



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 8, 2020 – 11:08 AM BST

PDB ID : 4AYJ
Title : Molecular structure of a metal-independent bacterial glycosyltransferase that catalyzes the synthesis of histo-blood group A antigen
Authors : Thiyagarajan, N.; Pham, T.T.K.; Stinson, B.; Sundriyal, A.; Tumbale, P.; Lizotte-Waniewskib, M.; Brewb, K.; Acharya, K.R.
Deposited on : 2012-06-21
Resolution : 3.00 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

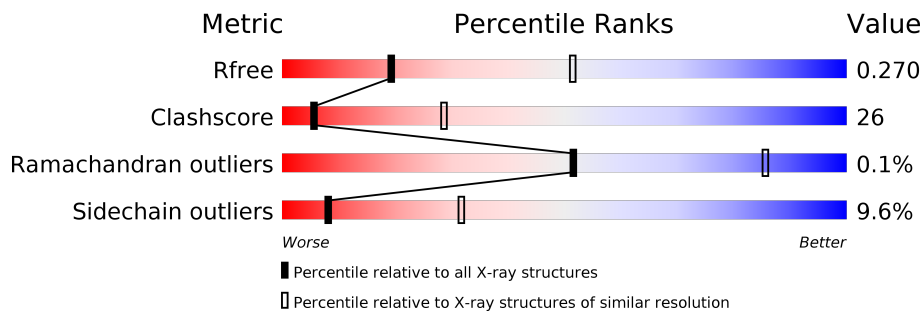
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.00 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	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	246	52% (green), 38% (yellow), 6% (orange), 4% (grey)
1	B	246	46% (green), 45% (yellow), 5% (orange), 4% (grey)
1	C	246	47% (green), 43% (yellow), 5% (orange), 5% (grey)
1	D	246	46% (green), 44% (yellow), 6% (orange), 4% (grey)
2	E	3	33% (yellow), 67% (orange)
2	F	3	67% (yellow), 33% (orange)
2	G	3	33% (yellow), 67% (orange)

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	H	3	 100%

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
2	BGC	G	1	-	-	X	-

2 Entry composition [i](#)

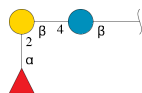
There are 2 unique types of molecules in this entry. The entry contains 7939 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 BOGT - METAL-INDEPENDENT GLYCOSYLTRANSFERASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	231	Total 1928	C 1262	N 311	O 348	S 7	4	0	0
1	B	235	Total 1961	C 1282	N 318	O 354	S 7	0	0	0
1	C	233	Total 1945	C 1272	N 315	O 351	S 7	4	0	0
1	D	236	Total 1973	C 1291	N 319	O 356	S 7	0	0	0

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



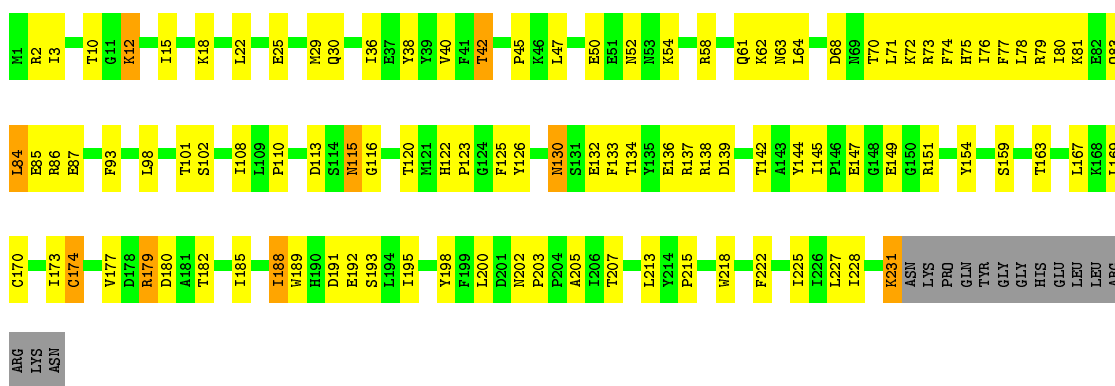
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
			Total	C	O			
2	E	3	Total 33	C 18	O 15	0	0	0
2	F	3	Total 33	C 18	O 15	0	0	0
2	G	3	Total 33	C 18	O 15	0	0	0
2	H	3	Total 33	C 18	O 15	0	0	0

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.

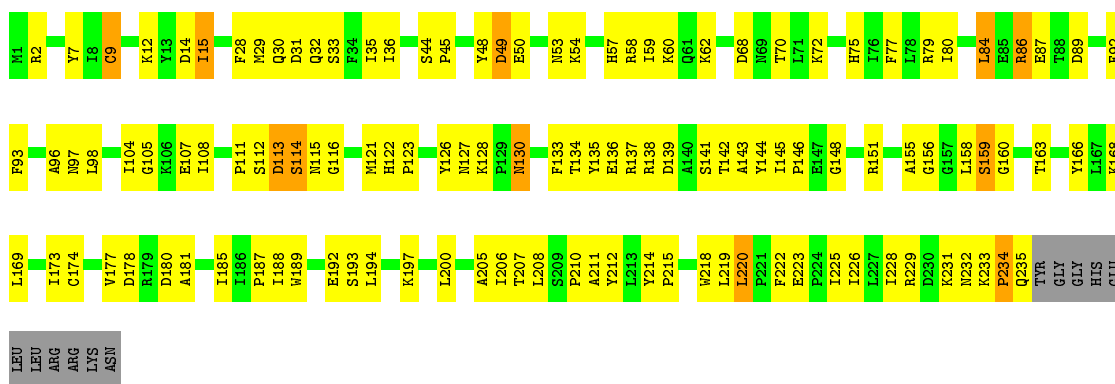
- Molecule 1: BOGT - METAL-INDEPENDENT GLYCOSYLTRANSFERASE

Chain A: 



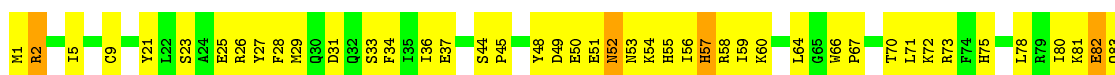
- Molecule 1: BOGT - METAL-INDEPENDENT GLYCOSYLTRANSFERASE

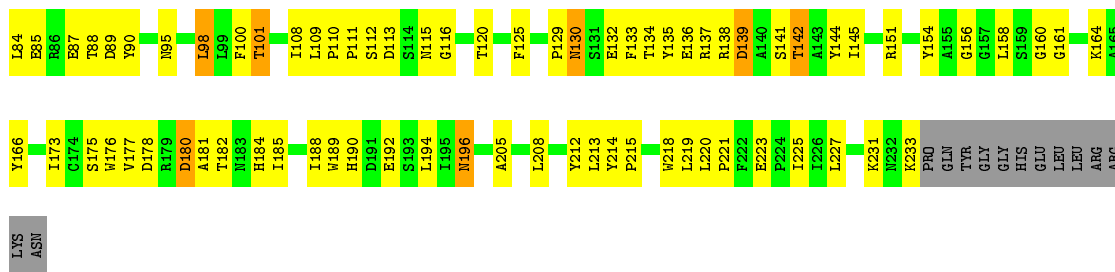
Chain B: 



- Molecule 1: BOGT - METAL-INDEPENDENT GLYCOSYLTRANSFERASE

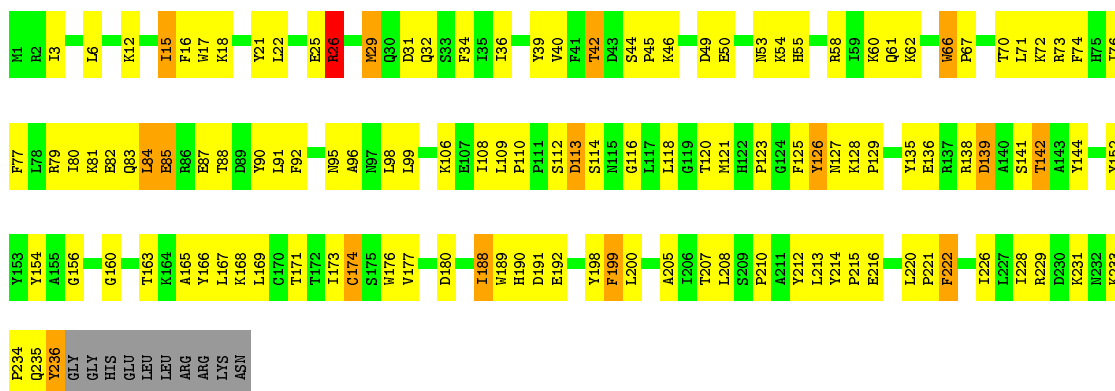
Chain C: 





- Molecule 1: BOGT - METAL-INDEPENDENT GLYCOSYLTRANSFERASE

Chain D: 46% 44% 6%



- Molecule 2: alpha-L-fucopyranose-(1-2)-beta-D-galactopyranose-(1-4)-beta-D-glucopyranos e

Chain E: 33% 67%



- Molecule 2: alpha-L-fucopyranose-(1-2)-beta-D-galactopyranose-(1-4)-beta-D-glucopyranos e

Chain F: 67% 33%



- Molecule 2: alpha-L-fucopyranose-(1-2)-beta-D-galactopyranose-(1-4)-beta-D-glucopyranos e

Chain G: 33% 67%



- Molecule 2: alpha-L-fucopyranose-(1-2)-beta-D-galactopyranose-(1-4)-beta-D-glucopyranos e

Chain H:  100%

E001
GALL2
F003

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	70.85Å 93.87Å 75.51Å 90.00° 93.82° 90.00°	Depositor
Resolution (Å)	70.69 – 3.00 70.69 – 3.00	Depositor EDS
% Data completeness (in resolution range)	93.7 (70.69-3.00) 93.7 (70.69-3.00)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	0.25 (at 3.01Å)	Xtrriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.181 , 0.262 0.443 , 0.270	Depositor DCC
R_{free} test set	951 reflections (2.55%)	wwPDB-VP
Wilson B-factor (Å ²)	46.4	Xtrriage
Anisotropy	0.144	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 45.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.29$, $\langle L^2 \rangle = 0.12$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.59	EDS
Total number of atoms	7939	wwPDB-VP
Average B, all atoms (Å ²)	36.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.98% 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: BGC, GAL, FUC

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	0.50	0/1991	0.72	0/2704
1	B	0.47	0/2025	0.68	1/2750 (0.0%)
1	C	0.47	0/2008	0.65	0/2726
1	D	0.47	0/2038	0.67	1/2768 (0.0%)
All	All	0.48	0/8062	0.68	2/10948 (0.0%)

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

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	26	ARG	NE-CZ-NH1	5.31	122.96	120.30
1	B	220	LEU	CA-CB-CG	5.21	127.29	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	115	ASN	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	1928	0	1869	89	0
1	B	1961	0	1903	93	1
1	C	1945	0	1888	101	0
1	D	1973	0	1911	116	1
2	E	33	0	30	9	0
2	F	33	0	30	2	0
2	G	33	0	30	7	0
2	H	33	0	30	3	0
All	All	7939	0	7691	399	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 26.

The worst 5 of 399 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:D:128:LYS:NZ	2:G:1:BGC:O5	1.79	1.15
1:C:2:ARG:NH1	1:C:89:ASP:H	1.50	1.08
1:C:2:ARG:O	1:C:2:ARG:NH1	1.93	1.02
1:A:134:THR:HG21	1:A:189:TRP:HE1	1.28	0.98
1:B:229:ARG:HD3	1:C:223:GLU:OE2	1.64	0.97

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:B:12:LYS:NZ	1:D:200:LEU:O[2_645]	2.18	0.02

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	229/246 (93%)	225 (98%)	4 (2%)	0	100	100
1	B	233/246 (95%)	221 (95%)	11 (5%)	1 (0%)	34	72
1	C	231/246 (94%)	219 (95%)	12 (5%)	0	100	100
1	D	234/246 (95%)	223 (95%)	11 (5%)	0	100	100
All	All	927/984 (94%)	888 (96%)	38 (4%)	1 (0%)	51	85

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	234	PRO

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	208/221 (94%)	191 (92%)	17 (8%)	11	39
1	B	212/221 (96%)	190 (90%)	22 (10%)	7	27
1	C	210/221 (95%)	189 (90%)	21 (10%)	7	29
1	D	213/221 (96%)	192 (90%)	21 (10%)	8	30
All	All	843/884 (95%)	762 (90%)	81 (10%)	8	32

5 of 81 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	188	ILE
1	C	98	LEU
1	D	174	CYS
1	B	219	LEU
1	C	52	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	130	ASN
1	D	190	HIS
1	C	196	ASN
1	B	75	HIS
1	C	75	HIS

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](#)

12 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	E	1	2	12,12,12	1.32	2 (16%)	17,17,17	1.38	3 (17%)
2	GAL	E	2	2	11,11,12	1.88	1 (9%)	15,15,17	1.58	3 (20%)
2	FUC	E	3	2	10,10,11	2.35	4 (40%)	14,14,16	2.28	4 (28%)
2	BGC	F	1	2	12,12,12	1.23	2 (16%)	17,17,17	1.51	3 (17%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
2	GAL	F	2	2	11,11,12	1.84	1 (9%)	15,15,17	1.85	3 (20%)
2	FUC	F	3	2	10,10,11	2.54	5 (50%)	14,14,16	2.39	3 (21%)
2	BGC	G	1	2	12,12,12	1.05	2 (16%)	17,17,17	1.89	5 (29%)
2	GAL	G	2	2	11,11,12	1.82	1 (9%)	15,15,17	1.88	6 (40%)
2	FUC	G	3	2	10,10,11	2.36	5 (50%)	14,14,16	2.60	4 (28%)
2	BGC	H	1	2	12,12,12	1.45	2 (16%)	17,17,17	0.93	0
2	GAL	H	2	2	11,11,12	2.23	1 (9%)	15,15,17	2.06	3 (20%)
2	FUC	H	3	2	10,10,11	2.69	5 (50%)	14,14,16	2.38	2 (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
2	BGC	E	1	2	-	2/2/22/22	0/1/1/1
2	GAL	E	2	2	-	2/2/19/22	0/1/1/1
2	FUC	E	3	2	-	-	0/1/1/1
2	BGC	F	1	2	-	2/2/22/22	0/1/1/1
2	GAL	F	2	2	-	0/2/19/22	0/1/1/1
2	FUC	F	3	2	-	-	0/1/1/1
2	BGC	G	1	2	-	2/2/22/22	0/1/1/1
2	GAL	G	2	2	-	0/2/19/22	0/1/1/1
2	FUC	G	3	2	-	-	0/1/1/1
2	BGC	H	1	2	-	2/2/22/22	0/1/1/1
2	GAL	H	2	2	-	0/2/19/22	0/1/1/1
2	FUC	H	3	2	-	-	0/1/1/1

The worst 5 of 31 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	H	2	GAL	C2-C3	6.71	1.62	1.52
2	E	2	GAL	C2-C3	5.73	1.61	1.52
2	G	2	GAL	C2-C3	5.37	1.60	1.52
2	F	2	GAL	C2-C3	5.20	1.60	1.52
2	H	3	FUC	O5-C5	4.88	1.54	1.43

The worst 5 of 39 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	G	3	FUC	C1-C2-C3	7.62	119.03	109.67
2	H	3	FUC	C1-C2-C3	7.29	118.62	109.67
2	F	3	FUC	C1-C2-C3	6.83	118.07	109.67
2	H	2	GAL	C1-C2-C3	-6.70	101.43	109.67
2	E	3	FUC	C1-C2-C3	6.40	117.53	109.67

There are no chirality outliers.

5 of 10 torsion outliers are listed below:

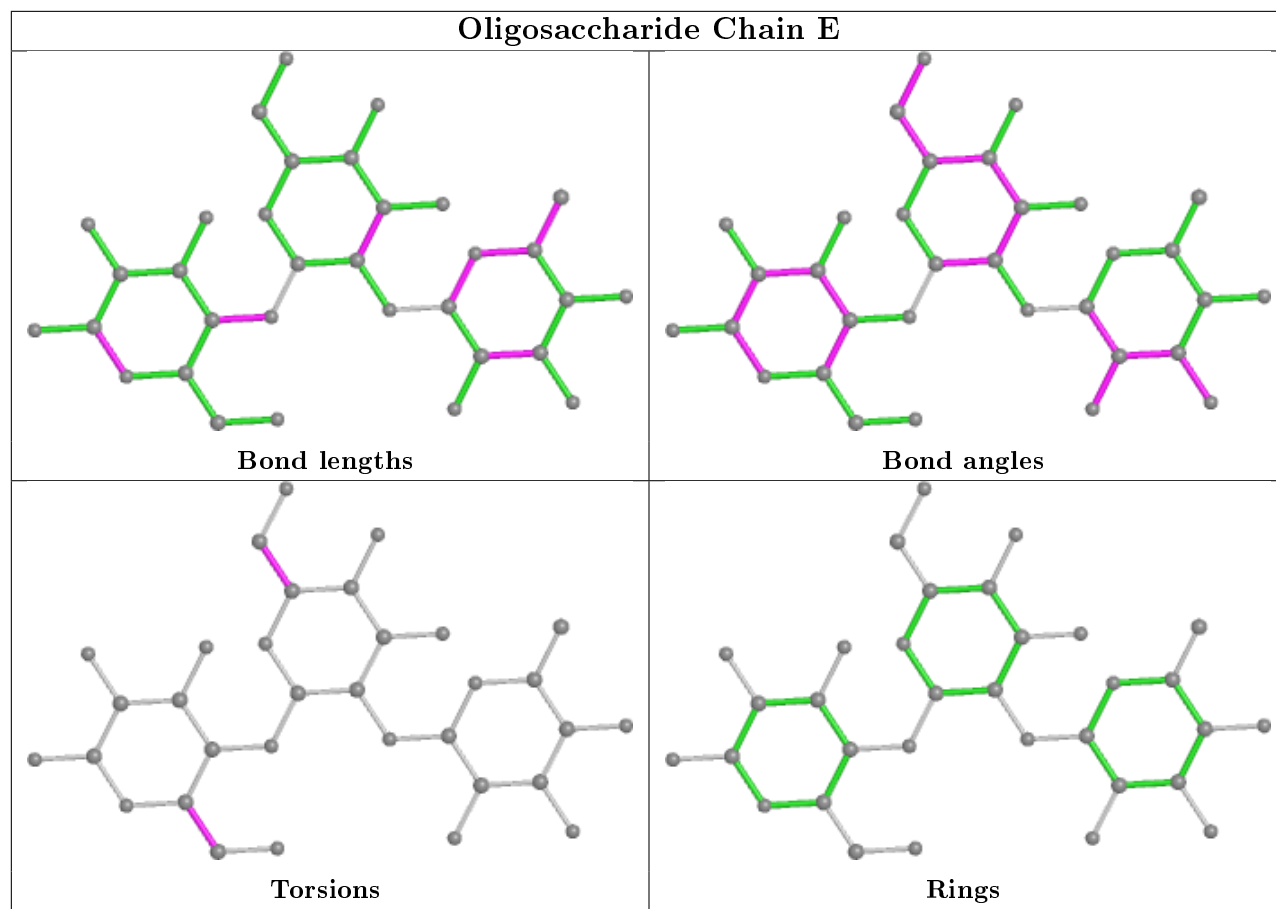
Mol	Chain	Res	Type	Atoms
2	G	1	BGC	O5-C5-C6-O6
2	E	1	BGC	O5-C5-C6-O6
2	G	1	BGC	C4-C5-C6-O6
2	H	1	BGC	O5-C5-C6-O6
2	E	2	GAL	C4-C5-C6-O6

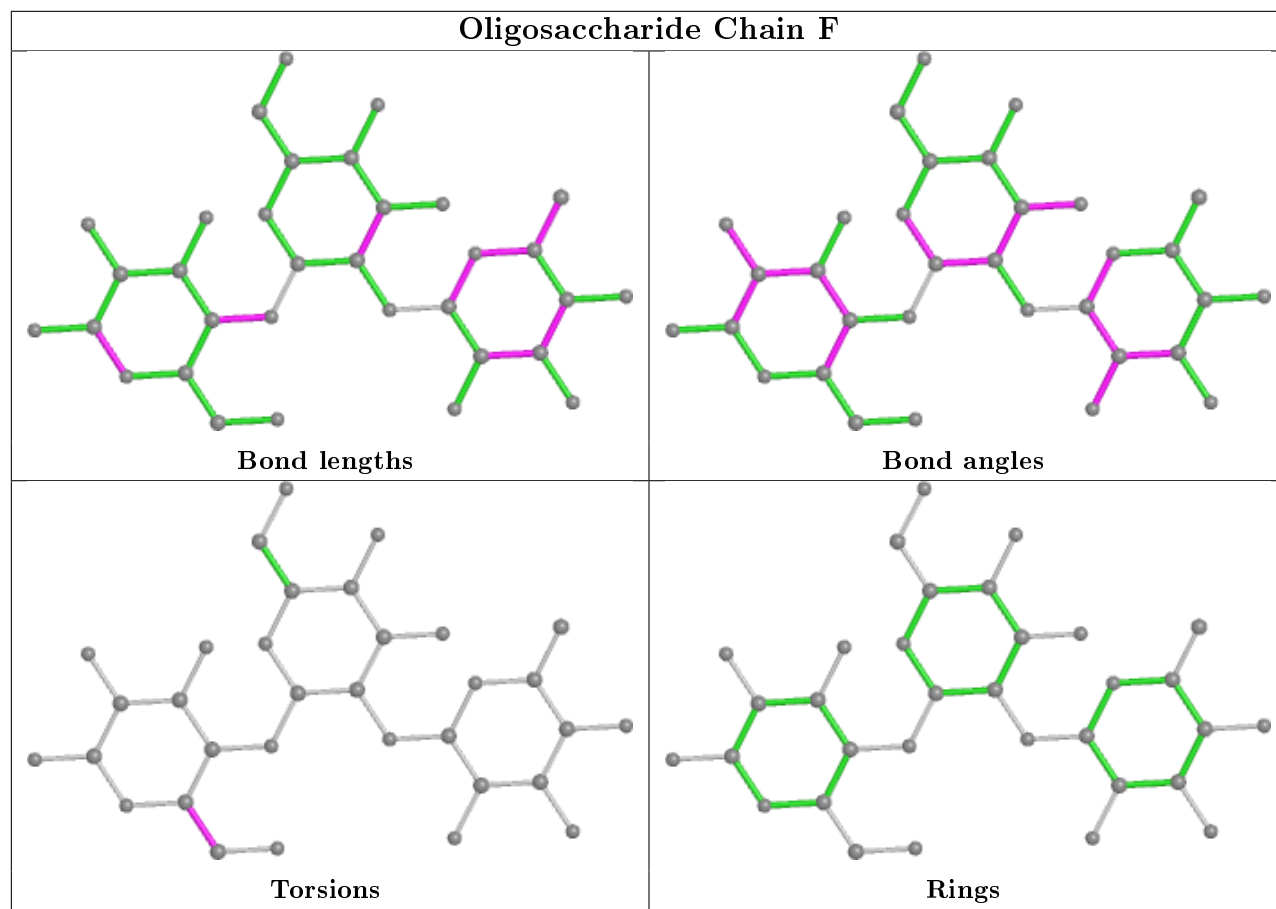
There are no ring outliers.

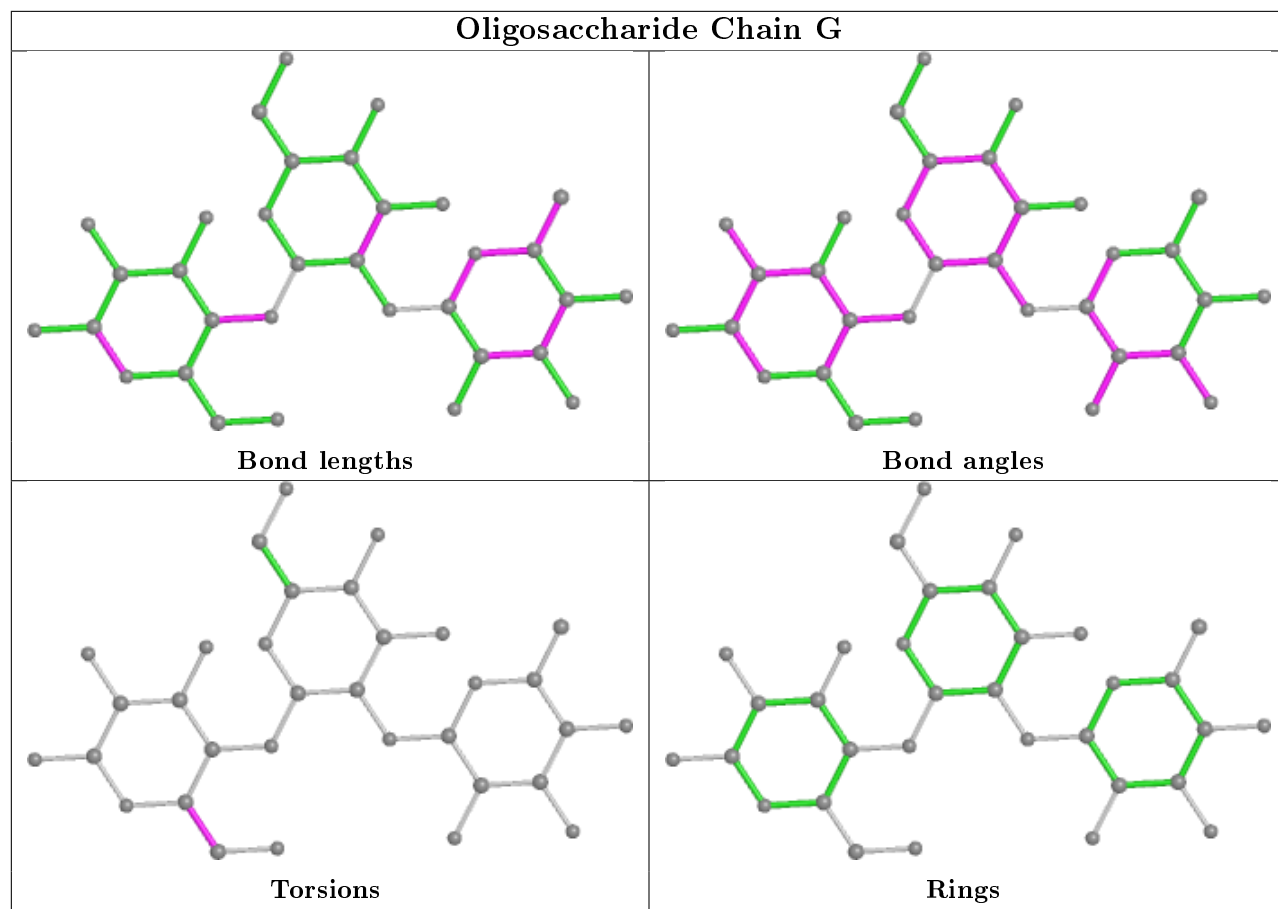
8 monomers are involved in 21 short contacts:

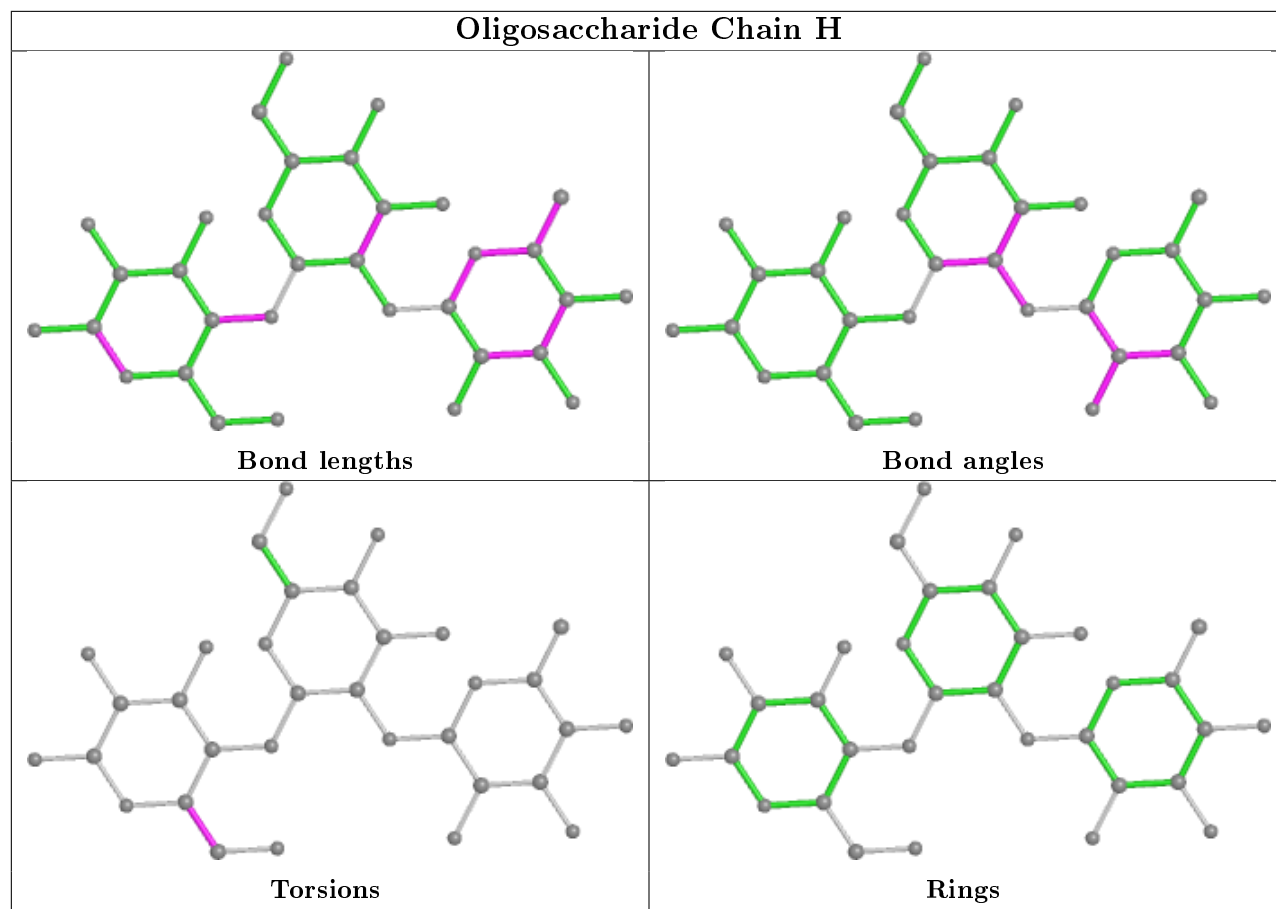
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	E	2	GAL	4	0
2	G	2	GAL	4	0
2	H	1	BGC	1	0
2	E	1	BGC	5	0
2	G	1	BGC	6	0
2	F	2	GAL	2	0
2	H	3	FUC	1	0
2	H	2	GAL	2	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.









5.6 Ligand geometry [i](#)

There are no ligands in this entry.

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 depositor's R factor - this section is therefore empty.

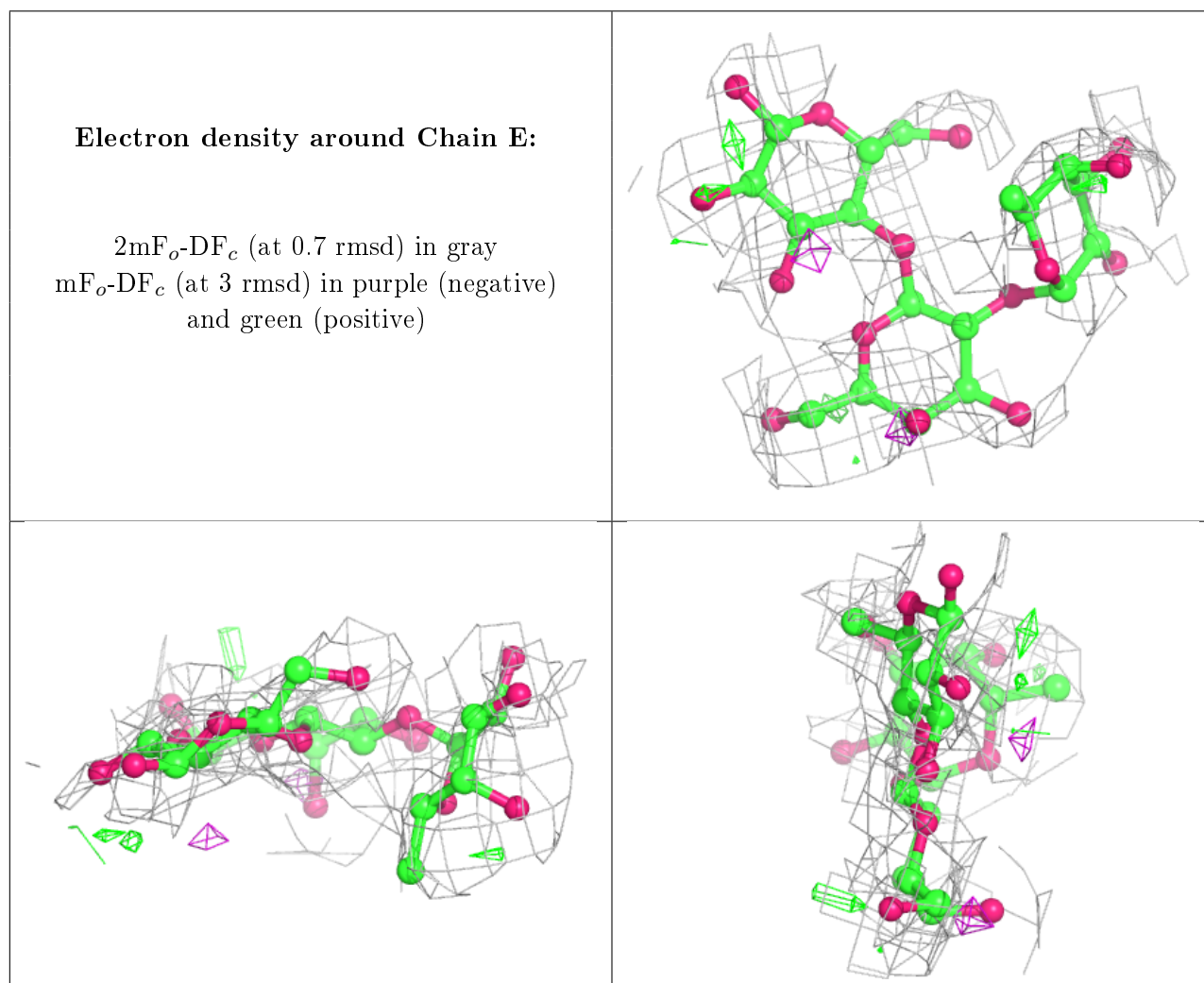
6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

Unable to reproduce the depositor's R factor - this section is therefore empty.

6.3 Carbohydrates [i](#)

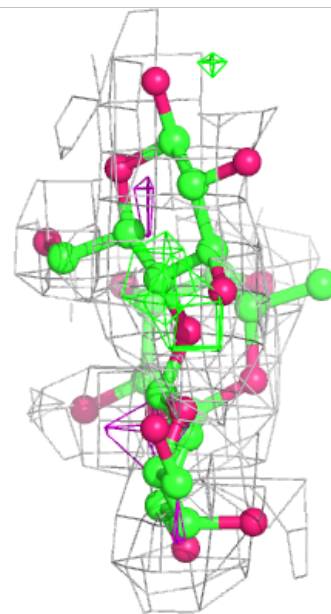
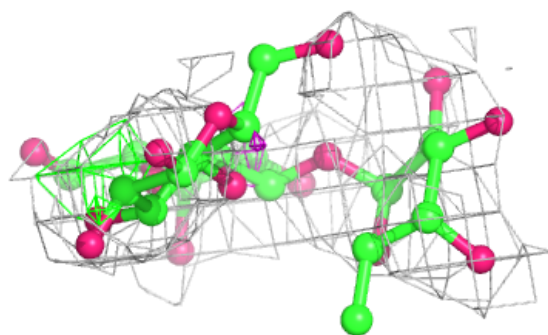
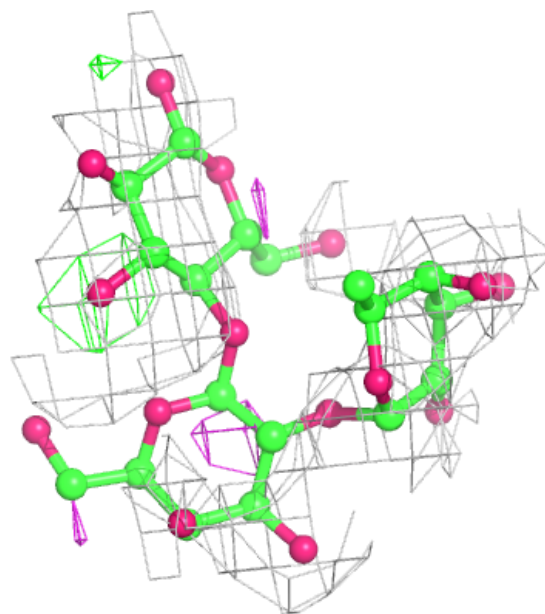
Unable to reproduce the depositor's R factor - this section is therefore empty.

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.



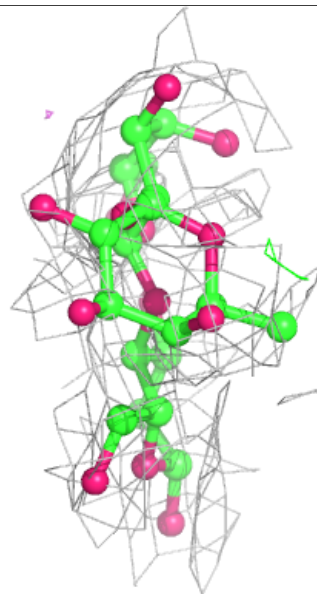
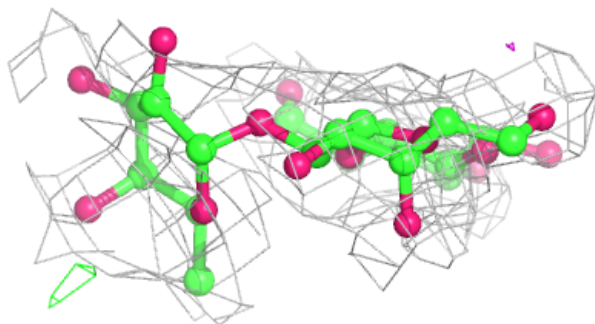
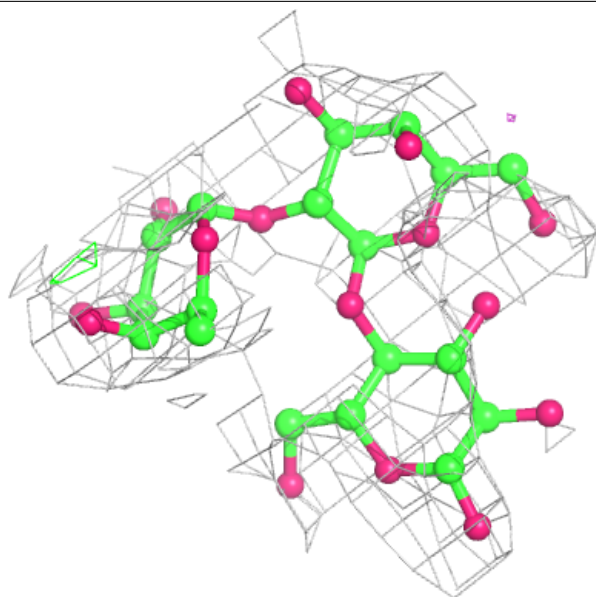
Electron density around Chain F:

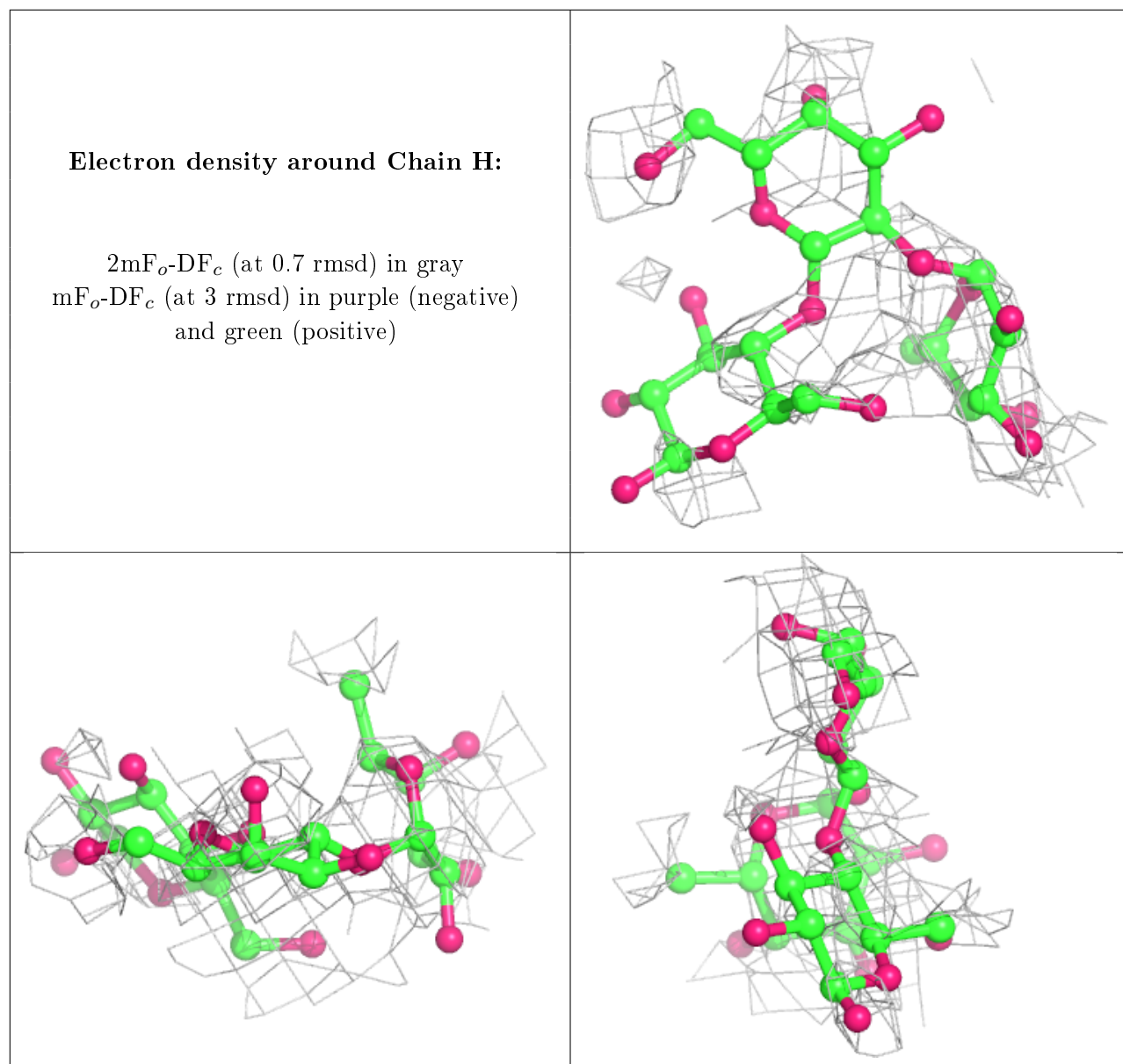
$2mF_o-DF_c$ (at 0.7 rmsd) in gray
 mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)



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)





6.4 Ligands [i](#)

Unable to reproduce the depositor's R factor - this section is therefore empty.

6.5 Other polymers [i](#)

Unable to reproduce the depositor's R factor - this section is therefore empty.