Full wwPDB X-ray Structure Validation Report (i)

Sep 18, 2023 – 05:36 PM EDT

PDB ID	:	5C0R
Title	:	Crystal Structure of a Generation 3 Influenza Hemagglutinin Stabilized Stem Complexed with the Broadly Neutralizing Antibody C179
Authors	:	Boyington, J.C.; kwong, P.D.; Nabel, G.J.; Mascola, J.R.
Deposited on	:	2015-06-12
Resolution	:	3.19 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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:

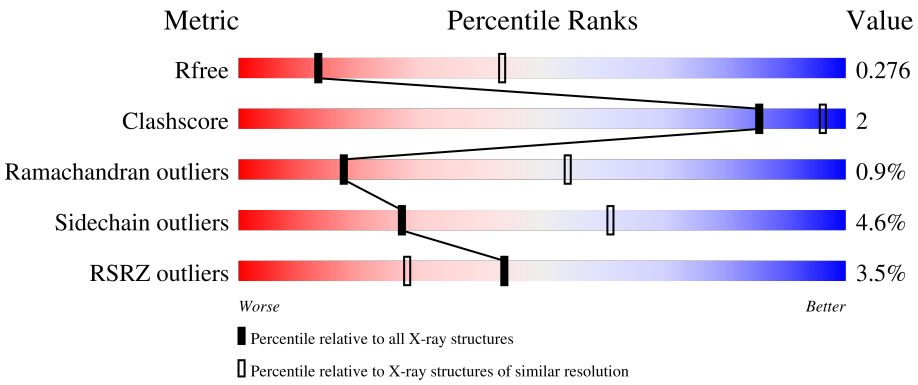
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 3.19 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	1467 (3.20-3.16)
Clashscore	141614	1599 (3.20-3.16)
Ramachandran outliers	138981	1574 (3.20-3.16)
Sidechain outliers	138945	1573 (3.20-3.16)
RSRZ outliers	127900	1423 (3.20-3.16)

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	305	3%, 80%, 7% • 11%
2	L	214	93%, 6% •
3	H	233	6%, 87%, 7% • 5%
4	B	2	50%, 50%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 10886 atoms, of which 5338 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 Hemagglutinin, Envelope glycoprotein, Fibritin fusion protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	270	4238	1355	2071	376	427	9	0	0	0

There are 35 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	33	GLY	-	linker	UNP Q6WG00
A	34	TRP	-	linker	UNP Q6WG00
A	35	GLY	-	linker	UNP Q6WG00
A	47	GLN	SER	conflict	UNP Q6WG00
A	47A	ARG	ILE	conflict	UNP Q6WG00
A	47B	GLU	GLN	conflict	UNP Q6WG00
A	47C	THR	SER	conflict	UNP Q6WG00
A	108	GLY	-	linker	UNP Q6WG00
A	109	GLY	-	linker	UNP Q6WG00
A	123	ILE	LEU	conflict	UNP P04578
A	125	TYR	HIS	conflict	UNP P04578
A	137	ASN	-	linker	UNP P04578
A	138	GLY	-	linker	UNP P04578
A	139	THR	-	linker	UNP P04578
A	140	GLY	-	linker	UNP P04578
A	141	GLY	-	linker	UNP P04578
A	142	GLY	-	linker	UNP P04578
A	257	PRO	-	linker	UNP Q6WG00
A	258	GLY	-	linker	UNP Q6WG00
A	259	SER	-	linker	UNP Q6WG00
A	288	GLY	-	expression tag	UNP D9IEJ2
A	289	ARG	-	expression tag	UNP D9IEJ2
A	290	LEU	-	expression tag	UNP D9IEJ2
A	291	VAL	-	expression tag	UNP D9IEJ2
A	292	PRO	-	expression tag	UNP D9IEJ2
A	293	ARG	-	expression tag	UNP D9IEJ2
A	294	GLY	-	expression tag	UNP D9IEJ2

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	295	SER	-	expression tag	UNP D9IEJ2
A	296	GLY	-	expression tag	UNP D9IEJ2
A	297	HIS	-	expression tag	UNP D9IEJ2
A	298	HIS	-	expression tag	UNP D9IEJ2
A	299	HIS	-	expression tag	UNP D9IEJ2
A	300	HIS	-	expression tag	UNP D9IEJ2
A	301	HIS	-	expression tag	UNP D9IEJ2
A	302	HIS	-	expression tag	UNP D9IEJ2

- Molecule 2 is a protein called C179 Fab light chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
2	L	214	3231	1033	1582	272	337	7	0	0	0

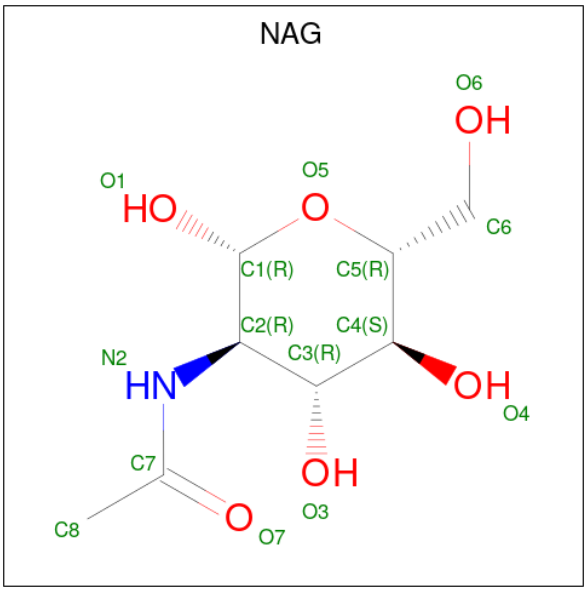
- Molecule 3 is a protein called C179 Fab heavy chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
3	H	222	3327	1063	1645	279	332	8	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	H	N	O			
4	B	2	55	16	27	2	10	0	0	0

- Molecule 5 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	H	N	O		
5	A	1	27	8	13	1	5	0	0

- Molecule 6 is water.

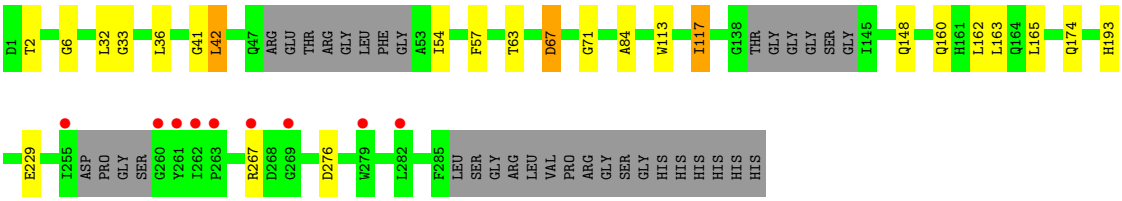
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	A	6	6	6	0	0
			Total	O		
6	L	1	1	1	0	0
			Total	O		
6	H	1	1	1	0	0

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.

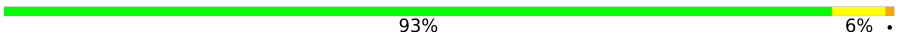
• Molecule 1: Hemagglutinin, Envelope glycoprotein, Fibritin fusion protein

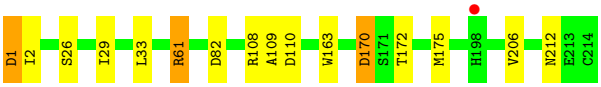
Chain A:



• Molecule 2: C179 Fab light chain

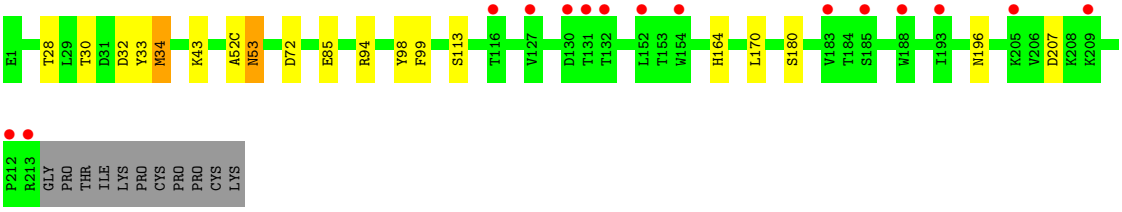
Chain L:





• Molecule 3: C179 Fab heavy chain

Chain H:



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

Chain B:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α, β, γ	112.04Å 112.04Å 205.12Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	32.34 – 3.19 32.34 – 3.19	Depositor EDS
% Data completeness (in resolution range)	98.2 (32.34-3.19) 98.3 (32.34-3.19)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.47 (at 3.18Å)	Xtriage
Refinement program	PHENIX 1.8.4_1496	Depositor
R, R_{free}	0.199 , 0.251 0.217 , 0.276	Depositor DCC
R_{free} test set	783 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å^2)	94.7	Xtriage
Anisotropy	0.337	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 56.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.027 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	10886	wwPDB-VP
Average B, all atoms (Å^2)	92.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.46% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: 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	0.23	0/2207	0.38	0/2984
2	L	0.22	0/1688	0.39	0/2292
3	H	0.22	0/1725	0.39	0/2356
All	All	0.22	0/5620	0.39	0/7632

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 (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	2167	2071	2069	11	0
2	L	1649	1582	1581	6	0
3	H	1682	1645	1643	6	0
4	B	28	27	25	3	0
5	A	14	13	13	0	0
6	A	6	0	0	2	0
6	H	1	0	0	0	0
6	L	1	0	0	0	0
All	All	5548	5338	5331	24	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 2.

All (24) 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:H:30:THR:O	3:H:94:ARG:NH2	2.19	0.76
2:L:108:ARG:NH1	2:L:109:ALA:O	2.30	0.64
2:L:1:ASP:N	2:L:1:ASP:OD1	2.33	0.62
3:H:196:ASN:ND2	3:H:207:ASP:OD1	2.34	0.60
1:A:33:GLY:N	6:A:501:HOH:O	2.34	0.60
2:L:61:ARG:NH1	2:L:82:ASP:OD2	2.36	0.58
2:L:163:TRP:NE1	2:L:175:MET:SD	2.75	0.57
2:L:170:ASP:N	2:L:170:ASP:OD1	2.36	0.57
1:A:229:GLU:OE2	4:B:2:NAG:N2	2.43	0.52
3:H:52(C):ALA:O	3:H:53:ASN:ND2	2.47	0.48
2:L:170:ASP:OD2	2:L:172:THR:OG1	2.32	0.47
3:H:85:GLU:OE2	3:H:85:GLU:N	2.48	0.46
1:A:32:LEU:HB2	1:A:36:LEU:HB2	1.98	0.46
3:H:32:ASP:O	3:H:34:MET:N	2.48	0.45
4:B:2:NAG:O6	4:B:2:NAG:O3	2.32	0.45
1:A:229:GLU:OE2	4:B:2:NAG:O4	2.30	0.44
1:A:67:ASP:HB2	1:A:84:ALA:HB3	2.00	0.43
1:A:42:LEU:HD23	1:A:42:LEU:N	2.34	0.43
3:H:164:HIS:O	3:H:180:SER:N	2.47	0.43
1:A:36:LEU:N	6:A:501:HOH:O	2.48	0.42
1:A:41:GLY:O	1:A:193:HIS:NE2	2.52	0.42
1:A:71:GLY:HA3	1:A:84:ALA:HA	2.01	0.41
1:A:6:GLY:HA2	1:A:57:PHE:HB3	2.02	0.41
1:A:113:TRP:CE2	1:A:117:ILE:HD12	2.56	0.40

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	262/305 (86%)	245 (94%)	15 (6%)	2 (1%)	19	56
2	L	212/214 (99%)	195 (92%)	16 (8%)	1 (0%)	29	66
3	H	220/233 (94%)	195 (89%)	22 (10%)	3 (1%)	11	44
All	All	694/752 (92%)	635 (92%)	53 (8%)	6 (1%)	17	54

All (6) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	H	113	SER
3	H	33	TYR
1	A	54	ILE
1	A	276	ASP
3	H	99	PHE
2	L	212	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	235/260 (90%)	223 (95%)	12 (5%)	24	57
2	L	187/187 (100%)	178 (95%)	9 (5%)	25	60
3	H	190/200 (95%)	183 (96%)	7 (4%)	34	67
All	All	612/647 (95%)	584 (95%)	28 (5%)	27	61

All (28) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	2	THR
1	A	42	LEU
1	A	63	THR
1	A	67	ASP
1	A	117	ILE
1	A	148	GLN
1	A	160	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	162	LEU
1	A	163	LEU
1	A	165	LEU
1	A	174	GLN
1	A	267	ARG
2	L	1	ASP
2	L	2	ILE
2	L	26	SER
2	L	29	ILE
2	L	33	LEU
2	L	61	ARG
2	L	110	ASP
2	L	170	ASP
2	L	206	VAL
3	H	28	THR
3	H	34	MET
3	H	43	LYS
3	H	53	ASN
3	H	72	ASP
3	H	98	TYR
3	H	170	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

2 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	$\#\|Z\| > 2$
4	NAG	B	1	4,1	14,14,15	0.42	0	17,19,21	0.42	0
4	NAG	B	2	4	14,14,15	0.28	0	17,19,21	0.39	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	1	4,1	-	2/6/23/26	0/1/1/1
4	NAG	B	2	4	-	3/6/23/26	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	1	NAG	O5-C5-C6-O6
4	B	1	NAG	C4-C5-C6-O6
4	B	2	NAG	C4-C5-C6-O6
4	B	2	NAG	O5-C5-C6-O6
4	B	2	NAG	C1-C2-N2-C7

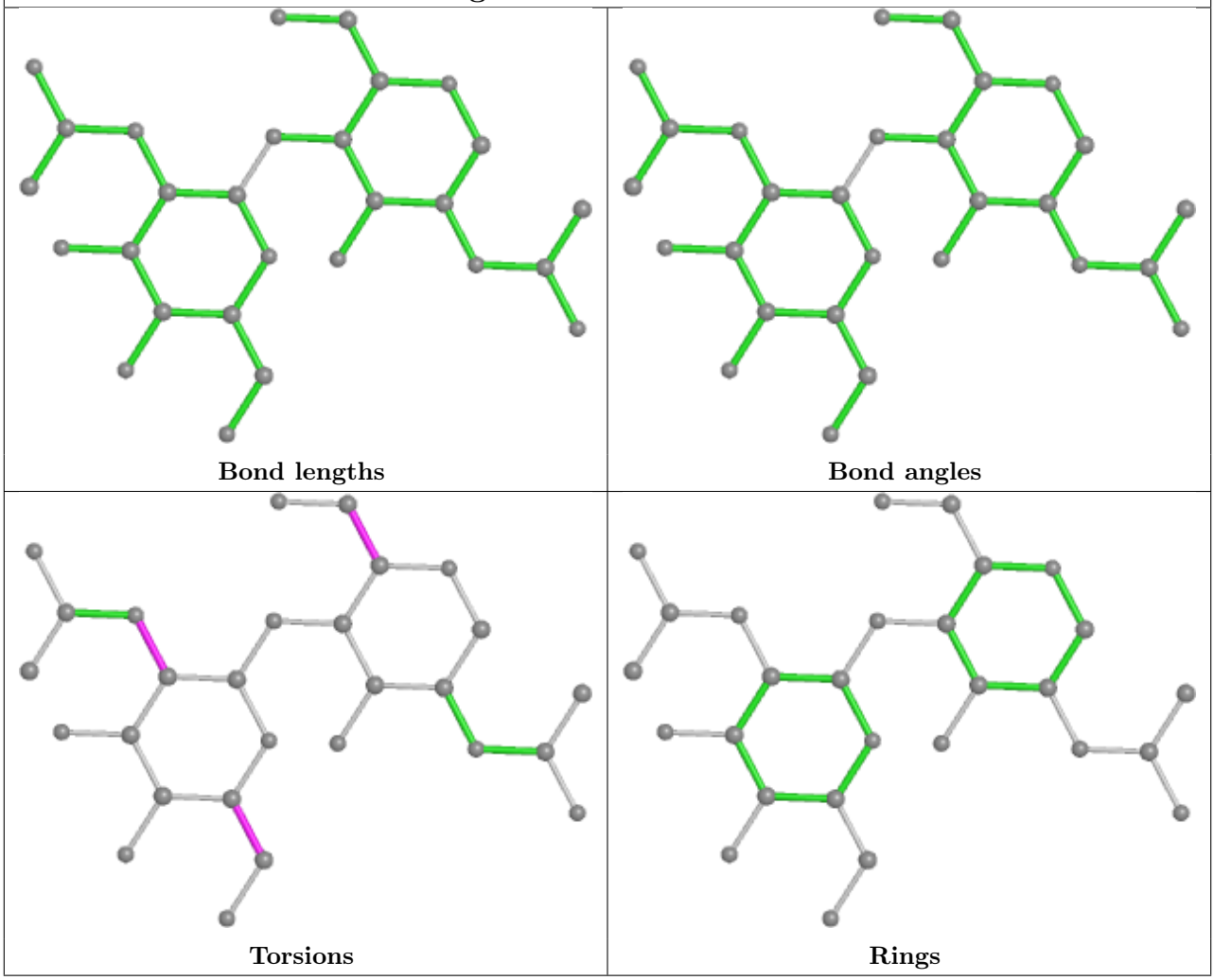
There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	2	NAG	3	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.

Oligosaccharide Chain B



5.6 Ligand geometry (i)

1 ligand is 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	$\#\|Z\| > 2$
5	NAG	A	401	1	14,14,15	0.20	0	17,19,21	0.42	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
5	NAG	A	401	1	-	2/6/23/26	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	401	NAG	O5-C5-C6-O6
5	A	401	NAG	C4-C5-C6-O6

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	270/305 (88%)	0.16	9 (3%)	46	30	60, 95, 197, 323	0
2	L	214/214 (100%)	0.21	1 (0%)	91	86	73, 133, 190, 223	0
3	H	222/233 (95%)	0.42	15 (6%)	17	9	70, 127, 240, 303	0
All	All	706/752 (93%)	0.26	25 (3%)	44	28	60, 120, 213, 323	0

All (25) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	H	212	PRO	6.8
1	A	267	ARG	5.4
1	A	255	ILE	5.0
1	A	263	PRO	4.9
3	H	132	THR	4.3
3	H	131	THR	4.2
1	A	282	LEU	4.1
3	H	130	ASP	4.1
1	A	262	ILE	3.7
3	H	209	LYS	3.6
1	A	261	TYR	3.5
3	H	205	LYS	3.3
1	A	260	GLY	3.0
3	H	193	ILE	2.8
1	A	269	GLY	2.7
2	L	198	HIS	2.7
3	H	183	VAL	2.6
3	H	154	TRP	2.6
3	H	188	TRP	2.5
3	H	152	LEU	2.4
3	H	116	THR	2.3
3	H	127	VAL	2.3
3	H	213	ARG	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	279	TRP	2.0
3	H	185	SER	2.0

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)

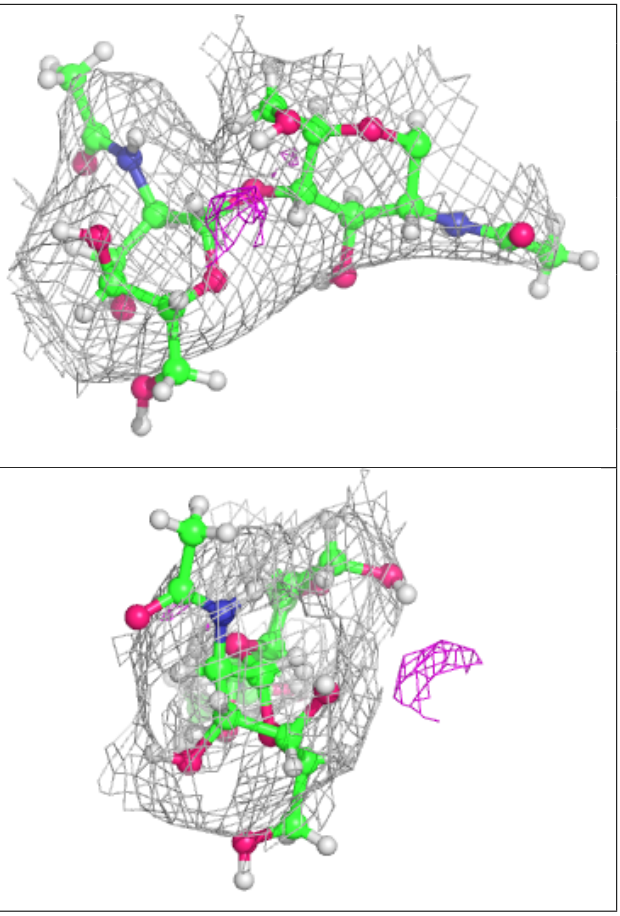
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists 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	NAG	B	2	14/15	0.83	0.33	47,65,208,209	0
4	NAG	B	1	14/15	0.86	0.23	50,64,146,151	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 (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
5	NAG	A	401	14/15	0.88	0.36	51,63,154,157	0

6.5 Other polymers (i)

There are no such residues in this entry.