wwPDB X-ray Structure Validation Summary Report (i)

Oct 4, 2021 – 10:04 am BST

PDB ID	:	7A9D
Title	:	Crystal structure of H12 Haemagglutinin
Authors	:	Xiong, X.; Walker, P.; Zhang, J.; Gamblin, S.; Skehel, J.J.
Deposited on	:	2020-09-01
Resolution	:	2.70 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 following versions of software and data (see references (1)) were used in the production of this report:

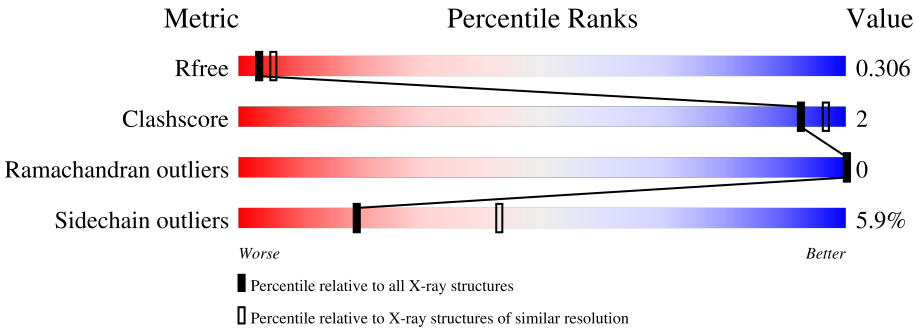
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.70 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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)

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	323	93% 6% •
1	C	323	91% 7% •
2	B	167	85% 7% • 7%
2	D	167	92% 7% •
3	E	2	50% 50%
3	F	2	50% 50%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 7816 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 Hemagglutinin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	322	Total	C	N	O	S	0	0	0
			2544	1600	439	490	15			
1	C	322	Total	C	N	O	S	0	0	0
			2544	1600	439	490	15			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	ASP	-	expression tag	UNP C6KJK3
A	0	PRO	-	expression tag	UNP C6KJK3
C	-1	ASP	-	expression tag	UNP C6KJK3
C	0	PRO	-	expression tag	UNP C6KJK3

- Molecule 2 is a protein called Hemagglutinin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	155	Total	C	N	O	S	0	0	0
			1249	771	226	246	6			
2	D	167	Total	C	N	O	S	0	0	0
			1339	824	242	266	7			

- Molecule 3 is an oligosaccharide called 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose.

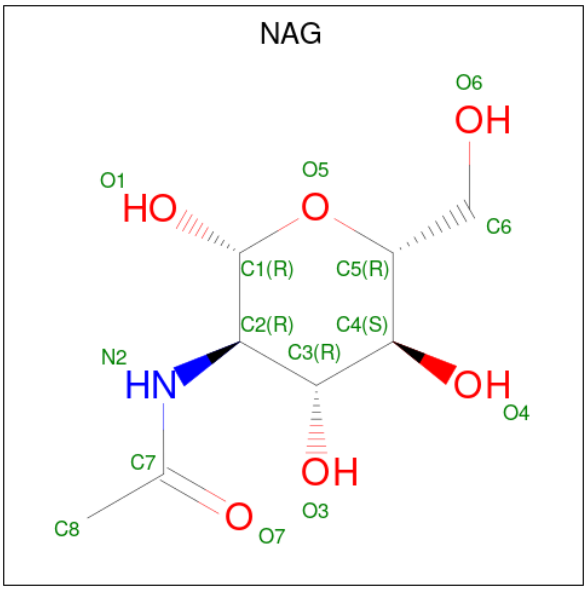
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	E	2	Total	C	N	O	0	0	0
			28	16	2	10			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
3	F	2	Total C N O 28 16 2 10	0	0	0

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



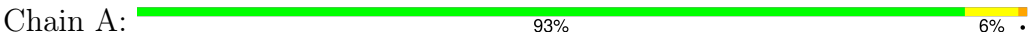
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C N O 14 8 1 5	0	0
4	A	1	Total C N O 14 8 1 5	0	0
4	C	1	Total C N O 14 8 1 5	0	0
4	C	1	Total C N O 14 8 1 5	0	0
4	C	1	Total C N O 14 8 1 5	0	0
4	D	1	Total C N O 14 8 1 5	0	0

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: Hemagglutinin

Chain A:



• Molecule 1: Hemagglutinin

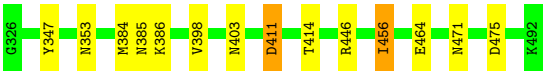
Chain C:

• Molecule 2: Hemagglutinin

Chain B:

• Molecule 2: Hemagglutinin

Chain D:



• Molecule 3: 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain E:

• Molecule 3: 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain F:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, α, β, γ	104.51Å 104.51Å 694.83Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	89.75 – 2.70 80.27 – 2.70	Depositor EDS
% Data completeness (in resolution range)	96.7 (89.75-2.70) 96.7 (80.27-2.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.66 (at 2.69Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
R, R_{free}	0.268 , 0.303 0.272 , 0.306	Depositor DCC
R_{free} test set	1980 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å^2)	61.1	Xtriage
Anisotropy	0.431	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	7816	wwPDB-VP
Average B, all atoms (Å^2)	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.64% 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: 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	0.66	0/2602	0.71	0/3530
1	C	0.64	0/2602	0.71	0/3530
2	B	0.67	0/1268	0.70	0/1710
2	D	0.65	0/1361	0.70	0/1836
All	All	0.65	0/7833	0.70	0/10606

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	2544	0	2503	5	0
1	C	2544	0	2502	9	0
2	B	1249	0	1175	4	0
2	D	1339	0	1259	5	0
3	E	28	0	25	0	0
3	F	28	0	25	0	0
4	A	28	0	26	0	0
4	C	42	0	39	0	0
4	D	14	0	13	0	0
All	All	7816	0	7567	23	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.

The worst 5 of 23 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:C:242:THR:HG21	1:C:246:LEU:HD22	1.74	0.70
2:B:353:ASN:HD21	2:B:471:ASN:ND2	1.94	0.65
1:C:291:GLN:HE21	1:C:293:THR:H	1.48	0.59
2:B:353:ASN:HD21	2:B:471:ASN:HD21	1.51	0.57
2:D:353:ASN:HD21	2:D:471:ASN:HD22	1.58	0.52

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	320/323 (99%)	308 (96%)	12 (4%)	0	100	100
1	C	320/323 (99%)	307 (96%)	13 (4%)	0	100	100
2	B	149/167 (89%)	143 (96%)	6 (4%)	0	100	100
2	D	165/167 (99%)	159 (96%)	6 (4%)	0	100	100
All	All	954/980 (97%)	917 (96%)	37 (4%)	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	291/292 (100%)	277 (95%)	14 (5%)	25	53
1	C	291/292 (100%)	273 (94%)	18 (6%)	18	40
2	B	133/143 (93%)	123 (92%)	10 (8%)	13	31
2	D	143/143 (100%)	134 (94%)	9 (6%)	18	40
All	All	858/870 (99%)	807 (94%)	51 (6%)	19	43

5 of 51 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	298	ILE
2	B	375	ASN
2	D	456	ILE
1	C	301	CYS
2	B	347	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	403	ASN
2	D	375	ASN
2	D	372	ASN
2	D	385	ASN
1	C	192	ASN

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)

4 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	$\#\|Z\| > 2$
3	NAG	E	1	3,1	14,14,15	0.49	0	17,19,21	1.58	3 (17%)
3	NAG	E	2	3	14,14,15	0.35	0	17,19,21	0.57	0
3	NAG	F	1	3,1	14,14,15	0.32	0	17,19,21	1.30	3 (17%)
3	NAG	F	2	3	14,14,15	0.30	0	17,19,21	0.67	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	NAG	E	1	3,1	-	0/6/23/26	0/1/1/1
3	NAG	E	2	3	-	0/6/23/26	0/1/1/1
3	NAG	F	1	3,1	-	1/6/23/26	0/1/1/1
3	NAG	F	2	3	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	E	1	NAG	C4-C3-C2	4.84	118.11	111.02
3	F	1	NAG	O5-C1-C2	-2.86	106.77	111.29
3	E	1	NAG	C3-C4-C5	2.81	115.25	110.24
3	F	1	NAG	C4-C3-C2	2.58	114.80	111.02
3	E	1	NAG	O4-C4-C3	-2.31	105.00	110.35

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	F	1	NAG	C4-C5-C6-O6

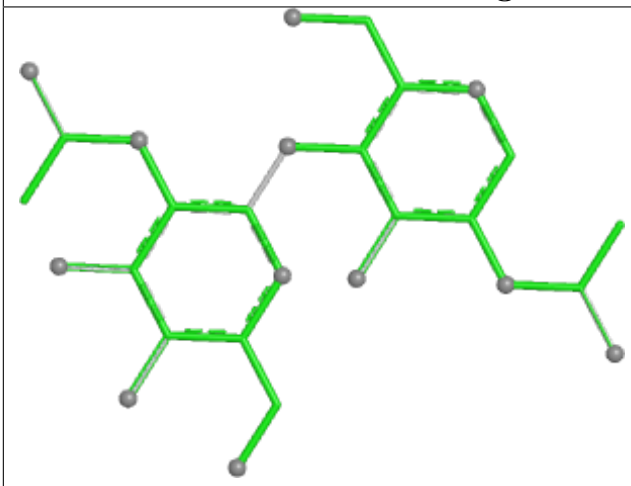
There are no ring outliers.

No monomer is involved in short contacts.

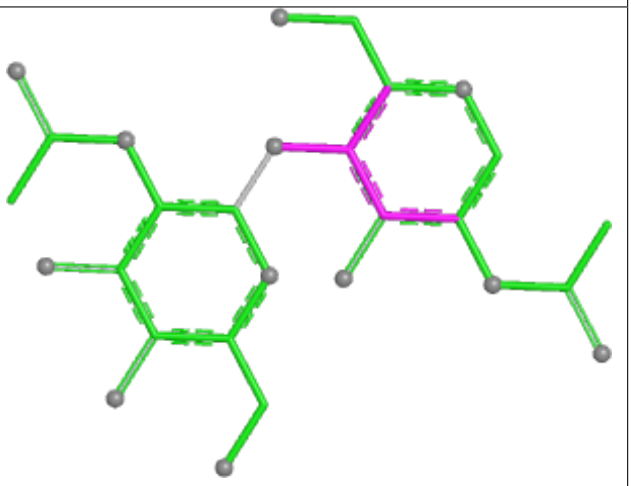
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths,

bond angles, torsion angles, and ring geometry for oligosaccharide.

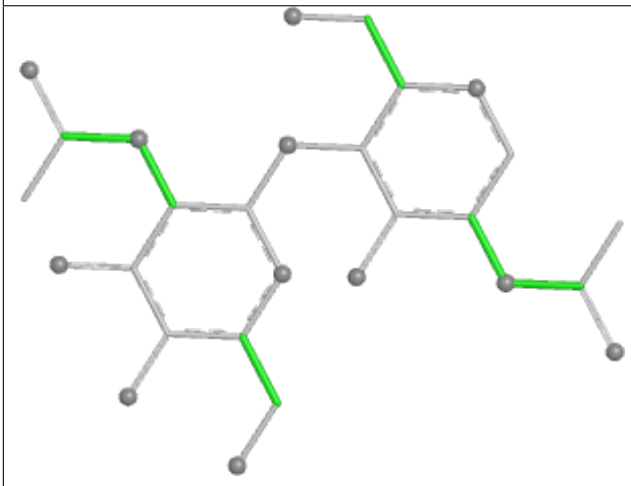
Oligosaccharide Chain E



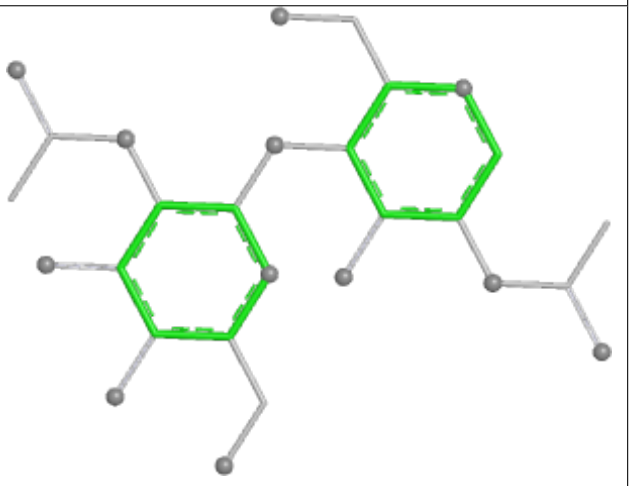
Bond lengths



Bond angles

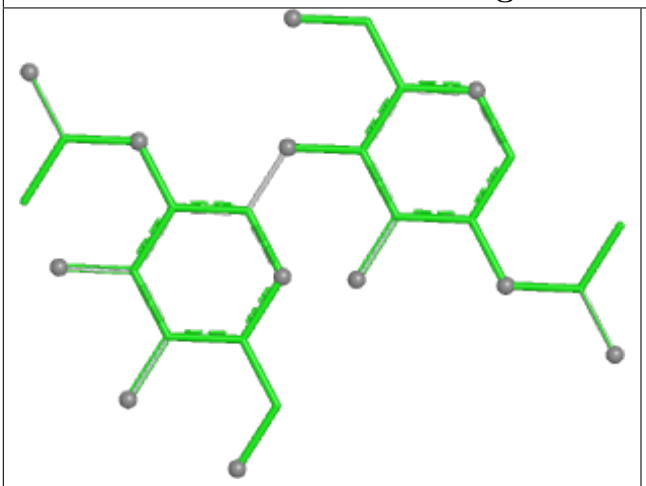


Torsions

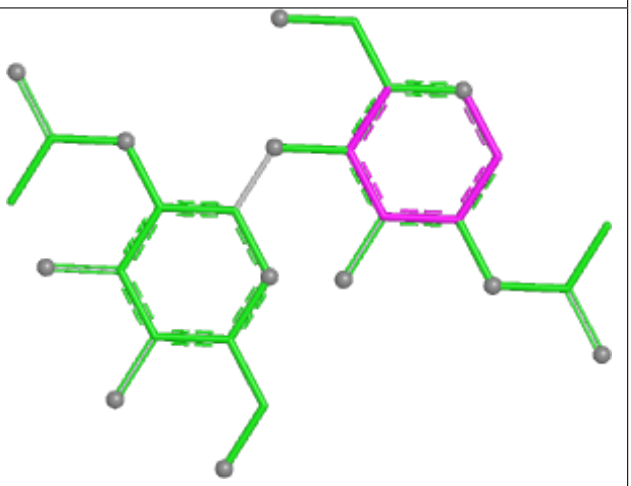


Rings

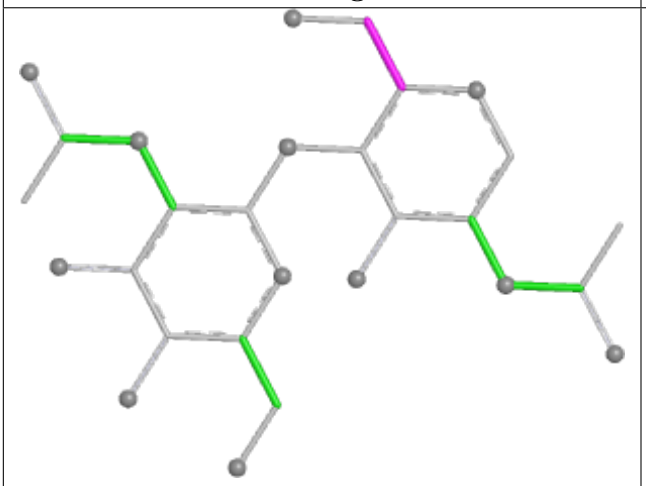
Oligosaccharide Chain F



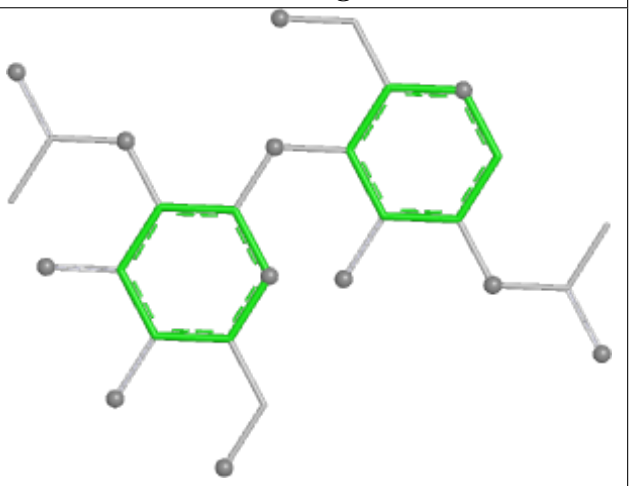
Bond lengths



Bond angles



Torsions



Rings

5.6 Ligand geometry

6 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	A	401	1	14,14,15	0.25	0	17,19,21	0.79	1 (5%)
4	NAG	D	501	2	14,14,15	0.40	0	17,19,21	0.79	0
4	NAG	C	403	1	14,14,15	0.34	0	17,19,21	0.96	1 (5%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
4	NAG	C	402	1	14,14,15	0.31	0	17,19,21	0.97	1 (5%)
4	NAG	A	402	1	14,14,15	0.34	0	17,19,21	0.53	0
4	NAG	C	401	1	14,14,15	0.33	0	17,19,21	1.21	2 (11%)

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	A	401	1	-	1/6/23/26	0/1/1/1
4	NAG	D	501	2	-	3/6/23/26	0/1/1/1
4	NAG	C	403	1	-	3/6/23/26	0/1/1/1
4	NAG	C	402	1	-	2/6/23/26	0/1/1/1
4	NAG	A	402	1	-	0/6/23/26	0/1/1/1
4	NAG	C	401	1	-	1/6/23/26	0/1/1/1

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	C	401	NAG	O5-C1-C2	-2.58	107.21	111.29
4	C	403	NAG	C1-C2-N2	-2.31	106.54	110.49
4	C	401	NAG	C2-N2-C7	2.24	126.09	122.90
4	C	402	NAG	O5-C5-C6	2.14	110.56	107.20
4	A	401	NAG	O5-C1-C2	-2.06	108.04	111.29

There are no chirality outliers.

5 of 10 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	C	401	NAG	C3-C2-N2-C7
4	D	501	NAG	O5-C5-C6-O6
4	C	402	NAG	O5-C5-C6-O6
4	D	501	NAG	C4-C5-C6-O6
4	C	402	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

6.1 Protein, DNA and RNA chains

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

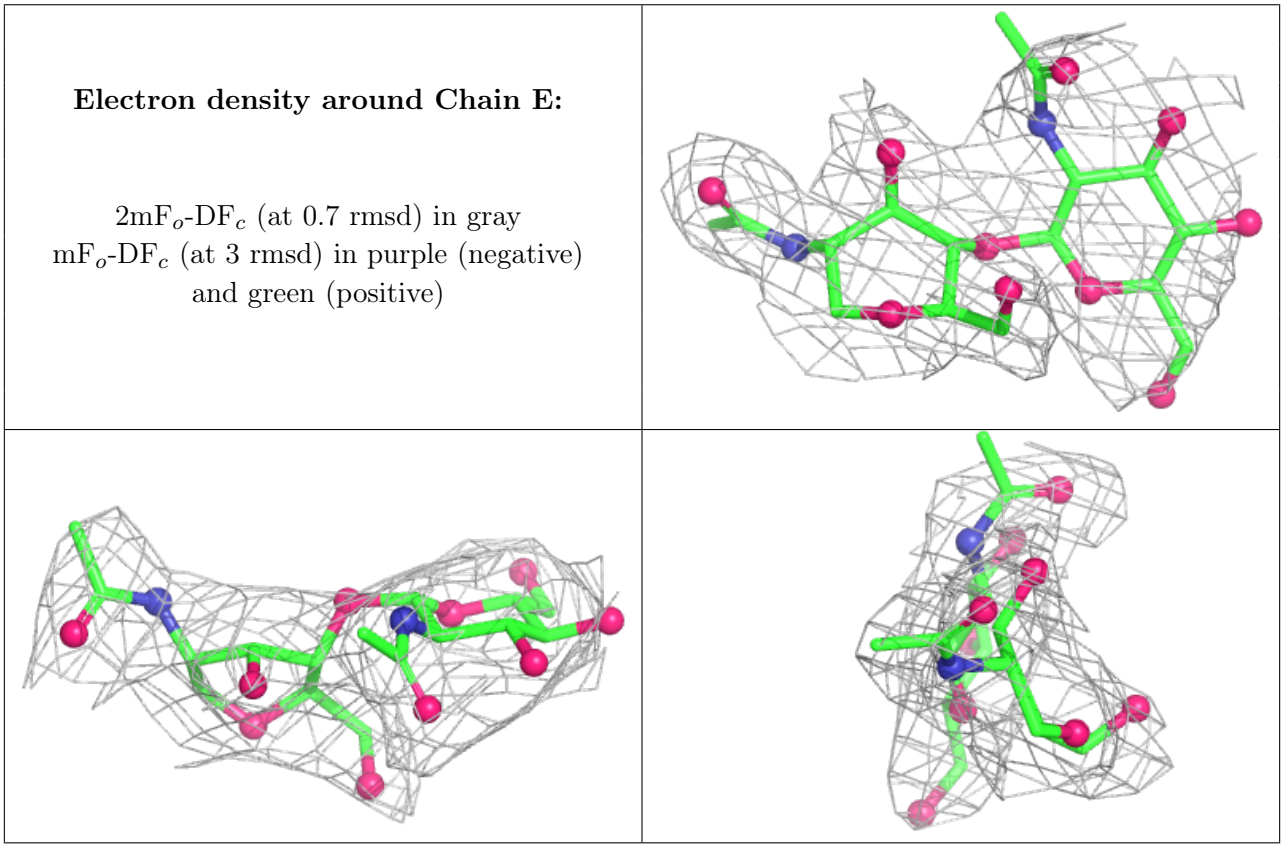
6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

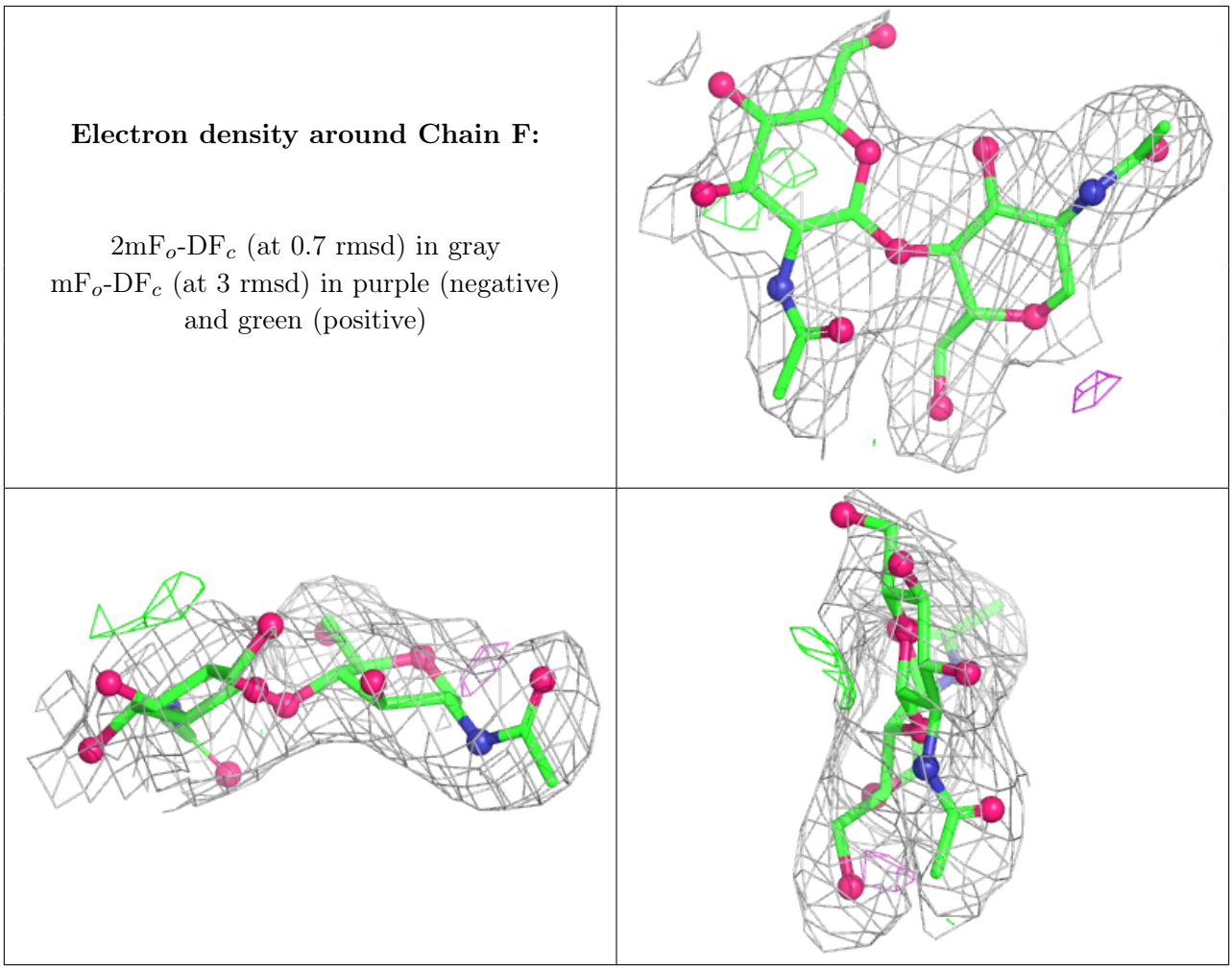
Unable to reproduce the depositors R factor - this section is therefore empty.

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 F:

$2mF_o$-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.