



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 9, 2020 – 11:11 AM BST

PDB ID : 2YA1
Title : Product complex of a multi-modular glycogen-degrading pneumococcal virulence factor SpuA
Authors : Lammerts van Bueren, A.; Ficko-Blean, E.; Pluinage, B.; Hehemann, J.H.; Higgins, M.A.; Deng, L.; Ogunniyi, A.D.; Stroeder, U.H.; Warry, N.E.; Burke, R.D.; Czjzek, M.; Paton, J.C.; Vocadlo, D.J.; Boraston, A.B.
Deposited on : 2011-02-17
Resolution : 2.25 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

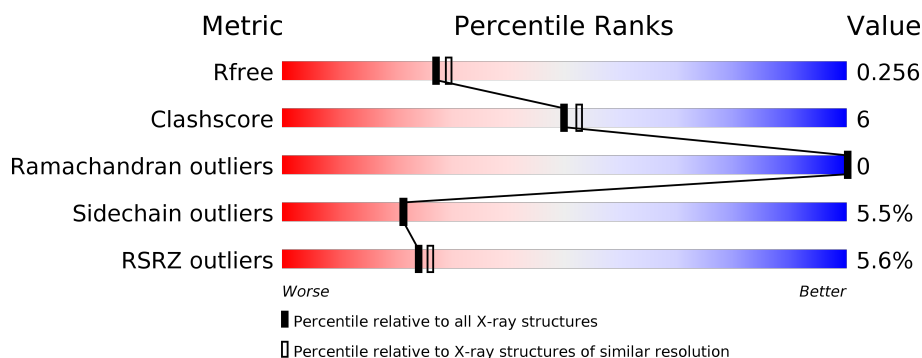
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.25 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	1377 (2.26-2.26)
Clashscore	141614	1487 (2.26-2.26)
Ramachandran outliers	138981	1449 (2.26-2.26)
Sidechain outliers	138945	1450 (2.26-2.26)
RSRZ outliers	127900	1356 (2.26-2.26)

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 $\leq 5\%$. 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	1014	<div> <div>5%</div> <div> <div></div> <div>82%</div> <div>12%</div> <div>• •</div> </div> </div>
2	B	4	<div> <div>25%</div> <div>75%</div> </div>
3	C	4	<div> <div>25%</div> <div>75%</div> </div>
4	D	3	<div> <div>33%</div> <div>67%</div> </div>

2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 8725 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 PUTATIVE ALKALINE AMYLOPULLULANASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	981	7880	4997	1338	1531	14	36	2	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	130	ALA	-	expression tag	UNP Q97SQ7
A	131	HIS	-	expression tag	UNP Q97SQ7
A	132	MET	-	expression tag	UNP Q97SQ7
A	133	ALA	-	expression tag	UNP Q97SQ7
A	134	SER	-	expression tag	UNP Q97SQ7
A	954	ASN	ASP	conflict	UNP Q97SQ7
A	958	ARG	LYS	conflict	UNP Q97SQ7

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



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

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



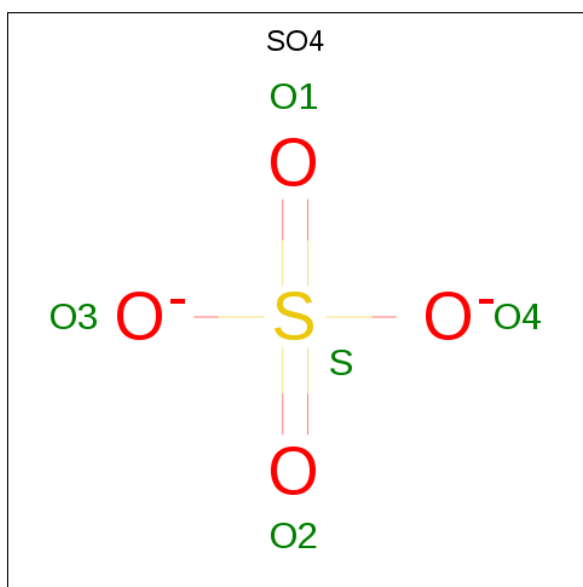
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
3	C	4	Total	C	O	0	0	0
			45	24	21			

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



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

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	O	S	0	0
			5	4	1		

- Molecule 6 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	5	Total	Na	0	0
			5	5		

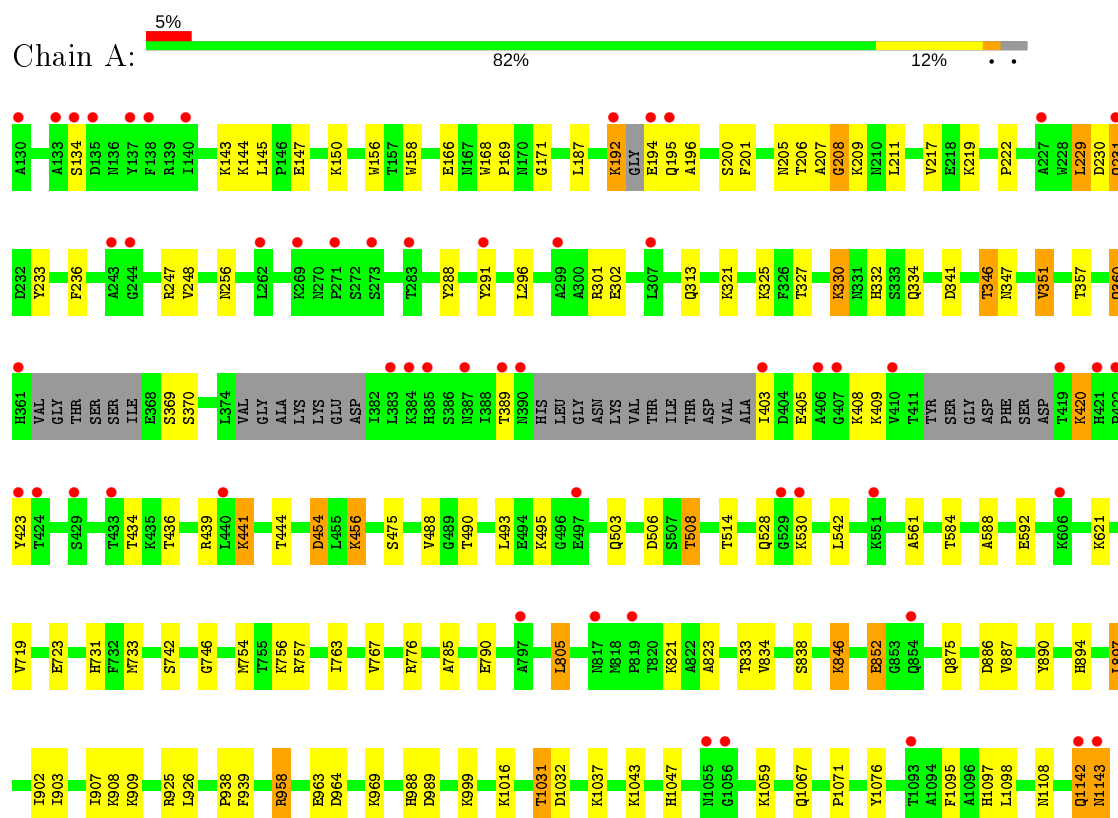
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	711	Total 711	O 711	0	0

3 Residue-property plots

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: PUTATIVE ALKALINE AMYLOPULLULANASE



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



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





- Molecule 4: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose

Chain D:  33% 67%



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	80.14Å 86.43Å 193.22Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.25 19.85 – 2.25	Depositor EDS
% Data completeness (in resolution range)	92.5 (20.00-2.25) 92.5 (19.85-2.25)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.59 (at 2.26Å)	Xtriage
Refinement program	REFMAC 5.5.0072	Depositor
R, R_{free}	0.205 , 0.265 0.205 , 0.256	Depositor DCC
R_{free} test set	3025 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å ²)	36.3	Xtriage
Anisotropy	0.069	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 50.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	8725	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 3.54% 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, NA, BGC, SO4

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	0.68	3/8075 (0.0%)	0.68	7/10939 (0.1%)

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

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	423	TYR	CB-CG	-29.55	1.07	1.51
1	A	409	LYS	CB-CG	-7.05	1.33	1.52
1	A	420	LYS	CA-CB	-5.56	1.41	1.53

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	423	TYR	CA-CB-CG	11.29	134.85	113.40
1	A	925	ARG	NE-CZ-NH2	-6.19	117.20	120.30
1	A	805	LEU	CA-CB-CG	5.97	129.03	115.30
1	A	360	GLN	CA-CB-CG	5.95	126.48	113.40
1	A	958	ARG	NE-CZ-NH1	5.62	123.11	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	207	ALA	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	7880	0	7607	91	0
2	B	45	0	39	0	0
3	C	45	0	39	0	0
4	D	34	0	30	0	0
5	A	5	0	0	0	0
6	A	5	0	0	0	0
7	A	711	0	0	36	1
All	All	8725	0	7715	91	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 6.

The worst 5 of 91 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:875:GLN:HG2	7:A:3502:HOH:O	1.58	1.01
1:A:194:GLU:HB3	1:A:195:GLN:HA	1.58	0.85
1:A:909:LYS:HE2	7:A:3521:HOH:O	1.75	0.85
1:A:327:THR:HG22	7:A:3151:HOH:O	1.78	0.82
1:A:302:GLU:HG2	7:A:3149:HOH:O	1.80	0.80

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)
7:A:3114:HOH:O	7:A:3262:HOH:O[3_555]	2.09	0.11

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	971/1014 (96%)	924 (95%)	47 (5%)	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	844/869 (97%)	798 (94%)	46 (6%)	21	21

5 of 46 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	405	GLU
1	A	444	THR
1	A	999	LYS
1	A	408	LYS
1	A	434	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	844	ASN
1	A	905	GLN
1	A	1029	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	728	ASN
1	A	1067	GLN

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 ⓘ

11 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	0.79	0	17,17,17	1.28	2 (11%)
2	GLC	B	2	2	11,11,12	0.79	0	15,15,17	1.03	0
2	GLC	B	3	2	11,11,12	0.83	0	15,15,17	1.23	2 (13%)
2	GLC	B	4	2	11,11,12	0.61	0	15,15,17	1.43	2 (13%)
3	GLC	C	1	3	12,12,12	0.66	0	17,17,17	0.74	0
3	GLC	C	2	3	11,11,12	0.67	0	15,15,17	1.09	1 (6%)
3	GLC	C	3	3	11,11,12	0.64	0	15,15,17	1.04	1 (6%)
3	GLC	C	4	3	11,11,12	0.87	0	15,15,17	1.05	1 (6%)
4	GLC	D	1	4	12,12,12	0.62	0	17,17,17	0.90	0
4	GLC	D	2	4	11,11,12	0.62	0	15,15,17	0.88	1 (6%)
4	GLC	D	3	4	11,11,12	0.63	0	15,15,17	1.44	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	-	0/2/22/22	0/1/1/1
2	GLC	B	2	2	-	0/2/19/22	0/1/1/1
2	GLC	B	3	2	-	0/2/19/22	0/1/1/1
2	GLC	B	4	2	-	0/2/19/22	0/1/1/1
3	GLC	C	1	3	-	2/2/22/22	0/1/1/1
3	GLC	C	2	3	-	0/2/19/22	0/1/1/1
3	GLC	C	3	3	-	0/2/19/22	0/1/1/1
3	GLC	C	4	3	-	0/2/19/22	0/1/1/1
4	GLC	D	1	4	-	0/2/22/22	0/1/1/1
4	GLC	D	2	4	-	0/2/19/22	0/1/1/1
4	GLC	D	3	4	-	2/2/19/22	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(°)	Ideal(°)
3	C	2	GLC	C1-O5-C5	3.37	116.75	112.19
2	B	3	GLC	C1-O5-C5	3.29	116.66	112.19
4	D	3	GLC	O5-C1-C2	-2.97	106.18	110.77
4	D	3	GLC	O5-C5-C6	2.89	111.73	107.20
2	B	1	BGC	O5-C1-C2	-2.73	105.42	110.28

There are no chirality outliers.

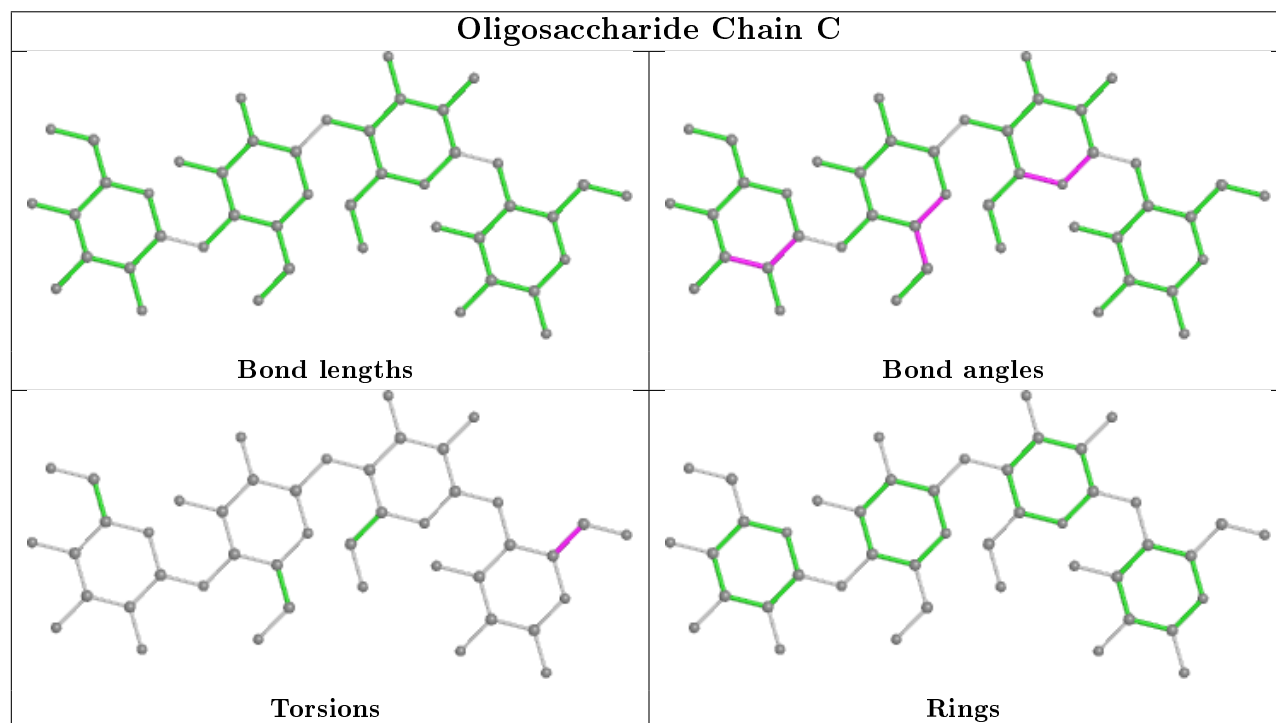
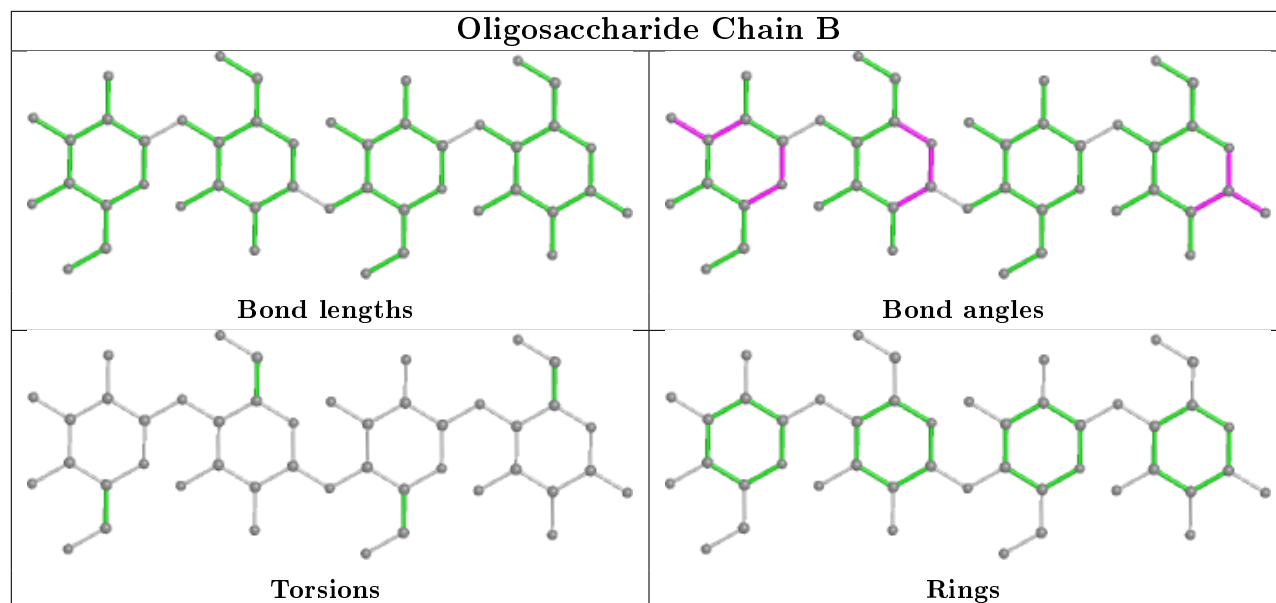
All (4) torsion outliers are listed below:

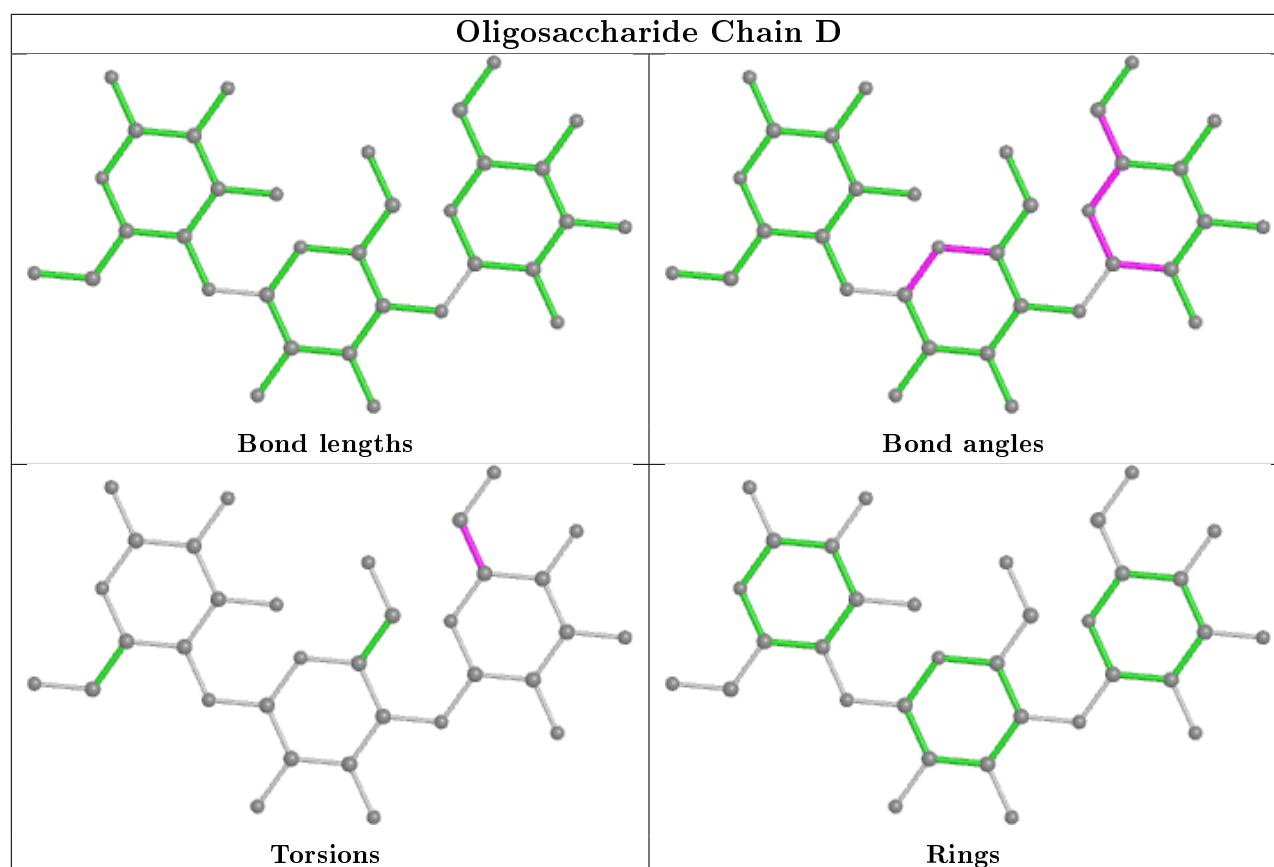
Mol	Chain	Res	Type	Atoms
4	D	3	GLC	O5-C5-C6-O6
4	D	3	GLC	C4-C5-C6-O6
3	C	1	GLC	C4-C5-C6-O6
3	C	1	GLC	O5-C5-C6-O6

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.





5.6 Ligand geometry [i](#)

Of 6 ligands modelled in this entry, 5 are monoatomic - leaving 1 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	$\# Z > 2$
5	SO4	A	2152	-	4,4,4	0.14	0	6,6,6	0.46	0

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 [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](#)

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, 95th 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	981/1014 (96%)	0.10	55 (5%) 24 26	18, 32, 60, 92	14 (1%)

The worst 5 of 55 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	421	HIS	5.5
1	A	361	HIS	5.4
1	A	429	SER	5.2
1	A	195	GLN	5.0
1	A	1143	ASN	4.9

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

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, 95th 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.

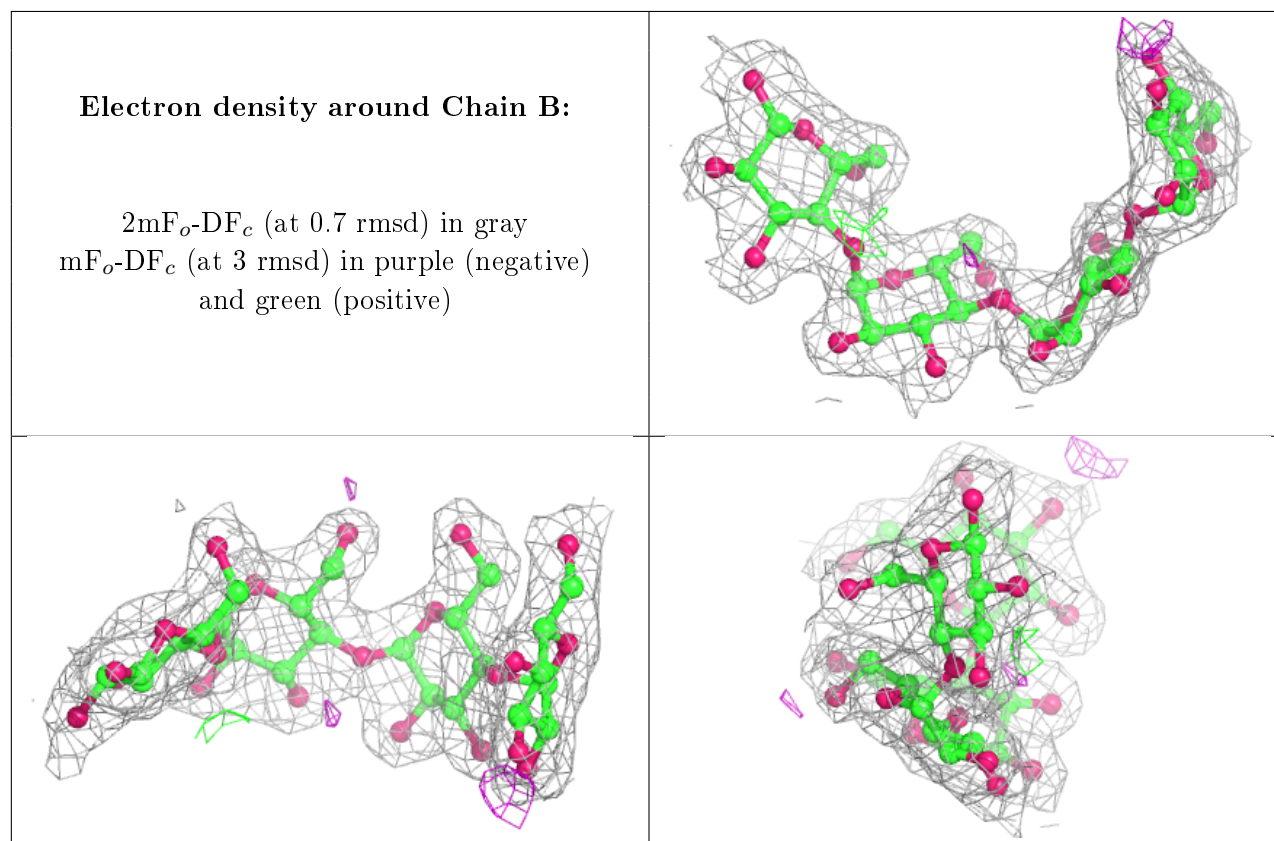
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
4	GLC	D	3	11/12	0.85	0.32	51,56,58,58	0
3	GLC	C	1	12/12	0.91	0.26	32,39,41,43	0
4	GLC	D	2	11/12	0.93	0.12	35,38,41,47	0
3	GLC	C	4	11/12	0.94	0.09	26,31,35,39	0
4	GLC	D	1	12/12	0.94	0.10	33,36,38,38	0
2	GLC	B	3	11/12	0.94	0.12	26,26,27,30	0
3	GLC	C	3	11/12	0.96	0.09	28,29,30,31	0

Continued on next page...

Continued from previous page...

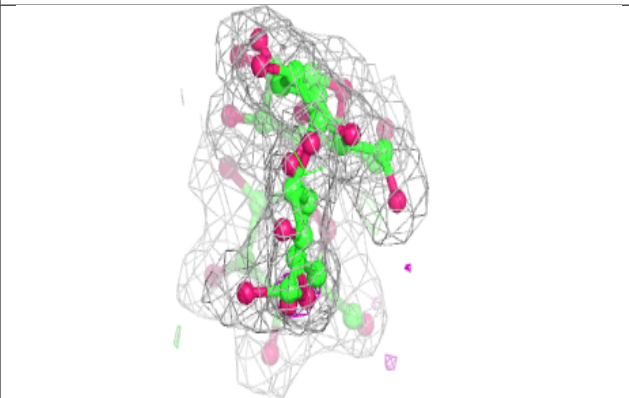
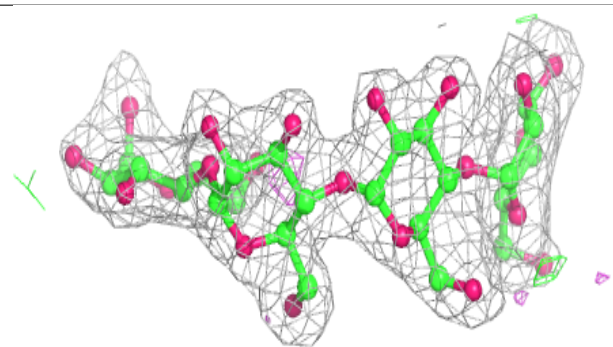
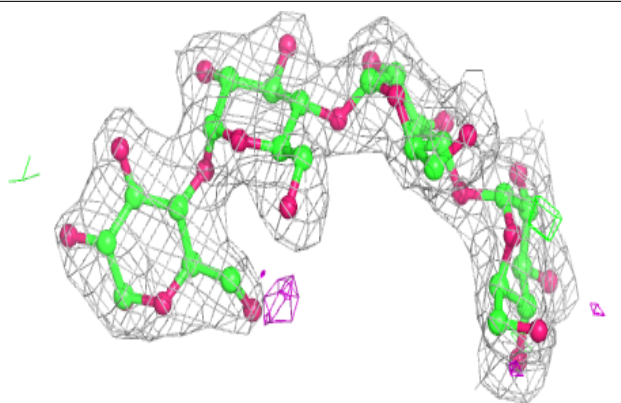
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
3	GLC	C	2	11/12	0.97	0.09	25,28,31,31	0
2	GLC	B	2	11/12	0.97	0.10	25,28,30,33	0
2	BGC	B	1	12/12	0.97	0.10	28,29,30,31	0
2	GLC	B	4	11/12	0.97	0.09	26,28,31,31	0

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.

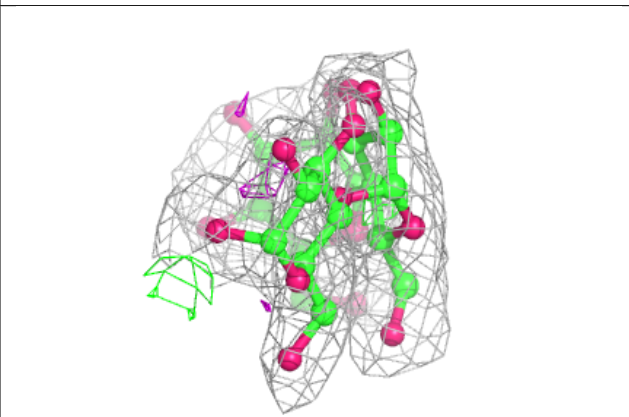
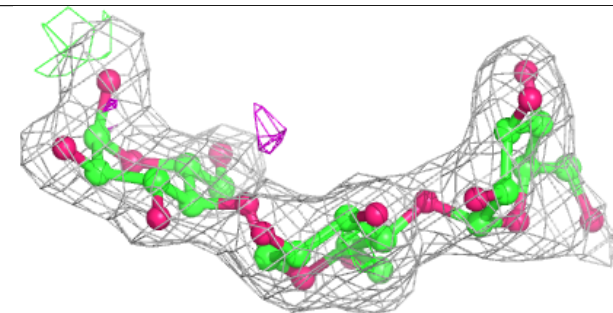
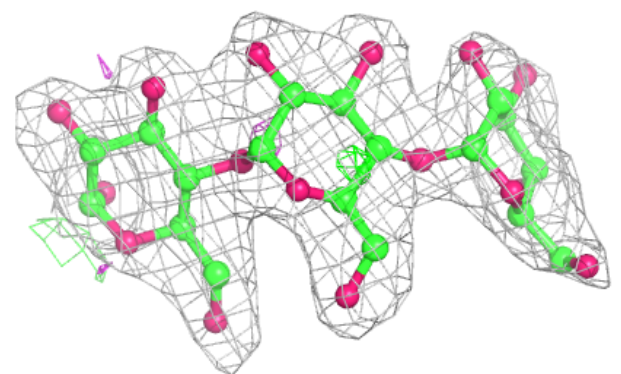


Electron density around Chain C:

$2mF_o-DF_c$ (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 [i](#)

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, 95th 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(\AA^2)	Q<0.9
6	NA	A	2156	1/1	0.81	0.15	49,49,49,49	0
6	NA	A	2157	1/1	0.91	0.12	54,54,54,54	0
6	NA	A	2154	1/1	0.94	0.25	33,33,33,33	0
6	NA	A	2155	1/1	0.95	0.26	35,35,35,35	0
5	SO4	A	2152	5/5	0.98	0.11	36,39,39,40	0
6	NA	A	2153	1/1	0.99	0.21	7,7,7,7	0

6.5 Other polymers [i](#)

There are no such residues in this entry.