Full wwPDB X-ray Structure Validation Report ⓘ

Aug 16, 2020 – 08:30 PM BST

PDB ID	:	6ZJX
Title	:	Cold-adapted beta-D-galactosidase from Arthrobacter sp. 32cB mutant D207A in complex with saccharose
Authors	:	Rutkiewicz, M.; Bujacz, A.; Bujacz, G.
Deposited on	:	2020-06-29
Resolution	:	2.21 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:

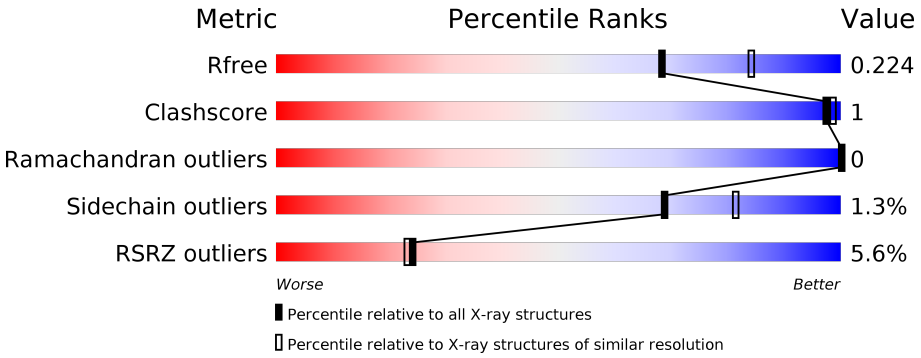
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13.1

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 2.21 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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 on 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	1010	5% 95%
2	B	2	50% 50%
2	C	2	50% 50%

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
2	FRU	C	2	-	-	-	X
5	MLI	A	1103	-	-	-	X

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 8210 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.

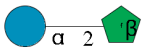
- Molecule 1 is a protein called Beta-galactosidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	989	Total	C	N	O	S	0	3	0
			7622	4796	1363	1445	18			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	207	ALA	ASP	engineered mutation	UNP A0A023UGN9

- Molecule 2 is an oligosaccharide called beta-D-fructofuranose-(2-1)-alpha-D-glucopyranose.

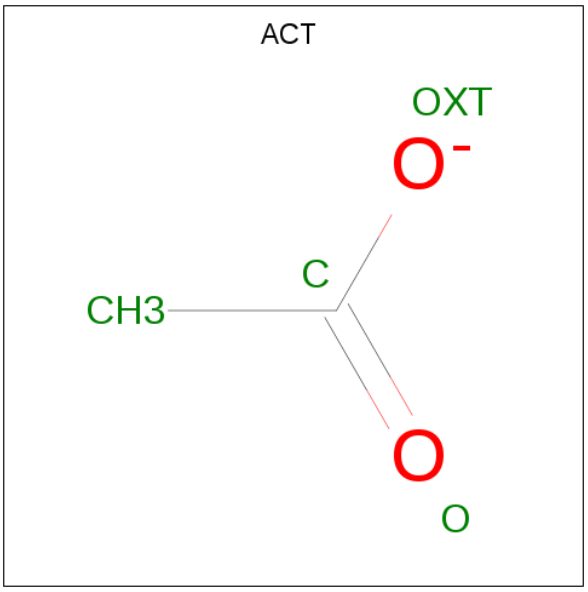


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	B	2	Total	C	O		0	0	0
			23	12	11				
2	C	2	Total	C	O		0	0	0
			23	12	11				

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

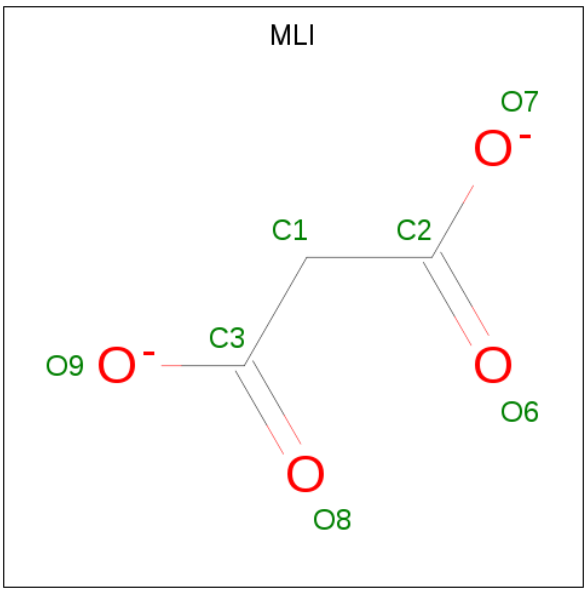
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Na	0	0
			1	1		

- Molecule 4 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).



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

- Molecule 5 is MALONATE ION (three-letter code: MLI) (formula: $C_3H_2O_4$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total C O 7 3 4	0	0

- Molecule 6 is water.

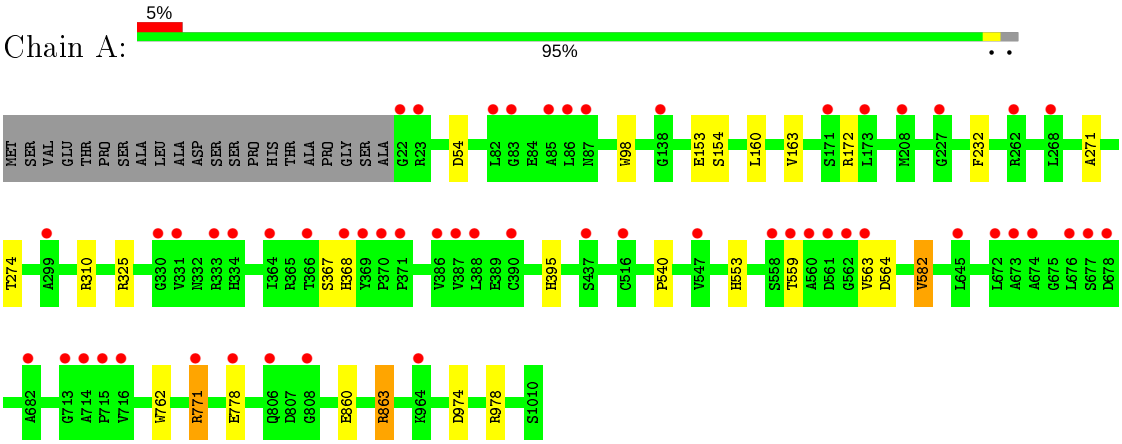
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	530	Total	O	0	0
			530	530		

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: Beta-galactosidase

Chain A:



- Molecule 2: beta-D-fructofuranose-(2-1)-alpha-D-glucopyranose

Chain B:

- Molecule 2: beta-D-fructofuranose-(2-1)-alpha-D-glucopyranose

Chain C:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α, β, γ	137.57Å 137.57Å 126.81Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	46.62 – 2.21 46.62 – 2.21	Depositor EDS
% Data completeness (in resolution range)	99.7 (46.62-2.21) 99.8 (46.62-2.21)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.12 (at 2.20Å)	Xtriage
Refinement program	PHENIX 1.12_2829	Depositor
R, R_{free}	0.194 , 0.224 0.197 , 0.224	Depositor DCC
R_{free} test set	1118 reflections (1.60%)	wwPDB-VP
Wilson B-factor (Å^2)	42.2	Xtriage
Anisotropy	0.509	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 34.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.025 for -h,-k,l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	8210	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.16% 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, GLC, FRU, MLI, ACT

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	0.29	0/7830	0.48	0/10672

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	7622	0	7321	12	0
2	B	23	0	21	0	0
2	C	23	0	21	1	0
3	A	1	0	0	0	0
4	A	4	0	3	0	0
5	A	7	0	2	0	0
6	A	530	0	0	0	0
All	All	8210	0	7368	12	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.

All (12) 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:54:ASP:HB2	2:C:2:FRU:O3	1.81	0.81
1:A:860:GLU:OE1	1:A:978:ARG:NH1	2.21	0.66
1:A:553:HIS:HB2	1:A:582:VAL:HG22	1.82	0.62
1:A:563:VAL:HG12	1:A:564:ASP:N	2.26	0.50
1:A:325:ARG:HD2	1:A:540:PRO:O	2.13	0.47
1:A:153:GLU:HA	1:A:154:SER:HA	1.75	0.46
1:A:563:VAL:HG12	1:A:564:ASP:H	1.81	0.45
1:A:271:ALA:O	1:A:274:THR:HG22	2.18	0.44
1:A:367:SER:HA	1:A:368:HIS:HA	1.80	0.43
1:A:160:LEU:O	1:A:163:VAL:HG12	2.18	0.43
1:A:863:ARG:HD3	1:A:974:ASP:OD1	2.21	0.40
1:A:771:ARG:CZ	1:A:778:GLU:HB3	2.52	0.40

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	990/1010 (98%)	960 (97%)	30 (3%)	0	100 100

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	778/792 (98%)	768 (99%)	10 (1%)	69 81

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	98	TRP
1	A	172	ARG
1	A	232	PHE
1	A	310	ARG
1	A	395	HIS
1	A	559	THR
1	A	582	VAL
1	A	762	TRP
1	A	771	ARG
1	A	863	ARG

Some 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	GLC	B	1	2	11,11,12	0.66	0	15,15,17	0.90	1 (6%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
2	FRU	B	2	2	11,12,12	0.50	0	10,18,18	0.62	0
2	GLC	C	1	2	11,11,12	0.27	0	15,15,17	0.62	0
2	FRU	C	2	2	11,12,12	0.56	0	10,18,18	0.75	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	B	1	2	-	0/2/19/22	0/1/1/1
2	FRU	B	2	2	-	2/5/24/24	0/1/1/1
2	GLC	C	1	2	-	0/2/19/22	0/1/1/1
2	FRU	C	2	2	-	0/5/24/24	0/1/1/1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	1	GLC	C1-O5-C5	2.35	115.37	112.19

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	2	FRU	O1-C1-C2-O2
2	B	2	FRU	O1-C1-C2-O5

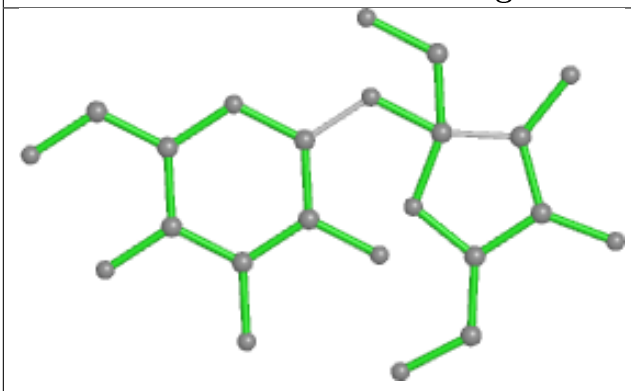
There are no ring outliers.

1 monomer is involved in 1 short contact:

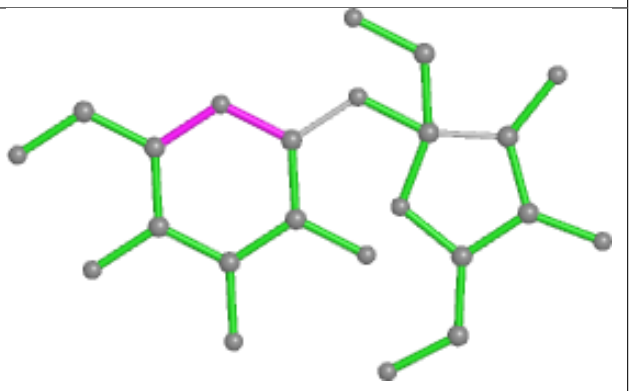
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	C	2	FRU	1	0

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

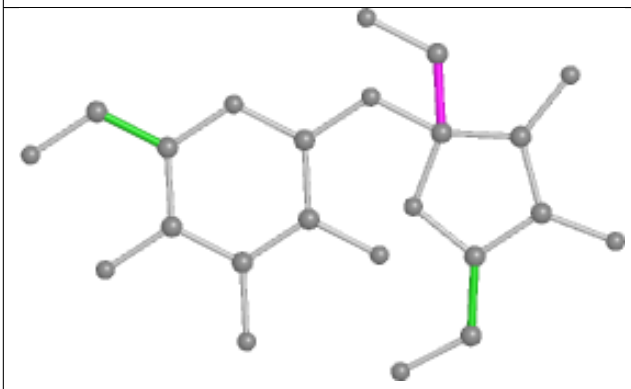
Oligosaccharide Chain B



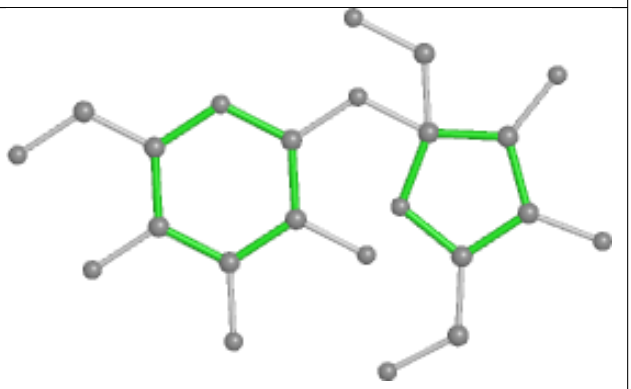
Bond lengths



Bond angles

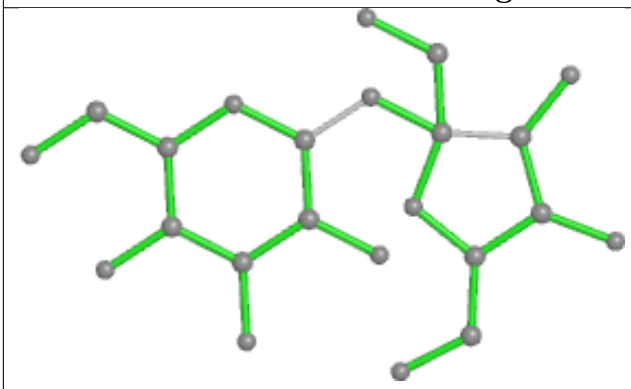


Torsions

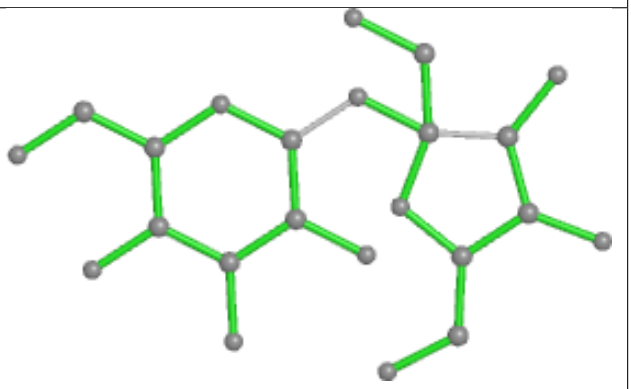


Rings

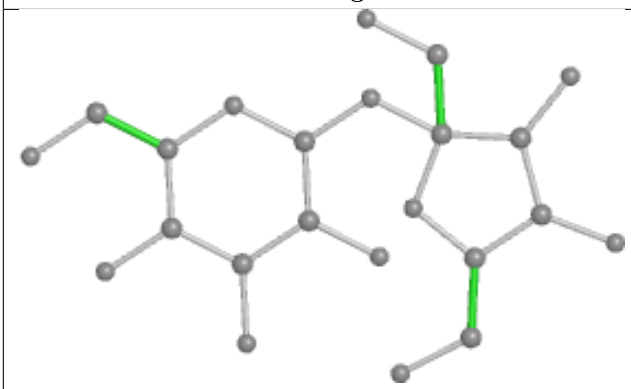
Oligosaccharide Chain C



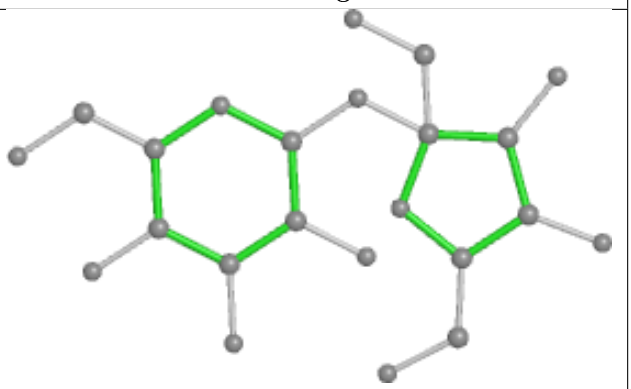
Bond lengths



Bond angles



Torsions



Rings

5.6 Ligand geometry (i)

Of 3 ligands modelled in this entry, 1 is monoatomic - leaving 2 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	$\#\|Z\| > 2$
5	MLI	A	1103	-	0,6,6	0.00	-	0,7,7	0.00	-
4	ACT	A	1102	-	1,3,3	6.66	1 (100%)	0,3,3	0.00	-

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
5	MLI	A	1103	-	-	0/0/4/4	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	1102	ACT	CH3-C	6.66	1.57	1.48

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

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	989/1010 (97%)	0.20	55 (5%) 24 23	33, 52, 84, 130	0

All (55) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	560	ALA	7.6
1	A	561	ASP	6.5
1	A	559	THR	5.1
1	A	563	VAL	4.5
1	A	716	VAL	4.3
1	A	713	GLY	4.3
1	A	673	ALA	4.2
1	A	299	ALA	4.0
1	A	674	ALA	3.9
1	A	85	ALA	3.7
1	A	558	SER	3.7
1	A	676	LEU	3.7
1	A	715	PRO	3.4
1	A	388	LEU	3.2
1	A	677	SER	3.2
1	A	678	ASP	3.1
1	A	808	GLY	3.1
1	A	682	ALA	3.1
1	A	83	GLY	3.0
1	A	516	CYS	3.0
1	A	778	GLU	2.9
1	A	547	VAL	2.9
1	A	208	MET	2.9
1	A	369	TYR	2.8
1	A	82	LEU	2.8
1	A	87	ASN	2.8
1	A	23	ARG	2.8

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	964	LYS	2.8
1	A	562	GLY	2.7
1	A	173	LEU	2.6
1	A	331	VAL	2.6
1	A	86	LEU	2.6
1	A	806	GLN	2.6
1	A	390	CYS	2.5
1	A	138	GLY	2.5
1	A	714	ALA	2.5
1	A	366	THR	2.5
1	A	171	SER	2.5
1	A	334	HIS	2.5
1	A	672	LEU	2.4
1	A	268	LEU	2.4
1	A	370	PRO	2.3
1	A	371	PRO	2.3
1	A	364	ILE	2.2
1	A	645	LEU	2.2
1	A	22	GLY	2.2
1	A	386	VAL	2.1
1	A	368	HIS	2.1
1	A	387	VAL	2.1
1	A	771	ARG	2.1
1	A	262	ARG	2.1
1	A	227	GLY	2.1
1	A	437	SER	2.0
1	A	333	ARG	2.0
1	A	330	GLY	2.0

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)

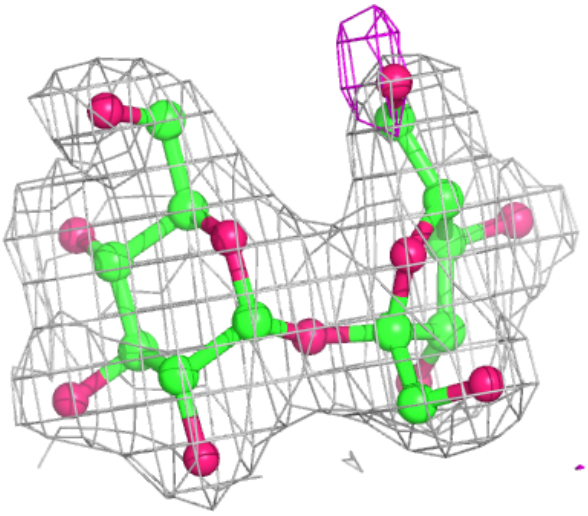
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

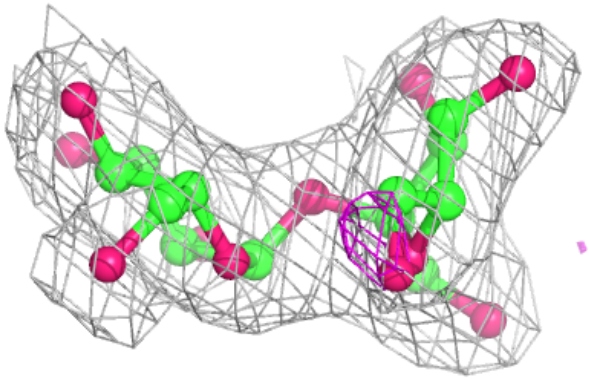
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	FRU	C	2	12/12	0.51	0.50	46,47,48,50	12
2	GLC	C	1	11/12	0.61	0.29	44,46,47,47	11
2	FRU	B	2	12/12	0.94	0.10	46,47,49,50	0
2	GLC	B	1	11/12	0.96	0.09	46,46,47,48	0

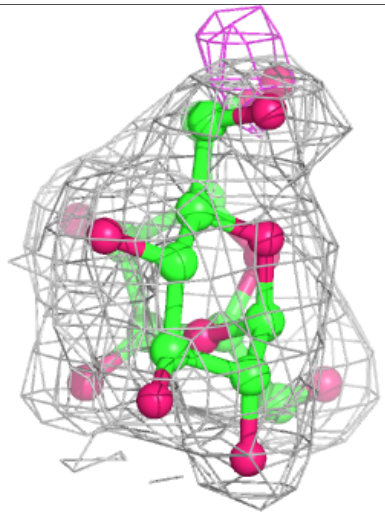
The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.

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)

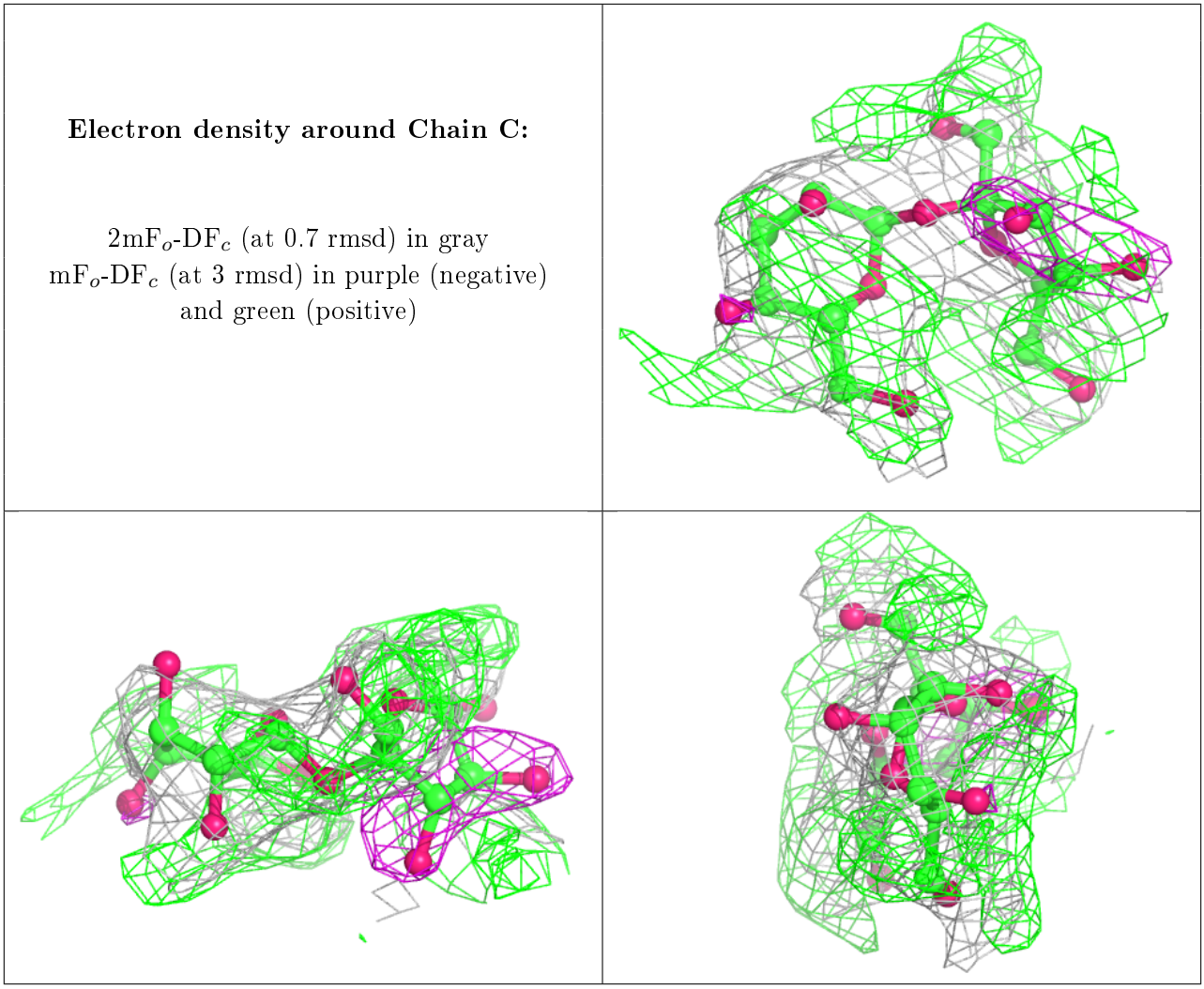






Electron density around Chain C:

$2mF_o$-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
5	MLI	A	1103	7/7	0.74	0.40	49,51,51,51	0
4	ACT	A	1102	4/4	0.89	0.16	46,47,48,48	0
3	NA	A	1101	1/1	0.99	0.12	45,45,45,45	0

6.5 Other polymers

There are no such residues in this entry.