wwPDB X-ray Structure Validation Summary Report (i)

Nov 2, 2023 – 04:36 AM EDT

PDB ID	:	3HTO
Title	:	the hemagglutinin structure of an avian H1N1 influenza A virus
Authors	:	Wang, G.; Li, A.; Zhang, Q.; Wu, C.; Zhang, R.; Cai, Q.; Song, W.; Yuen, K.-Y.
Deposited on	:	2009-06-12
Resolution	:	2.95 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:

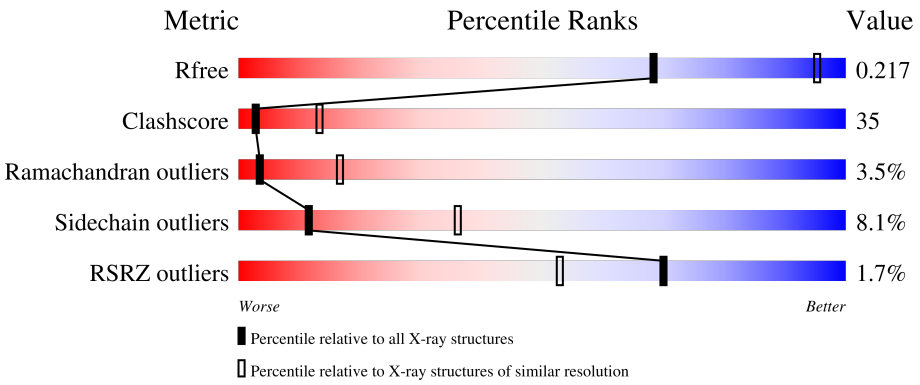
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.95 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	3104 (3.00-2.92)
Clashscore	141614	3462 (3.00-2.92)
Ramachandran outliers	138981	3340 (3.00-2.92)
Sidechain outliers	138945	3343 (3.00-2.92)
RSRZ outliers	127900	2986 (3.00-2.92)

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	324	2%, 50%, 42%, 8%
2	B	160	%, 55%, 42%
3	C	2	50%, 50%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	NAG	C	1	X	-	-	-

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 4132 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
			Total	C	N	O	S			
1	A	324	2530	1596	432	491	11	0	0	0

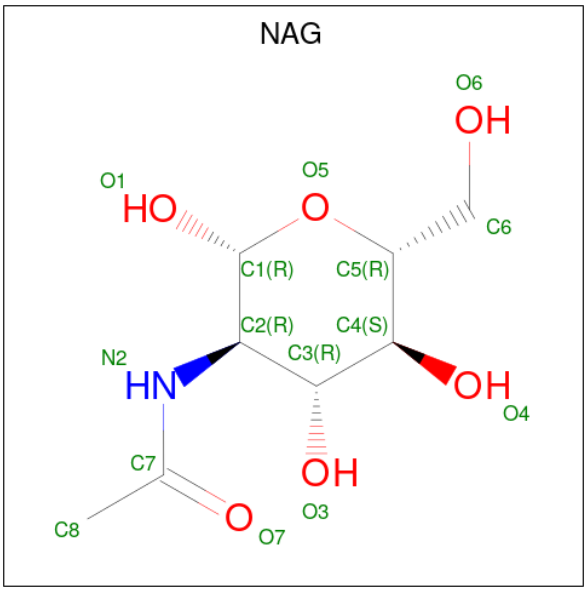
- Molecule 2 is a protein called Hemagglutinin HA2 chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	160	1286	803	220	256	7	0	0	0

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	C	2	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
			Total	C	N	O		
4	A	1	14	8	1	5	0	0
			Total	C	N	O		
4	A	1	14	8	1	5	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	161	161	161	0	0
			Total	O		
5	B	99	99	99	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

Chain A: 2% 50% 42% 8%

D5 T6 I9 N15 S16 T17 V23 H32 N35 L36 L37 E38 N42 G43 K44 L45 C46 K47 L48 A52 P53 K58 W64 L65 L66 G67 N68 P69 E70 C71 D72 L73 L74 L75 N78 S79 W80 S81 Y82 I83 I84 E85 T86 S87 N88 C94 Y95 P96 G97

E98 F99 I100 E103 E104 V112 S113 S114 F115 E116 R117 F118 E119 I120 F121 P122 K123 A124 S125 S126 W127 P128 N129 H130 E131 T132 T132A K133 G134 V135 T136 A137 A138 C139 S140 Y141 F142 G143 A144 S145 S146 F147 Y148 R149 N150 L151 L152 W153 I154 T155 K156 K157 G158 T159 S160 Y161 L164

S167 Y168 T169 N170 N171 K172 E175 W180 P185 P186 T187 T188 S189 E190 Q191 Q192 T193 L194 Y195 Q196 N197 T198 D199 A200 Y201 V204 G205 S206 S207 K208 Y209 N210 R211 R212 F213 T214 I217 K222 V223 R224 G225 Q226 A227 G228 R229 M230 Y233 W234 T235 T242 I243

E246 A247 T248 G249 P254 W255 Y256 A257 F258 A259 L260 N261 K262 G263 S264 D265 I268 I269 T270 N277 C278 D279 T280 K281 T284 P285 H286 G287 A288 I289 P294 F295 Q296 N297 V298 H299 P300 I301 T302 I303 G304 P307 K308 Y309 V310 K311 L321 R322 N323 I324 P325 S326

I327

• Molecule 2: Hemagglutinin HA2 chain

Chain B: % 55% 42%

G501 L502 F503 G504 A505 M506 A507 G508 F509 I510 T515 G516 M517 G520 W521 Y522 G523 H526 Q527 Q530 G531 S532 A536 D537 Q538 Q542 K551 V552 N553 E557 N560 T561 Q562 F563 T564 A565 K568 E569 N572 L573 R576 L580 D585 L589 D590

T593 E597 L602 E603 S613 N614 V615 R616 Y619 E620 K621 V622 K623 S624 Q625 L626 R627 N628 N629 E632 I633 G634 N635 G636 C637 F638 E639 F640 Y641 H642 E647 C648 M649 V652 K653 N654 G655 T656 Y657 D658 Y659 P660

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

Chain C: 50% 50%

NAG1 NDG2

4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 3	Depositor
Cell constants a, b, c, α, β, γ	198.48Å 198.48Å 198.48Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.62 – 2.95 49.62 – 2.95	Depositor EDS
% Data completeness (in resolution range)	99.6 (49.62-2.95) 99.9 (49.62-2.95)	Depositor EDS
R_{merge}	0.18	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.45 (at 2.96Å)	Xtriage
Refinement program	CNS 1.21	Depositor
R, R_{free}	0.215 , 0.260 0.225 , 0.217	Depositor DCC
R_{free} test set	1374 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	59.3	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 61.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.020 for -l,-k,-h	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	4132	wwPDB-VP
Average B, all atoms (Å^2)	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.46% 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: NDG, 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	0.37	0/2594	0.64	0/3534
2	B	0.41	0/1312	0.65	0/1766
All	All	0.38	0/3906	0.64	0/5300

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	2530	0	2443	193	0
2	B	1286	0	1197	95	0
3	C	28	0	24	1	0
4	A	28	0	26	3	0
5	A	161	0	0	37	0
5	B	99	0	0	19	0
All	All	4132	0	3690	267	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 35.

The worst 5 of 267 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:157:LYS:HA	1:A:157:LYS:HE3	1.36	1.05
1:A:78:ASN:HA	5:A:436:HOH:O	1.57	1.03
2:B:517:MET:HE1	2:B:523:GLY:HA3	1.42	1.00
1:A:134:GLY:HA3	1:A:153:TRP:HB3	1.47	0.96
2:B:635:ASN:HD21	2:B:637:CYS:HB2	1.36	0.90

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	322/324 (99%)	274 (85%)	35 (11%)	13 (4%)	3 14
2	B	158/160 (99%)	139 (88%)	15 (10%)	4 (2%)	5 25
All	All	480/484 (99%)	413 (86%)	50 (10%)	17 (4%)	3 17

5 of 17 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	80	TRP
1	A	197	ASN
1	A	199	ASP
2	B	634	GLY
1	A	79	SER

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	282/282 (100%)	258 (92%)	24 (8%)	10	34
2	B	136/136 (100%)	126 (93%)	10 (7%)	13	41
All	All	418/418 (100%)	384 (92%)	34 (8%)	11	36

5 of 34 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	572	ASN
2	B	580	LEU
2	B	635	ASN
1	A	214	THR
1	A	196	GLN

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	543	ASN
2	B	572	ASN
2	B	560	ASN
2	B	625	GLN
1	A	226	GLN

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	$\#\|Z\| > 2$
3	NAG	C	1	1,3	14,14,15	0.79	1 (7%)	17,19,21	1.30	3 (17%)
3	NDG	C	2	3	14,14,15	0.59	0	17,19,21	0.57	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	C	1	1,3	1/1/5/7	2/6/23/26	0/1/1/1
3	NDG	C	2	3	-	5/6/23/26	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	1	NAG	C1-C2	2.27	1.55	1.52

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	1	NAG	C2-N2-C7	-2.59	119.21	122.90
3	C	1	NAG	O5-C1-C2	2.28	114.88	111.29
3	C	1	NAG	C3-C4-C5	-2.28	106.18	110.24

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
3	C	1	NAG	C1

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	C	1	NAG	C8-C7-N2-C2
3	C	1	NAG	O7-C7-N2-C2
3	C	2	NDG	C3-C2-N2-C7
3	C	2	NDG	C8-C7-N2-C2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
3	C	2	NDG	O7-C7-N2-C2

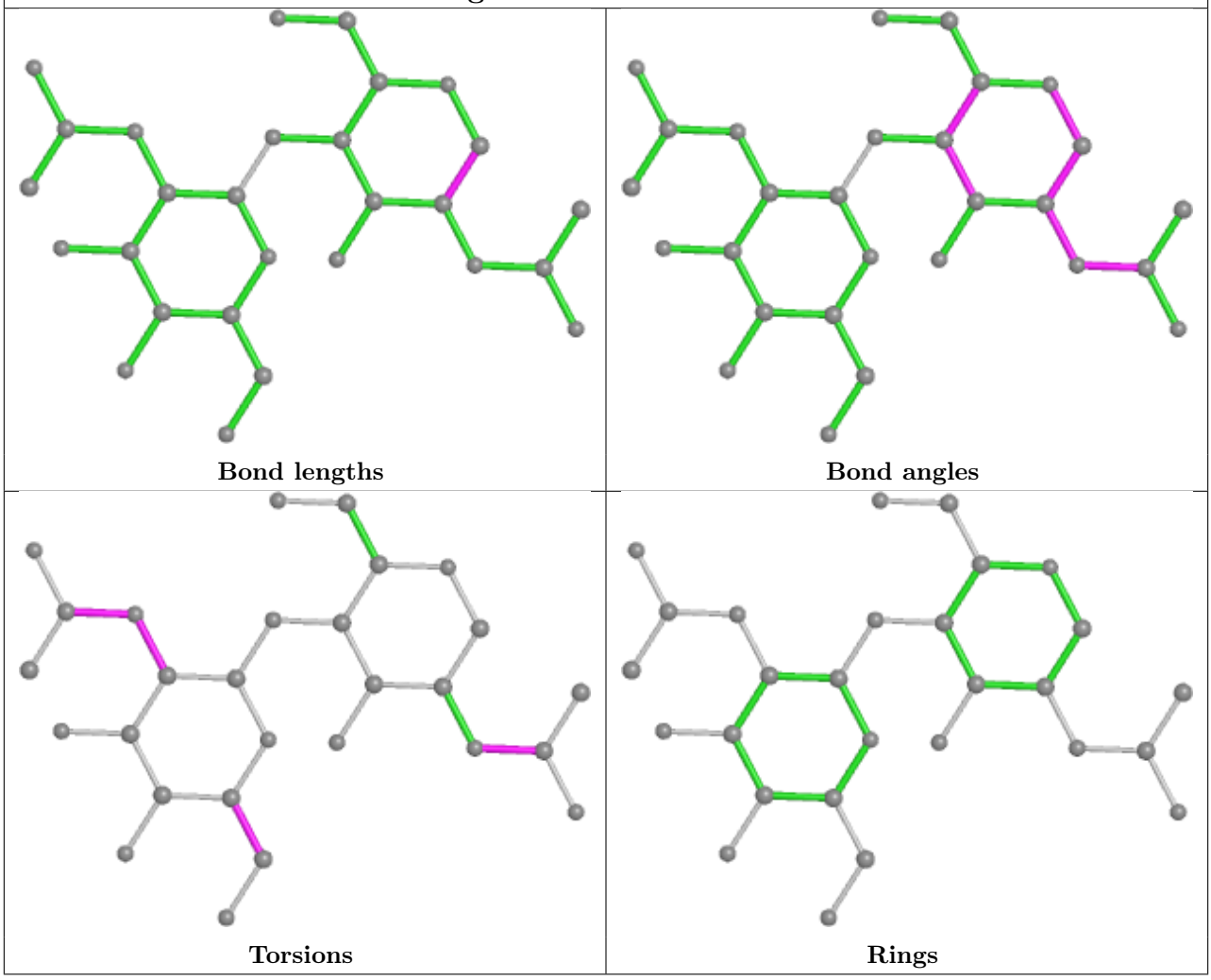
There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	1	NAG	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.

Oligosaccharide Chain C



5.6 Ligand geometry (i)

2 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	633	1	14,14,15	0.74	0	17,19,21	0.76	0
4	NAG	A	632	1	14,14,15	0.66	0	17,19,21	0.64	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	A	633	1	-	3/6/23/26	0/1/1/1
4	NAG	A	632	1	-	4/6/23/26	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	633	NAG	C8-C7-N2-C2
4	A	633	NAG	O7-C7-N2-C2
4	A	632	NAG	C8-C7-N2-C2
4	A	632	NAG	O7-C7-N2-C2
4	A	632	NAG	C4-C5-C6-O6

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	632	NAG	3	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)

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	324/324 (100%)	-0.05	7 (2%)	62	45	36, 62, 89, 100	0
2	B	160/160 (100%)	-0.20	1 (0%)	89	78	32, 54, 74, 93	0
All	All	484/484 (100%)	-0.10	8 (1%)	70	53	32, 59, 88, 100	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	263	GLY	3.3
2	B	660	PRO	3.2
1	A	327	ILE	2.5
1	A	158	GLY	2.4
1	A	142	PHE	2.4

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)

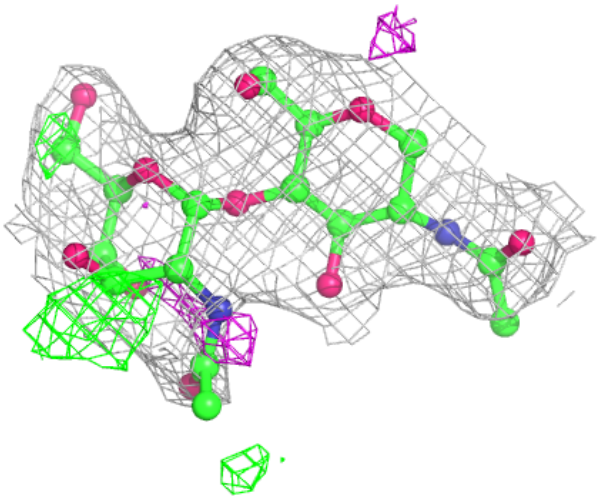
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

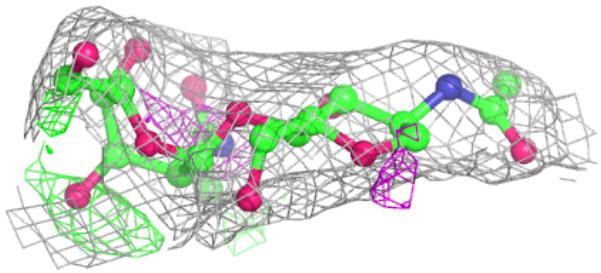
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	NDG	C	2	14/15	0.80	0.31	99,101,105,105	0
3	NAG	C	1	14/15	0.96	0.25	83,85,88,93	0

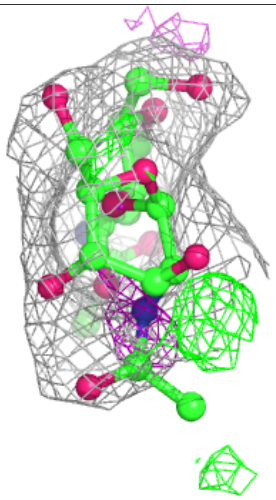
The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.

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)







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
4	NAG	A	633	14/15	0.83	0.28	91,95,98,98	0
4	NAG	A	632	14/15	0.85	0.42	93,97,99,100	0

6.5 Other polymers (i)

There are no such residues in this entry.