wwPDB X-ray Structure Validation Summary Report (i)

Dec 13, 2023 – 01:21 am GMT

PDB ID	:	2YGM
Title	:	THE X-RAY CRYSTAL STRUCTURE OF TANDEM CBM51 MODULES OF SP3GH98, THE FAMILY 98 GLYCOSIDE HYDROLASE FROM STREPTOCOCCUS PNEUMONIAE SP3-BS71, IN COMPLEX WITH THE BLOOD GROUP B ANTIGEN
Authors	:	Higgins, M.A.; Ficko-Blean, E.; Wright, C.; Meloncelli, P.J.; Lowary, T.L.; Boraston, A.B.
Deposited on	:	2011-04-19
Resolution	:	2.35 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:

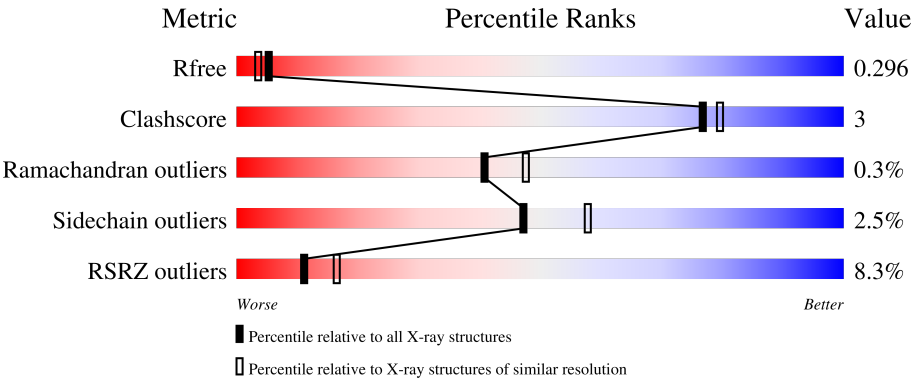
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)

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 2.35 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	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)
RSRZ outliers	127900	1150 (2.36-2.36)

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	A	354	10%, 91%, 7%
1	B	354	6%, 90%, 10%
2	C	4	25%, 75%
2	D	4	100%

Validation Pipeline (wwPDB-VP) : 2.36

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 5697 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.

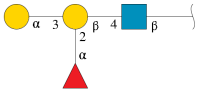
- Molecule 1 is a protein called BLOOD GROUP A-AND B-CLEAVING ENDO-BETA-GAL ACTOSIDASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	351	Total	C	N	O	S	0	1	0
			2712	1707	447	555	3			
1	B	352	Total	C	N	O	S	0	0	0
			2712	1706	447	556	3			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	60	GLY	-	expression tag	UNP C1CB04
A	61	SER	-	expression tag	UNP C1CB04
A	62	HIS	-	expression tag	UNP C1CB04
A	63	MET	-	expression tag	UNP C1CB04
A	64	ALA	-	expression tag	UNP C1CB04
A	65	SER	-	expression tag	UNP C1CB04
B	60	GLY	-	expression tag	UNP C1CB04
B	61	SER	-	expression tag	UNP C1CB04
B	62	HIS	-	expression tag	UNP C1CB04
B	63	MET	-	expression tag	UNP C1CB04
B	64	ALA	-	expression tag	UNP C1CB04
B	65	SER	-	expression tag	UNP C1CB04

- Molecule 2 is an oligosaccharide called alpha-L-fucopyranose-(1-2)-[alpha-D-galactopyranose-(1-3)]beta-D-galactopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose.



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	4	Total	C	N	O	0	0	0
			47	26	1	20			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
2	D	4	Total C N O 47 26 1 20	0	0	0

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Ca 1 1	0	0
3	B	1	Total Ca 1 1	0	0

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total Na 1 1	0	0
4	B	1	Total Na 1 1	0	0

- Molecule 5 is water.

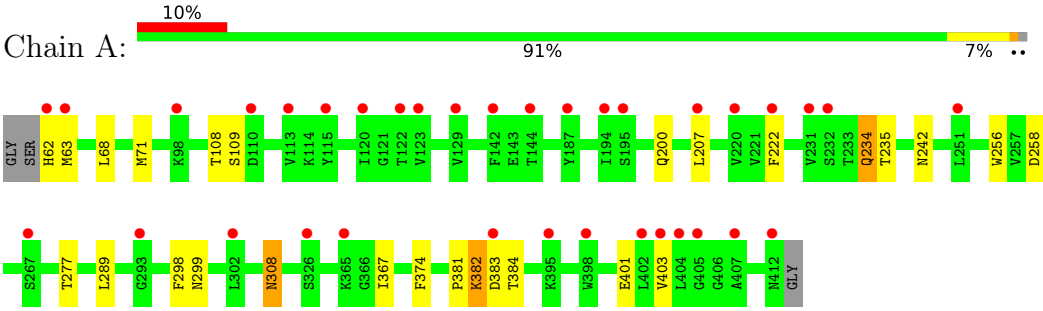
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	98	Total O 98 98	0	0
5	B	77	Total O 77 77	0	0

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.

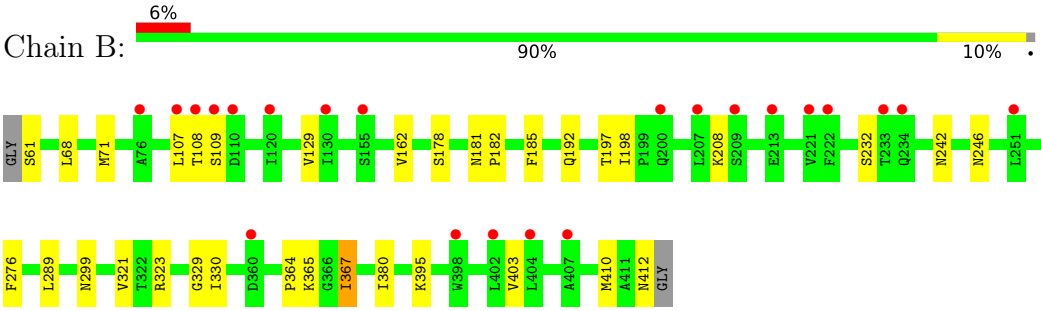
• Molecule 1: BLOOD GROUP A-AND B-CLEAVING ENDO-BETA-GALACTOSIDASE

Chain A:



• Molecule 1: BLOOD GROUP A-AND B-CLEAVING ENDO-BETA-GALACTOSIDASE

Chain B:

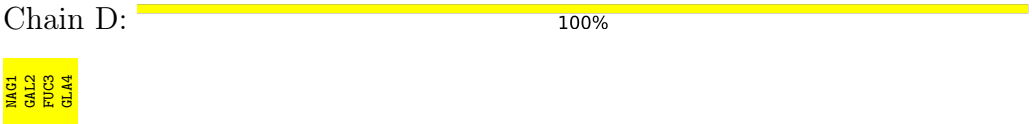


• Molecule 2: alpha-L-fucopyranose-(1-2)-[alpha-D-galactopyranose-(1-3)]beta-D-galactopyranose -(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain C:

• Molecule 2: alpha-L-fucopyranose-(1-2)-[alpha-D-galactopyranose-(1-3)]beta-D-galactopyranose -(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	75.70Å 78.50Å 131.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.00 – 2.35 29.25 – 2.35	Depositor EDS
% Data completeness (in resolution range)	99.9 (40.00-2.35) 100.0 (29.25-2.35)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.85 (at 2.36Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.245 , 0.297 0.245 , 0.296	Depositor DCC
R_{free} test set	1690 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	46.4	Xtriage
Anisotropy	0.052	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 33.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.027 for k,h,-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	5697	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 3.63% 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: CA, GLA, NA, FUC, GAL, 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	0.35	0/2768	0.49	0/3748
1	B	0.34	0/2765	0.49	0/3744
All	All	0.35	0/5533	0.49	0/7492

There are no bond length outliers.

There are no bond angle outliers.

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	2712	0	2644	16	0
1	B	2712	0	2641	17	0
2	C	47	0	42	0	0
2	D	47	0	42	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	A	98	0	0	1	0
5	B	77	0	0	0	0
All	All	5697	0	5369	33	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 3.

The worst 5 of 33 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:382:LYS:HD3	1:A:383:ASP:H	1.28	0.94
1:B:108:THR:HG22	1:B:109:SER:H	1.52	0.73
1:B:181:ASN:HD21	1:B:192:GLN:HG3	1.55	0.71
1:A:382:LYS:HD3	1:A:383:ASP:N	2.07	0.68
1:A:382:LYS:HA	1:A:382:LYS:HE2	1.81	0.62

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	350/354 (99%)	334 (95%)	15 (4%)	1 (0%)	41	47
1	B	350/354 (99%)	337 (96%)	12 (3%)	1 (0%)	41	47
All	All	700/708 (99%)	671 (96%)	27 (4%)	2 (0%)	41	47

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	276	PHE
1	A	308	ASN

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	303/303 (100%)	296 (98%)	7 (2%)	50	61
1	B	303/303 (100%)	295 (97%)	8 (3%)	46	56
All	All	606/606 (100%)	591 (98%)	15 (2%)	47	58

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	107	LEU
1	B	395	LYS
1	B	197	THR
1	B	412	ASN
1	B	274	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	335	ASN
1	A	412	ASN
1	B	412	ASN
1	B	242	ASN
1	B	261	HIS

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)

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	NAG	C	1	2	15,15,15	0.45	0	21,21,21	1.16	3 (14%)
2	GAL	C	2	2	11,11,12	0.72	0	15,15,17	1.11	2 (13%)
2	FUC	C	3	2	10,10,11	0.58	0	14,14,16	0.59	0
2	GLA	C	4	2	11,11,12	0.51	0	15,15,17	1.37	1 (6%)
2	NAG	D	1	2	15,15,15	0.44	0	21,21,21	0.97	1 (4%)
2	GAL	D	2	2	11,11,12	0.63	0	15,15,17	1.15	3 (20%)
2	FUC	D	3	2	10,10,11	0.66	0	14,14,16	0.85	1 (7%)
2	GLA	D	4	2	11,11,12	0.56	0	15,15,17	0.75	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	NAG	C	1	2	-	0/6/26/26	0/1/1/1
2	GAL	C	2	2	-	0/2/19/22	0/1/1/1
2	FUC	C	3	2	-	-	0/1/1/1
2	GLA	C	4	2	-	0/2/19/22	0/1/1/1
2	NAG	D	1	2	-	0/6/26/26	0/1/1/1
2	GAL	D	2	2	-	0/2/19/22	0/1/1/1
2	FUC	D	3	2	-	-	0/1/1/1
2	GLA	D	4	2	-	0/2/19/22	0/1/1/1

There are no bond length outliers.

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	4	GLA	C1-O5-C5	4.63	118.46	112.19
2	C	1	NAG	C4-C3-C2	2.85	114.51	110.34
2	D	2	GAL	C1-C2-C3	2.58	112.84	109.67
2	C	1	NAG	C3-C4-C5	2.44	114.59	110.24
2	D	1	NAG	C4-C3-C2	2.35	113.78	110.34

There are no chirality outliers.

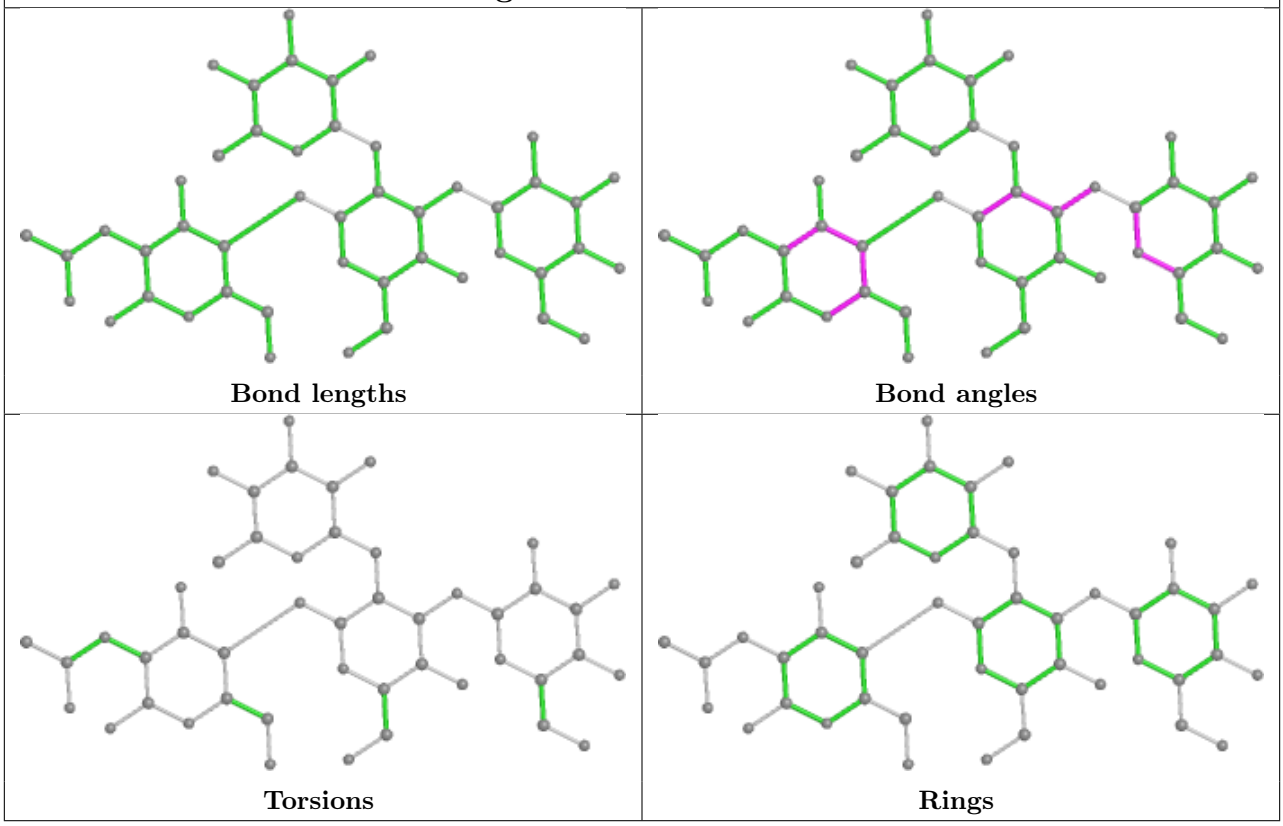
There are no torsion outliers.

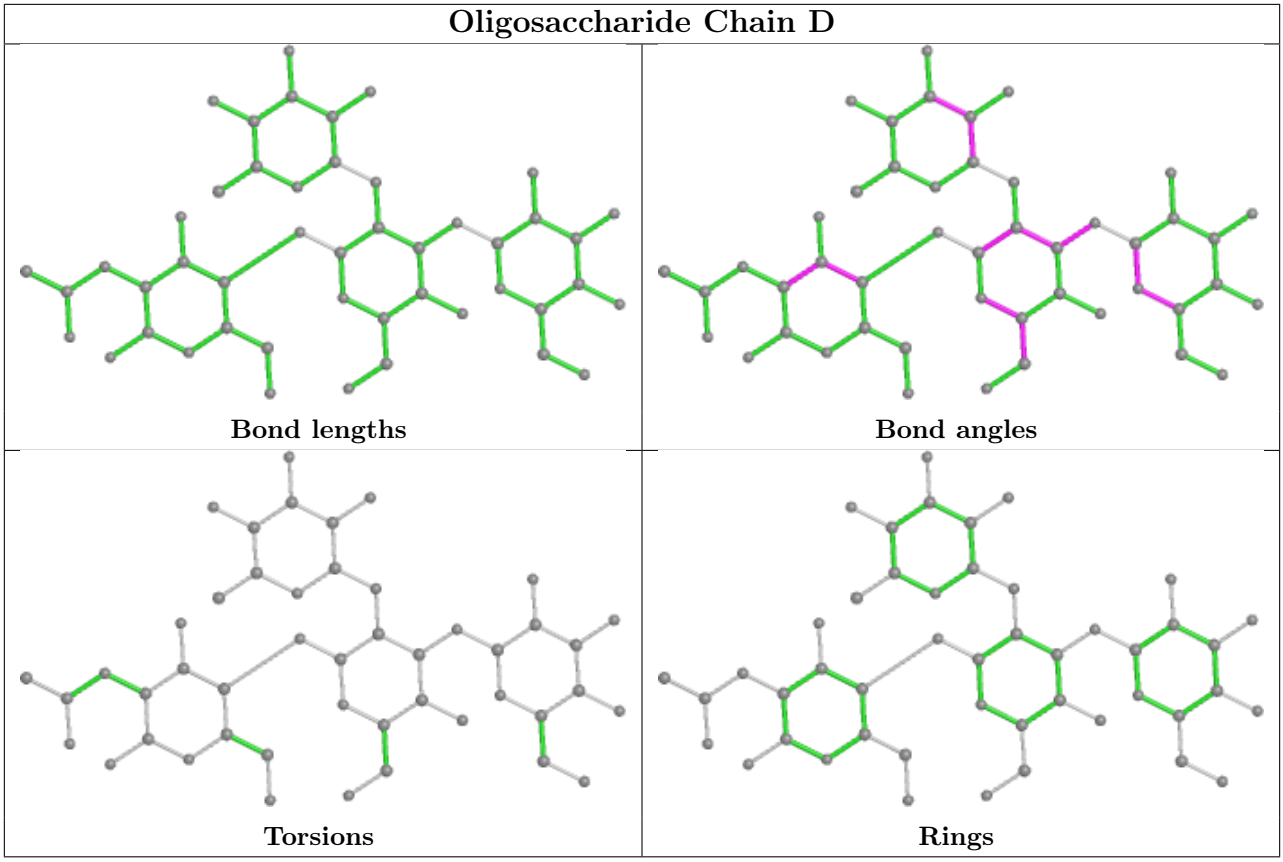
There are no ring outliers.

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 C





5.6 Ligand geometry

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

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	351/354 (99%)	0.62	35 (9%)	7	11	31, 45, 62, 67	0
1	B	352/354 (99%)	0.57	23 (6%)	18	27	33, 47, 58, 61	0
All	All	703/708 (99%)	0.59	58 (8%)	11	16	31, 46, 59, 67	0

The worst 5 of 58 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	120	ILE	4.4
1	A	398	TRP	4.3
1	B	404	LEU	4.2
1	B	200	GLN	3.9
1	B	110	ASP	3.9

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	NAG	C	1	15/15	0.81	0.18	58,59,60,60	0
2	FUC	C	3	10/11	0.88	0.18	57,58,58,58	0
2	NAG	D	1	15/15	0.88	0.17	53,55,55,55	0
2	FUC	D	3	10/11	0.91	0.17	50,51,51,51	0
2	GLA	C	4	11/12	0.92	0.14	53,53,54,55	0

Continued on next page...

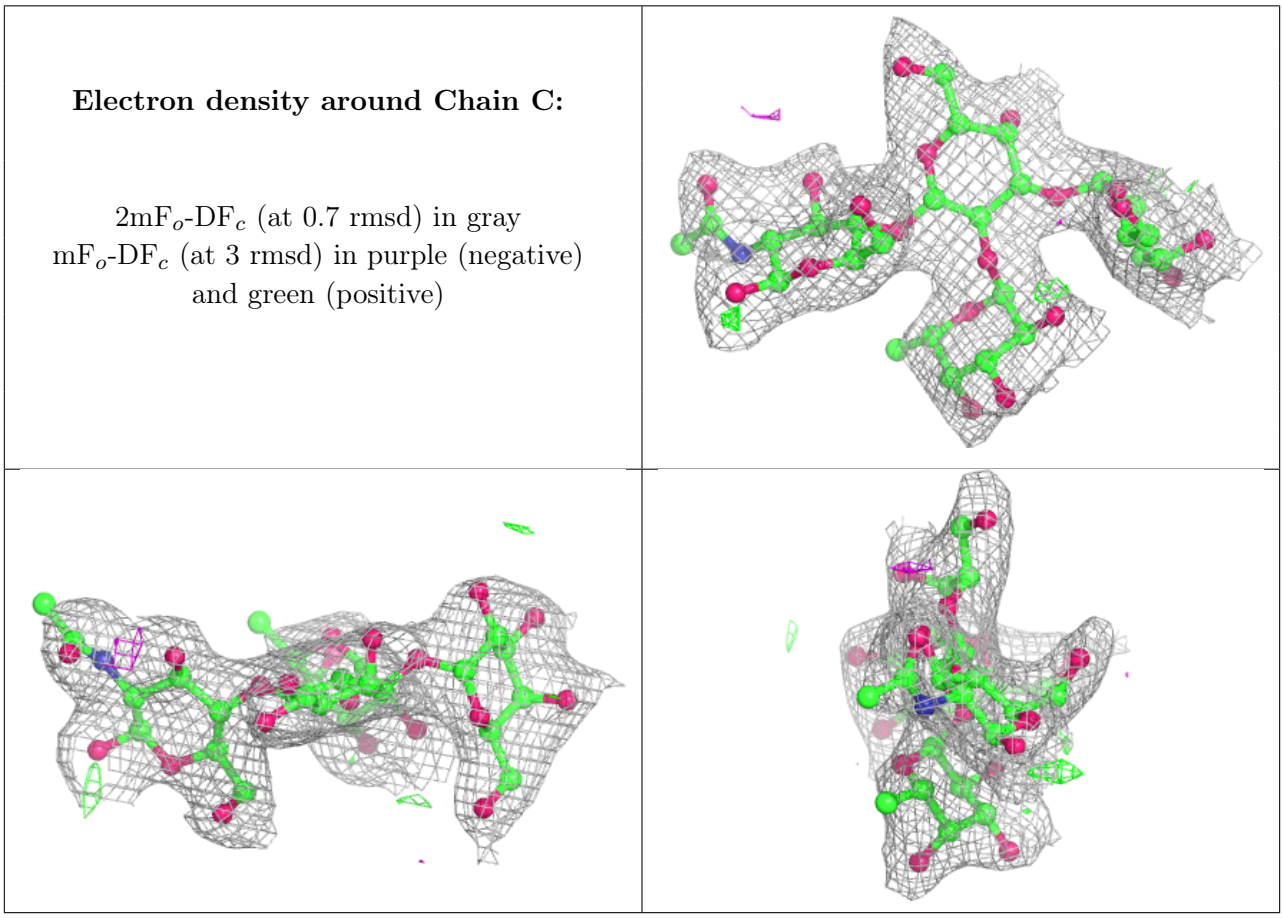
Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	GAL	D	2	11/12	0.94	0.14	51,52,53,53	0
2	GAL	C	2	11/12	0.95	0.10	55,56,56,57	0
2	GLA	D	4	11/12	0.95	0.10	50,50,51,51	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.

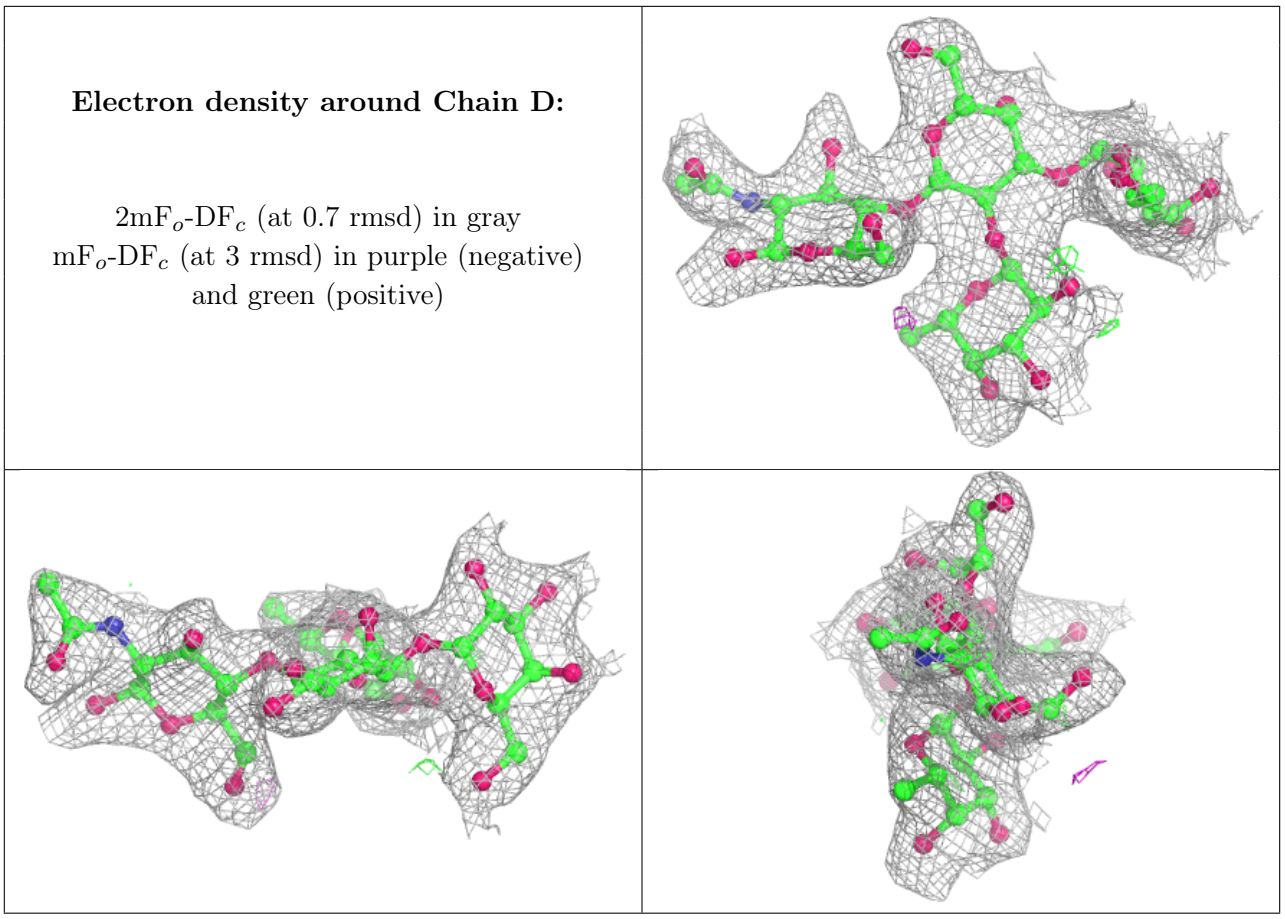
Electron density around Chain C:

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



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

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
4	NA	A	506	1/1	0.89	0.20	43,43,43,43	0
3	CA	A	505	1/1	0.96	0.05	44,44,44,44	0
4	NA	B	505	1/1	0.97	0.19	34,34,34,34	0
3	CA	B	506	1/1	0.98	0.14	53,53,53,53	0

6.5 Other polymers

There are no such residues in this entry.