wwPDB X-ray Structure Validation Summary Report (i)

Aug 10, 2020 – 06:53 AM BST

PDB ID	:	4TXT
Title	:	Crystal Structure of a GH48 cellobiohydrolase from Caldicellulosiruptor bescii
Authors	:	An, J.; Feng, Y.
Deposited on	:	2014-07-07
Resolution	:	2.00 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:

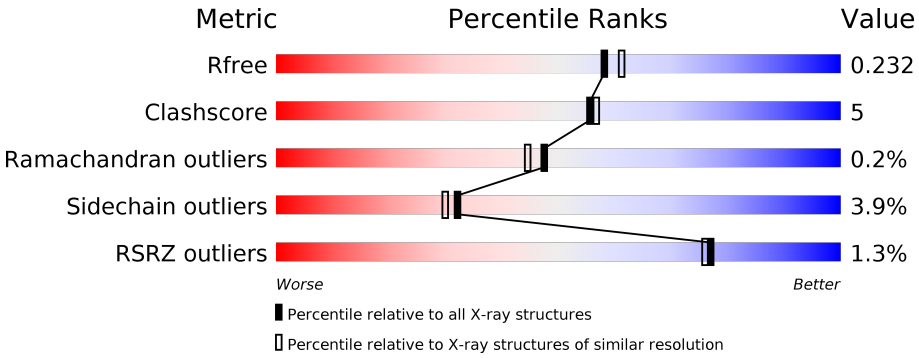
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry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

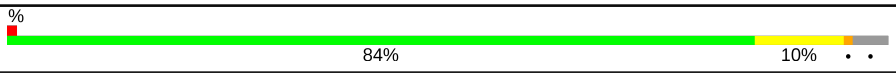
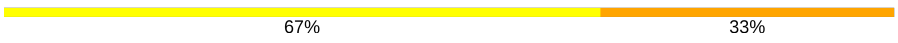
The reported resolution of this entry is 2.00 Å.

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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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	656	
2	B	3	

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
3	EDO	A	703	-	-	X	-

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 5317 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.

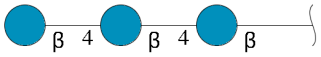
- Molecule 1 is a protein called Glycoside hydrolase family 48.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	627	Total	C	N	O	S	0	0	0
			5070	3281	832	940	17			

There are 18 discrepancies between the modelled and reference sequences:

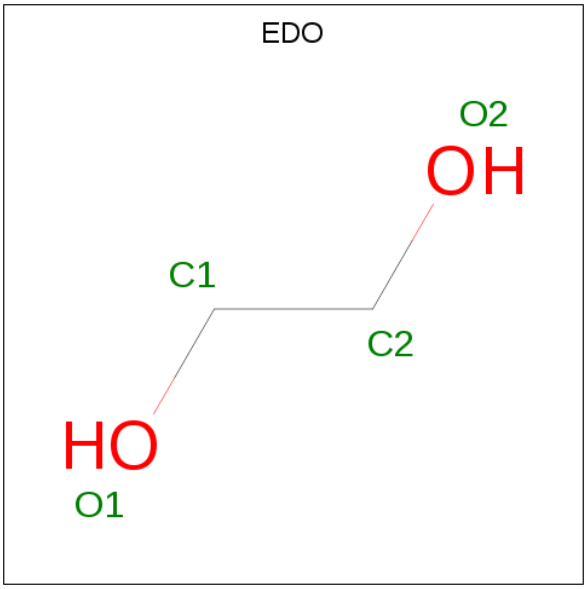
Chain	Residue	Modelled	Actual	Comment	Reference
A	640	SER	-	expression tag	UNP B9MKU7
A	641	SER	-	expression tag	UNP B9MKU7
A	642	SER	-	expression tag	UNP B9MKU7
A	643	VAL	-	expression tag	UNP B9MKU7
A	644	ASP	-	expression tag	UNP B9MKU7
A	645	LYS	-	expression tag	UNP B9MKU7
A	646	LEU	-	expression tag	UNP B9MKU7
A	647	ALA	-	expression tag	UNP B9MKU7
A	648	ALA	-	expression tag	UNP B9MKU7
A	649	ALA	-	expression tag	UNP B9MKU7
A	650	LEU	-	expression tag	UNP B9MKU7
A	651	GLU	-	expression tag	UNP B9MKU7
A	652	HIS	-	expression tag	UNP B9MKU7
A	653	HIS	-	expression tag	UNP B9MKU7
A	654	HIS	-	expression tag	UNP B9MKU7
A	655	HIS	-	expression tag	UNP B9MKU7
A	656	HIS	-	expression tag	UNP B9MKU7
A	657	HIS	-	expression tag	UNP B9MKU7

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



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

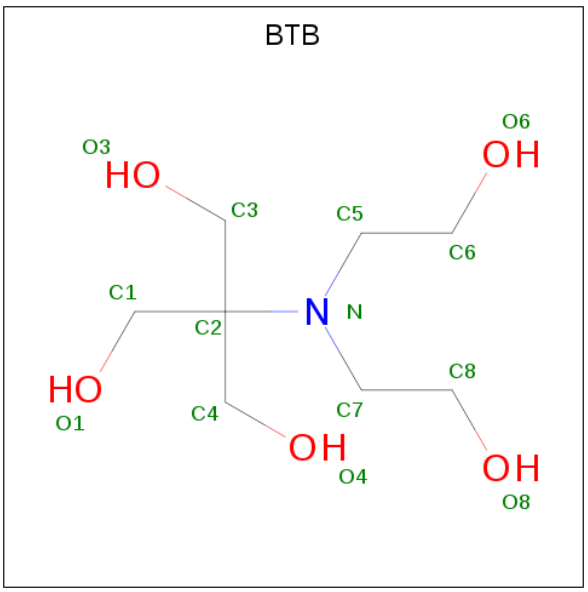
- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



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

- Molecule 4 is 2-[BIS-(2-HYDROXY-ETHYL)-AMINO]-2-HYDROXYMETHYL-PROPAN

E-1,3-DIOL (three-letter code: BTB) (formula: $C_8H_{19}NO_5$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 5 is water.

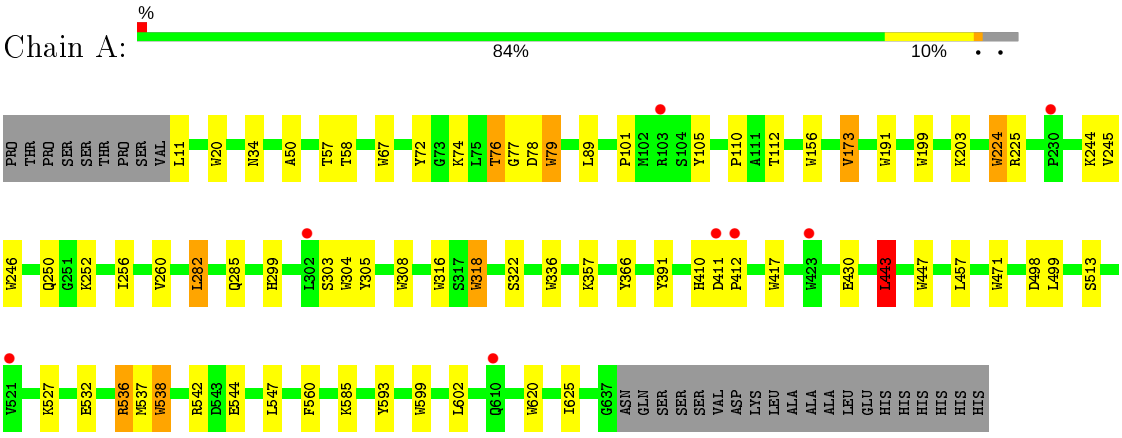
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	159	Total	O	0	0
			159	159		

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.

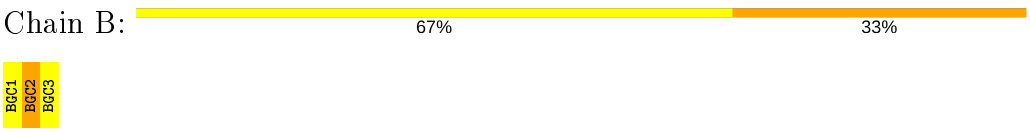
- Molecule 1: Glycoside hydrolase family 48

Chain A:



- Molecule 2: beta-D-glucopyranose-(1-4)-beta-D-glucopyranose-(1-4)-beta-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	116.35Å 57.35Å 105.60Å 90.00° 103.36° 90.00°	Depositor
Resolution (Å)	50.00 – 2.00 45.34 – 2.00	Depositor EDS
% Data completeness (in resolution range)	99.6 (50.00-2.00) 99.6 (45.34-2.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.79 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.197 , 0.229 0.200 , 0.232	Depositor DCC
R_{free} test set	2311 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	30.9	Xtriage
Anisotropy	0.034	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 38.6	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	0.96	EDS
Total number of atoms	5317	wwPDB-VP
Average B, all atoms (Å^2)	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.59% 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: BGC, EDO, BTB

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	0.70	16/5252 (0.3%)	0.64	2/7147 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

The worst 5 of 16 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	79	TRP	CD2-CE2	5.70	1.48	1.41
1	A	304	TRP	CD2-CE2	5.53	1.48	1.41
1	A	20	TRP	CD2-CE2	5.29	1.47	1.41
1	A	318	TRP	CD2-CE2	5.28	1.47	1.41
1	A	538	TRP	CD2-CE2	5.28	1.47	1.41

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	443	LEU	CA-CB-CG	6.75	130.82	115.30
1	A	173	VAL	CB-CA-C	-6.73	98.61	111.40

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	305	TYR	Peptide

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	5070	0	4735	47	0
2	B	34	0	29	2	0
3	A	40	0	60	5	0
4	A	14	0	18	4	0
5	A	159	0	0	0	0
All	All	5317	0	4842	51	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 5.

The worst 5 of 51 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:76:THR:HG22	1:A:78:ASP:N	1.60	1.16
1:A:76:THR:CG2	1:A:78:ASP:H	1.67	1.06
1:A:76:THR:HG21	1:A:78:ASP:HB3	1.31	1.05
1:A:537:MET:HE1	1:A:625:ILE:HG23	1.42	1.01
1:A:537:MET:HE1	1:A:625:ILE:CG2	1.92	0.99

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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	625/656 (95%)	606 (97%)	18 (3%)	1 (0%)	47 44

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	50	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	509/535 (95%)	489 (96%)	20 (4%)	32 30

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	245	VAL
1	A	260	VAL
1	A	536	ARG
1	A	225	ARG
1	A	244	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	23	ASN
1	A	183	GLN
1	A	250	GLN
1	A	285	GLN
1	A	474	GLN

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	BGC	B	1	2	12,12,12	0.74	0	17,17,17	1.38	3 (17%)
2	BGC	B	2	2	11,11,12	1.25	1 (9%)	15,15,17	2.29	4 (26%)
2	BGC	B	3	2	11,11,12	0.92	0	15,15,17	1.34	2 (13%)

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	BGC	B	1	2	-	1/2/22/22	0/1/1/1
2	BGC	B	2	2	-	2/2/19/22	0/1/1/1
2	BGC	B	3	2	-	0/2/19/22	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	2	BGC	O5-C1	-2.94	1.39	1.43

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	2	BGC	C1-O5-C5	6.54	121.05	112.19
2	B	2	BGC	C6-C5-C4	-3.53	104.75	113.00
2	B	3	BGC	C1-C2-C3	3.17	113.57	109.67
2	B	1	BGC	O5-C5-C4	3.10	115.33	109.69
2	B	2	BGC	O5-C5-C6	-2.96	102.56	107.20

There are no chirality outliers.

All (3) torsion outliers are listed below:

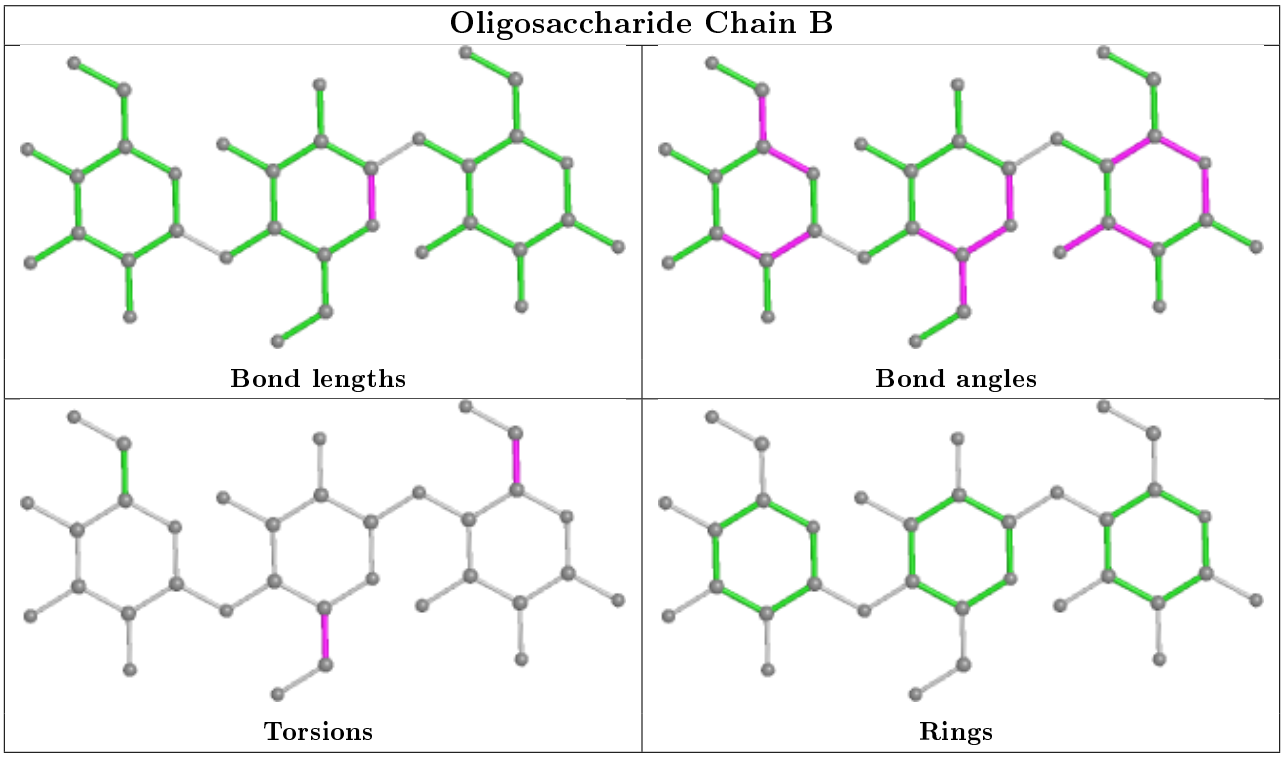
Mol	Chain	Res	Type	Atoms
2	B	2	BGC	O5-C5-C6-O6
2	B	2	BGC	C4-C5-C6-O6
2	B	1	BGC	C4-C5-C6-O6

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	2	BGC	2	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)

11 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	$\#\|Z\|>2$
4	BTB	A	711	-	13,13,13	1.41	2 (15%)	7,16,16	1.47	1 (14%)
3	EDO	A	701	-	3,3,3	0.52	0	2,2,2	0.39	0
3	EDO	A	708	-	3,3,3	0.47	0	2,2,2	0.34	0
3	EDO	A	705	-	3,3,3	0.57	0	2,2,2	0.33	0
3	EDO	A	710	-	3,3,3	0.51	0	2,2,2	0.47	0
3	EDO	A	704	-	3,3,3	0.47	0	2,2,2	0.45	0
3	EDO	A	703	-	3,3,3	0.60	0	2,2,2	0.14	0
3	EDO	A	702	-	3,3,3	0.35	0	2,2,2	0.50	0
3	EDO	A	707	-	3,3,3	0.55	0	2,2,2	0.22	0
3	EDO	A	709	-	3,3,3	0.59	0	2,2,2	0.11	0
3	EDO	A	706	-	3,3,3	0.55	0	2,2,2	0.59	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	BTB	A	711	-	-	8/21/21/21	-
3	EDO	A	701	-	-	0/1/1/1	-
3	EDO	A	708	-	-	1/1/1/1	-
3	EDO	A	705	-	-	1/1/1/1	-
3	EDO	A	710	-	-	0/1/1/1	-
3	EDO	A	704	-	-	1/1/1/1	-
3	EDO	A	703	-	-	0/1/1/1	-
3	EDO	A	702	-	-	1/1/1/1	-
3	EDO	A	707	-	-	1/1/1/1	-
3	EDO	A	709	-	-	1/1/1/1	-
3	EDO	A	706	-	-	0/1/1/1	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	711	BTB	C7-N	-3.08	1.43	1.48
4	A	711	BTB	C5-N	-2.79	1.43	1.48

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	711	BTB	C6-C5-N	-2.51	101.79	111.59

There are no chirality outliers.

5 of 14 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	711	BTB	C1-C2-N-C5
4	A	711	BTB	C1-C2-N-C7
4	A	711	BTB	C3-C2-N-C5
4	A	711	BTB	C3-C2-N-C7
4	A	711	BTB	C4-C2-N-C5

There are no ring outliers.

3 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	711	BTB	4	0
3	A	703	EDO	4	0
3	A	702	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	A	627/656 (95%)	-0.16	8 (1%)	77	76	20, 33, 48, 69	10 (1%)

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	521	VAL	3.0
1	A	610	GLN	2.9
1	A	103	ARG	2.9
1	A	412	PRO	2.9
1	A	423	TRP	2.6

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

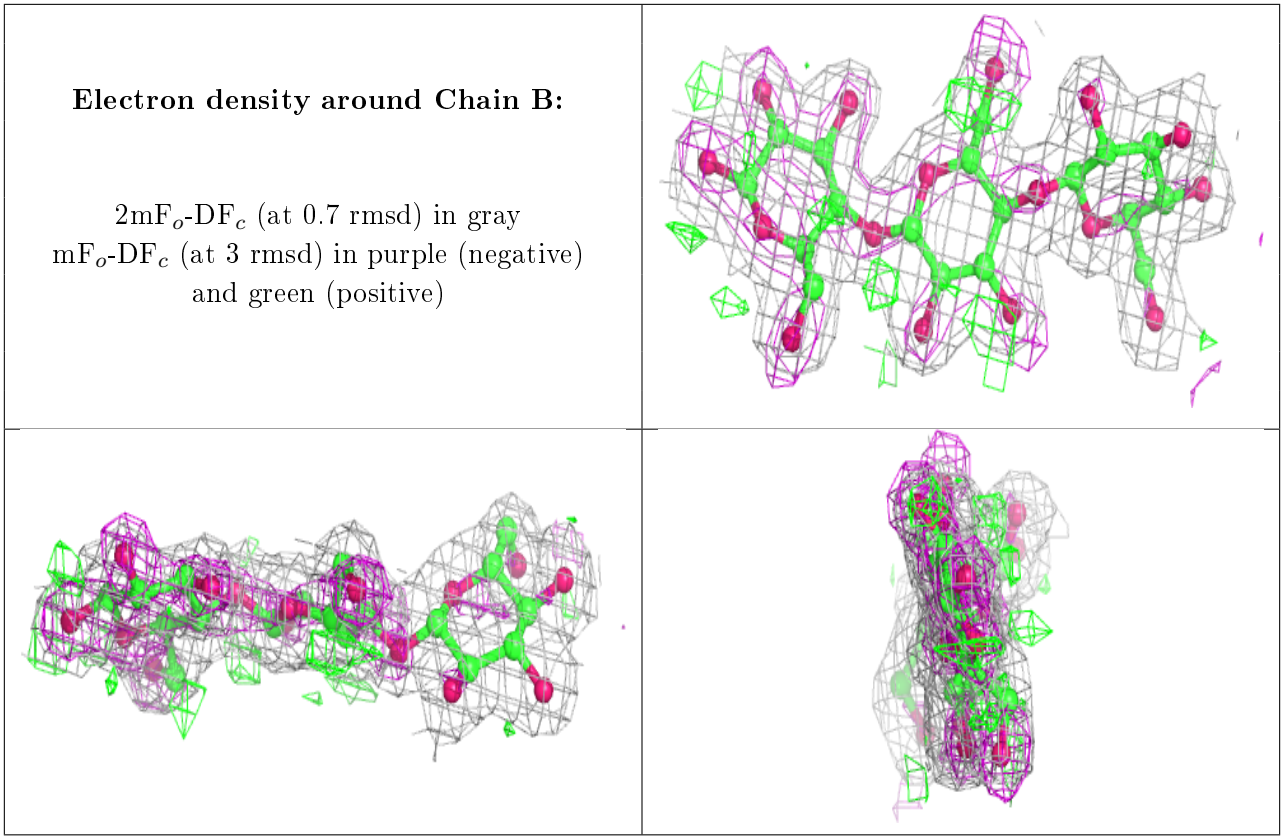
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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	BGC	B	1	12/12	0.87	0.27	20,20,20,20	0
2	BGC	B	2	11/12	0.91	0.12	20,20,20,20	0
2	BGC	B	3	11/12	0.96	0.08	20,20,20,20	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

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	A	705	4/4	0.68	0.22	45,46,49,51	0
3	EDO	A	709	4/4	0.74	0.23	45,46,46,48	0
3	EDO	A	703	4/4	0.80	0.21	41,44,46,46	0
3	EDO	A	710	4/4	0.83	0.26	39,40,40,40	0
3	EDO	A	704	4/4	0.84	0.33	47,47,48,49	0
3	EDO	A	708	4/4	0.85	0.26	54,55,56,58	0
3	EDO	A	707	4/4	0.87	0.18	43,48,52,56	0
4	BTB	A	711	14/14	0.87	0.15	20,20,20,20	0
3	EDO	A	701	4/4	0.94	0.08	32,35,40,44	0
3	EDO	A	702	4/4	0.95	0.11	32,33,34,35	0
3	EDO	A	706	4/4	0.96	0.09	30,32,33,34	0

6.5 Other polymers (i)

There are no such residues in this entry.