



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 8, 2020 – 12:36 AM BST

PDB ID : 5YSE
Title : Crystal structure of beta-1,2-glucooligosaccharide binding protein in complex with sophorotetraose
Authors : Abe, K.; Nakajima, M.; Taguchi, H.; Arakawa, T.; Fushinobu, S.
Deposited on : 2017-11-14
Resolution : 1.60 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 ⓘ 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

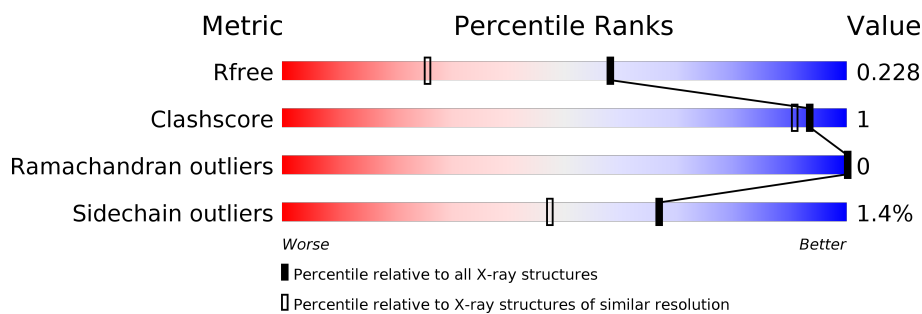
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.60 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	3398 (1.60-1.60)
Clashscore	141614	3665 (1.60-1.60)
Ramachandran outliers	138981	3564 (1.60-1.60)
Sidechain outliers	138945	3563 (1.60-1.60)

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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	397	
1	B	397	
2	C	4	
2	D	4	

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 6899 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.

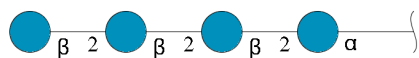
- Molecule 1 is a protein called Lin1841 protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	387	Total	C	N	O	S	0	3	0
			3081	1968	502	595	16			
1	B	387	Total	C	N	O	S	0	1	0
			3071	1961	502	592	16			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	26	MET	-	expression tag	UNP Q92AS8
A	415	LEU	-	expression tag	UNP Q92AS8
A	416	GLU	-	expression tag	UNP Q92AS8
A	417	HIS	-	expression tag	UNP Q92AS8
A	418	HIS	-	expression tag	UNP Q92AS8
A	419	HIS	-	expression tag	UNP Q92AS8
A	420	HIS	-	expression tag	UNP Q92AS8
A	421	HIS	-	expression tag	UNP Q92AS8
A	422	HIS	-	expression tag	UNP Q92AS8
B	26	MET	-	expression tag	UNP Q92AS8
B	415	LEU	-	expression tag	UNP Q92AS8
B	416	GLU	-	expression tag	UNP Q92AS8
B	417	HIS	-	expression tag	UNP Q92AS8
B	418	HIS	-	expression tag	UNP Q92AS8
B	419	HIS	-	expression tag	UNP Q92AS8
B	420	HIS	-	expression tag	UNP Q92AS8
B	421	HIS	-	expression tag	UNP Q92AS8
B	422	HIS	-	expression tag	UNP Q92AS8

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



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

- Molecule 3 is (4S)-2-METHYL-2,4-PENTANEDIOL (three-letter code: MPD) (formula: C₆H₁₄O₂).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			8	6	2		
3	A	1	Total	C	O	0	0
			8	6	2		
3	A	1	Total	C	O	0	0
			8	6	2		

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Mg	0	0
			1	1		

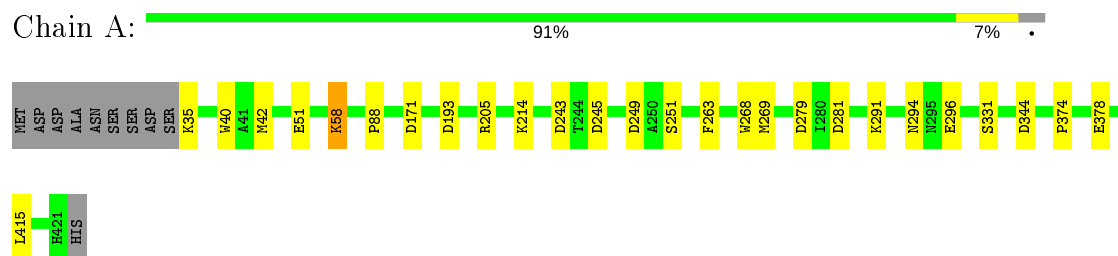
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	427	Total 427	O 427	0	0
5	B	205	Total 205	O 205	0	0

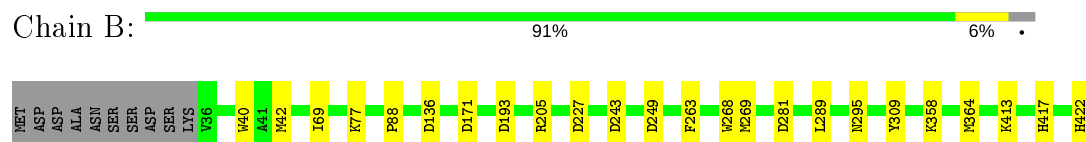
3 Residue-property plots [i](#)

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.

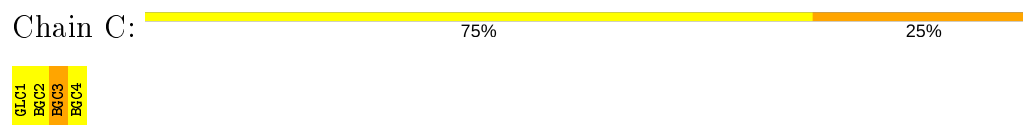
- Molecule 1: Lin1841 protein



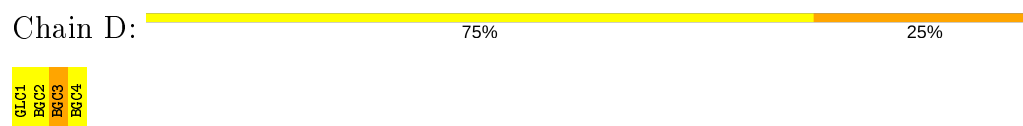
- Molecule 1: Lin1841 protein



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



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



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	36.01Å 63.28Å 87.45Å 89.99° 82.24° 86.08°	Depositor
Resolution (Å)	43.32 – 1.60 31.56 – 1.60	Depositor EDS
% Data completeness (in resolution range)	96.4 (43.32-1.60) 96.4 (31.56-1.60)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.85 (at 1.60Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
R, R_{free}	0.158 , 0.188 0.206 , 0.228	Depositor DCC
R_{free} test set	4783 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å ²)	19.7	Xtriage
Anisotropy	0.251	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 42.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	6899	wwPDB-VP
Average B, all atoms (Å ²)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.29% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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: GLC, MG, BGC, 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	1.11	5/3165 (0.2%)	1.03	13/4284 (0.3%)
1	B	0.84	2/3153 (0.1%)	0.94	11/4268 (0.3%)
All	All	0.98	7/6318 (0.1%)	0.99	24/8552 (0.3%)

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	331	SER	CB-OG	13.19	1.59	1.42
1	A	279	ASP	CB-CG	8.43	1.69	1.51
1	B	227	ASP	CG-OD2	6.36	1.40	1.25
1	A	51	GLU	CD-OE1	6.26	1.32	1.25
1	A	378	GLU	CD-OE2	5.81	1.32	1.25

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	279	ASP	CB-CG-OD1	14.11	131.00	118.30
1	B	227	ASP	CB-CG-OD2	-13.00	106.60	118.30
1	A	171	ASP	CB-CG-OD1	8.45	125.90	118.30
1	B	136	ASP	CB-CG-OD2	-8.31	110.83	118.30
1	B	136	ASP	CB-CG-OD1	7.27	124.85	118.30

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	3081	0	3002	9	0
1	B	3071	0	2986	7	0
2	C	45	0	39	1	0
2	D	45	0	39	1	0
3	A	24	0	42	2	0
4	A	1	0	0	0	0
5	A	427	0	0	3	1
5	B	205	0	0	0	0
All	All	6899	0	6108	17	1

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.

The worst 5 of 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)
3:A:503:MPD:H52	3:A:503:MPD:HM1	1.92	0.52
1:A:35:LYS:N	5:A:605:HOH:O	2.43	0.51
1:B:69:ILE:HD13	1:B:77:LYS:HD3	1.94	0.50
1:A:263:PHE:CE2	1:A:269:MET:HG2	2.48	0.48
1:A:58:LYS:NZ	5:A:603:HOH:O	2.40	0.48

All (1) 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)
5:A:931:HOH:O	5:A:938:HOH:O[1_565]	2.13	0.07

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	388/397 (98%)	380 (98%)	8 (2%)	0	100	100
1	B	386/397 (97%)	378 (98%)	8 (2%)	0	100	100
All	All	774/794 (98%)	758 (98%)	16 (2%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

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	335/341 (98%)	331 (99%)	4 (1%)	71	54
1	B	333/341 (98%)	328 (98%)	5 (2%)	65	44
All	All	668/682 (98%)	659 (99%)	9 (1%)	67	50

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	42	MET
1	B	417	HIS
1	B	364	MET
1	A	214	LYS
1	B	358	LYS

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	271	ASN

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 ⓘ

8 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	C	1	2	12,12,12	0.72	0	17,17,17	1.12	1 (5%)
2	BGC	C	2	2	11,11,12	0.80	0	15,15,17	1.18	1 (6%)
2	BGC	C	3	2	11,11,12	0.81	0	15,15,17	0.98	1 (6%)
2	BGC	C	4	2	11,11,12	1.09	0	15,15,17	1.12	1 (6%)
2	GLC	D	1	2	12,12,12	0.85	0	17,17,17	1.05	1 (5%)
2	BGC	D	2	2	11,11,12	1.04	0	15,15,17	1.25	1 (6%)
2	BGC	D	3	2	11,11,12	0.98	0	15,15,17	0.71	1 (6%)
2	BGC	D	4	2	11,11,12	1.08	1 (9%)	15,15,17	0.73	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	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
2	GLC	D	1	2	-	0/2/22/22	0/1/1/1
2	BGC	D	2	2	-	0/2/19/22	0/1/1/1
2	BGC	D	3	2	-	0/2/19/22	0/1/1/1
2	BGC	D	4	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	D	4	BGC	O5-C1	-2.33	1.40	1.43

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	1	GLC	C3-C4-C5	-2.60	105.61	110.24
2	C	1	GLC	O6-C6-C5	-2.50	102.71	111.29
2	C	2	BGC	C1-C2-C3	-2.48	106.62	109.67
2	D	2	BGC	C6-C5-C4	-2.45	107.25	113.00
2	C	3	BGC	C1-O5-C5	2.16	115.12	112.19

There are no chirality outliers.

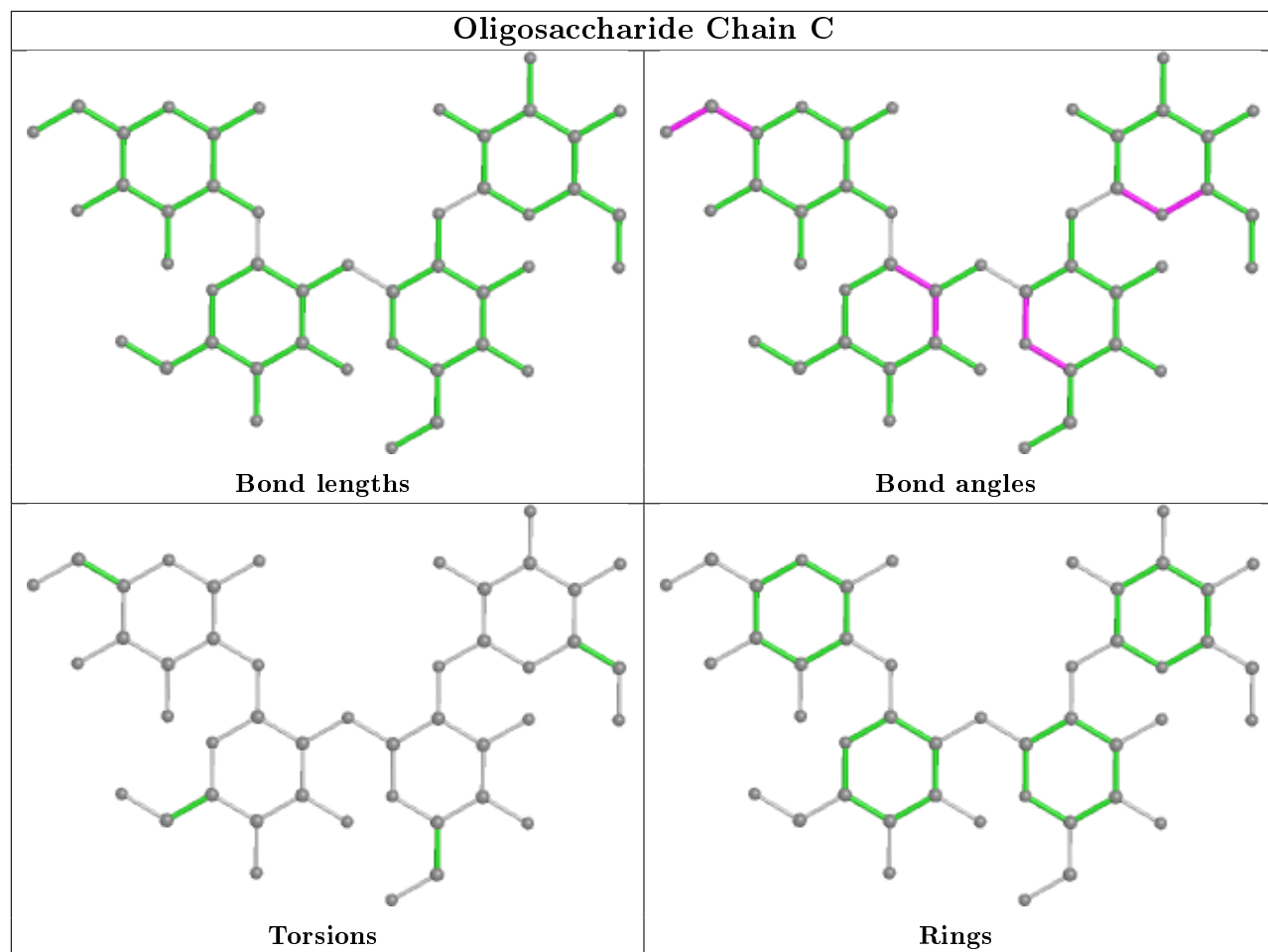
There are no torsion outliers.

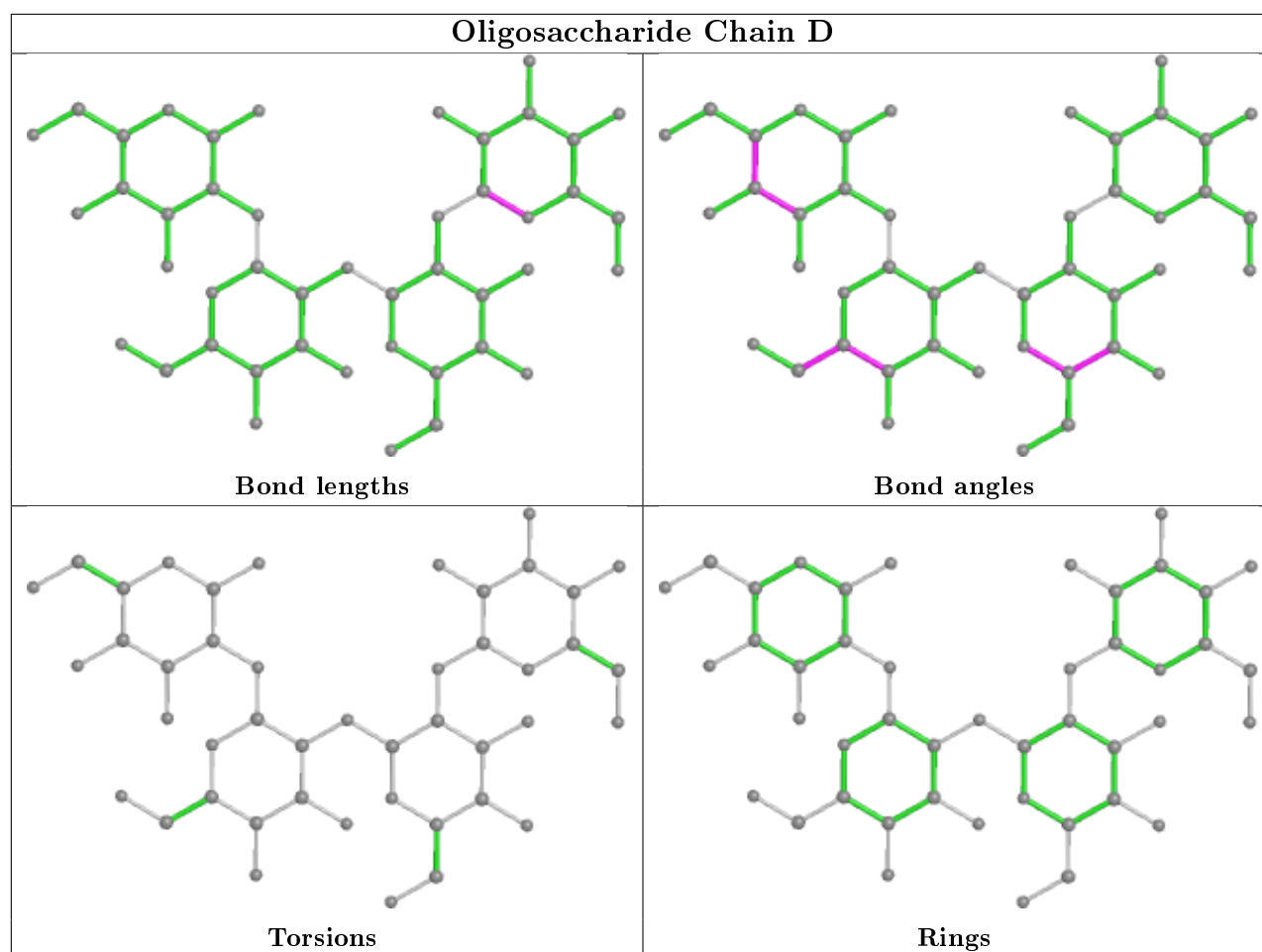
There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	D	3	BGC	1	0
2	C	3	BGC	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](#)

Of 4 ligands modelled in this entry, 1 is monoatomic - leaving 3 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	$\# Z > 2$
3	MPD	A	502	-	7,7,7	0.46	0	9,10,10	0.55	0
3	MPD	A	503	-	7,7,7	0.44	0	9,10,10	1.18	1 (11%)
3	MPD	A	501	-	7,7,7	0.54	0	9,10,10	0.66	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	MPD	A	502	-	-	1/5/5/5	-
3	MPD	A	503	-	-	1/5/5/5	-
3	MPD	A	501	-	-	2/5/5/5	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	503	MPD	CM-C2-C1	-2.25	105.89	110.57

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	502	MPD	C2-C3-C4-C5
3	A	503	MPD	C2-C3-C4-C5
3	A	501	MPD	CM-C2-C3-C4
3	A	501	MPD	O2-C2-C3-C4

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	503	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](#)

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

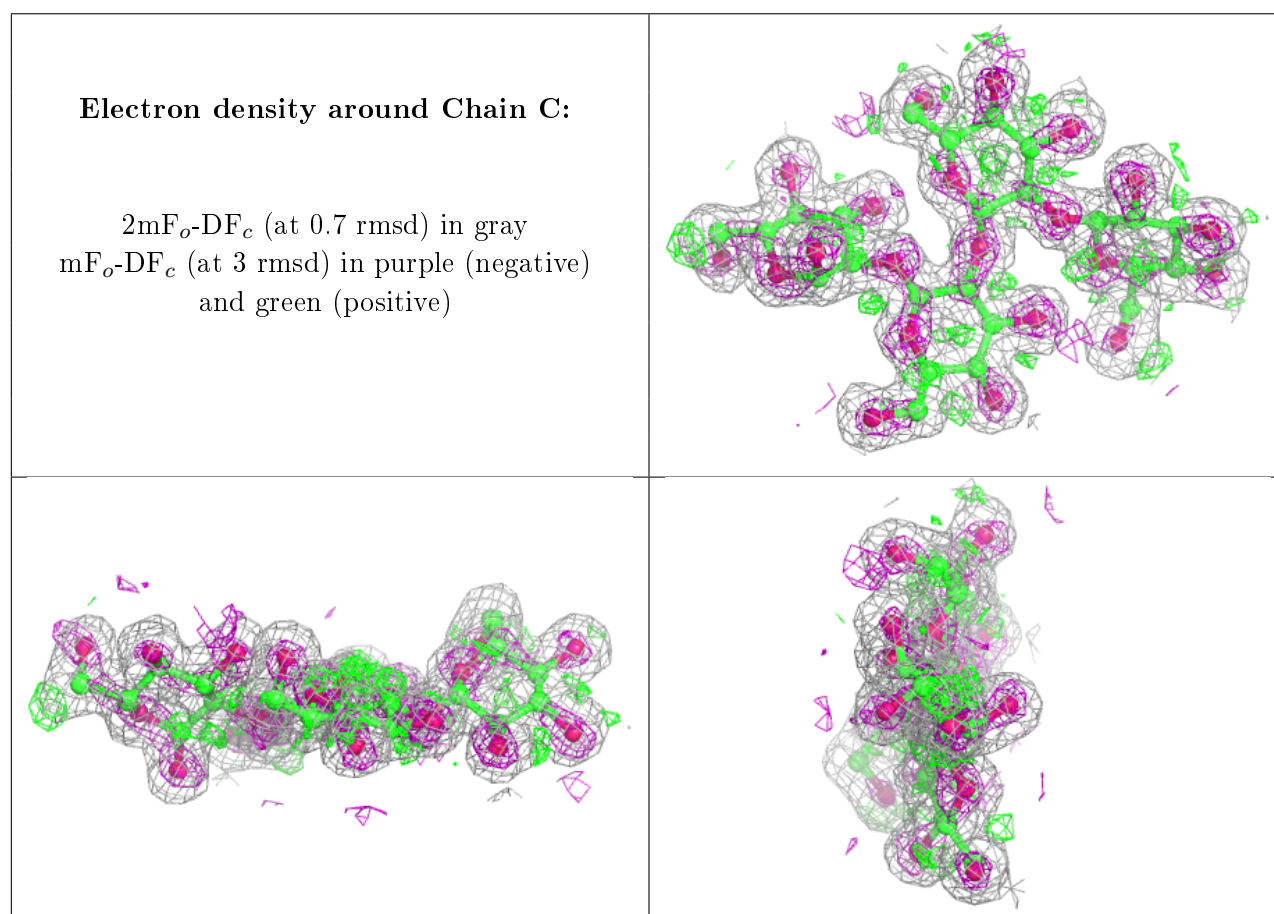
6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

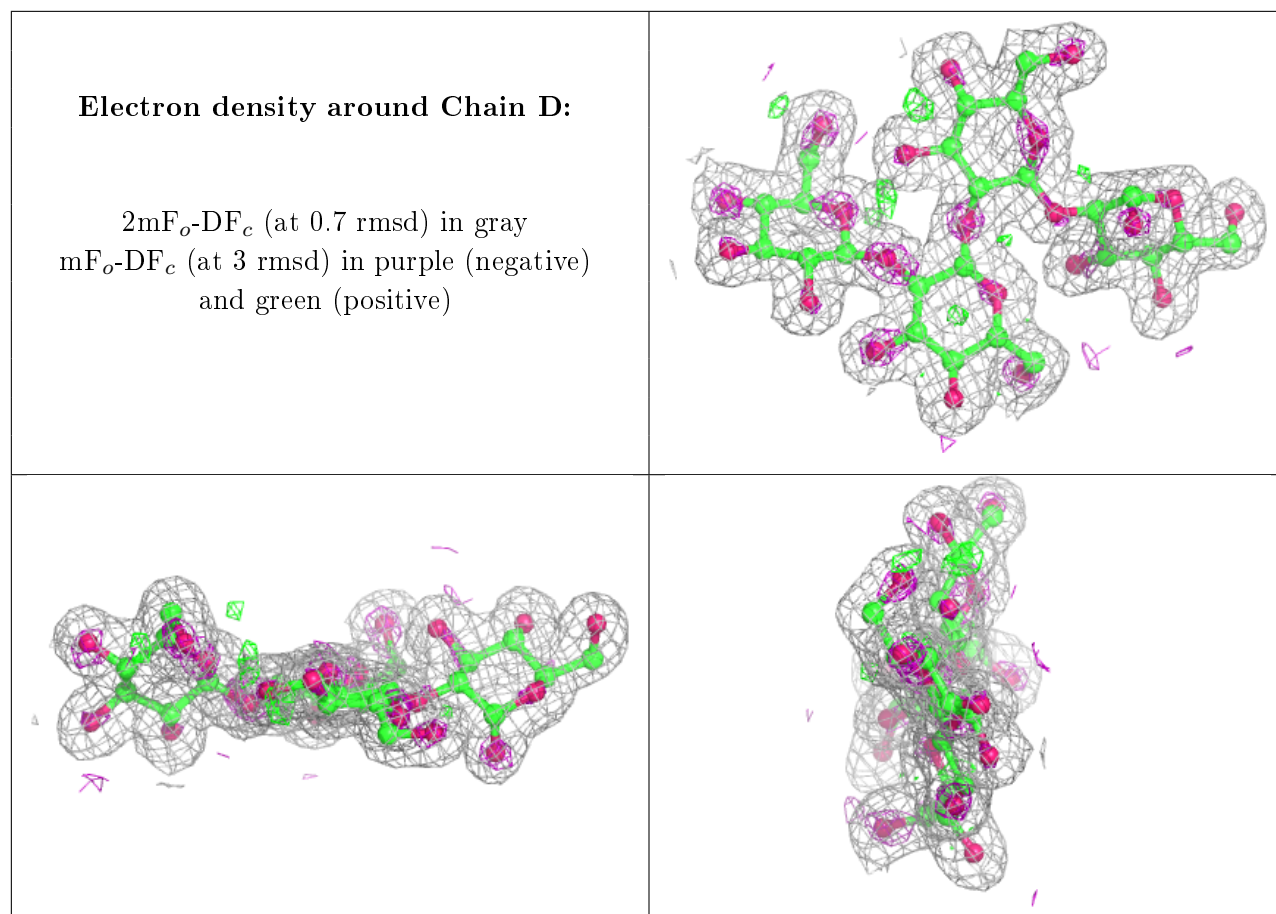
Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates [i](#)

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.





6.4 Ligands [i](#)

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

6.5 Other polymers [i](#)

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