wwPDB X-ray Structure Validation Summary Report (i)

Aug 16, 2023 – 08:10 PM EDT

PDB ID	:	2BVW
Title	:	CELLOBIOHYDROLASE II (CEL6A) FROM HUMICOLA INSOLENS IN COMPLEX WITH GLUCOSE AND CELLOTETRAOSE
Authors	:	Varrot, A.; Davies, G.J.; Schulein, M.
Deposited on	:	1999-02-18
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 (i)) were used in the production of this report:

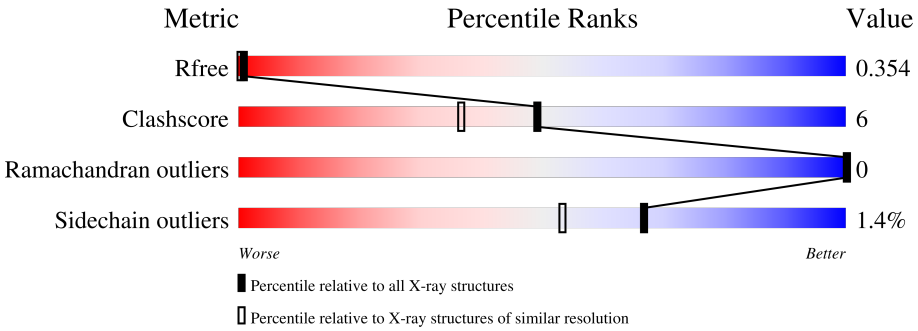
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 (#Entries)	Similar resolution (#Entries, resolution range(Å))
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)

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%

Mol	Chain	Length	Quality of chain
1	A	362	88% 11% .
1	B	362	86% 11% ..
2	C	4	25% 75%
3	D	3	100%

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	GLC	D	1	X	-	-	-

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 6474 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.

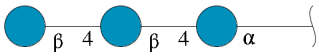
- Molecule 1 is a protein called CELLOBIOHYDROLASE II.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	361	Total	C	N	O	S	0	8	0
			2850	1803	502	535	10			
1	B	360	Total	C	N	O	S	0	7	0
			2834	1795	493	536	10			

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

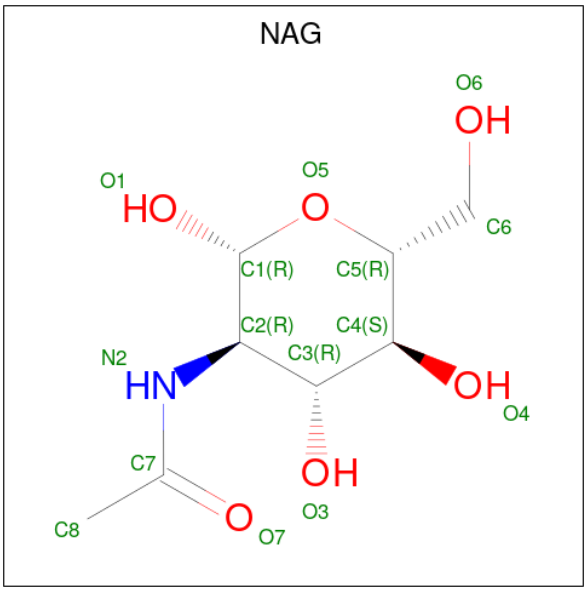
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
2	C	4	Total	C	O	0	0	0
			45	24	21			

- Molecule 3 is an oligosaccharide called beta-D-glucopyranose-(1-4)-beta-D-glucopyranose-(1-4)-alpha-D-glucopyranose.



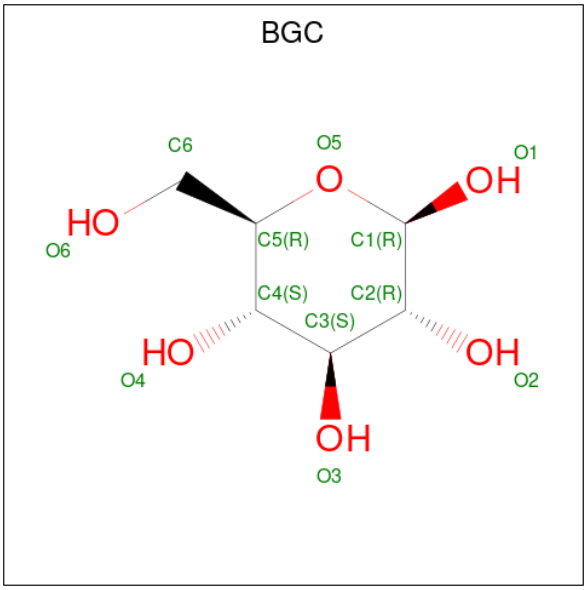
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
3	D	3	Total	C	O	0	0	0
			34	18	16			

- Molecule 4 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



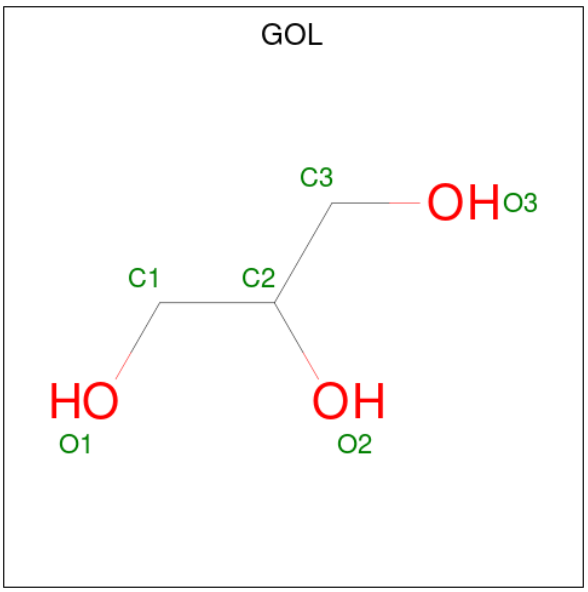
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C N O 14 8 1 5	0	0
4	B	1	Total C N O 14 8 1 5	0	0

- Molecule 5 is beta-D-glucopyranose (three-letter code: BGC) (formula: $C_6H_{12}O_6$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total C O 12 6 6	0	0
5	B	1	Total C O 12 6 6	0	0

- Molecule 6 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

- Molecule 7 is water.

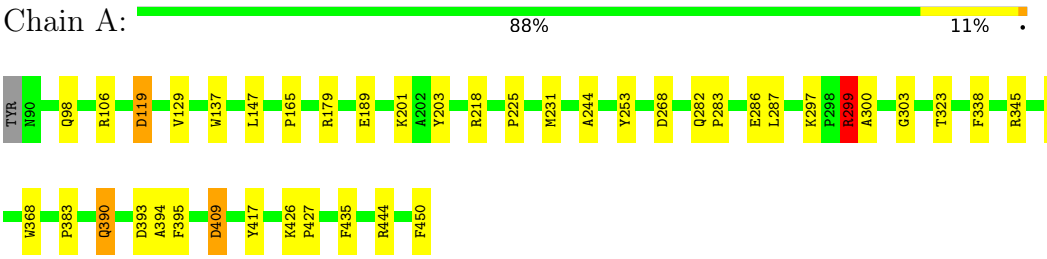
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	355	Total O 355 355	0	0
7	B	262	Total O 262 262	0	0

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. 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. 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.

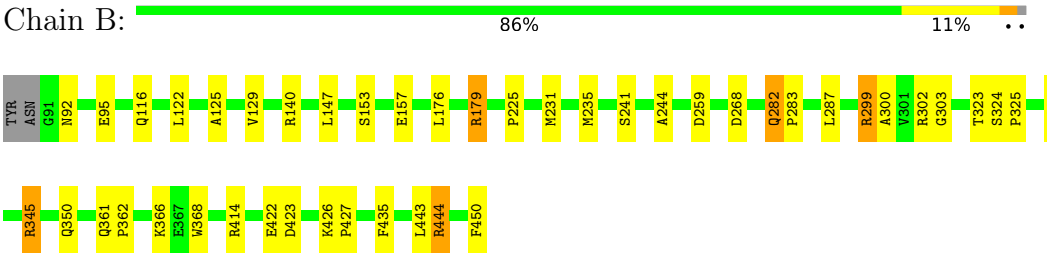
• Molecule 1: CELLOBIOHYDROLASE II

Chain A:



• Molecule 1: CELLOBIOHYDROLASE II

Chain B:

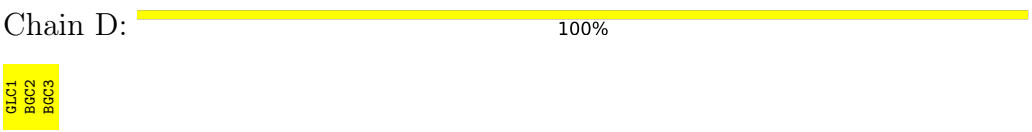


• Molecule 2: beta-D-glucopyranose-(1-4)-beta-D-glucopyranose-(1-4)-beta-D-glucopyranose-(1-4)-beta-D-glucopyranose

Chain C:

• Molecule 3: beta-D-glucopyranose-(1-4)-beta-D-glucopyranose-(1-4)-alpha-D-glucopyranose

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	48.56Å 154.43Å 51.04Å 90.00° 119.31° 90.00°	Depositor
Resolution (Å)	20.00 – 1.70 19.80 – 1.67	Depositor EDS
% Data completeness (in resolution range)	89.6 (20.00-1.70) 86.3 (19.80-1.67)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	0.08	Depositor
$< I/\sigma(I) >$ [1]	1.91 (at 1.67Å)	Xtriage
Refinement program	CCP4	Depositor
R, R_{free}	0.175 , 0.226 0.336 , 0.354	Depositor DCC
R_{free} test set	3313 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å^2)	13.7	Xtriage
Anisotropy	0.570	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 45.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.031 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.82	EDS
Total number of atoms	6474	wwPDB-VP
Average B, all atoms (Å^2)	18.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.83% 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, GOL, GLC, NAG

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	0.44	0/2972	1.03	8/4063 (0.2%)
1	B	0.44	0/2951	1.02	11/4035 (0.3%)
All	All	0.44	0/5923	1.03	19/8098 (0.2%)

There are no bond length outliers.

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	119	ASP	CB-CG-OD1	8.05	125.55	118.30
1	A	299[A]	ARG	NE-CZ-NH1	7.15	123.88	120.30
1	A	299[B]	ARG	NE-CZ-NH1	7.15	123.88	120.30
1	B	330	ASP	CB-CG-OD1	7.14	124.72	118.30
1	B	345	ARG	NE-CZ-NH1	6.83	123.72	120.30

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	2850	0	2721	35	4
1	B	2834	0	2697	32	3
2	C	45	0	39	0	0
3	D	34	0	30	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	14	0	13	0	0
4	B	14	0	13	0	0
5	A	12	0	12	1	0
5	B	12	0	12	0	0
6	A	36	0	48	3	0
6	B	6	0	8	0	0
7	A	355	0	0	3	7
7	B	262	0	0	7	1
All	All	6474	0	5593	68	8

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 6.

The worst 5 of 68 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:444[B]:ARG:NH1	7:A:722:HOH:O	1.81	1.14
1:A:129[A]:VAL:HG22	1:A:450:PHE:CZ	1.95	0.99
1:A:129[A]:VAL:HG22	1:A:450:PHE:CE2	2.10	0.87
1:B:299[B]:ARG:NH2	7:B:798:HOH:O	2.10	0.83
1:B:299[A]:ARG:NH2	7:B:858:HOH:O	2.12	0.83

The worst 5 of 8 symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:253:TYR:OH	7:A:824:HOH:O[1_454]	1.62	0.58
1:B:116:GLN:OE1	7:A:920:HOH:O[1_454]	1.97	0.23
7:A:860:HOH:O	7:B:711:HOH:O[1_656]	2.02	0.18
1:B:157[B]:GLU:CG	7:A:753:HOH:O[1_454]	2.04	0.16
1:A:119:ASP:OD1	7:A:694:HOH:O[1_656]	2.07	0.13

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	367/362 (101%)	356 (97%)	11 (3%)	0	100	100
1	B	365/362 (101%)	353 (97%)	12 (3%)	0	100	100
All	All	732/724 (101%)	709 (97%)	23 (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	296/289 (102%)	291 (98%)	5 (2%)	60	46
1	B	294/289 (102%)	289 (98%)	5 (2%)	60	46
All	All	590/578 (102%)	580 (98%)	10 (2%)	67	46

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	299[A]	ARG
1	B	299[B]	ARG
1	B	435	PHE
1	A	390	GLN
1	A	435	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	230	ASN
1	B	234	ASN
1	A	445	ASN
1	B	116	GLN
1	B	136	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)

7 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	C	1	2	12,12,12	0.75	0	17,17,17	0.95	0
2	BGC	C	2	2	11,11,12	0.68	0	15,15,17	1.37	2 (13%)
2	BGC	C	3	2	11,11,12	0.43	0	15,15,17	1.22	2 (13%)
2	BGC	C	4	2	11,11,12	0.91	0	15,15,17	1.12	1 (6%)
3	GLC	D	1	3	12,12,12	0.51	0	17,17,17	0.90	0
3	BGC	D	2	3	11,11,12	0.79	0	15,15,17	1.17	1 (6%)
3	BGC	D	3	3	11,11,12	0.52	0	15,15,17	0.88	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	BGC	C	1	2	-	0/2/22/22	0/1/1/1
2	BGC	C	2	2	-	0/2/19/22	0/1/1/1
2	BGC	C	3	2	-	0/2/19/22	0/1/1/1
2	BGC	C	4	2	-	0/2/19/22	0/1/1/1
3	GLC	D	1	3	1/1/5/5	0/2/22/22	0/1/1/1
3	BGC	D	2	3	-	0/2/19/22	0/1/1/1
3	BGC	D	3	3	-	0/2/19/22	0/1/1/1

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	2	BGC	O3-C3-C2	-3.24	103.79	109.99
2	C	2	BGC	O5-C1-C2	-2.62	106.73	110.77
2	C	3	BGC	O3-C3-C2	-2.36	105.48	109.99
3	D	2	BGC	C1-O5-C5	2.27	115.27	112.19
2	C	3	BGC	O5-C1-C2	-2.25	107.30	110.77

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
3	D	1	GLC	C1

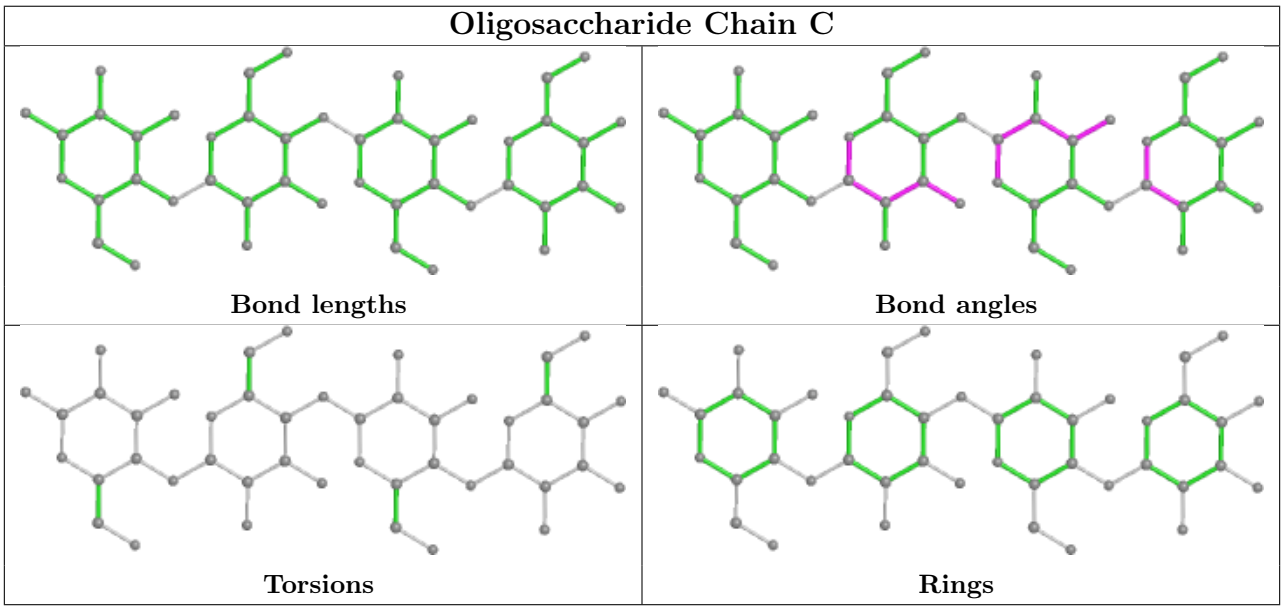
There are no torsion outliers.

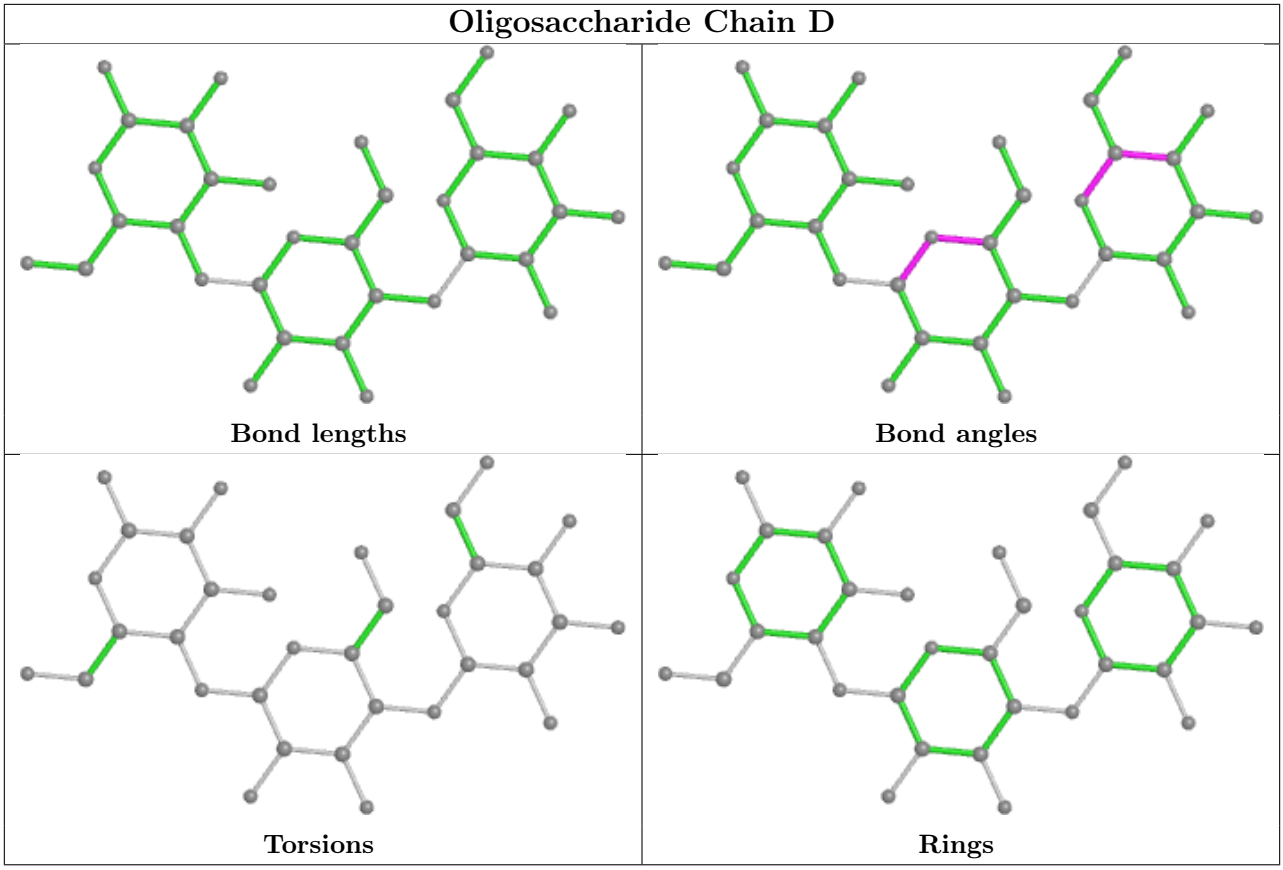
There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	D	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)

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	NAG	B	500	1	14,14,15	1.39	1 (7%)	17,19,21	1.19	1 (5%)
6	GOL	A	610	-	5,5,5	0.57	0	5,5,5	0.58	0
6	GOL	B	608	-	5,5,5	0.50	0	5,5,5	0.47	0
4	NAG	A	500	1	14,14,15	1.11	1 (7%)	17,19,21	0.97	0
6	GOL	A	605	-	5,5,5	0.42	0	5,5,5	0.79	0
6	GOL	A	606	-	5,5,5	0.65	0	5,5,5	0.51	0
5	BGC	A	602	-	12,12,12	0.67	0	17,17,17	1.02	0
6	GOL	A	607	-	5,5,5	0.58	0	5,5,5	0.44	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
5	BGC	B	603	-	12,12,12	0.75	0	17,17,17	0.86	0
6	GOL	A	604	-	5,5,5	0.67	0	5,5,5	0.54	0
6	GOL	A	609	-	5,5,5	0.74	0	5,5,5	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	NAG	B	500	1	-	0/6/23/26	0/1/1/1
6	GOL	A	610	-	-	2/4/4/4	-
6	GOL	B	608	-	-	4/4/4/4	-
4	NAG	A	500	1	-	0/6/23/26	0/1/1/1
6	GOL	A	605	-	-	2/4/4/4	-
6	GOL	A	606	-	-	0/4/4/4	-
5	BGC	A	602	-	-	0/2/22/22	0/1/1/1
6	GOL	A	607	-	-	4/4/4/4	-
5	BGC	B	603	-	-	0/2/22/22	0/1/1/1
6	GOL	A	604	-	-	2/4/4/4	-
6	GOL	A	609	-	-	2/4/4/4	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	500	NAG	O7-C7	-3.61	1.15	1.23
4	A	500	NAG	O7-C7	-3.47	1.15	1.23

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	B	500	NAG	O5-C1-C2	-2.12	107.94	111.29

There are no chirality outliers.

5 of 16 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	A	607	GOL	O1-C1-C2-C3
6	A	610	GOL	C1-C2-C3-O3
6	B	608	GOL	O1-C1-C2-C3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
6	B	608	GOL	C1-C2-C3-O3
6	A	607	GOL	O1-C1-C2-O2

There are no ring outliers.

4 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	A	610	GOL	1	0
6	A	605	GOL	1	0
5	A	602	BGC	1	0
6	A	609	GOL	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

Unable to reproduce the depositors R factor - this section is therefore empty.

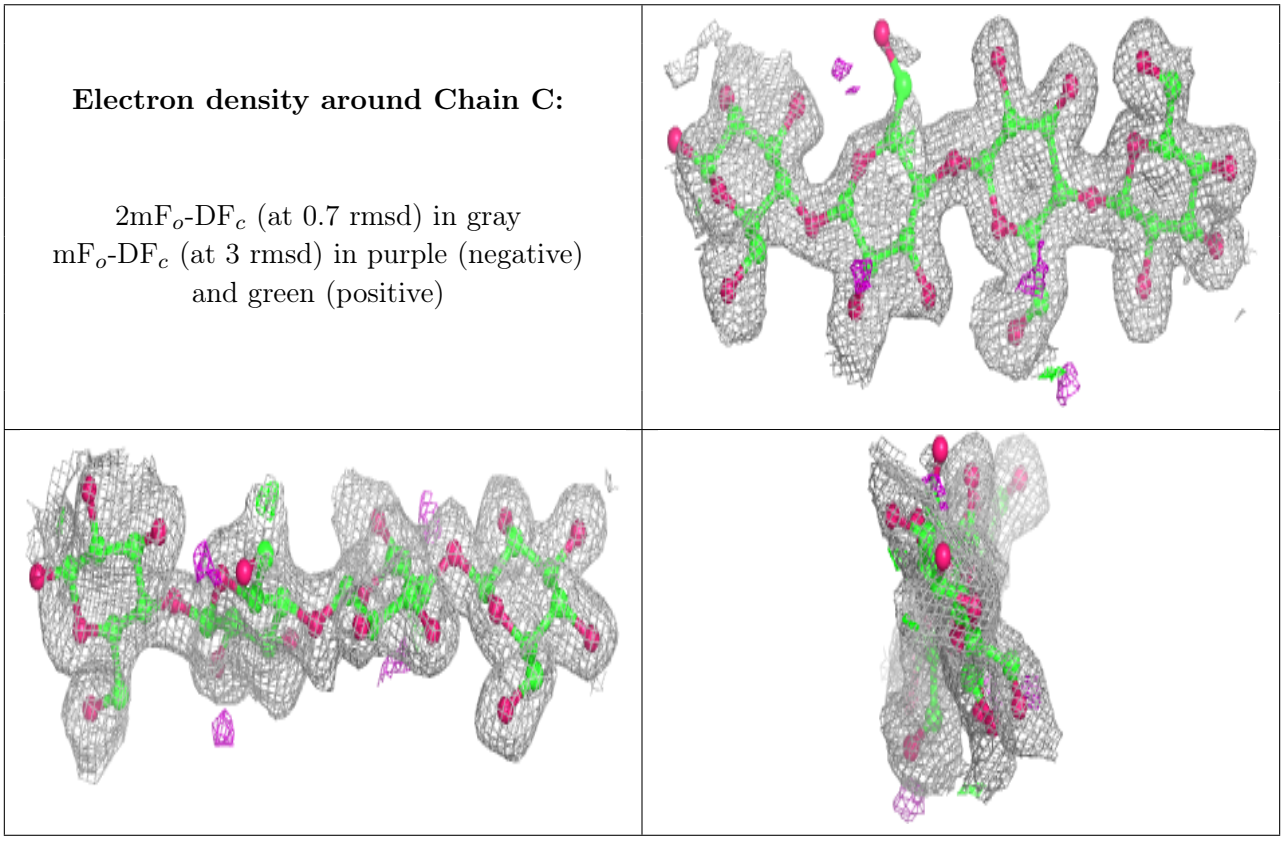
6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

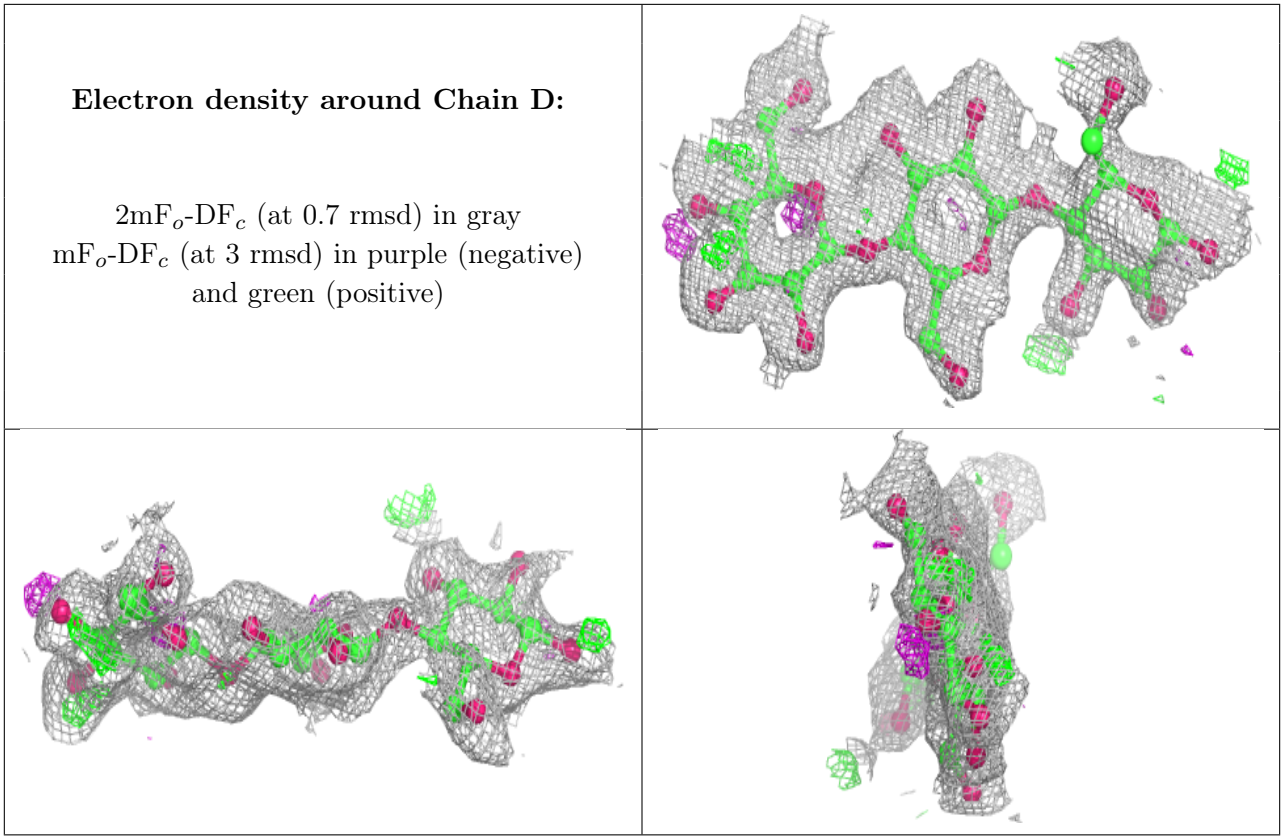
Unable to reproduce the depositors R factor - this section is therefore empty.

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 D:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)



6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.