W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Aug 6, 2020 – 05:56 PM BST

PDB ID : 4BH1

Title : H5 (tyTy) Influenza Virus Haemagglutinin in Complex with Avian Receptor Analogue 3'-SLN

Authors : Xiong, X.; Coombs, P.J.; Martin, S.R.; Liu, J.; Xiao, H.; McCauley, J.W.; Locher, K.; Walker, P.A.; Collins, P.J.; Kawaoka, Y.; Skehel, J.J.; Gamblin, S.J.

Deposited on : 2013-03-29

Resolution : 2.15 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 (i)) were used in the production of this report:
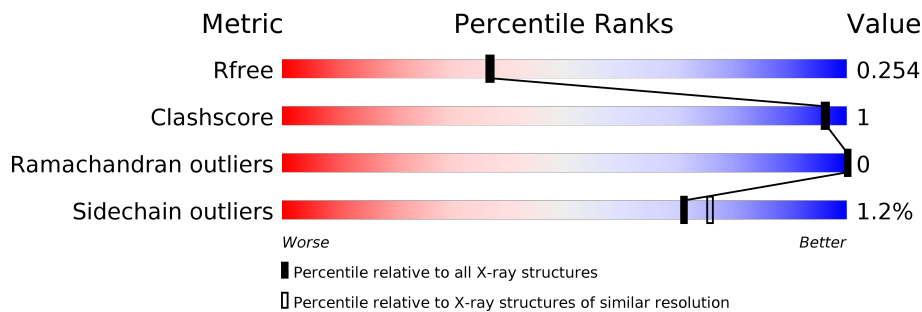
1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.15 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.
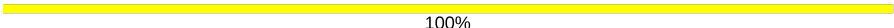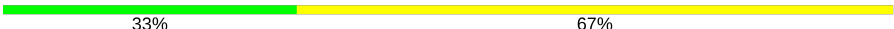
Metric	Percentile Ranks	Value
Rfree		0.254
Clashscore		1
Ramachandran outliers		0
Sidechain outliers		1.2%

Worse   *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1479 (2.16-2.16)
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)

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 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 <=5%

Mol	Chain	Length	Quality of chain
1	A	326	94% · ·
1	C	326	94% · ·
1	E	326	93% · ·
2	B	166	86% · 13%
2	D	166	84% · 12%
2	F	166	86% · 13%
3	G	3	33% 67%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	H	3	100%
3	I	3	33% 67%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 11980 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.

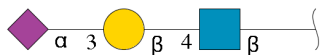| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	319	Total 2526	C 1592	N 439	O 481	S 14	0	0	0
1	C	319	Total 2519	C 1587	N 439	O 479	S 14	0	0	0
1	E	319	Total 2519	C 1587	N 439	O 479	S 14	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	323	ARG	-	expression tag	UNP Q207Z6
A	324	GLU	-	expression tag	UNP Q207Z6
A	325	THR	-	expression tag	UNP Q207Z6
A	326	ARG	-	expression tag	UNP Q207Z6
C	323	ARG	-	expression tag	UNP Q207Z6
C	324	GLU	-	expression tag	UNP Q207Z6
C	325	THR	-	expression tag	UNP Q207Z6
C	326	ARG	-	expression tag	UNP Q207Z6
E	323	ARG	-	expression tag	UNP Q207Z6
E	324	GLU	-	expression tag	UNP Q207Z6
E	325	THR	-	expression tag	UNP Q207Z6
E	326	ARG	-	expression tag	UNP Q207Z6

- Molecule 2 is a protein called HEMAGGLUTININ.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	145	Total 1129	C 697	N 200	O 224	S 8	0	0	0
2	D	146	Total 1130	C 698	N 201	O 223	S 8	0	0	0
2	F	145	Total 1129	C 697	N 200	O 224	S 8	0	0	0

- Molecule 3 is an oligosaccharide called N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galacto pyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose.

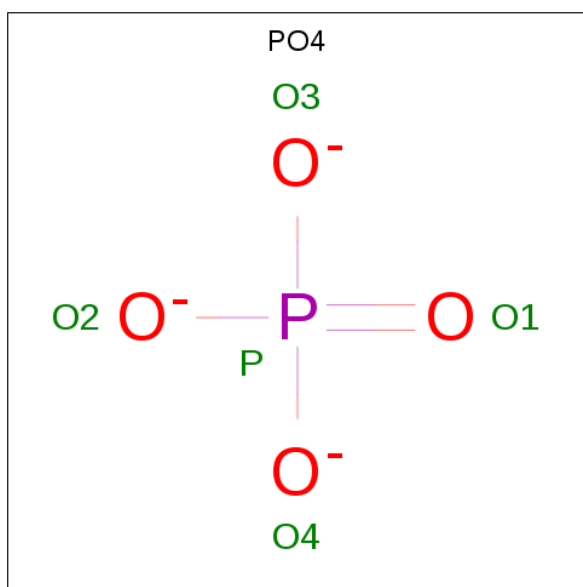| Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	G	3	Total 46	C 25	N 2	O 19	0	0	0
3	H	3	Total 46	C 25	N 2	O 19	0	0	0
3	I	3	Total 46	C 25	N 2	O 19	0	0	0

- Molecule 4 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).
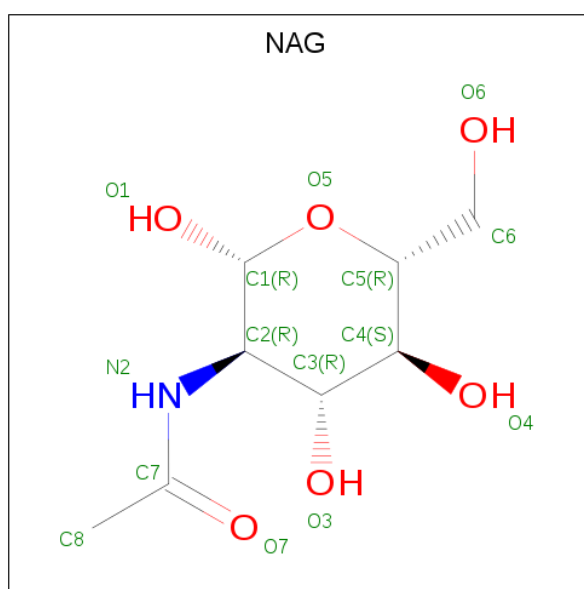
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total 14	C 8	N 1	O 5	0	0
4	C	1	Total 14	C 8	N 1	O 5	0	0
4	E	1	Total 14	C 8	N 1	O 5	0	0

- Molecule 5 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total 5	O 4	P 1	0	0
5	A	1	Total 5	O 4	P 1	0	0
5	C	1	Total 5	O 4	P 1	0	0
5	E	1	Total 5	O 4	P 1	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	266	Total 266	O 266	0	0
6	B	36	Total 36	O 36	0	0
6	C	243	Total 243	O 243	0	0
6	D	35	Total 35	O 35	0	0
6	E	218	Total 218	O 218	0	0
6	F	30	Total 30	O 30	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 1: HEMAGGLUTININ

Chain E:

- Molecule 2: HEMAGGLUTININ

Chain B:

- Molecule 2: HEMAGGLUTININ

Chain D:

- Molecule 2: HEMAGGLUTININ

Chain F:

GLY LEU PHE GLY ALA ILE ALA GLY PHE **I10** **Y24** **R75** **C137** **N154** GLY THR TYR ASP TYR PRO GLN TYR SER GLU GLU ALA

• Molecule 3: N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galactopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain G:

33%   67%

NAG1 GAL2 SIA3

• Molecule 3: N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galactopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain H:

100%

NAG1 GAL2 SIA3

• Molecule 3: N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galactopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain I:

33%   67%

NAG1 GAL2 SIA3

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	70.86Å 228.29Å 71.92Å 90.00° 113.71° 90.00°	Depositor
Resolution (Å)	62.48 – 2.15 43.13 – 2.16	Depositor EDS
% Data completeness (in resolution range)	98.0 (62.48-2.15) 97.8 (43.13-2.16)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.02 (at 2.16Å)	Xtriage
Refinement program	REFMAC 5.7.0032	Depositor
R, R_{free}	0.205 , 0.224 0.261 , 0.254	Depositor DCC
R_{free} test set	5489 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	35.4	Xtriage
Anisotropy	0.700	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 54.9	EDS
L-test for twinning[2]	$<|L|> = 0.47, <L^2> = 0.30$	Xtriage
Estimated twinning fraction	0.035 for l,-k,h	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	11980	wwPDB-VP
Average B, all atoms (Å^2)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.31% 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: PO4, GAL, NAG, SIA

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	0.32	0/2586	0.55	2/3515 (0.1%)
1	C	0.32	0/2579	0.55	2/3506 (0.1%)
1	E	0.31	0/2579	0.55	1/3506 (0.0%)
2	B	0.29	0/1149	0.44	0/1551
2	D	0.30	0/1150	0.44	0/1552
2	F	0.29	0/1149	0.43	0/1551
All	All	0.31	0/11192	0.52	5/15181 (0.0%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	216	ARG	NE-CZ-NH1	5.59	123.10	120.30
1	C	216	ARG	NE-CZ-NH2	-5.57	117.51	120.30
1	C	216	ARG	NE-CZ-NH1	5.46	123.03	120.30
1	E	216	ARG	NE-CZ-NH1	5.42	123.01	120.30
1	A	216	ARG	NE-CZ-NH2	-5.18	117.71	120.30

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	2526	0	2460	7	0
1	C	2519	0	2447	7	0
1	E	2519	0	2447	6	1
2	B	1129	0	1007	1	0
2	D	1130	0	1008	2	0
2	F	1129	0	1007	1	0
3	G	46	0	40	0	0
3	H	46	0	40	0	0
3	I	46	0	40	0	0
4	A	14	0	13	0	0
4	C	14	0	13	0	0
4	E	14	0	13	0	0
5	A	10	0	0	0	0
5	C	5	0	0	0	0
5	E	5	0	0	0	0
6	A	266	0	0	2	1
6	B	36	0	0	0	0
6	C	243	0	0	3	0
6	D	35	0	0	0	0
6	E	218	0	0	0	0
6	F	30	0	0	0	0
All	All	11980	0	10535	22	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 1.

The worst 5 of 22 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:162:ARG:HD2	6:C:2149:HOH:O	2.03	0.58
1:C:84:ASN:ND2	6:C:2069:HOH:O	2.36	0.57
1:A:7:TYR:HB2	1:A:317:LEU:CD2	2.43	0.48
1:A:130:GLY:HA3	1:A:149:TRP:HB3	1.96	0.47
1:E:130:GLY:HA3	1:E:149:TRP:HB3	1.95	0.47

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)
1:E:169:GLN:NE2	6:A:2135:HOH:O[1_655]	2.09	0.11

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	317/326 (97%)	309 (98%)	8 (2%)	0	100	100
1	C	317/326 (97%)	309 (98%)	8 (2%)	0	100	100
1	E	317/326 (97%)	309 (98%)	8 (2%)	0	100	100
2	B	143/166 (86%)	137 (96%)	6 (4%)	0	100	100
2	D	144/166 (87%)	137 (95%)	7 (5%)	0	100	100
2	F	143/166 (86%)	137 (96%)	6 (4%)	0	100	100
All	All	1381/1476 (94%)	1338 (97%)	43 (3%)	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	284/292 (97%)	282 (99%)	2 (1%)	84	89
1	C	282/292 (97%)	280 (99%)	2 (1%)	84	89
1	E	282/292 (97%)	279 (99%)	3 (1%)	73	78
2	B	113/141 (80%)	112 (99%)	1 (1%)	78	83
2	D	112/141 (79%)	108 (96%)	4 (4%)	35	33
2	F	113/141 (80%)	111 (98%)	2 (2%)	59	63
All	All	1186/1299 (91%)	1172 (99%)	14 (1%)	71	76

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	126	LEU
2	D	137	CYS
1	E	261	ASP
2	D	75	ARG
1	E	107	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
2	D	15	GLN
1	E	110	HIS
1	E	2	GLN
1	C	2	GLN
1	E	87	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)

9 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	G	1	3	15,15,15	0.37	0	21,21,21	1.36	4 (19%)
3	GAL	G	2	3	11,11,12	0.42	0	15,15,17	0.93	0
3	SIA	G	3	3	17,20,21	0.53	0	21,28,31	1.17	2 (9%)
3	NAG	H	1	3	15,15,15	0.44	0	21,21,21	1.24	2 (9%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
3	GAL	H	2	3	11,11,12	0.33	0	15,15,17	1.15	2 (13%)
3	SIA	H	3	3	17,20,21	0.56	0	21,28,31	1.13	2 (9%)
3	NAG	I	1	3	15,15,15	0.40	0	21,21,21	1.26	2 (9%)
3	GAL	I	2	3	11,11,12	0.26	0	15,15,17	0.88	0
3	SIA	I	3	3	17,20,21	0.37	0	21,28,31	1.14	3 (14%)

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	G	1	3	-	0/6/26/26	0/1/1/1
3	GAL	G	2	3	-	1/2/19/22	0/1/1/1
3	SIA	G	3	3	-	0/14/34/38	0/1/1/1
3	NAG	H	1	3	-	0/6/26/26	0/1/1/1
3	GAL	H	2	3	-	1/2/19/22	0/1/1/1
3	SIA	H	3	3	-	0/14/34/38	0/1/1/1
3	NAG	I	1	3	-	0/6/26/26	0/1/1/1
3	GAL	I	2	3	-	0/2/19/22	0/1/1/1
3	SIA	I	3	3	-	0/14/34/38	0/1/1/1

There are no bond length outliers.

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	H	1	NAG	C1-C2-N2	-4.25	105.80	110.73
3	G	1	NAG	C1-C2-N2	-3.61	106.55	110.73
3	I	1	NAG	C1-C2-N2	-3.52	106.65	110.73
3	H	3	SIA	C6-O6-C2	2.94	117.62	111.34
3	I	3	SIA	C6-O6-C2	2.73	117.19	111.34

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	H	2	GAL	C4-C5-C6-O6
3	G	2	GAL	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 G

Bond lengths

Bond angles

Torsions

Rings

Oligosaccharide Chain H

Bond lengths

Bond angles

Torsions

Rings

Oligosaccharide Chain I

Bond lengths

Bond angles

Torsions

Rings
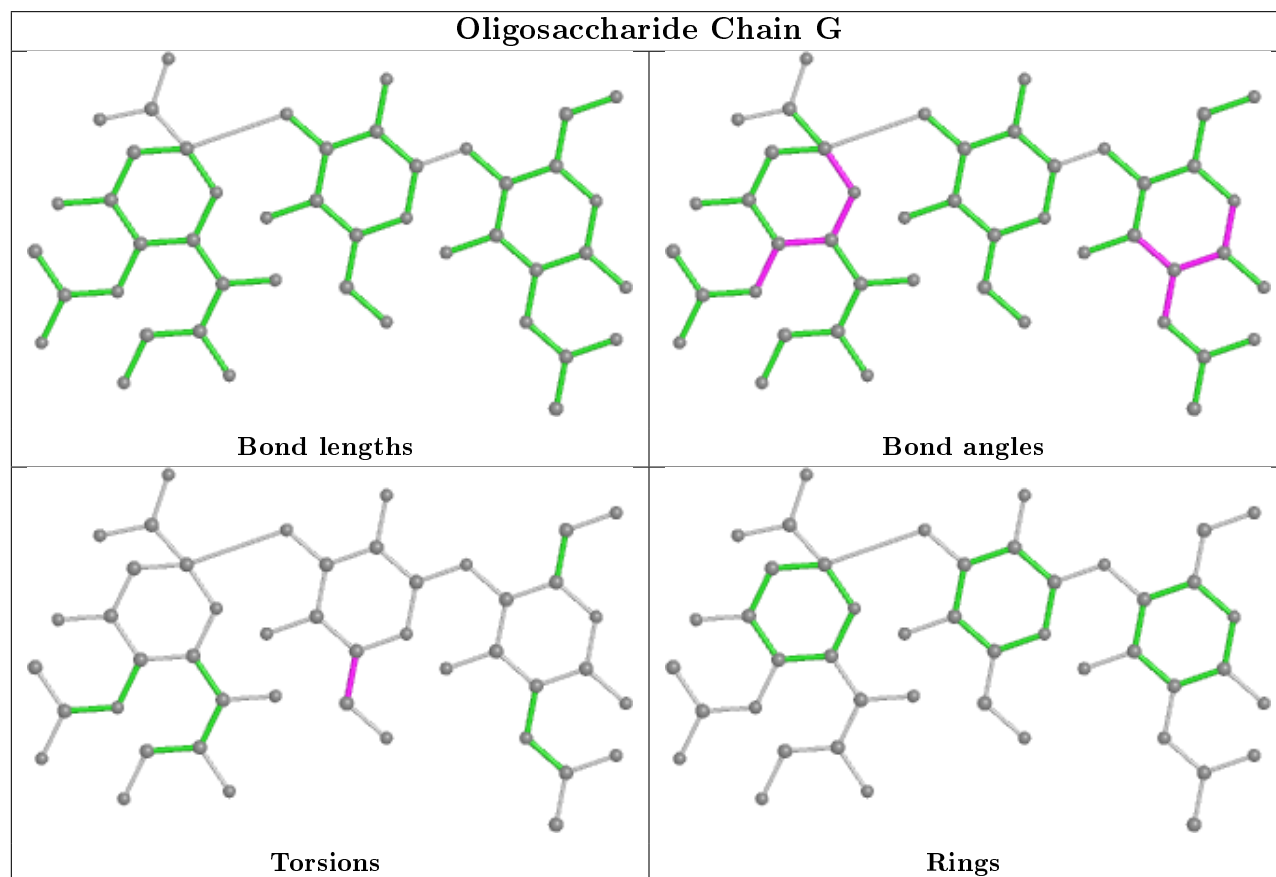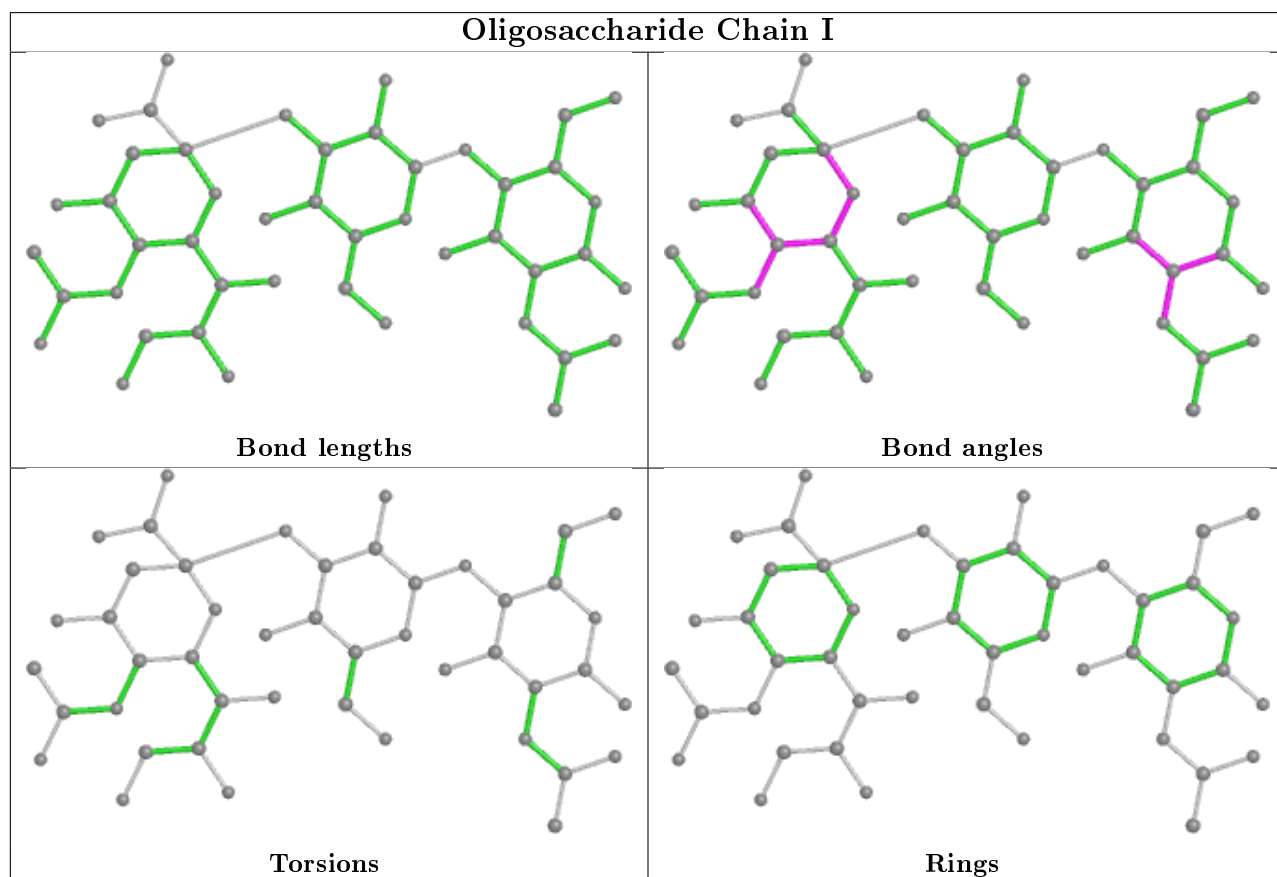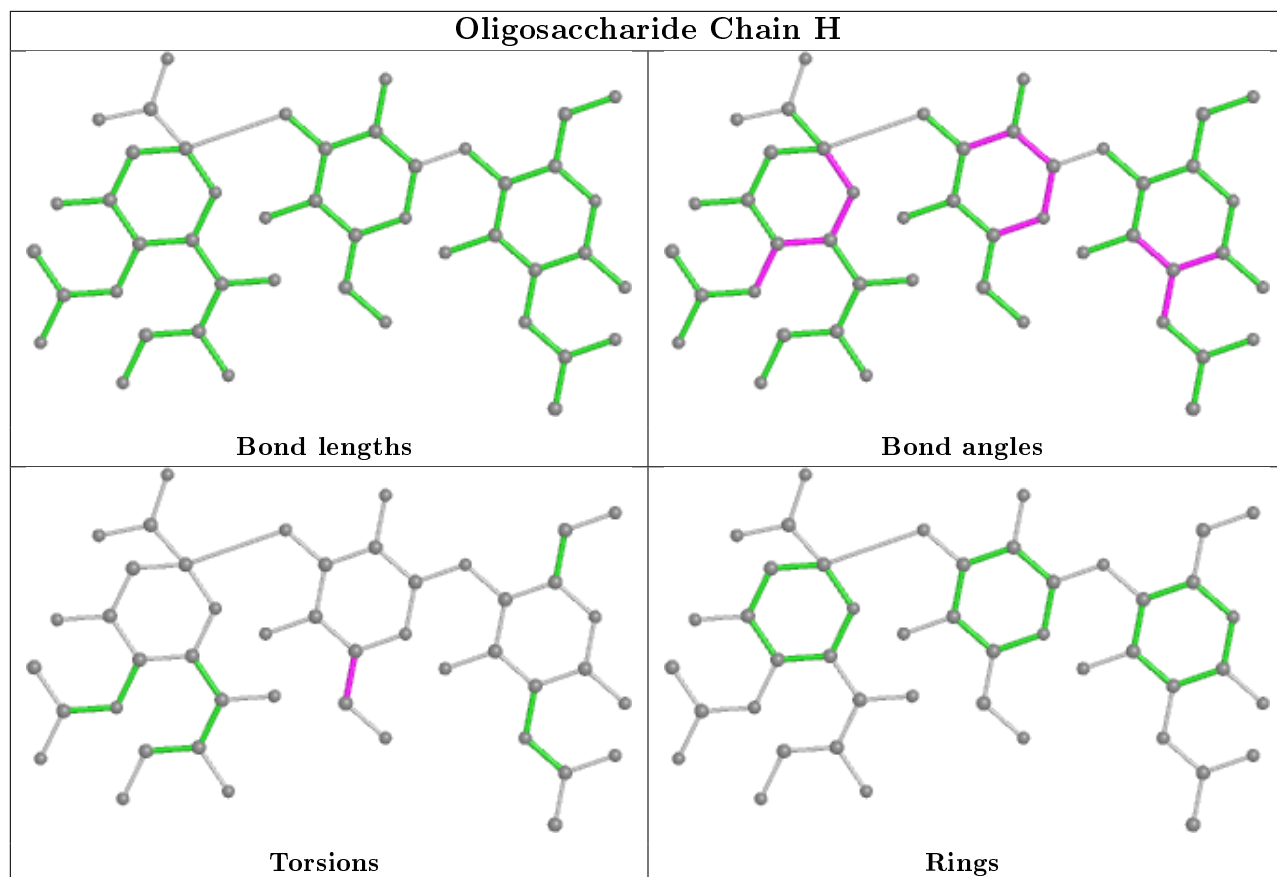
5.6 Ligand geometry (i)

7 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	1320	1	14,14,15	0.47	0	17,19,21	1.35	2 (11%)
4	NAG	C	1320	1	14,14,15	0.53	0	17,19,21	1.41	4 (23%)
4	NAG	E	1320	1	14,14,15	0.38	0	17,19,21	1.05	1 (5%)
5	PO4	A	1325	-	4,4,4	0.78	0	6,6,6	0.57	0
5	PO4	C	1324	-	4,4,4	0.94	0	6,6,6	0.51	0
5	PO4	A	1324	-	4,4,4	0.87	0	6,6,6	0.54	0
5	PO4	E	1324	-	4,4,4	0.95	0	6,6,6	0.53	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	C	1320	1	-	0/6/23/26	0/1/1/1
4	NAG	E	1320	1	-	0/6/23/26	0/1/1/1
4	NAG	A	1320	1	-	2/6/23/26	0/1/1/1

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	C	1320	NAG	C1-O5-C5	3.65	117.14	112.19
4	A	1320	NAG	C1-O5-C5	3.50	116.93	112.19
4	E	1320	NAG	C1-O5-C5	2.57	115.68	112.19
4	C	1320	NAG	O5-C5-C6	2.45	111.05	107.20
4	A	1320	NAG	C3-C4-C5	-2.13	106.44	110.24

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	1320	NAG	C4-C5-C6-O6
4	A	1320	NAG	O5-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)

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

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

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

6.3 Carbohydrates (i)

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 G:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
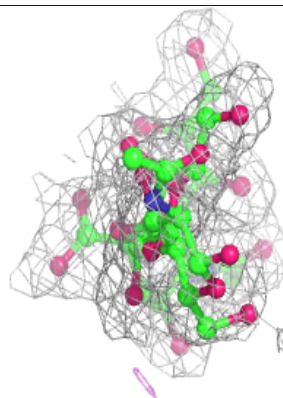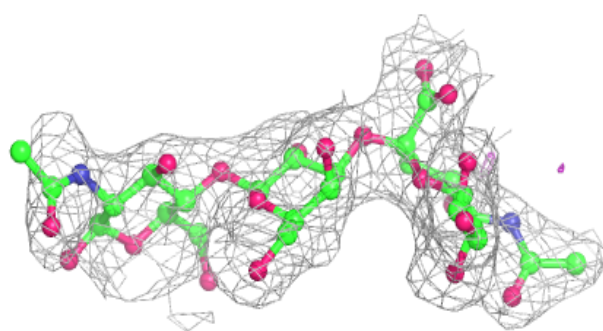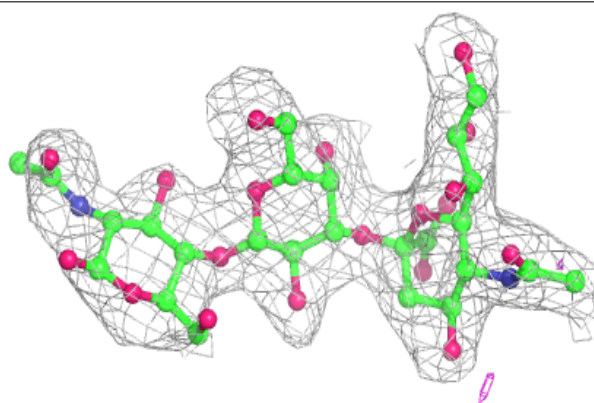mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)
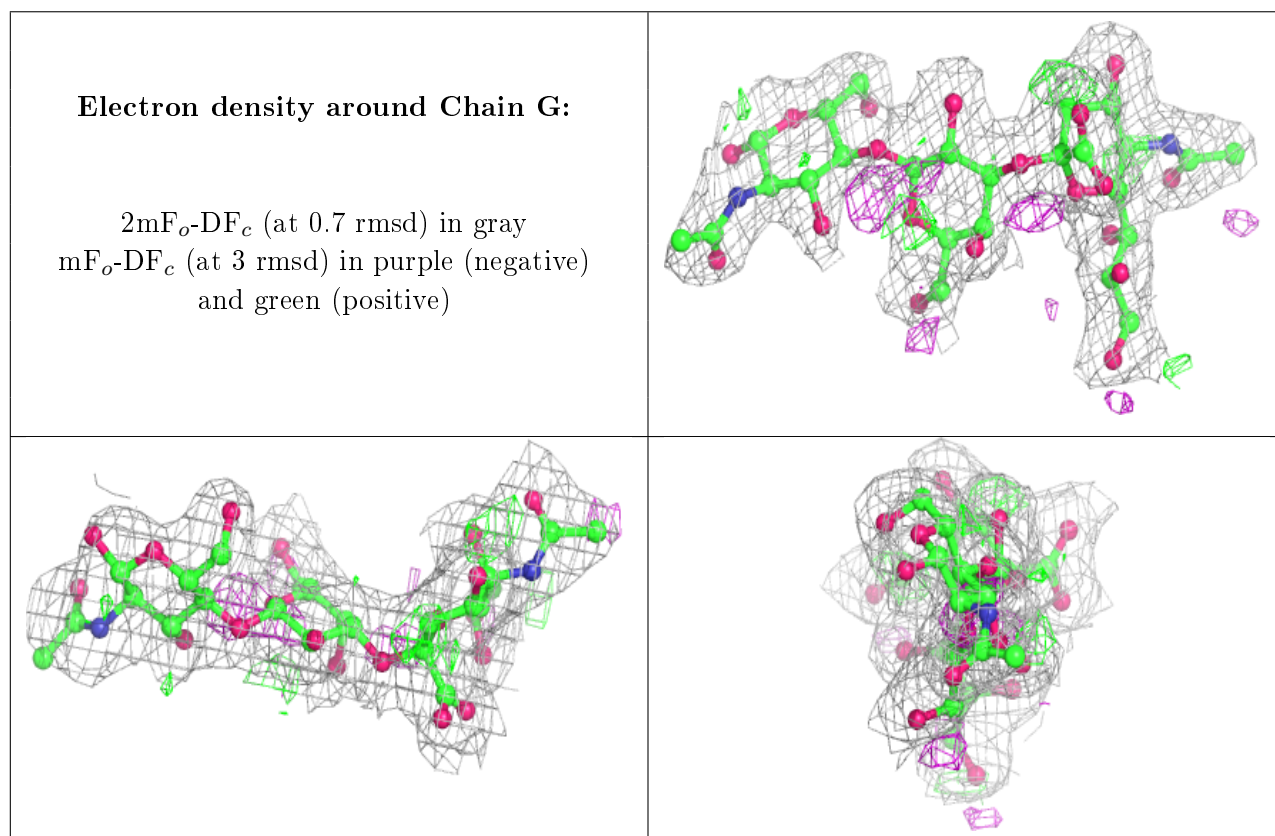
Electron density around Chain H:

$2mF_o-DF_c$ (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)
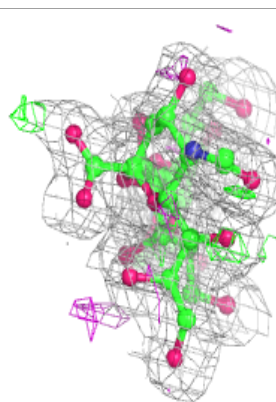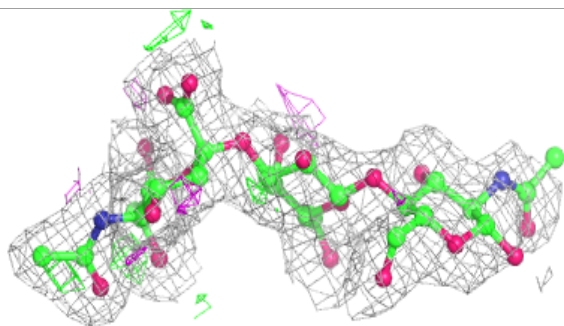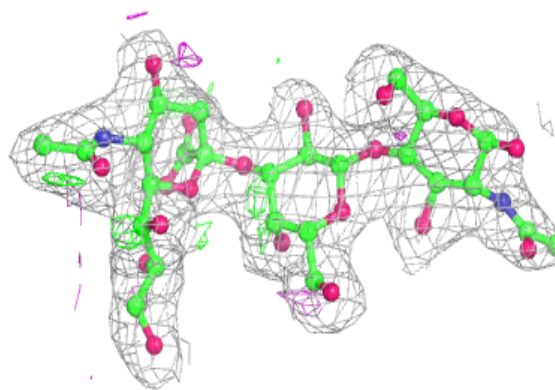
Electron density around Chain I:

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

6.4 Ligands (i)

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

6.5 Other polymers (i)

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