



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 10, 2026 – 12:53 PM UTC

PDB ID : 4FNU / pdb_00004fnu
Title : Crystal structure of GH36 alpha-galactosidase AgaA A355E D478A from *Geobacillus stearothermophilus* in complex with stachyose
Authors : Merceron, R.; Foucault, M.; Haser, R.; Mattes, R.; Watzlawick, H.; Gouet, P.
Deposited on : 2012-06-20
Resolution : 3.60 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4-5-2 with Phenix2.0
Mogul : 2022.3.0, CSD as543be (2022)
Xtriage (Phenix) : 2.0
EDS : 3.0
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4 : 9.0.010 (Gargrove)
Density-Fitness : 1.0.12
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.49

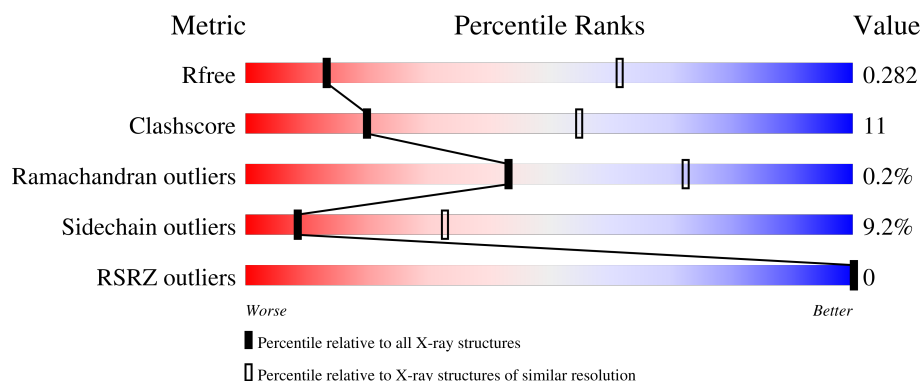
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.60 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}	180053	1747 (3.70-3.50)
Clashscore	190562	1827 (3.70-3.50)
Ramachandran outliers	187476	1773 (3.70-3.50)
Sidechain outliers	187428	1772 (3.70-3.50)
RSRZ outliers	180081	1745 (3.70-3.50)

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 $\leq 5\%$. 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	729	 65% 29% . .
1	B	729	 67% 27% . .
1	C	729	 68% 28% . .
1	D	729	 69% 26% . .
2	E	4	 50% 50%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	4	 100%
2	G	4	 50%50%
2	H	4	 100%

2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 23332 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.

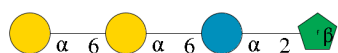
- Molecule 1 is a protein called Alpha-galactosidase AgaA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	718	Total	C	N	O	S	0	0	0
			5788	3675	1020	1070	23			
1	B	718	Total	C	N	O	S	0	0	0
			5788	3675	1020	1070	23			
1	C	718	Total	C	N	O	S	0	0	0
			5788	3675	1020	1070	23			
1	D	718	Total	C	N	O	S	0	0	0
			5788	3675	1020	1070	23			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	355	GLU	ALA	engineered mutation	UNP Q9ALJ4
A	478	ALA	ASP	engineered mutation	UNP Q9ALJ4
B	355	GLU	ALA	engineered mutation	UNP Q9ALJ4
B	478	ALA	ASP	engineered mutation	UNP Q9ALJ4
C	355	GLU	ALA	engineered mutation	UNP Q9ALJ4
C	478	ALA	ASP	engineered mutation	UNP Q9ALJ4
D	355	GLU	ALA	engineered mutation	UNP Q9ALJ4
D	478	ALA	ASP	engineered mutation	UNP Q9ALJ4

- Molecule 2 is an oligosaccharide called beta-D-fructofuranose-(2-1)-alpha-D-glucopyranose-(1-6)-alpha-D-galactopyranose-(1-6)-alpha-D-galactopyranose.



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
2	E	4	Total	C	O	0	0	0
			45	24	21			

Continued on next page...

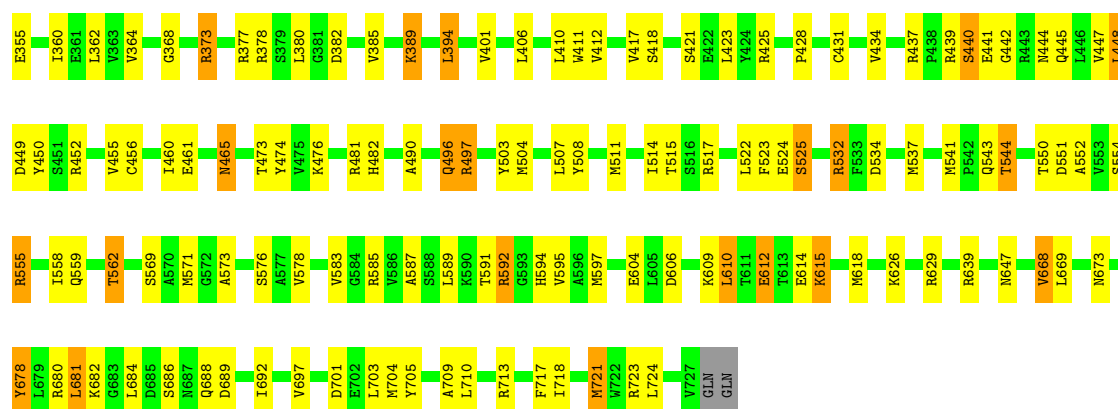
Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
2	F	4	Total	C	O	0	0	0
			45	24	21			
2	G	4	Total	C	O	0	0	0
			45	24	21			
2	H	4	Total	C	O	0	0	0
			45	24	21			

- Molecule 1: Alpha-galactosidase AgaA

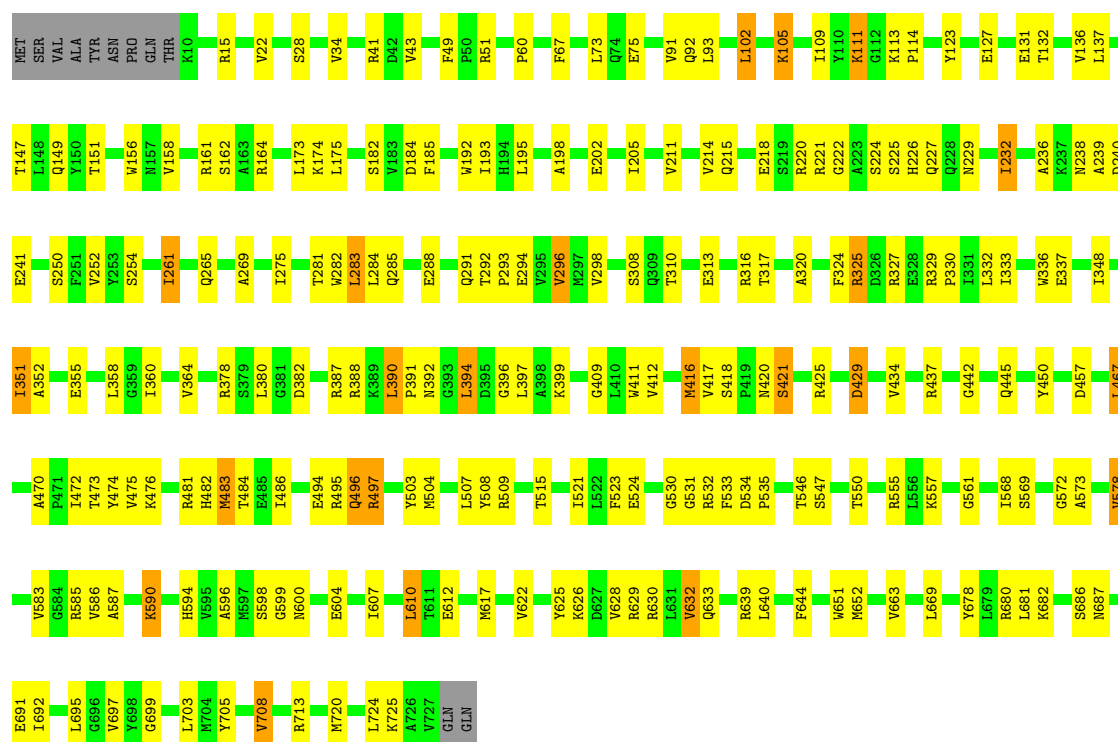
- Molecule 1: Alpha-galactosidase AgaA





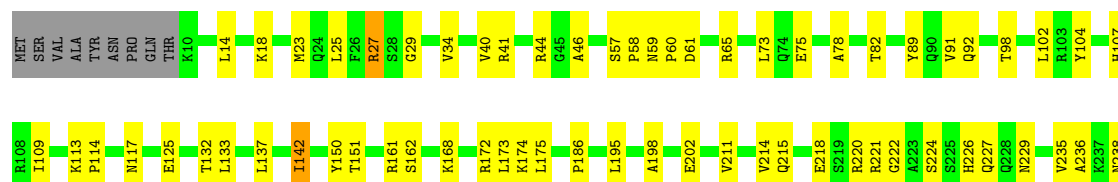
• Molecule 1: Alpha-galactosidase AgaA

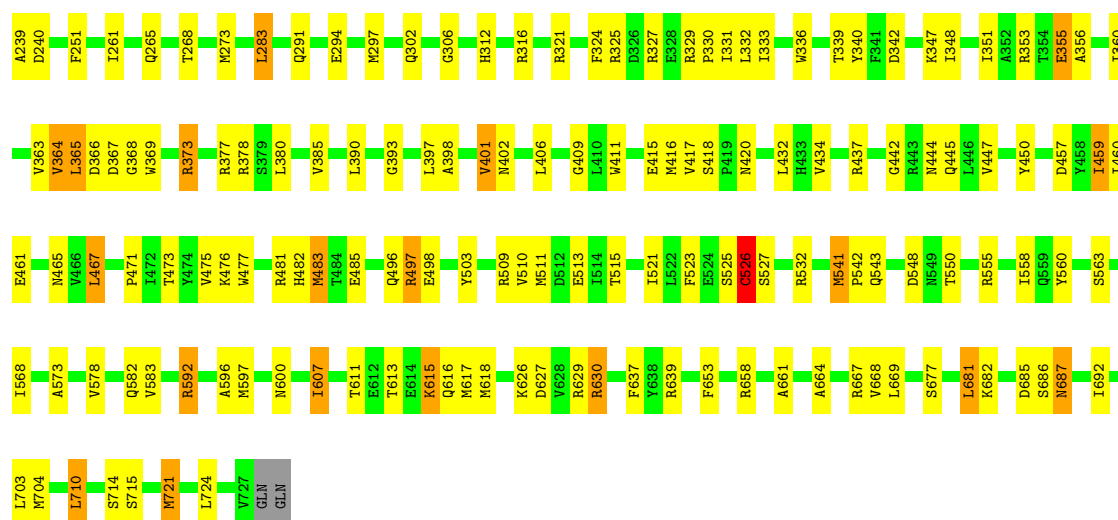
Chain C: 68% 28%



• Molecule 1: Alpha-galactosidase AgaA

Chain D: 69% 26%



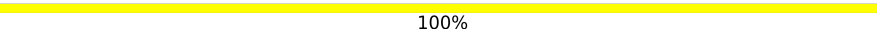


- Molecule 2: beta-D-fructofuranose-(2-1)-alpha-D-glucopyranose-(1-6)-alpha-D-galactopyranose-(1-6)-alpha-D-galactopyranose

Chain E: 



- Molecule 2: beta-D-fructofuranose-(2-1)-alpha-D-glucopyranose-(1-6)-alpha-D-galactopyranose-(1-6)-alpha-D-galactopyranose

Chain F: 

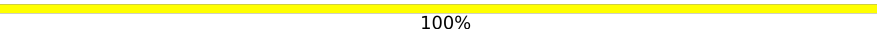


- Molecule 2: beta-D-fructofuranose-(2-1)-alpha-D-glucopyranose-(1-6)-alpha-D-galactopyranose-(1-6)-alpha-D-galactopyranose

Chain G: 



- Molecule 2: beta-D-fructofuranose-(2-1)-alpha-D-glucopyranose-(1-6)-alpha-D-galactopyranose-(1-6)-alpha-D-galactopyranose

Chain H: 



4 Data and refinement statistics

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α , β , γ	154.07Å 154.07Å 238.02Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	19.92 – 3.60 19.92 – 3.60	Depositor EDS
% Data completeness (in resolution range)	99.7 (19.92-3.60) 99.0 (19.92-3.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.07 (at 3.62Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, R_{free}	0.238 , 0.282 0.238 , 0.282	Depositor DCC
R_{free} test set	1906 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	93.0	Xtriage
Anisotropy	0.223	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.24 , 19.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.024 for -h,-k,l	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	23332	wwPDB-VP
Average B, all atoms (Å ²)	101.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.35% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GLC, FRU, GLA

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	0.28	0/5935	0.73	2/8046 (0.0%)
1	B	0.27	0/5935	0.73	3/8046 (0.0%)
1	C	0.28	0/5935	0.73	2/8046 (0.0%)
1	D	0.28	0/5935	0.73	7/8046 (0.1%)
All	All	0.28	0/23740	0.73	14/32184 (0.0%)

There are no bond length outliers.

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	525	SER	N-CA-C	8.35	122.72	112.54
1	D	526	CYS	N-CA-C	5.90	123.37	110.80
1	B	27	ARG	CB-CA-C	-5.69	110.03	116.63
1	B	541	MET	CA-C-N	5.48	125.19	119.82
1	B	541	MET	C-N-CA	5.48	125.19	119.82

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	5788	0	5621	154	0
1	B	5788	0	5621	139	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	5788	0	5621	143	0
1	D	5788	0	5621	128	0
2	E	45	0	39	4	0
2	F	45	0	39	0	0
2	G	45	0	39	1	0
2	H	45	0	39	0	0
All	All	23332	0	22640	525	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 11.

The worst 5 of 525 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:B:681:LEU:HD11	1:B:703:LEU:HB3	1.61	0.81
1:B:113:LYS:HD2	1:B:151:THR:HG21	1.66	0.78
1:A:113:LYS:HD2	1:A:151:THR:HG21	1.66	0.77
1:C:348:ILE:HD11	1:C:397:LEU:HD11	1.65	0.76
1:D:667:ARG:NH2	1:D:715:SER:O	2.20	0.75

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	716/729 (98%)	683 (95%)	32 (4%)	1 (0%)	48 79
1	B	716/729 (98%)	685 (96%)	30 (4%)	1 (0%)	48 79
1	C	716/729 (98%)	683 (95%)	32 (4%)	1 (0%)	48 79
1	D	716/729 (98%)	682 (95%)	32 (4%)	2 (0%)	36 65
All	All	2864/2916 (98%)	2733 (95%)	126 (4%)	5 (0%)	43 72

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	496	GLN
1	B	496	GLN
1	C	496	GLN
1	D	496	GLN
1	D	526	CYS

5.3.2 Protein sidechains ⓘ

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	608/618 (98%)	543 (89%)	65 (11%)	6	28
1	B	608/618 (98%)	545 (90%)	63 (10%)	7	29
1	C	608/618 (98%)	560 (92%)	48 (8%)	11	37
1	D	608/618 (98%)	561 (92%)	47 (8%)	12	38
All	All	2432/2472 (98%)	2209 (91%)	223 (9%)	8	33

5 of 223 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	610	LEU
1	D	710	LEU
1	C	284	LEU
1	D	687	ASN
1	D	417	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	149	GLN
1	C	194	HIS
1	D	580	ASN
1	C	600	ASN
1	A	302	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

16 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	GLA	E	1	2	11,11,12	2.48	2 (18%)	15,15,17	0.90	0
2	GLA	E	2	2	11,11,12	2.42	2 (18%)	15,15,17	1.56	2 (13%)
2	GLC	E	3	2	11,11,12	2.30	2 (18%)	15,15,17	1.19	1 (6%)
2	FRU	E	4	2	11,12,12	1.71	4 (36%)	10,18,18	1.65	2 (20%)
2	GLA	F	1	2	11,11,12	2.53	3 (27%)	15,15,17	1.09	0
2	GLA	F	2	2	11,11,12	2.38	3 (27%)	15,15,17	2.18	7 (46%)
2	GLC	F	3	2	11,11,12	2.51	4 (36%)	15,15,17	1.42	3 (20%)
2	FRU	F	4	2	11,12,12	1.72	4 (36%)	10,18,18	1.96	4 (40%)
2	GLA	G	1	2	11,11,12	2.52	2 (18%)	15,15,17	1.19	1 (6%)
2	GLA	G	2	2	11,11,12	2.39	2 (18%)	15,15,17	1.96	5 (33%)
2	GLC	G	3	2	11,11,12	2.25	2 (18%)	15,15,17	1.94	6 (40%)
2	FRU	G	4	2	11,12,12	1.69	2 (18%)	10,18,18	1.90	4 (40%)
2	GLA	H	1	2	11,11,12	2.47	3 (27%)	15,15,17	1.20	1 (6%)
2	GLA	H	2	2	11,11,12	2.29	3 (27%)	15,15,17	2.04	6 (40%)
2	GLC	H	3	2	11,11,12	2.35	3 (27%)	15,15,17	1.60	3 (20%)
2	FRU	H	4	2	11,12,12	1.73	3 (27%)	10,18,18	2.03	4 (40%)

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	GLA	E	1	2	-	2/2/19/22	0/1/1/1
2	GLA	E	2	2	-	2/2/19/22	0/1/1/1
2	GLC	E	3	2	-	2/2/19/22	0/1/1/1
2	FRU	E	4	2	-	3/5/24/24	0/1/1/1
2	GLA	F	1	2	-	1/2/19/22	0/1/1/1
2	GLA	F	2	2	-	2/2/19/22	0/1/1/1
2	GLC	F	3	2	-	2/2/19/22	0/1/1/1
2	FRU	F	4	2	-	3/5/24/24	0/1/1/1
2	GLA	G	1	2	-	0/2/19/22	0/1/1/1
2	GLA	G	2	2	-	0/2/19/22	0/1/1/1
2	GLC	G	3	2	-	1/2/19/22	0/1/1/1
2	FRU	G	4	2	-	5/5/24/24	0/1/1/1
2	GLA	H	1	2	-	1/2/19/22	0/1/1/1
2	GLA	H	2	2	-	1/2/19/22	0/1/1/1
2	GLC	H	3	2	-	0/2/19/22	0/1/1/1
2	FRU	H	4	2	-	5/5/24/24	0/1/1/1

The worst 5 of 44 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	G	1	GLA	C2-C3	-7.08	1.41	1.52
2	F	1	GLA	C2-C3	-7.05	1.41	1.52
2	E	1	GLA	C2-C3	-6.97	1.41	1.52
2	E	2	GLA	C2-C3	-6.84	1.42	1.52
2	H	1	GLA	C2-C3	-6.75	1.42	1.52

The worst 5 of 49 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	G	3	GLC	C1-C2-C3	4.38	116.02	109.64
2	F	2	GLA	C1-C2-C3	4.12	115.65	109.64
2	H	2	GLA	C1-O5-C5	3.96	117.50	112.19
2	H	4	FRU	O1-C1-C2	3.96	120.43	111.67
2	F	4	FRU	O1-C1-C2	3.89	120.28	111.67

There are no chirality outliers.

5 of 30 torsion outliers are listed below:

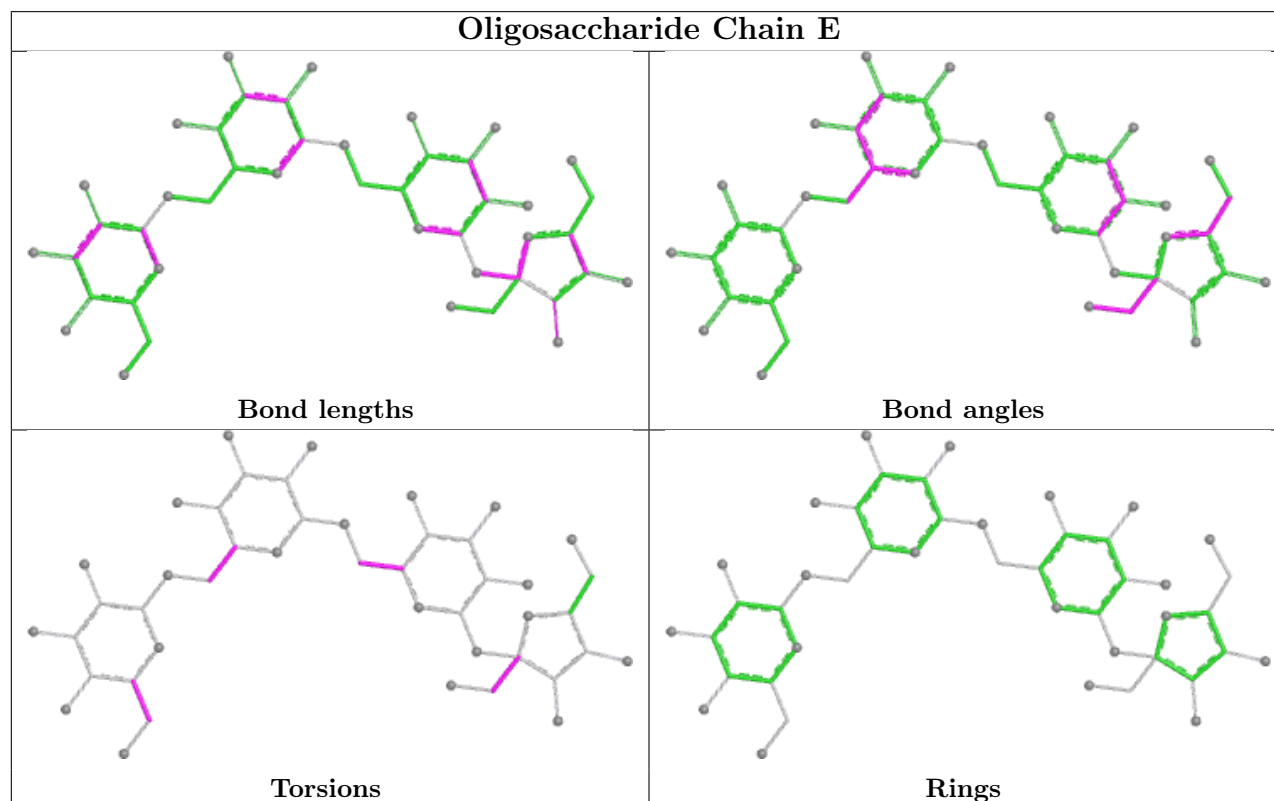
Mol	Chain	Res	Type	Atoms
2	E	4	FRU	O1-C1-C2-C3
2	E	4	FRU	O1-C1-C2-O2
2	E	4	FRU	O1-C1-C2-O5
2	G	4	FRU	O1-C1-C2-C3
2	G	4	FRU	O1-C1-C2-O2

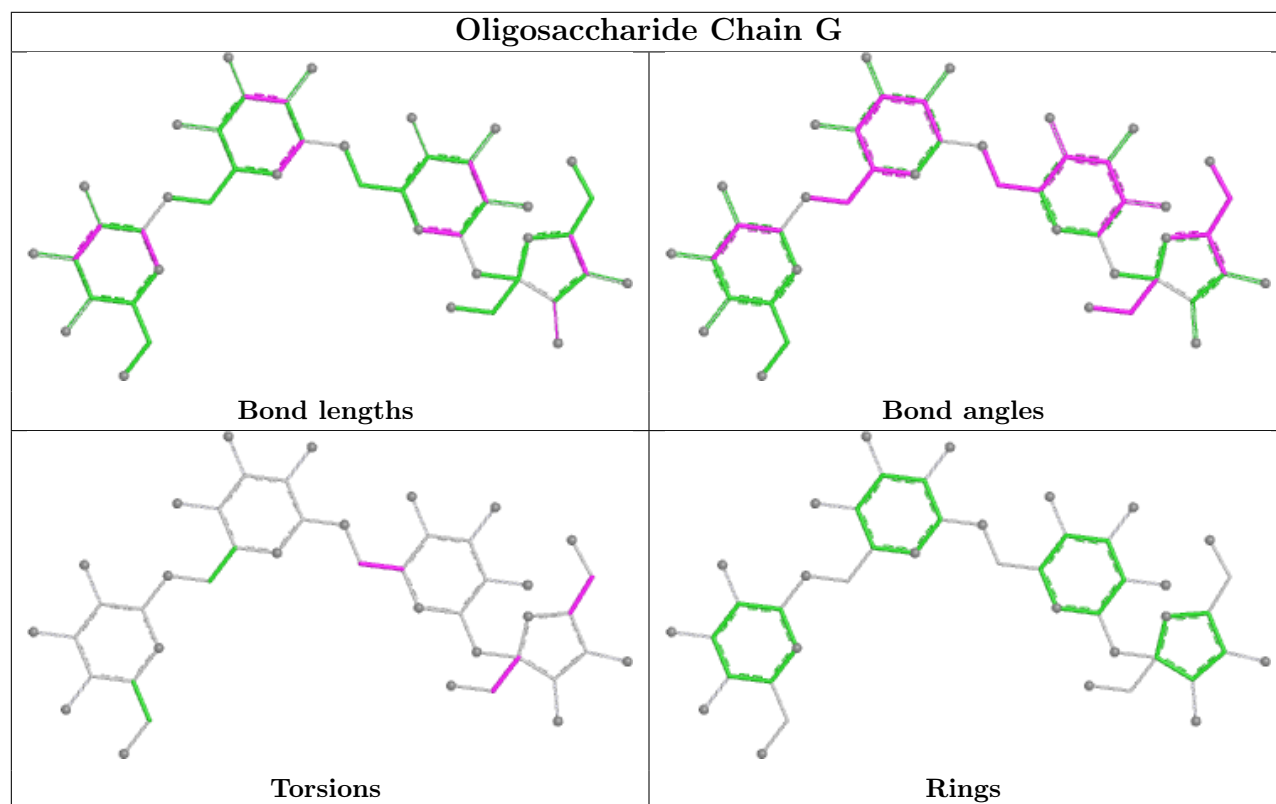
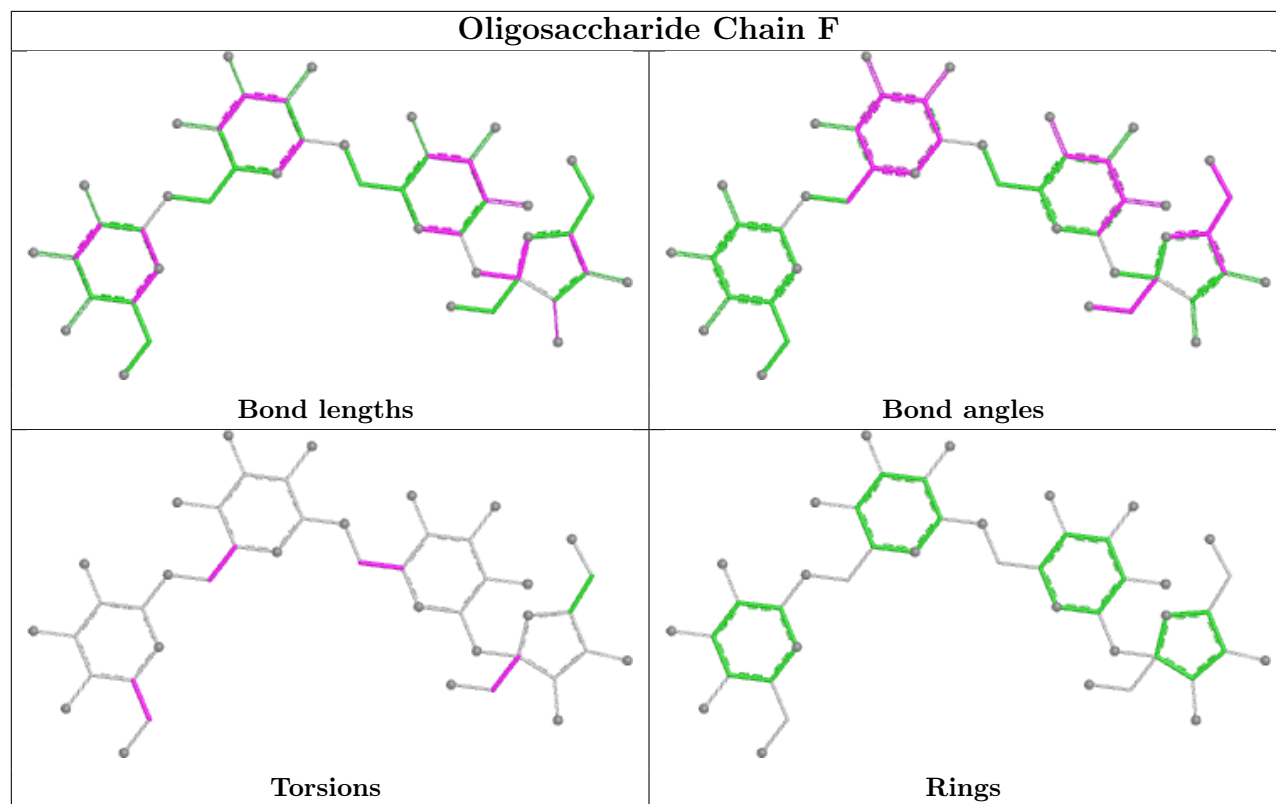
There are no ring outliers.

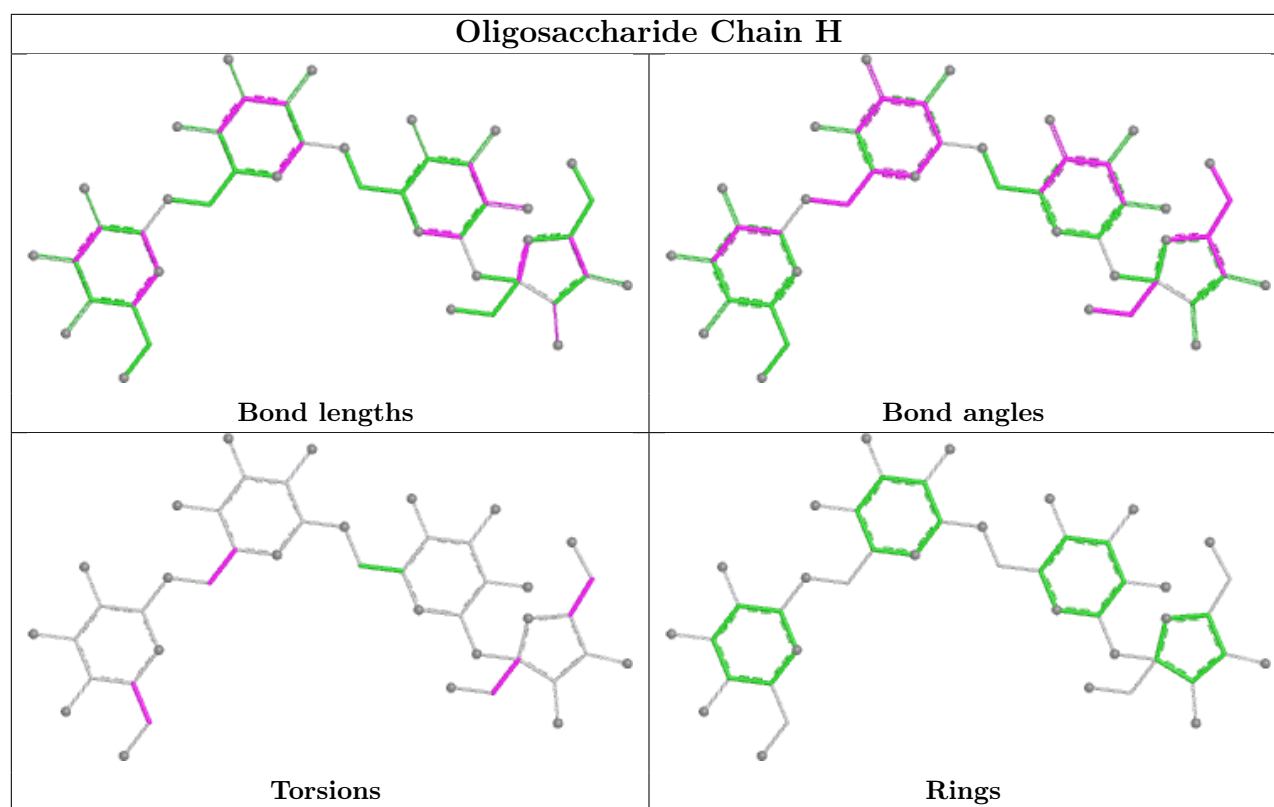
4 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	E	2	GLA	2	0
2	G	2	GLA	1	0
2	E	1	GLA	2	0
2	G	4	FRU	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.







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](#)

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	718/729 (98%)	-0.32	0 100 100	61, 101, 145, 173	0
1	B	718/729 (98%)	-0.44	0 100 100	56, 90, 127, 144	0
1	C	718/729 (98%)	-0.30	0 100 100	71, 113, 139, 163	0
1	D	718/729 (98%)	-0.39	0 100 100	62, 93, 126, 151	0
All	All	2872/2916 (98%)	-0.36	0 100 100	56, 100, 136, 173	0

There are no RSRZ outliers to report.

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.

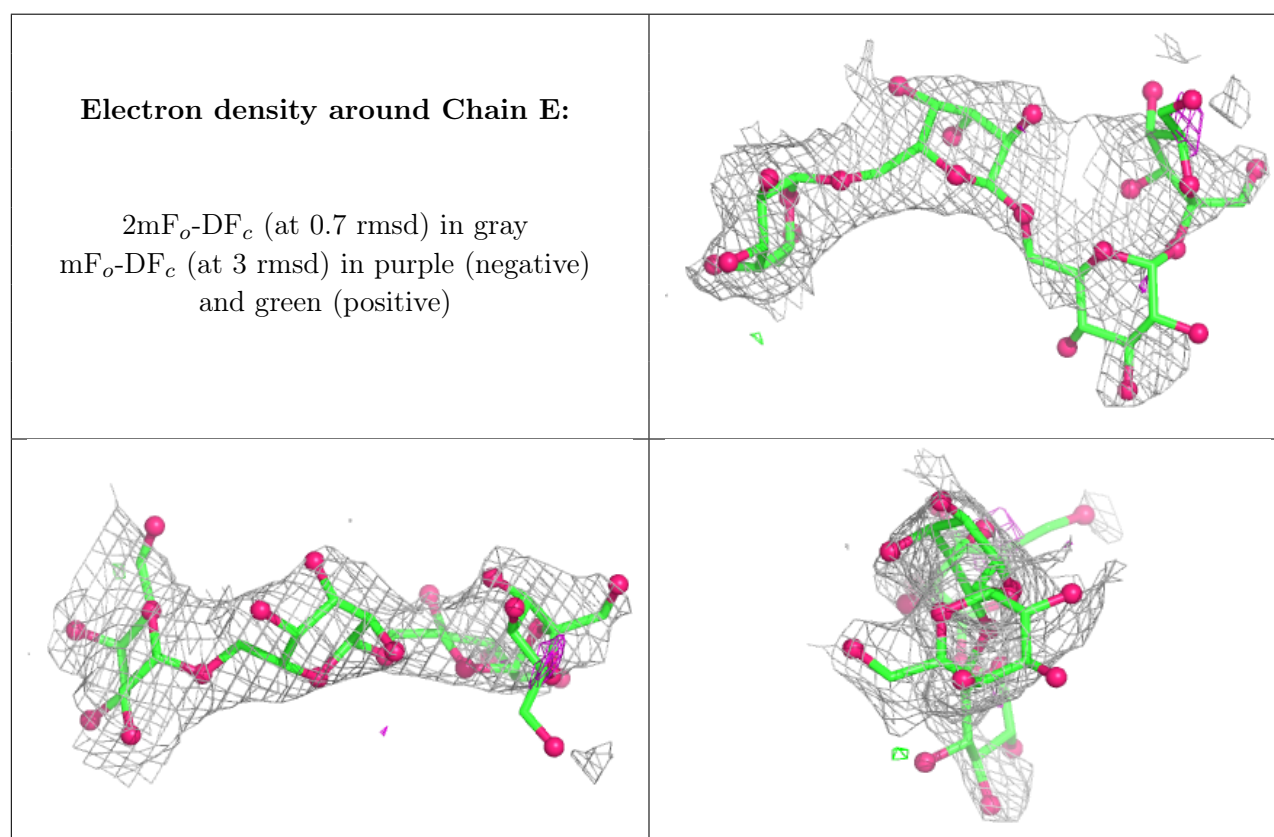
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
2	FRU	G	4	12/12	0.42	0.11	99,131,142,146	0
2	FRU	F	4	12/12	0.52	0.10	104,131,137,140	0
2	GLC	G	3	11/12	0.56	0.11	124,139,148,150	0
2	GLC	E	3	11/12	0.61	0.10	125,141,146,151	0
2	GLA	E	1	11/12	0.67	0.09	105,123,130,131	0
2	FRU	E	4	12/12	0.70	0.12	117,138,149,157	0
2	GLA	G	1	11/12	0.70	0.11	114,119,123,125	0
2	FRU	H	4	12/12	0.71	0.10	102,121,135,143	0
2	GLC	H	3	11/12	0.73	0.10	109,129,137,138	0

Continued on next page...

Continued from previous page...

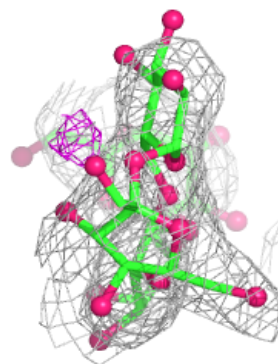
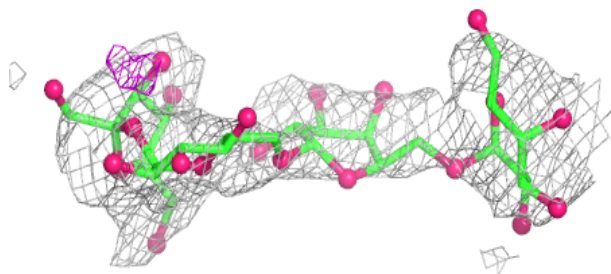
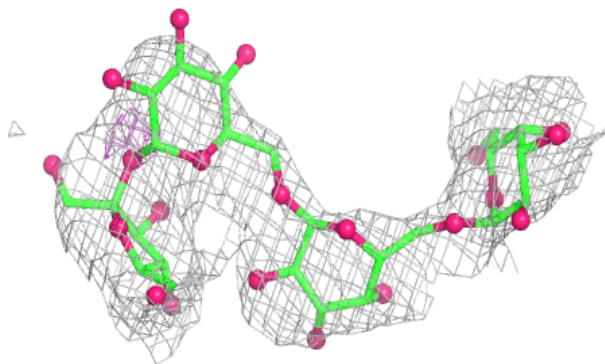
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($\text{\AA}^2$)	Q<0.9
2	GLA	E	2	11/12	0.78	0.10	106,116,119,121	0
2	GLA	F	1	11/12	0.80	0.10	95,103,109,110	0
2	GLC	F	3	11/12	0.80	0.09	122,129,136,138	0
2	GLA	F	2	11/12	0.82	0.12	104,109,124,127	0
2	GLA	G	2	11/12	0.87	0.11	99,110,115,119	0
2	GLA	H	2	11/12	0.87	0.10	94,102,110,110	0
2	GLA	H	1	11/12	0.88	0.08	87,98,102,103	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.

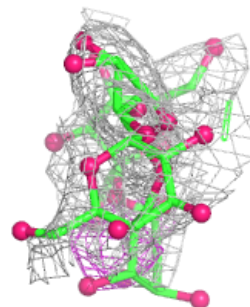
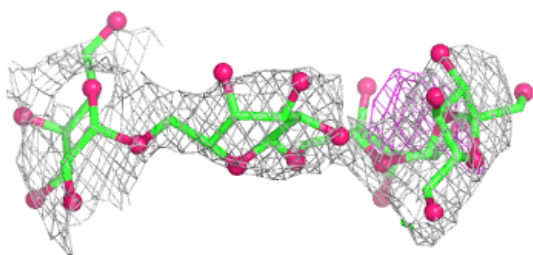
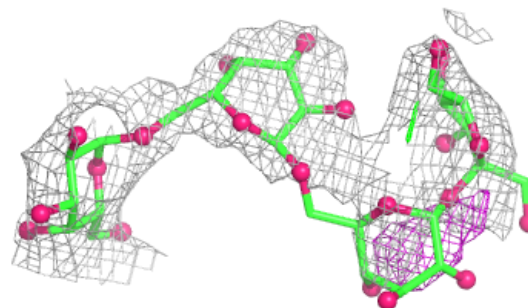


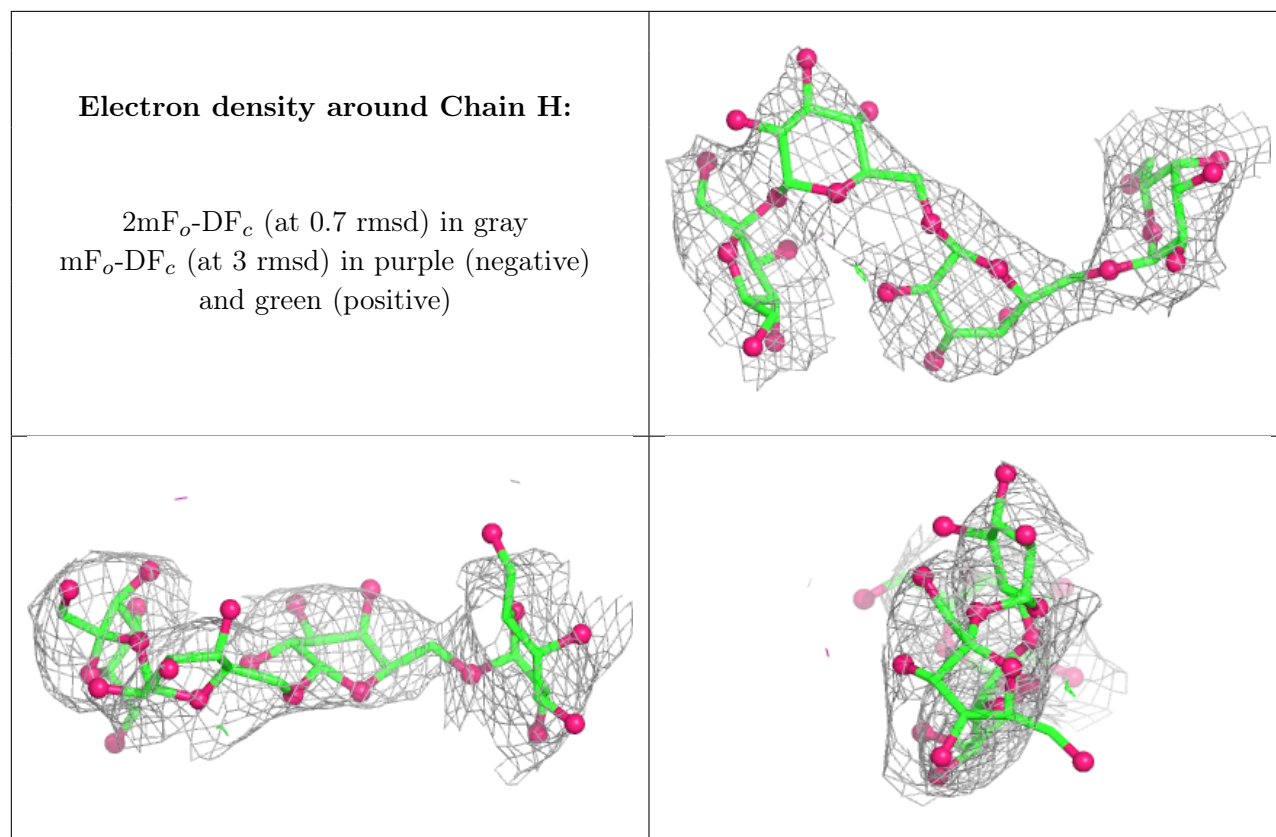
Electron density around Chain F:

$2mF_o - DF_c$ (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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.