wwPDB X-ray Structure Validation Summary Report (i)

Sep 12, 2022 – 06:15 pm BST

PDB ID	:	8AGK
Title	:	Botulinum neurotoxin subtype A6 cell binding domain in complex with GD1a ganglioside
Authors	:	Gregory, K.S.; Acharya, K.R.; Liu, S.M.; Newell, A.R.; Mojanaga, O.O.
Deposited on	:	2022-07-20
Resolution	:	1.50 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 (i)) were used in the production of this report:

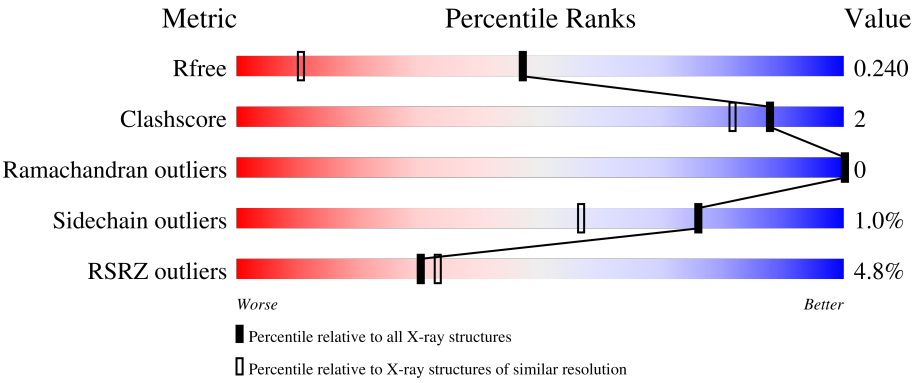
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.30
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.30

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 1.50 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	2936 (1.50-1.50)
Clashscore	141614	3144 (1.50-1.50)
Ramachandran outliers	138981	3066 (1.50-1.50)
Sidechain outliers	138945	3064 (1.50-1.50)
RSRZ outliers	127900	2884 (1.50-1.50)

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	AAA	433	5% 90% 6% 5%
2	BBB	5	20% 40% 40%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3729 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.

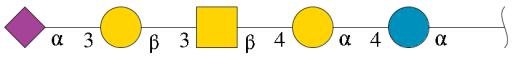
- Molecule 1 is a protein called Bont/A1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	AAA	413	3423	2195	580	635	13	0	4	0

There are 7 discrepancies between the modelled and reference sequences:

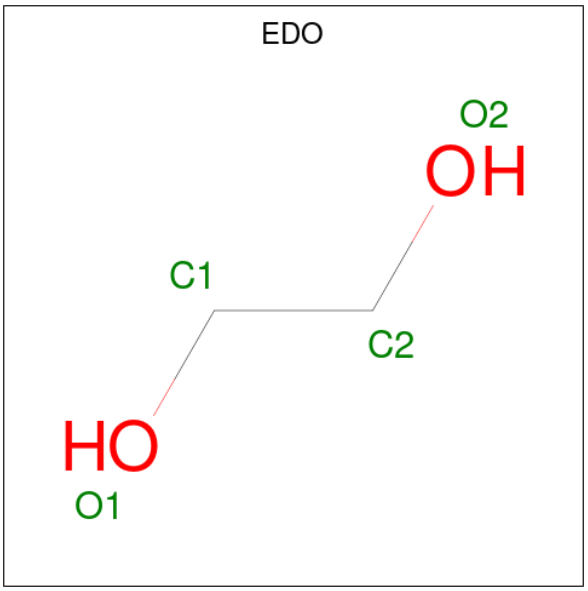
Chain	Residue	Modelled	Actual	Comment	Reference
AAA	864	MET	-	initiating methionine	UNP C9WWY7
AAA	865	HIS	-	expression tag	UNP C9WWY7
AAA	866	HIS	-	expression tag	UNP C9WWY7
AAA	867	HIS	-	expression tag	UNP C9WWY7
AAA	868	HIS	-	expression tag	UNP C9WWY7
AAA	869	HIS	-	expression tag	UNP C9WWY7
AAA	870	HIS	-	expression tag	UNP C9WWY7

- Molecule 2 is an oligosaccharide called N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galacto pyranose-(1-3)-2-acetamido-2-deoxy-beta-D-galactopyranose-(1-4)-alpha-D-galactopyranose -(1-4)-alpha-D-glucopyranose.



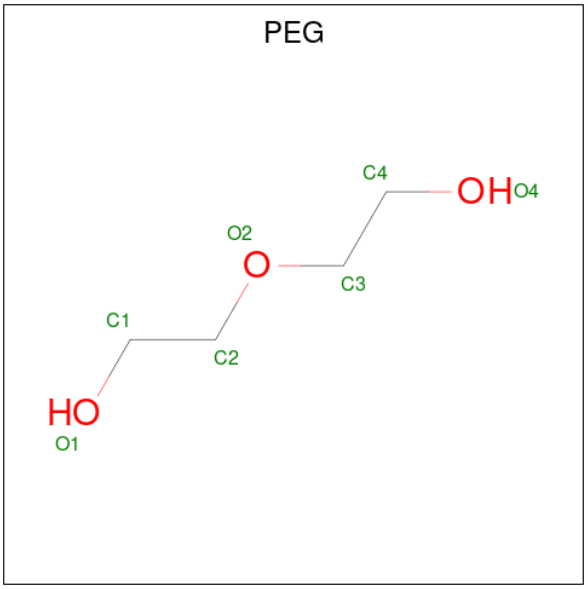
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	BBB	5	68	37	2	29	0	0	0

- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



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

- Molecule 4 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



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

- Molecule 5 is water.

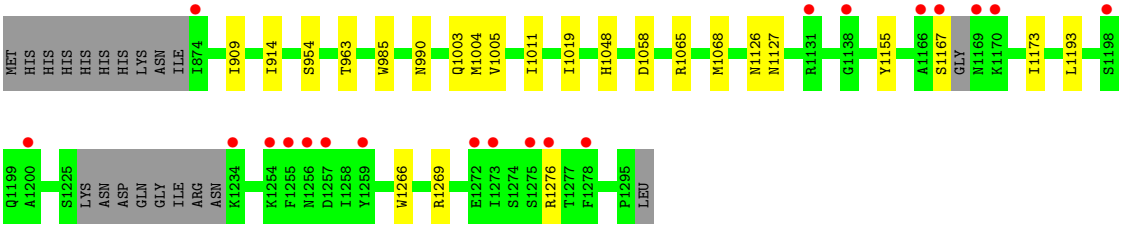
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	AAA	197	Total O 197 197	0	0

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.

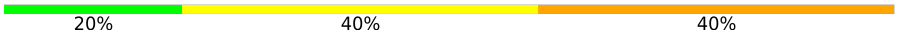
• Molecule 1: Bont/A1

Chain AAA:



• Molecule 2: N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galactopyranose-(1-3)-2-acetamido-2-deoxy-beta-D-galactopyranose-(1-4)-alpha-D-galactopyranose-(1-4)-alpha-D-glucopyranose

Chain BBB:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	44.31Å 83.64Å 58.31Å 90.00° 98.66° 90.00°	Depositor
Resolution (Å)	57.65 – 1.50 47.47 – 1.50	Depositor EDS
% Data completeness (in resolution range)	99.9 (57.65-1.50) 99.9 (47.47-1.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.14 (at 1.50Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.203 , 0.236 0.210 , 0.240	Depositor DCC
R_{free} test set	3343 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å^2)	17.3	Xtriage
Anisotropy	0.270	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	3729	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 6.95% 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: PEG, GLA, NGA, SIA, GAL, EDO, 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	AAA	0.70	0/3503	0.84	1/4730 (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	AAA	1155	TYR	CB-CG-CD2	-5.00	118.00	121.00

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	AAA	3423	0	3408	13	0
2	BBB	68	0	58	2	0
3	AAA	20	0	30	1	0
4	AAA	21	0	29	2	0
5	AAA	197	0	0	1	0
All	All	3729	0	3525	14	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.

The worst 5 of 14 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:AAA:909:ILE:HG23	1:AAA:1173:ILE:HD11	1.78	0.63
4:AAA:1306:PEG:H42	5:AAA:1500:HOH:O	2.02	0.58
1:AAA:1266:TRP:CE3	2:BBB:4:GAL:H5	2.43	0.54
1:AAA:1276:ARG:NH1	2:BBB:5:SIA:O9	2.40	0.54
1:AAA:1003:GLN:HA	1:AAA:1011:ILE:HD11	1.90	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	AAA	411/433 (95%)	397 (97%)	14 (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	AAA	386/400 (96%)	382 (99%)	4 (1%)	76 57

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

Mol	Chain	Res	Type
1	AAA	1004	MET
1	AAA	1065	ARG
1	AAA	1167	SER
1	AAA	1269	ARG

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)

5 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	BBB	1	2	12,12,12	0.81	0	17,17,17	0.91	1 (5%)
2	GLA	BBB	2	2	11,11,12	0.82	0	15,15,17	2.63	7 (46%)
2	NGA	BBB	3	2	14,14,15	0.38	0	17,19,21	0.86	0
2	GAL	BBB	4	2	11,11,12	0.42	0	15,15,17	1.24	2 (13%)
2	SIA	BBB	5	2	20,20,21	0.69	0	24,28,31	1.00	1 (4%)

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	BBB	1	2	-	0/2/22/22	0/1/1/1
2	GLA	BBB	2	2	-	2/2/19/22	0/1/1/1
2	NGA	BBB	3	2	-	0/6/23/26	0/1/1/1
2	GAL	BBB	4	2	-	0/2/19/22	0/1/1/1
2	SIA	BBB	5	2	-	2/18/34/38	0/1/1/1

There are no bond length outliers.

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	BBB	2	GLA	C1-C2-C3	5.93	116.96	109.67
2	BBB	2	GLA	O5-C1-C2	4.56	117.81	110.77
2	BBB	2	GLA	C1-O5-C5	4.12	117.78	112.19
2	BBB	2	GLA	O5-C5-C6	3.06	112.01	107.20
2	BBB	2	GLA	O2-C2-C3	-2.78	104.56	110.14

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	BBB	5	SIA	C11-C10-N5-C5
2	BBB	5	SIA	O10-C10-N5-C5
2	BBB	2	GLA	O5-C5-C6-O6
2	BBB	2	GLA	C4-C5-C6-O6

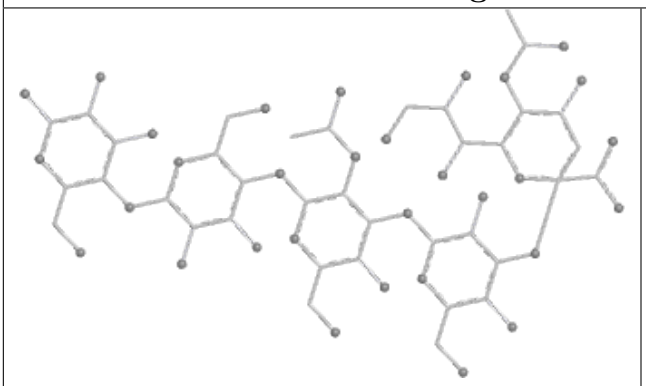
There are no ring outliers.

2 monomers are involved in 2 short contacts:

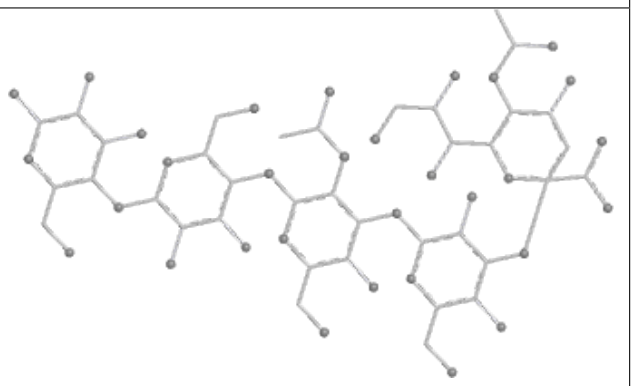
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	BBB	5	SIA	1	0
2	BBB	4	GAL	1	0

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

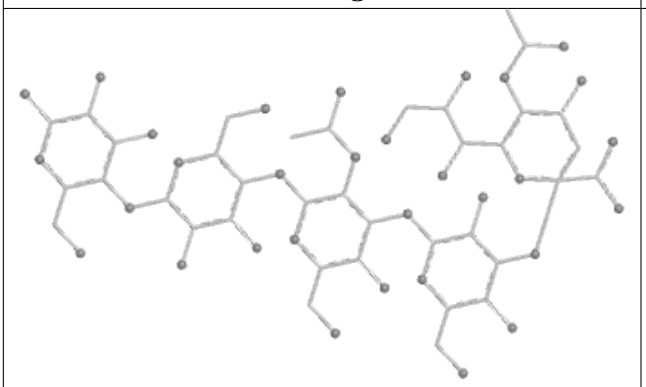
Oligosaccharide Chain BBB



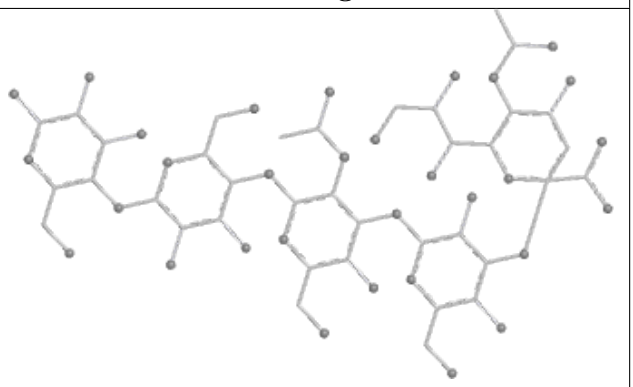
Bond lengths



Bond angles



Torsions



Rings

5.6 Ligand geometry (i)

8 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	$\#\|Z\| > 2$
3	EDO	AAA	1305	-	3,3,3	0.10	0	2,2,2	0.11	0
3	EDO	AAA	1308	-	3,3,3	0.16	0	2,2,2	0.10	0
4	PEG	AAA	1303	-	6,6,6	0.39	0	5,5,5	0.16	0
4	PEG	AAA	1307	-	6,6,6	0.22	0	5,5,5	0.33	0
4	PEG	AAA	1306	-	6,6,6	1.09	0	5,5,5	0.71	0
3	EDO	AAA	1302	-	3,3,3	0.47	0	2,2,2	0.18	0
3	EDO	AAA	1304	-	3,3,3	0.10	0	2,2,2	0.26	0
3	EDO	AAA	1309	-	3,3,3	0.23	0	2,2,2	0.74	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
3	EDO	AAA	1305	-	-	1/1/1/1	-
3	EDO	AAA	1308	-	-	1/1/1/1	-
4	PEG	AAA	1303	-	-	1/4/4/4	-
4	PEG	AAA	1307	-	-	2/4/4/4	-
4	PEG	AAA	1306	-	-	1/4/4/4	-
3	EDO	AAA	1302	-	-	1/1/1/1	-
3	EDO	AAA	1304	-	-	1/1/1/1	-
3	EDO	AAA	1309	-	-	0/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	AAA	1307	PEG	O1-C1-C2-O2
3	AAA	1302	EDO	O1-C1-C2-O2
3	AAA	1304	EDO	O1-C1-C2-O2
3	AAA	1308	EDO	O1-C1-C2-O2
4	AAA	1307	PEG	C4-C3-O2-C2

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	AAA	1306	PEG	2	0
3	AAA	1309	EDO	1	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

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	AAA	413/433 (95%)	0.13	20 (4%) 30 33	15, 25, 50, 85	0

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	1257	ASP	5.8
1	AAA	1276	ARG	4.4
1	AAA	1273	ILE	4.2
1	AAA	1200	ALA	3.9
1	AAA	874	ILE	3.6

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

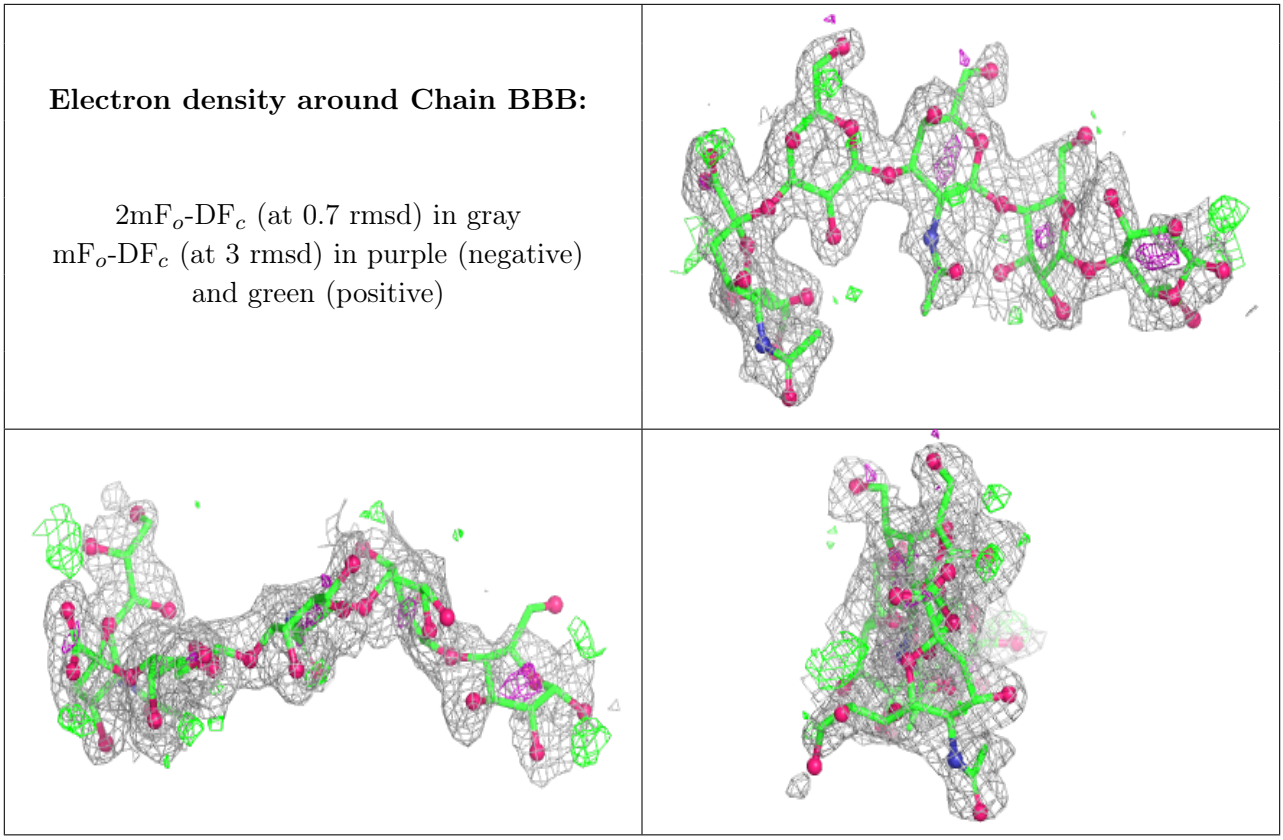
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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	GLC	BBB	1	12/12	0.58	0.25	66,75,77,78	0
2	GLA	BBB	2	11/12	0.76	0.12	57,61,65,67	0
2	NGA	BBB	3	14/15	0.82	0.18	36,47,51,52	0
2	SIA	BBB	5	20/21	0.86	0.18	38,49,64,65	0
2	GAL	BBB	4	11/12	0.89	0.09	31,36,38,39	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 BBB:

$2mF_o$-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	EDO	AAA	1304	4/4	0.83	0.16	45,47,48,49	0
4	PEG	AAA	1303	7/7	0.83	0.11	35,41,48,53	0
4	PEG	AAA	1307	7/7	0.84	0.18	29,44,52,52	0
3	EDO	AAA	1308	4/4	0.85	0.25	39,49,55,64	0
3	EDO	AAA	1302	4/4	0.85	0.15	29,29,34,40	0
3	EDO	AAA	1305	4/4	0.85	0.12	55,56,58,61	0
4	PEG	AAA	1306	7/7	0.87	0.15	24,31,34,36	0
3	EDO	AAA	1309	4/4	0.91	0.16	29,30,31,32	0

6.5 Other polymers

There are no such residues in this entry.