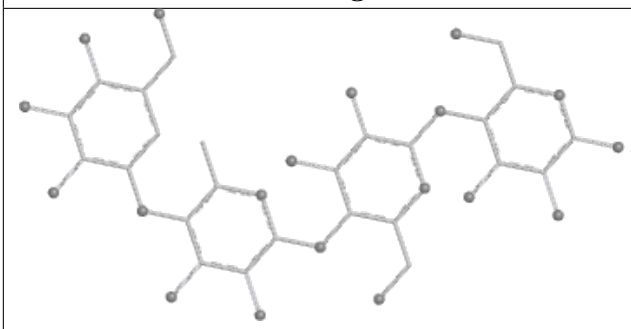
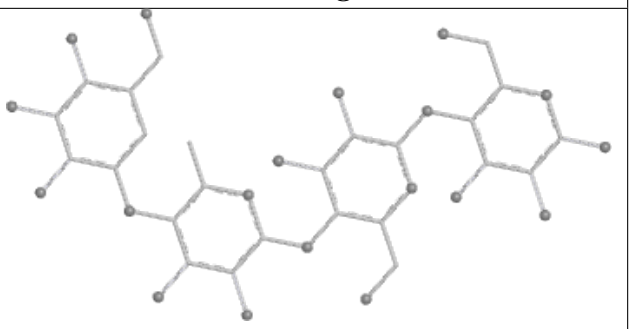
There are no chirality outliers.

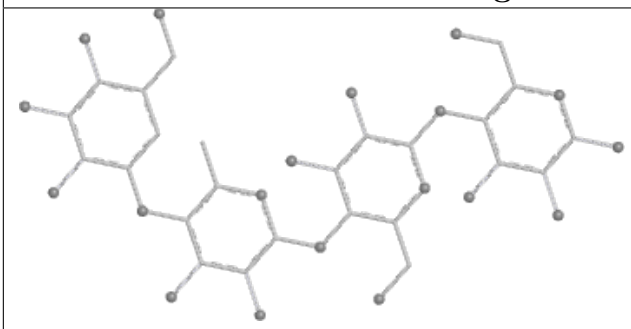
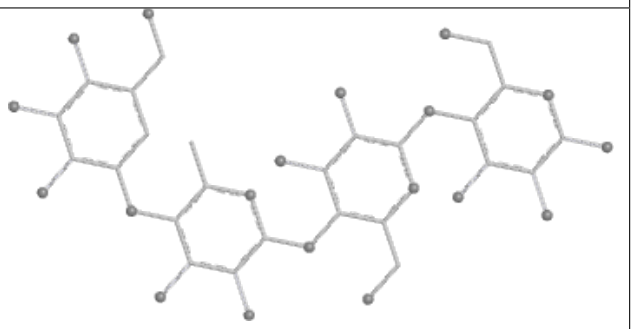
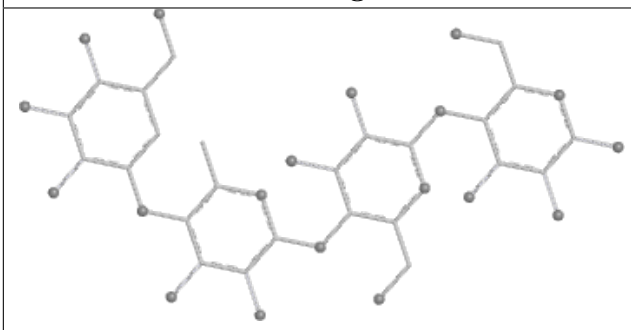
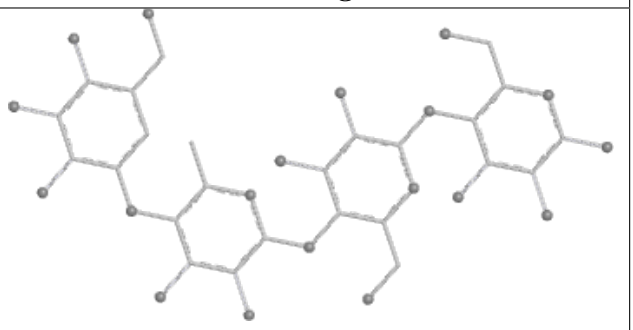
There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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 D	
 Bond lengths	 Bond angles
 Torsions	 Rings

Oligosaccharide Chain G	
	
Bond lengths	**Bond angles**
	
Torsions	**Rings**

5.6 Ligand geometry

Of 3 ligands modelled in this entry, 2 are monoatomic - leaving 1 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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

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	A	832/881 (94%)	0.95	68 (8%) 11 9	30, 41, 66, 102	0
1	B	832/881 (94%)	0.91	58 (6%) 16 12	29, 41, 56, 96	0
All	All	1664/1762 (94%)	0.93	126 (7%) 13 10	29, 41, 62, 102	0

The worst 5 of 126 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	808	ALA	8.6
1	A	1275	TYR	6.4
1	A	806	TYR	5.4
1	B	810	TYR	5.4
1	A	810	TYR	5.2

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

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
2	BGC	D	1	12/12	0.70	0.34	51,53,53,54	0
2	BGC	G	1	12/12	0.81	0.31	46,47,47,47	0
2	AC1	D	3	21/22	0.82	0.28	45,46,47,47	0
2	AC1	G	3	21/22	0.82	0.26	45,45,46,46	0
2	GLC	D	2	11/12	0.83	0.22	47,49,49,49	0

Continued on next page...

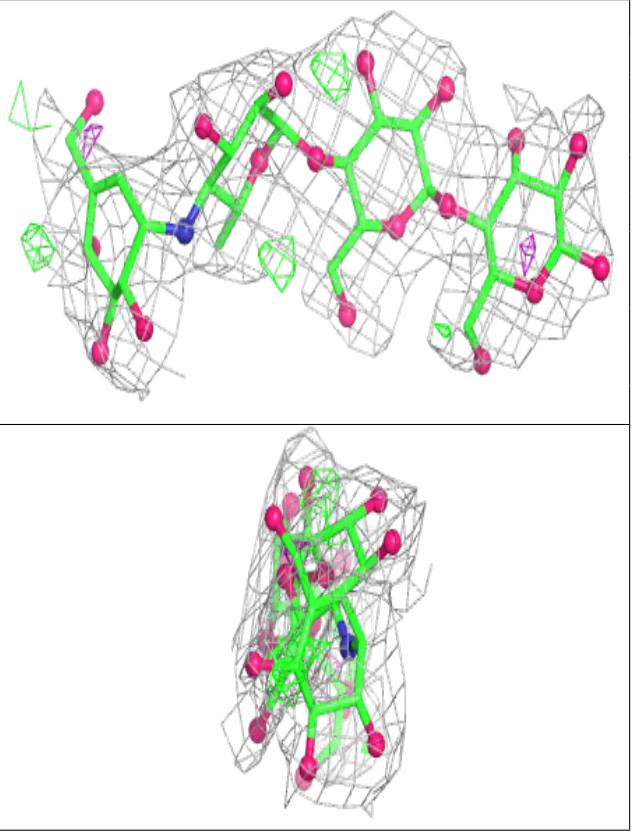
Continued from previous page...

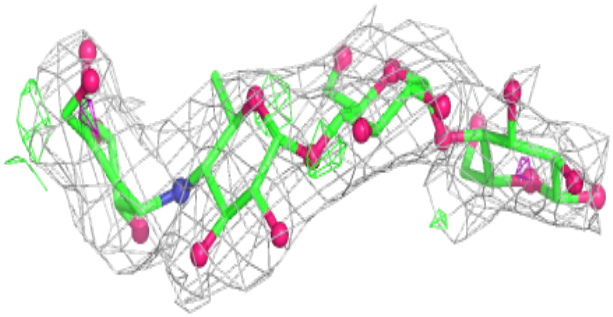
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	GLC	G	2	11/12	0.87	0.22	43,45,45,45	0

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 D:

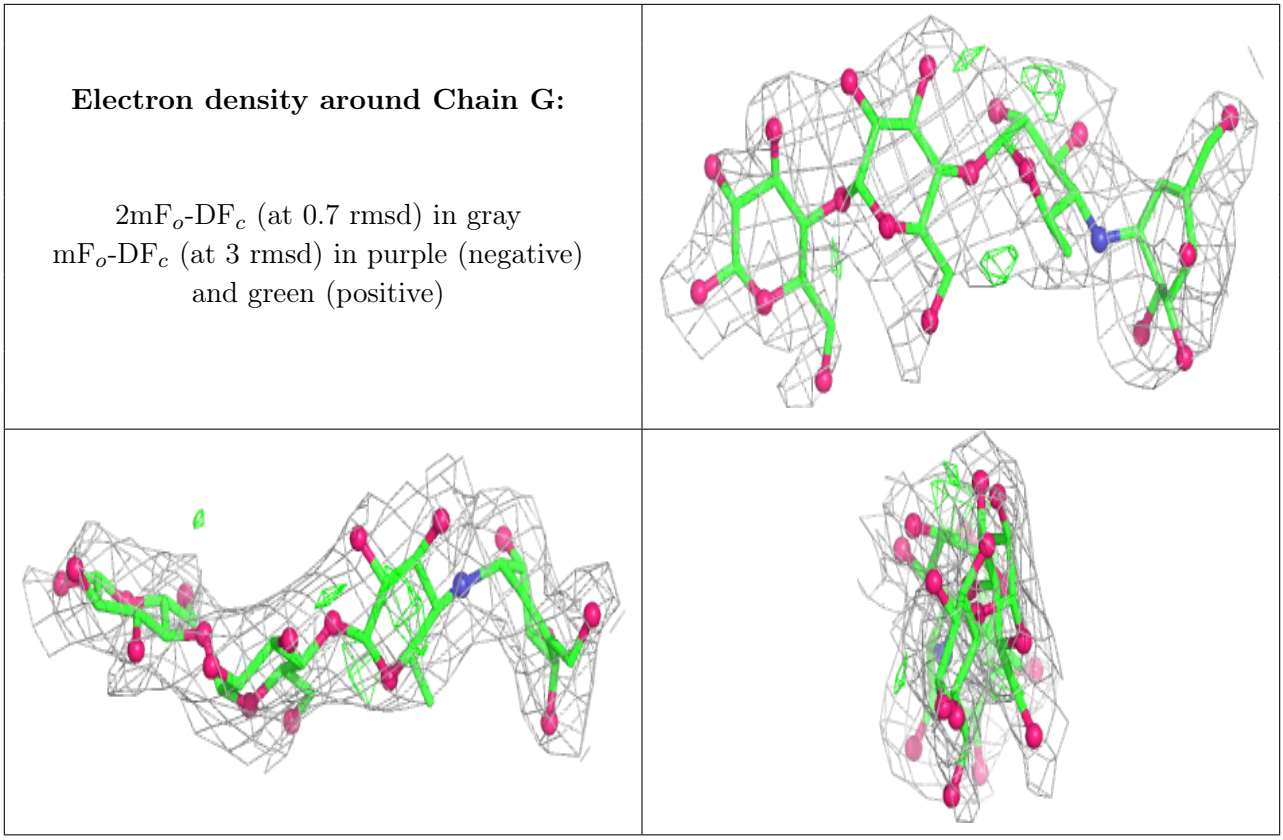
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)





Electron density around Chain G:

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



6.4 Ligands

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	CA	B	1301	1/1	0.69	0.14	27,27,27,27	0
4	SO4	A	1302	5/5	0.74	0.34	72,73,73,73	0
3	CA	A	1301	1/1	0.79	0.14	29,29,29,29	0

6.5 Other polymers

There are no such residues in this entry.