

wwPDB X-ray Structure Validation Summary Report (i)

Sep 20, 2023 – 11:21 PM EDT

PDB ID : 5F7U

Title: Cycloalternan-forming enzyme from Listeria monocytogenes in complex with

pentasaccharide substrate

Authors: Light, S.H.; Anderson, W.F.; Center for Structural Genomics of Infectious

Diseases (CSGID)

Deposited on : 2015-12-08

Resolution : 1.70 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 (1)) were used in the production of this report:

 $Mol Probity \quad : \quad 4.02b\text{--}467$

Mogul: 1.8.5 (274361), CSD as541be (2020)

Xtriage (Phenix) : 1.13 EDS : 2.35.1

Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)

 $Refmac \quad : \quad 5.8.0158$

CCP4 : 7.0.044 (Gargrove)

Ideal geometry (proteins) : Engh & Huber (2001) Ideal geometry (DNA, RNA) : Parkinson et al. (1996)

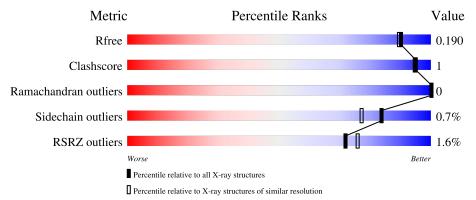
Validation Pipeline (wwPDB-VP) : 2.35.1

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 1.70 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 $(\# \mathrm{Entries})$	$\begin{array}{c} {\rm Similar\ resolution} \\ (\#{\rm Entries},{\rm resolution\ range}(\mathring{\rm A})) \end{array}$
R_{free}	130704	4298 (1.70-1.70)
Clashscore	141614	4695 (1.70-1.70)
Ramachandran outliers	138981	4610 (1.70-1.70)
Sidechain outliers	138945	4610 (1.70-1.70)
RSRZ outliers	127900	4222 (1.70-1.70)

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	1063	2%	97%		•
2	В	4	25%	50%	25%	
3	С	2		100%		
3	D	2	50%		50%	
4	Е	2		100%		

Continued on next page...



Continued from previous page...

Mol	Chain	Length		Quality of chain	
5	F	3	33%	67%	-



2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 9918 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.

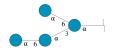
• Molecule 1 is a protein called Cycloalternan-forming enzyme.

Mol	Chain	Residues		A	toms			ZeroOcc	AltConf	Trace
1	A	1059	Total 8405	C 5277	N 1342	O 1755	S 31	0	10	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	29	SER	-	expression tag	UNP Q8Y4J2
A	30	ASN	-	expression tag	UNP Q8Y4J2
A	565	ASN	ASP	engineered mutation	UNP Q8Y4J2

• Molecule 2 is an oligosaccharide called alpha-D-glucopyranose-(1-6)-alpha-D-glucopyranose-(1-3)-[alpha-D-glucopyranose-(1-6)]alpha-D-glucopyranose.



Mol	Chain	Residues	At	oms	ı	ZeroOcc	AltConf	Trace
2	В	4	Total 45	C 24	O 21	0	0	0

• Molecule 3 is an oligosaccharide called alpha-D-glucopyranose-(1-6)-alpha-D-glucopyranose.



Mol	Chain	Residues	At	\overline{oms}		ZeroOcc	AltConf	Trace
3	С	2	Total 23	C 12	O 11	0	0	0

Continued on next page...



Continued from previous page...

Mol	Chain	Residues	At	oms		ZeroOcc	AltConf	Trace
3	D	2	Total 23	C 12	O 11	0	0	0

• Molecule 4 is an oligosaccharide called alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose.



Mol	Chain	Residues	At	oms		ZeroOcc	AltConf	Trace
4	E	2	Total 23	C 12	O 11	0	0	0

• Molecule 5 is an oligosaccharide called alpha-D-glucopyranose-(1-6)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose.



Mol	Chain	Residues	At	oms		ZeroOcc	AltConf	Trace
5	F	3	Total 34	C 18	O 16	0	0	0

• Molecule 6 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

\mathbf{M}	ol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	j	A	5	Total Mg 5 5	0	0

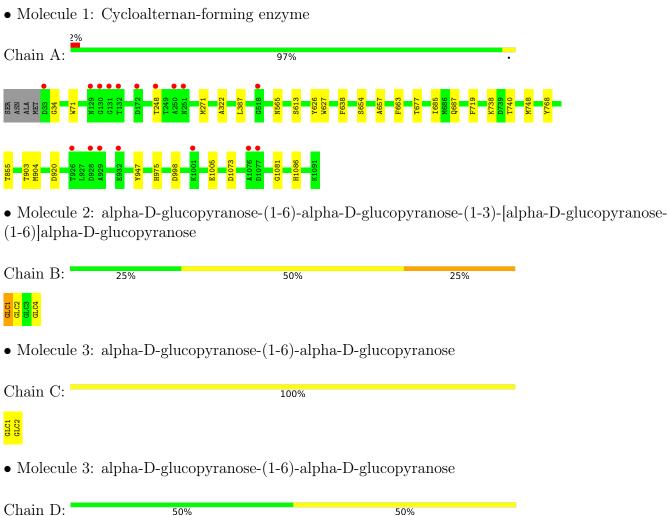
• Molecule 7 is water.

\mathbf{Mol}	Chain	Residues	${f Atoms}$	ZeroOcc	AltConf
7	A	1360	Total O 1360 1360	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.



50% 50%

• Molecule 4: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose

Chain E: 100%





 \bullet Molecule 5: alpha-D-glucopyranose-(1-6)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose e

Chain F: 33% 67%





4 Data and refinement statistics (i)

Property	Value	Source	
Space group	C 1 2 1	Depositor	
Cell constants	168.22Å 100.89Å 71.81Å	Depositor	
a, b, c, α , β , γ	90.00° 104.01° 90.00°	Depositor	
Resolution (Å)	30.00 - 1.70	Depositor	
resolution (A)	29.56 - 1.70	EDS	
% Data completeness	99.8 (30.00-1.70)	Depositor	
(in resolution range)	99.9 (29.56-1.70)	EDS	
R_{merge}	0.07	Depositor	
R_{sym}	(Not available)	Depositor	
$< I/\sigma(I) > 1$	2.44 (at 1.70Å)	Xtriage	
Refinement program	REFMAC 5.8.0135	Depositor	
R, R_{free}	0.159 , 0.190	Depositor	
it, it free	0.159 , 0.190	DCC	
R_{free} test set	6444 reflections (5.05%)	wwPDB-VP	
Wilson B-factor (Å ²)	18.3	Xtriage	
Anisotropy	0.342	Xtriage	
Bulk solvent $k_{sol}(e/Å^3)$, $B_{sol}(Å^2)$	0.36, 41.9	EDS	
L-test for twinning ²	$ < L > = 0.50, < L^2> = 0.34$	Xtriage	
Estimated twinning fraction	No twinning to report.	Xtriage	
F_o, F_c correlation	0.97	EDS	
Total number of atoms	9918	wwPDB-VP	
Average B, all atoms $(Å^2)$	22.0	wwPDB-VP	

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 3.82% of the height of the origin peak. No significant pseudotranslation is detected.

²Theoretical 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.



¹Intensities estimated from amplitudes.

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: MG, 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	Во	ond angles
		Chain	RMSZ	# Z > 5	RMSZ	# Z > 5
	1	A	0.54	0/8614	0.69	$1/11735 \ (0.0\%)$

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$Observed(^o)$	$\operatorname{Ideal}(^{o})$
1	A	638	PHE	CB-CG-CD2	-5.13	117.21	120.80

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (i)

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	8405	0	7725	17	0
2	В	45	0	39	1	0
3	С	23	0	21	0	0
3	D	23	0	21	0	0
4	Е	23	0	21	0	0
5	F	34	0	30	3	0
6	A	5	0	0	0	0
7	A	1360	0	0	2	0
All	All	9918	0	7857	19	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 1.

The worst 5 of 19 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	$\begin{array}{c} {\rm Interatomic} \\ {\rm distance} \ ({\rm \AA}) \end{array}$	$egin{aligned} ext{Clash} \ ext{overlap} & (ext{Å}) \end{aligned}$	
1:A:565:ASN:HD21	5:F:2:GLC:H2	1.13	1.10	
1:A:565:ASN:ND2	5:F:2:GLC:H2	1.94	0.82	
7:A:1506:HOH:O	2:B:1:GLC:H61	2.00	0.61	
1:A:748:MET:HG3	1:A:768:TYR:CD1	2.37	0.59	
1:A:248:THR:HB	1:A:271:MET:CE	2.37	0.54	

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	1067/1063 (100%)	1038 (97%)	29 (3%)	0	100 1	100	

There are no Ramachandran outliers to report.

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	900/893 (101%)	894 (99%)	6 (1%)	84 77		



_	CC	• 1	• . 1			• 1	1 .	1.	. 1	1 1
\circ	of b	residues	with a	non-rotam	eric	sideo	chain	are li	sted	below:

Mol	Chain	Res	Type
1	A	920	ASP
1	A	1073	ASP
1	A	1086	HIS
1	A	663	PHE
1	A	627	TRP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	509	GLN
1	A	565	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)

13 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		
MIOI	туре				Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z >2
2	GLC	В	1	2	12,12,12	0.87	0	17,17,17	1.91	5 (29%)
2	GLC	В	2	2	11,11,12	0.48	0	15,15,17	1.17	2 (13%)
2	GLC	В	3	2	11,11,12	0.50	0	15,15,17	0.75	0
2	GLC	В	4	2	11,11,12	0.82	0	15,15,17	2.06	3 (20%)
3	GLC	С	1	3	12,12,12	0.74	0	17,17,17	1.34	3 (17%)



Mol	Tuno	Chain	Res	Link	Вс	Bond lengths			Bond angles		
MIOI	Type		nes	Lilik	Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2	
3	GLC	С	2	3	11,11,12	0.46	0	15,15,17	1.22	1 (6%)	
3	GLC	D	1	3	12,12,12	0.65	0	17,17,17	0.77	0	
3	GLC	D	2	3	11,11,12	0.57	0	15,15,17	1.05	1 (6%)	
4	GLC	Е	1	4	12,12,12	0.63	0	17,17,17	0.90	1 (5%)	
4	GLC	Е	2	4	11,11,12	0.62	0	15,15,17	1.02	1 (6%)	
5	GLC	F	1	5	12,12,12	0.68	0	17,17,17	1.13	2 (11%)	
5	GLC	F	2	5	11,11,12	0.73	0	15,15,17	1.33	3 (20%)	
5	GLC	F	3	5	11,11,12	0.50	0	15,15,17	0.93	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
2	GLC	В	1	2	-	2/2/22/22	0/1/1/1
2	GLC	В	2	2	-	0/2/19/22	0/1/1/1
2	GLC	В	3	2	-	0/2/19/22	0/1/1/1
2	GLC	В	4	2	-	0/2/19/22	0/1/1/1
3	GLC	С	1	3	-	2/2/22/22	0/1/1/1
3	GLC	С	2	3	-	0/2/19/22	0/1/1/1
3	GLC	D	1	3	-	0/2/22/22	0/1/1/1
3	GLC	D	2	3	-	0/2/19/22	0/1/1/1
4	GLC	Ε	1	4	-	1/2/22/22	0/1/1/1
4	GLC	Ε	2	4	-	1/2/19/22	0/1/1/1
5	GLC	F	1	5	-	2/2/22/22	0/1/1/1
5	GLC	F	2	5	-	0/2/19/22	0/1/1/1
5	GLC	F	3	5	-	0/2/19/22	0/1/1/1

There are no bond length outliers.

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^o)$	$\operatorname{Ideal}({}^{o})$
2	В	4	GLC	C1-O5-C5	5.93	120.22	112.19
3	С	2	GLC	C1-O5-C5	4.17	117.84	112.19
2	В	1	GLC	O3-C3-C4	3.68	118.86	110.35
3	D	2	GLC	C1-O5-C5	3.49	116.92	112.19
2	В	1	GLC	C4-C3-C2	-3.38	104.93	110.82

There are no chirality outliers.



5 of 8 torsion outliers are listed below:

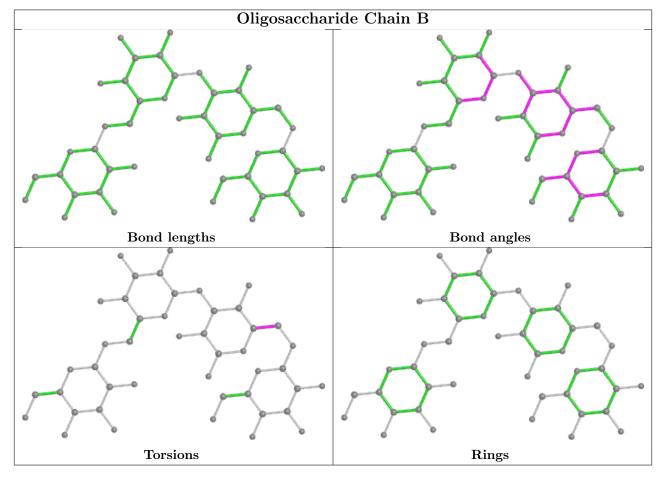
Mol	Chain	Res	Type	Atoms
2	В	1	GLC	O5-C5-C6-O6
5	F	1	GLC	O5-C5-C6-O6
3	С	1	GLC	O5-C5-C6-O6
2	В	1	GLC	C4-C5-C6-O6
3	С	1	GLC	C4-C5-C6-O6

There are no ring outliers.

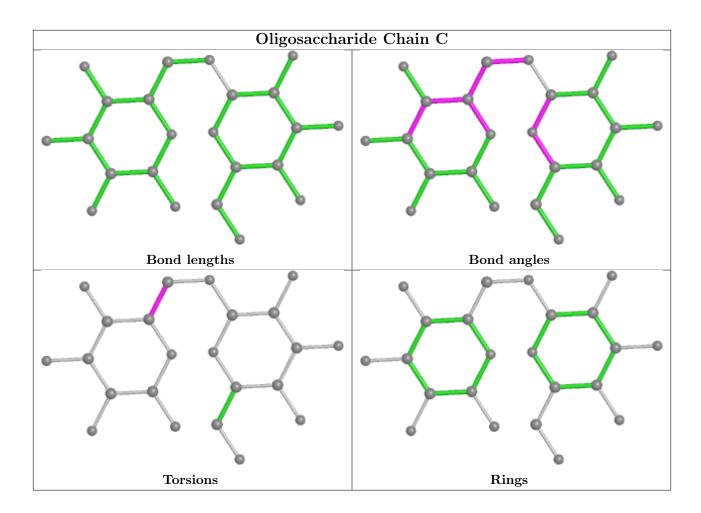
3 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	F	2	GLC	3	0
5	F	1	GLC	1	0
2	В	1	GLC	1	0

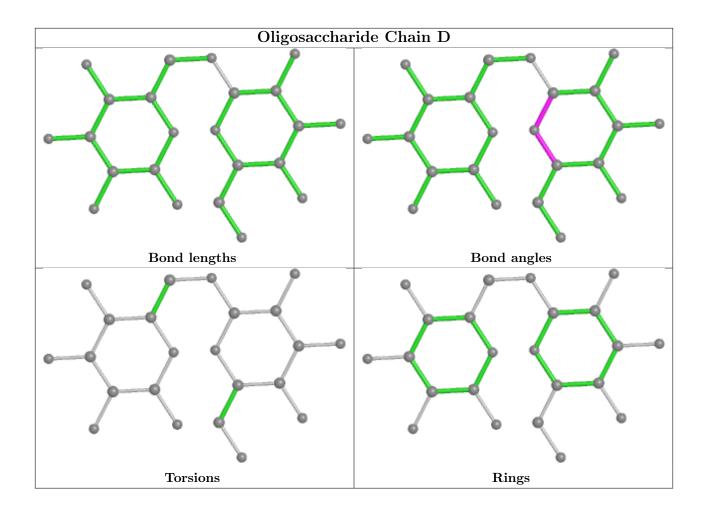
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.



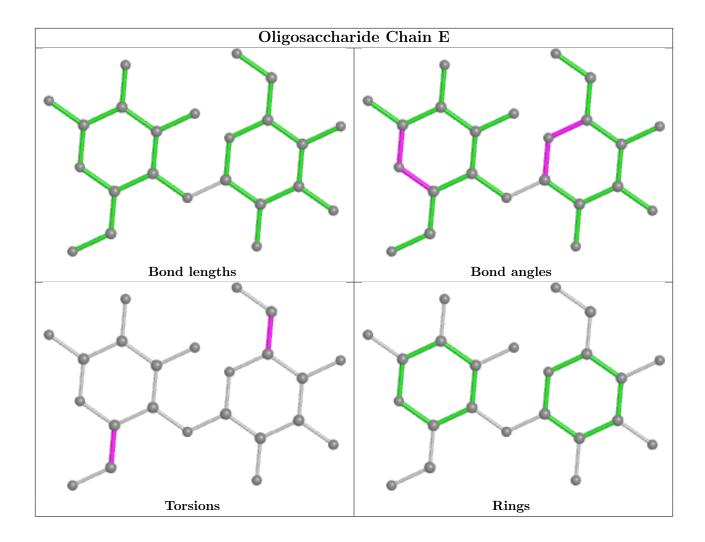




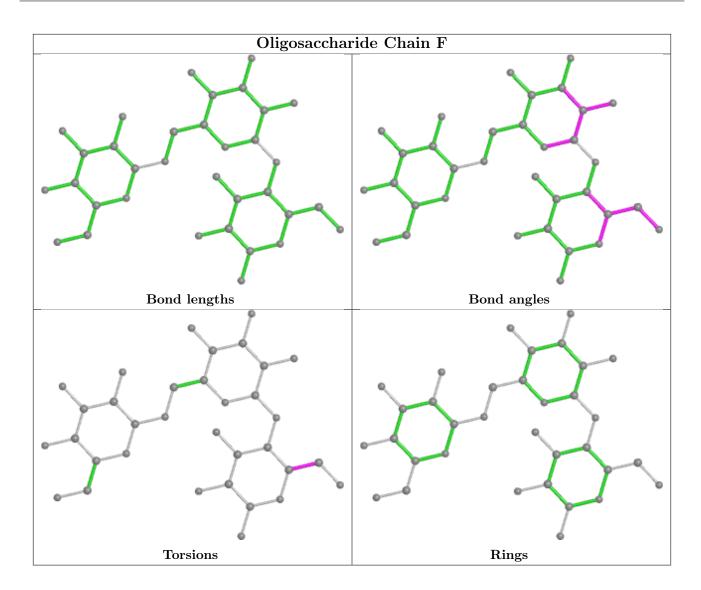












5.6 Ligand geometry (i)

Of 5 ligands modelled in this entry, 5 are monoatomic - leaving 0 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 (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^{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	${\bf Analysed} \qquad <\! {\bf RSRZ}\! >\!$		#RSRZ	i>2	$OWAB(Å^2)$	Q < 0.9
1	A	1059/1063 (99%)	-0.19	17 (1%) 72	76	10, 19, 34, 44	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	132	THR	4.8
1	A	250	ALA	4.4
1	A	129	ASN	4.0
1	A	251	ASN	3.4
1	A	932	GLU	3.0

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^{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	$\operatorname{B-factors}(\mathring{\mathbf{A}}^2)$	Q<0.9
2	GLC	В	4	11/12	0.59	0.29	43,45,46,47	0
3	GLC	С	1	12/12	0.71	0.27	35,51,55,56	0
2	GLC	В	1	12/12	0.75	0.21	30,38,42,44	0
5	GLC	F	1	12/12	0.75	0.20	27,36,38,43	0
3	GLC	D	2	11/12	0.82	0.16	34,37,39,39	0
4	GLC	Е	1	12/12	0.84	0.24	38,41,43,44	0
3	GLC	D	1	12/12	0.84	0.11	31,33,35,36	0

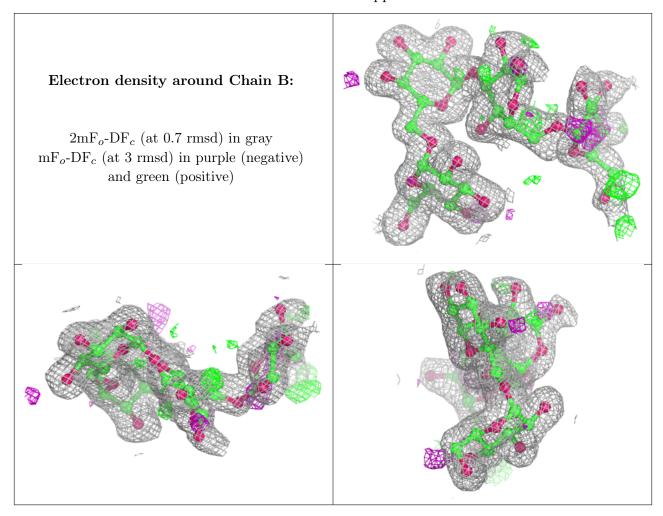
Continued on next page...



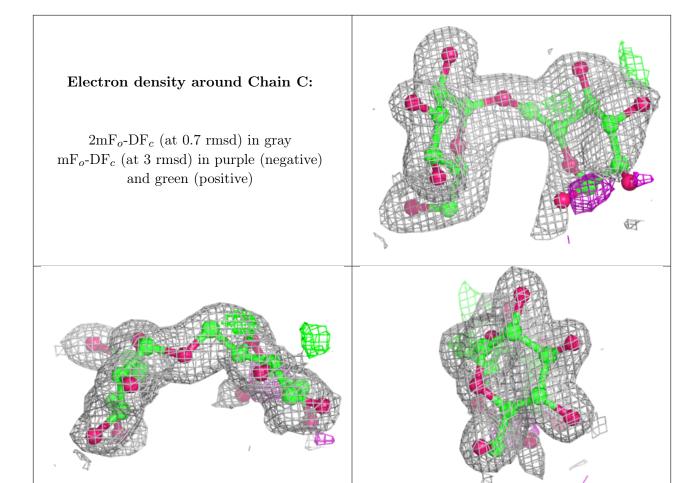
Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q < 0.9
4	GLC	Ε	2	11/12	0.86	0.14	40,40,41,41	0
5	GLC	F	3	11/12	0.90	0.10	23,25,28,29	0
3	GLC	С	2	11/12	0.92	0.12	27,30,31,32	0
5	GLC	F	2	11/12	0.93	0.09	19,22,24,24	0
2	GLC	В	2	11/12	0.95	0.08	18,23,26,26	0
2	GLC	В	3	11/12	0.97	0.06	17,17,18,18	0

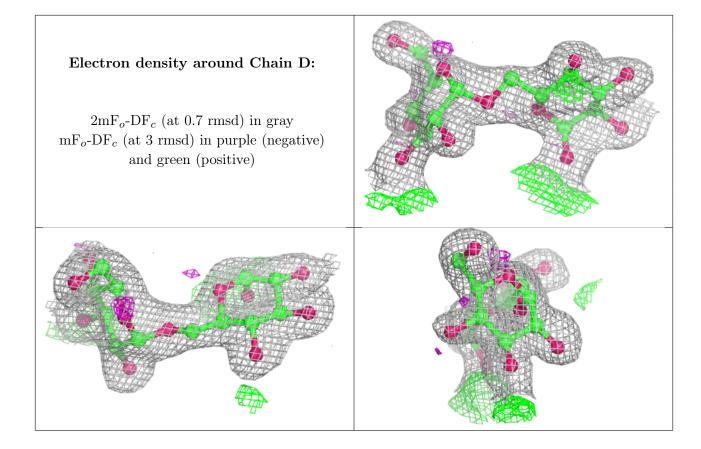
The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.



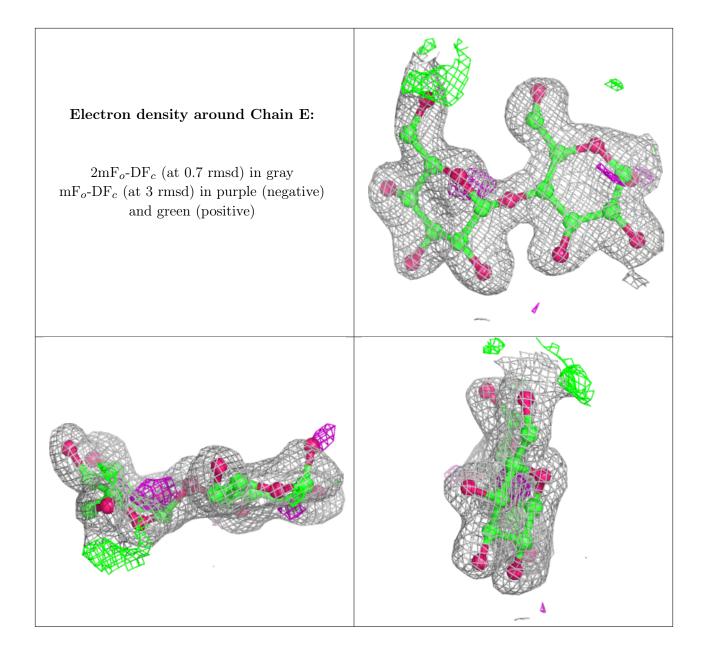




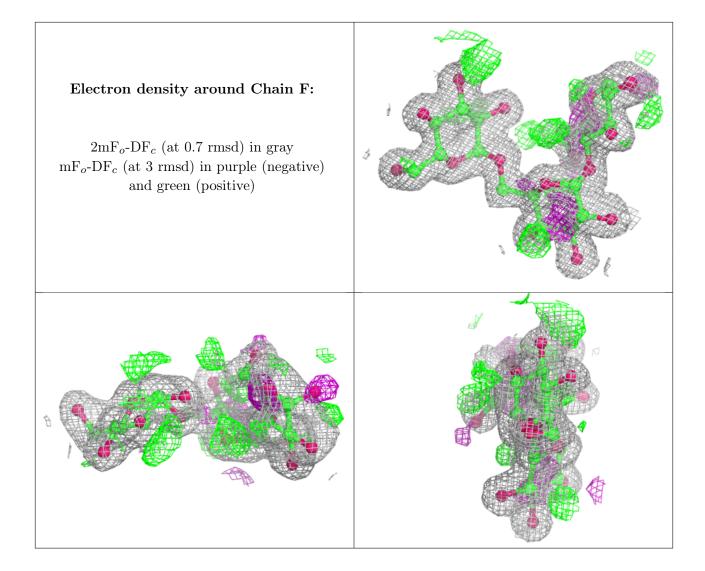












6.4 Ligands (i)

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^{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	${f B-factors}({f A}^2)$	Q<0.9
6	MG	A	1105	1/1	0.95	0.21	24,24,24,24	0
6	MG	A	1103	1/1	0.98	0.27	30,30,30,30	0
6	MG	A	1101	1/1	0.98	0.19	11,11,11,11	0
6	MG	A	1104	1/1	0.99	0.05	15,15,15,15	0
6	MG	A	1102	1/1	0.99	0.03	14,14,14,14	0



6.5 Other polymers (i)

There are no such residues in this entry.

