Full wwPDB X-ray Structure Validation Report (i)

Aug 6, 2020 – 10:38 PM BST

PDB ID	:	5IHJ
Title	:	Fusion of Maltose-binding Protein and PilA from Acinetobacter baumannii BIDMC57
Authors	:	Piepenbrink, K.H.; Sundberg, E.J.
Deposited on	:	2016-02-29
Resolution	:	2.20 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	:	2.13.1
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 ⓘ

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

The reported resolution of this entry is 2.20 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.
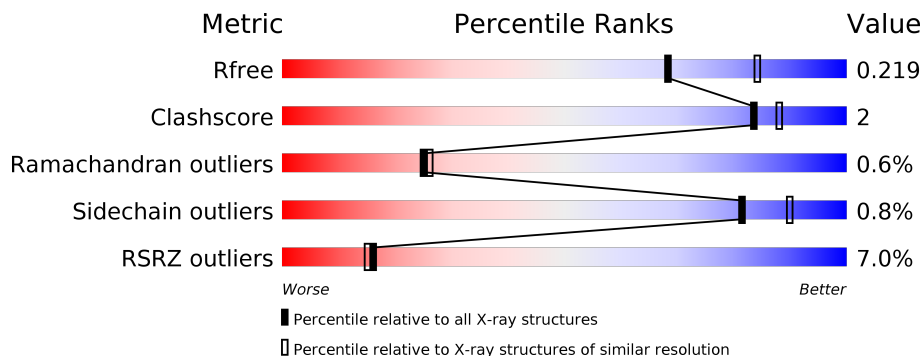
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	495	7% / 93% ...
2	B	3	67% / 33%

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
4	GOL	A	1204	-	-	-	X

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 3796 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 Maltose-binding periplasmic protein,Fimbrial protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	484	3654	2336	596	713	9	0	2	0

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	initiating methionine	UNP P0AEX9
A	82	ALA	ASP	engineered mutation	UNP P0AEX9
A	83	ALA	LYS	engineered mutation	UNP P0AEX9
A	87	ALA	ASP	engineered mutation	UNP P0AEX9
A	88	ALA	LYS	engineered mutation	UNP P0AEX9
A	172	ALA	GLU	engineered mutation	UNP P0AEX9
A	173	ALA	ASN	engineered mutation	UNP P0AEX9
A	239	ALA	LYS	engineered mutation	UNP P0AEX9
A	359	ALA	GLU	engineered mutation	UNP P0AEX9
A	362	ALA	LYS	engineered mutation	UNP P0AEX9
A	363	ALA	ASP	engineered mutation	UNP P0AEX9
A	367	ASN	-	linker	UNP P0AEX9
A	368	ALA	-	linker	UNP P0AEX9
A	369	ALA	-	linker	UNP P0AEX9
A	370	ALA	-	linker	UNP P0AEX9
A	1023	ALA	-	linker	UNP P0AEX9
A	1137	GLY	-	expression tag	UNP A0A140QW15
A	1138	SER	-	expression tag	UNP A0A140QW15
A	1139	LEU	-	expression tag	UNP A0A140QW15
A	1140	GLU	-	expression tag	UNP A0A140QW15
A	1141	HIS	-	expression tag	UNP A0A140QW15
A	1142	HIS	-	expression tag	UNP A0A140QW15
A	1143	HIS	-	expression tag	UNP A0A140QW15
A	1144	HIS	-	expression tag	UNP A0A140QW15
A	1145	HIS	-	expression tag	UNP A0A140QW15
A	1146	HIS	-	expression tag	UNP A0A140QW15

- Molecule 2 is an oligosaccharide called alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
2	B	3	Total 34	C 18	O 16	0	0	0

- Molecule 3 is (4S)-2-METHYL-2,4-PENTANEDIOL (three-letter code: MPD) (formula: $C_6H_{14}O_2$).
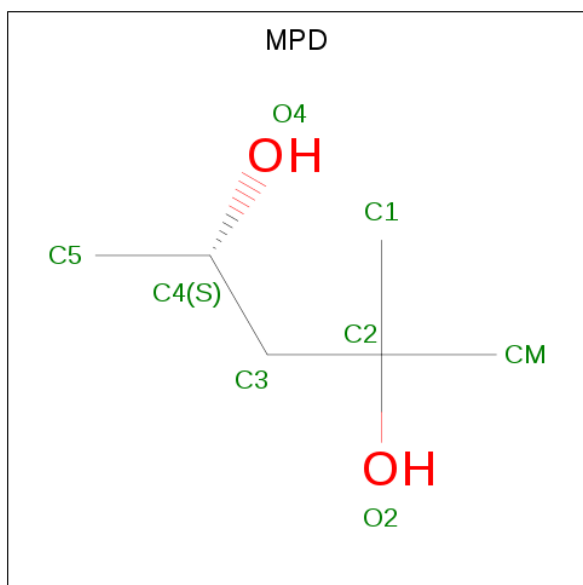
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 8	C 6	O 2	0	0

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

GOL

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total 6	C 3	O 3	0	0
4	A	1	Total 6	C 3	O 3	0	0
4	A	1	Total 6	C 3	O 3	0	0
4	A	1	Total 6	C 3	O 3	0	0

- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	3	Total 3	Ca 3	0	0

- Molecule 6 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	73	Total 73	O 73	0	0

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: Maltose-binding periplasmic protein,Fimbrial protein

Chain A:
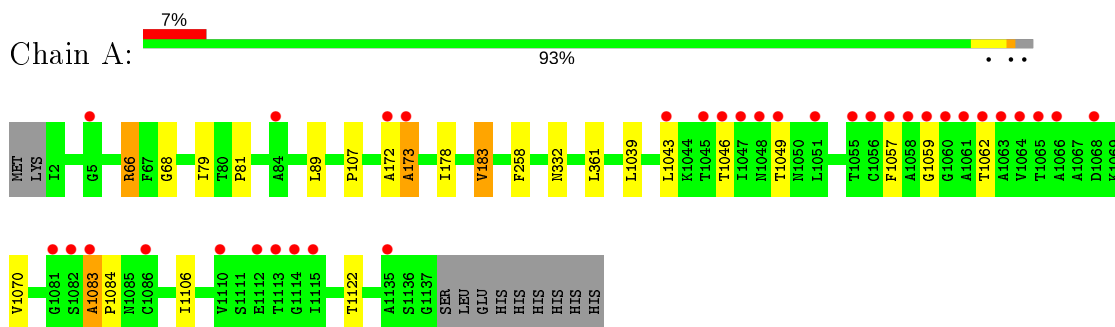
• Molecule 2: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose

Chain B:
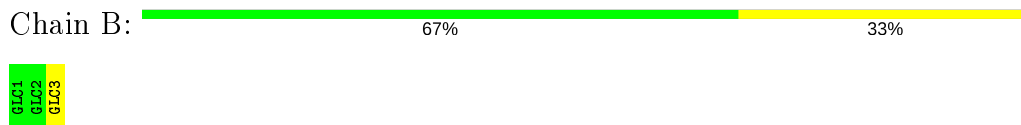
4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	175.07Å 56.64Å 50.00Å 90.00° 91.60° 90.00°	Depositor
Resolution (Å)	43.93 – 2.20 43.93 – 2.20	Depositor EDS
% Data completeness (in resolution range)	94.8 (43.93-2.20) 94.9 (43.93-2.20)	Depositor EDS
R_{merge}	0.35	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.59 (at 2.20Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.211 , 0.244 0.219 , 0.219	Depositor DCC
R_{free} test set	1211 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å^2)	34.5	Xtriage
Anisotropy	0.991	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 39.0	EDS
L-test for twinning[2]	$<|L|> = 0.48$, $<L^2> = 0.32$	Xtriage
Estimated twinning fraction	0.027 for -h,-k,l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3796	wwPDB-VP
Average B, all atoms (Å^2)	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.36% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, CA, GLC, MPD

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	0.48	0/3740	0.63	1/5087 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	66	ARG	NE-CZ-NH1	5.23	122.92	120.30

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	3654	0	3621	17	0
2	B	34	0	30	0	0
3	A	8	0	14	2	0
4	A	24	0	32	0	0
5	A	3	0	0	0	0
6	A	73	0	0	0	0
All	All	3796	0	3697	17	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 2.

All (17) 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:1039:LEU:O	1:A:1043:LEU:HD13	1.77	0.83
1:A:1043:LEU:HD11	1:A:1070:VAL:CG1	2.11	0.80
1:A:1043:LEU:HD11	1:A:1070:VAL:HG11	1.70	0.73
1:A:1043:LEU:HD11	1:A:1070:VAL:HG12	1.78	0.64
1:A:89:LEU:HD23	1:A:107:PRO:HG2	1.88	0.55
1:A:1083:ALA:HB1	1:A:1084:PRO:CD	2.39	0.53
1:A:183:VAL:HG13	1:A:361:LEU:HD22	1.93	0.51
1:A:178:ILE:HD12	3:A:1201:MPD:H52	1.93	0.51
1:A:1057:PHE:HB2	1:A:1062:THR:HG23	1.94	0.50
1:A:1106:ILE:HA	1:A:1122:THR:HG23	1.93	0.49
1:A:79:ILE:HD12	1:A:81:PRO:HD3	1.93	0.49
1:A:1046:THR:HA	1:A:1049:THR:HG22	1.95	0.48
1:A:178:ILE:HD12	3:A:1201:MPD:C5	2.43	0.48
1:A:1043:LEU:CD1	1:A:1070:VAL:HG11	2.43	0.47
1:A:172:ALA:O	1:A:173:ALA:O	2.35	0.45
1:A:1083:ALA:CB	1:A:1084:PRO:CD	2.96	0.44
1:A:68:GLY:HA3	1:A:332:ASN:O	2.20	0.41

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	484/495 (98%)	468 (97%)	13 (3%)	3 (1%)	25	26

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	173	ALA
1	A	1083	ALA

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	1059	GLY

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	374/383 (98%)	371 (99%)	3 (1%)	81 90

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	66	ARG
1	A	183	VAL
1	A	258	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1026	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)

3 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	12,12,12	0.48	0	17,17,17	0.60	0
2	GLC	B	2	2	11,11,12	0.34	0	15,15,17	0.93	0
2	GLC	B	3	2	11,11,12	0.26	0	15,15,17	0.81	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	GLC	B	1	2	-	1/2/22/22	0/1/1/1
2	GLC	B	2	2	-	0/2/19/22	0/1/1/1
2	GLC	B	3	2	-	0/2/19/22	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	3	GLC	C1-O5-C5	2.31	115.33	112.19

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	1	GLC	C4-C5-C6-O6

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 B

Bond lengths   Bond angles

Torsions   Rings
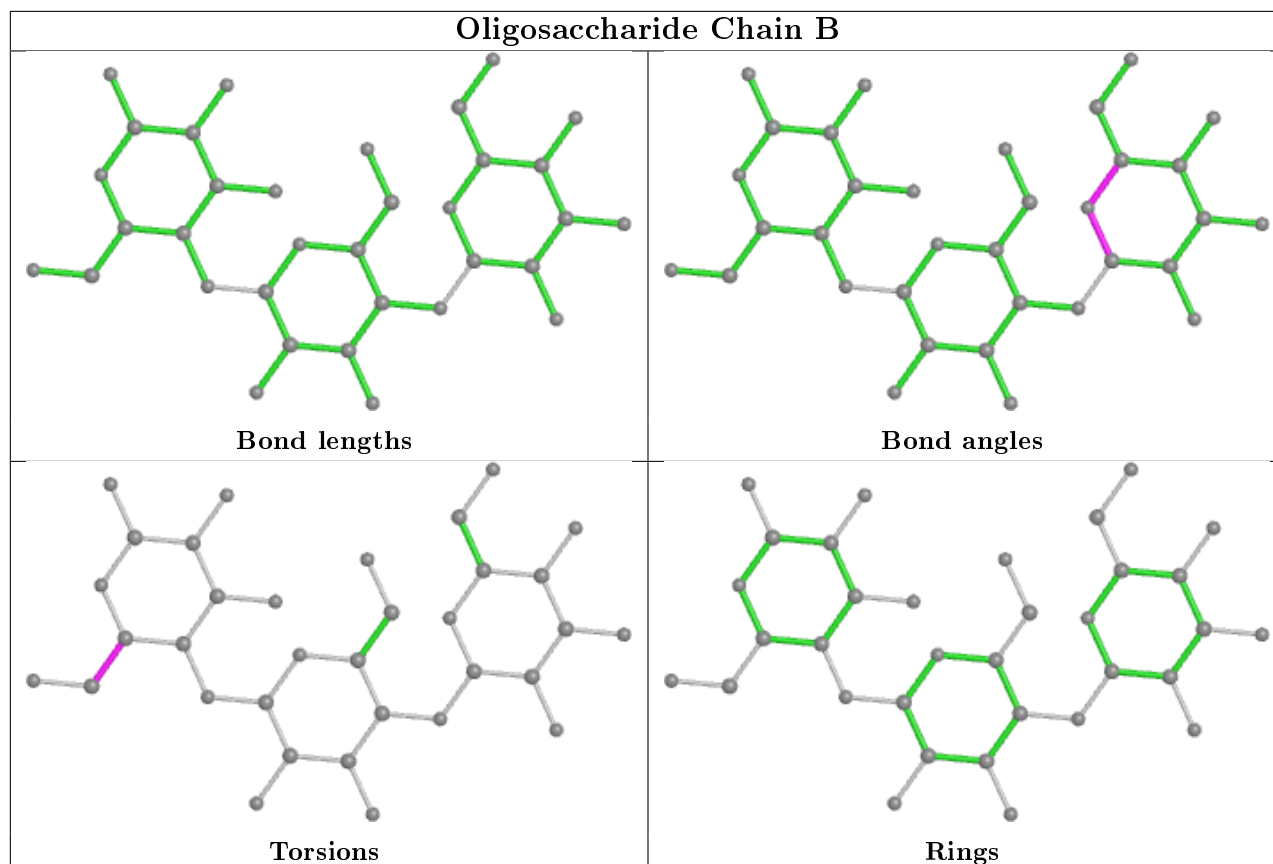
5.6 Ligand geometry (i)

Of 8 ligands modelled in this entry, 3 are monoatomic - leaving 5 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	GOL	A	1205	-	5,5,5	0.29	0	5,5,5	0.22	0
3	MPD	A	1201	-	7,7,7	0.30	0	9,10,10	0.43	0
4	GOL	A	1202	-	5,5,5	0.30	0	5,5,5	0.34	0
4	GOL	A	1204	-	5,5,5	0.44	0	5,5,5	0.30	0
4	GOL	A	1203	-	5,5,5	0.28	0	5,5,5	0.11	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	GOL	A	1205	-	-	2/4/4/4	-
3	MPD	A	1201	-	-	1/5/5/5	-
4	GOL	A	1202	-	-	0/4/4/4	-
4	GOL	A	1204	-	-	4/4/4/4	-
4	GOL	A	1203	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (9) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	1205	GOL	C1-C2-C3-O3
4	A	1204	GOL	O1-C1-C2-C3
4	A	1204	GOL	C1-C2-C3-O3
4	A	1203	GOL	C1-C2-C3-O3
4	A	1203	GOL	O2-C2-C3-O3
4	A	1204	GOL	O2-C2-C3-O3
4	A	1204	GOL	O1-C1-C2-O2
4	A	1205	GOL	O2-C2-C3-O3
3	A	1201	MPD	C2-C3-C4-C5

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1201	MPD	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, 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	484/495 (97%)	0.12	34 (7%)	16	15	13, 25, 47, 60	0

All (34) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1058	ALA	7.6
1	A	1061	ALA	7.3
1	A	1064	VAL	7.1
1	A	1059	GLY	6.3
1	A	1068	ASP	4.6
1	A	1113	THR	4.3
1	A	1047	ILE	4.1
1	A	172	ALA	4.0
1	A	1063	ALA	3.6
1	A	1083	ALA	3.4
1	A	1060	GLY	3.4
1	A	1112	GLU	3.3
1	A	1082	SER	3.3
1	A	1049	THR	3.3
1	A	1062	THR	3.3
1	A	1065	THR	3.2
1	A	1066	ALA	3.0
1	A	1057	PHE	3.0
1	A	1056	CYS	2.9
1	A	1045	THR	2.8
1	A	1115	ILE	2.8
1	A	5	GLY	2.6
1	A	173	ALA	2.6
1	A	1081	GLY	2.5
1	A	84	ALA	2.5
1	A	1110	VAL	2.5
1	A	1051	LEU	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	1048	ASN	2.4
1	A	1135	ALA	2.2
1	A	1043	LEU	2.2
1	A	1055	THR	2.2
1	A	1086	CYS	2.1
1	A	1114	GLY	2.0
1	A	1046	THR	2.0

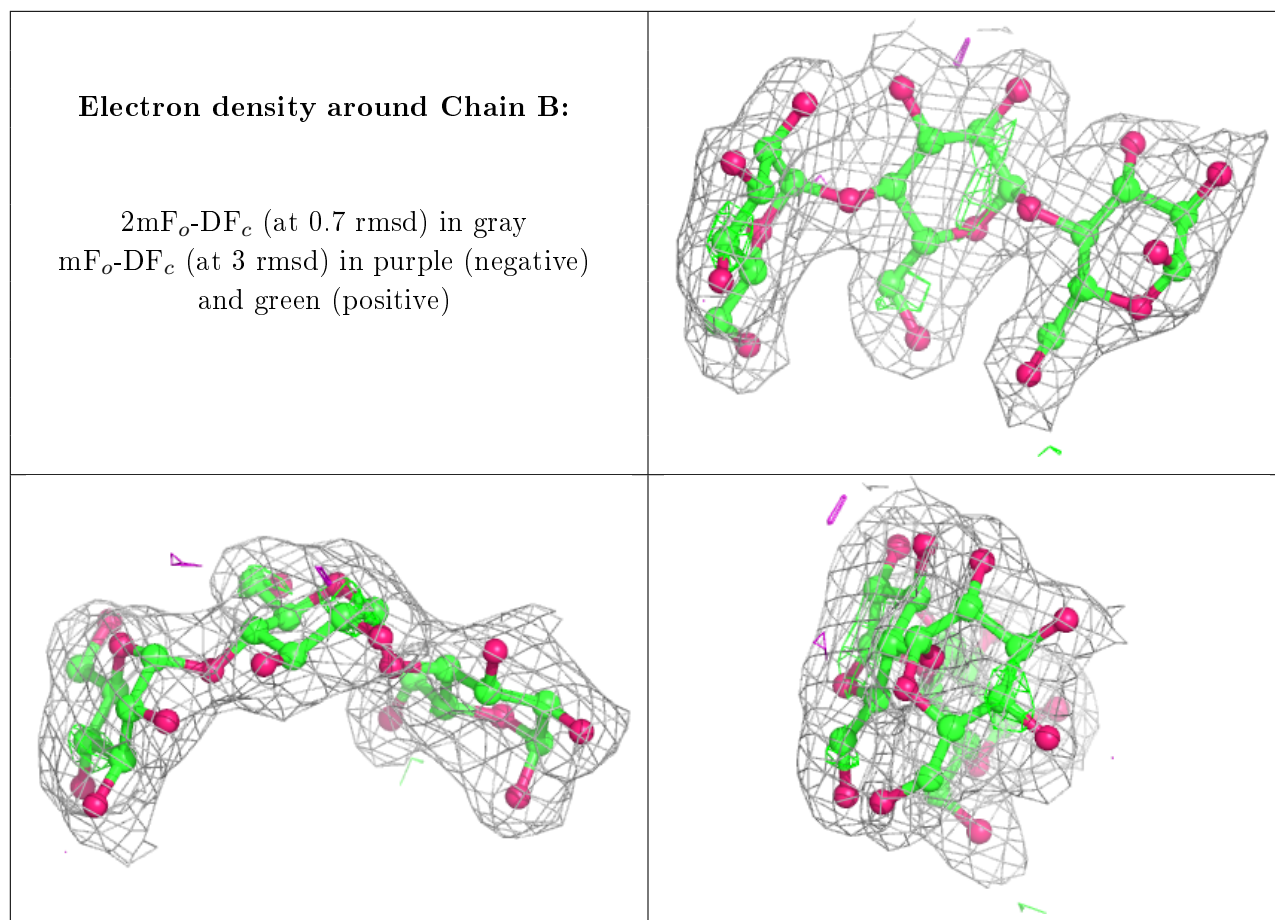## 6.2 Non-standard residues in protein, DNA, RNA chains (i)

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, 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	GLC	B	2	11/12	0.94	0.19	29,31,36,37	0
2	GLC	B	1	12/12	0.96	0.16	29,36,44,45	0
2	GLC	B	3	11/12	0.97	0.17	29,32,35,46	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(Å²)	Q<0.9
4	GOL	A	1204	6/6	0.48	0.43	61,80,104,105	0
5	CA	A	1208	1/1	0.73	0.08	71,71,71,71	0
4	GOL	A	1205	6/6	0.76	0.29	60,63,69,80	0
4	GOL	A	1203	6/6	0.78	0.16	80,85,86,93	0
4	GOL	A	1202	6/6	0.81	0.20	65,77,82,82	0
5	CA	A	1207	1/1	0.81	0.24	77,77,77,77	0
5	CA	A	1206	1/1	0.85	0.12	95,95,95,95	0
3	MPD	A	1201	8/8	0.93	0.26	34,38,51,65	0

6.5 Other polymers (i)

There are no such residues in this entry.