



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 8, 2020 – 02:09 AM BST

PDB ID : 1TEX  
Title : Mycobacterium smegmatis Stf0 Sulfotransferase with Trehalose  
Authors : Mougous, J.D.; Petzold, C.J.; Senaratne, R.H.; Lee, D.H.; Akey, D.L.; Lin, F.L.; Munchel, S.E.; Pratt, M.R.; Riley, L.W.; Leary, J.A.; Berger, J.M.; Bertozzi, C.R.  
Deposited on : 2004-05-25  
Resolution : 2.60 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](mailto: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

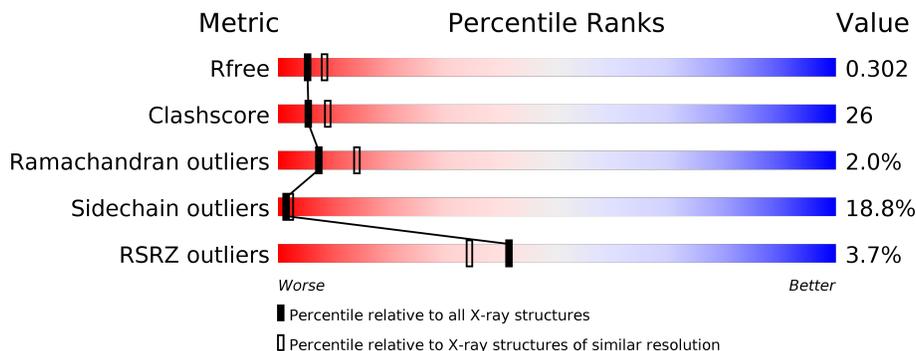
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.60 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.



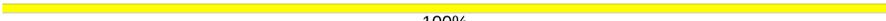
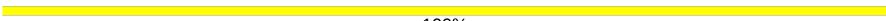
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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  $\geq 3$ , 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	287	 7% 43% 30% 11% 17%
1	B	287	 2% 44% 30% 10% 15%
1	C	287	 % 54% 24% 7% 15%
1	D	287	 2% 53% 28% 6% 14%
2	E	2	 50% 50%
2	F	2	 50% 50%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	G	2	 100%
2	H	2	 100%

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 7919 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 Stf0 Sulfotransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	239	Total 1794	C 1145	N 312	O 334	S 3	0	0	0
1	B	243	Total 1853	C 1183	N 327	O 340	S 3	0	0	0
1	C	244	Total 1883	C 1205	N 326	O 349	S 3	0	0	0
1	D	247	Total 1887	C 1212	N 329	O 343	S 3	0	0	0

- Molecule 2 is an oligosaccharide called alpha-D-glucopyranose-(1-1)-alpha-D-glucopyranose.



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
			Total	C	O			
2	E	2	Total 23	C 12	O 11	0	0	0
2	F	2	Total 23	C 12	O 11	0	0	0
2	G	2	Total 23	C 12	O 11	0	0	0
2	H	2	Total 23	C 12	O 11	0	0	0

- Molecule 3 is water.

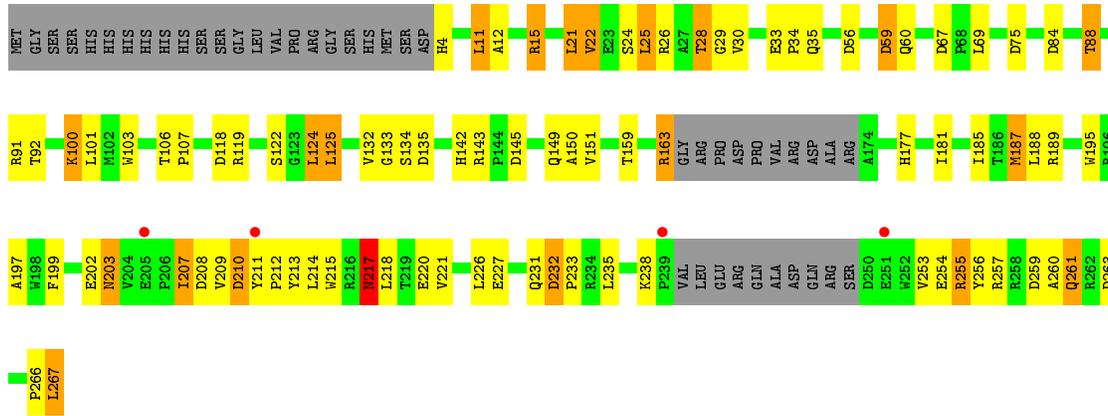
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	63	Total 63	O 63	0	0
3	B	113	Total 113	O 113	0	0

*Continued on next page...*

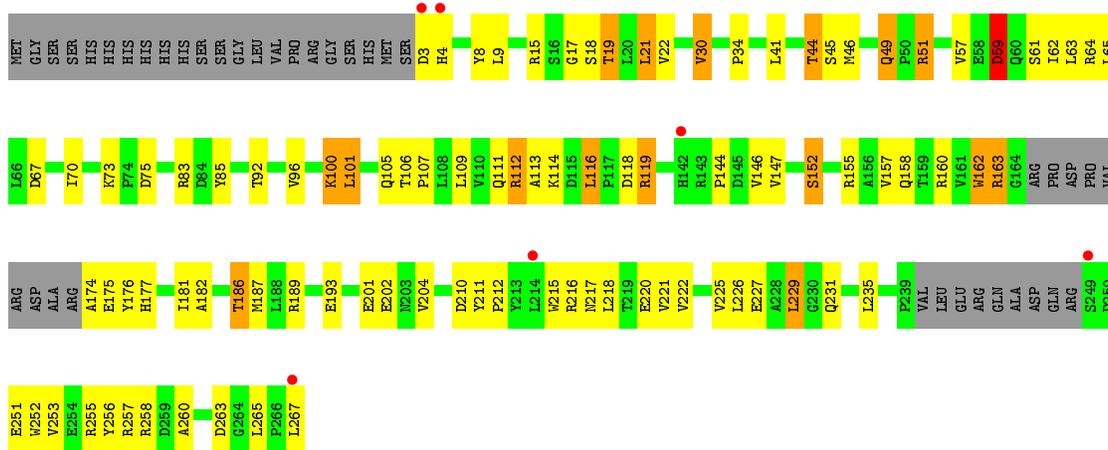
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
3	C	138	Total 138	O 138	0	0
3	D	96	Total 96	O 96	0	0





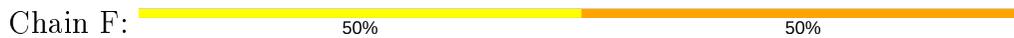
- Molecule 1: Stf0 Sulfotransferase



- Molecule 2: alpha-D-glucopyranose-(1-1)-alpha-D-glucopyranose



- Molecule 2: alpha-D-glucopyranose-(1-1)-alpha-D-glucopyranose



- Molecule 2: alpha-D-glucopyranose-(1-1)-alpha-D-glucopyranose



- Molecule 2: alpha-D-glucopyranose-(1-1)-alpha-D-glucopyranose

Chain H:

100%

GLC1  
GLC2

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	101.93Å 101.93Å 228.21Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	15.00 – 2.60 14.99 – 2.60	Depositor EDS
% Data completeness (in resolution range)	100.0 (15.00-2.60) 97.6 (14.99-2.60)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.44 (at 2.61Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, $R_{free}$	0.218 , 0.274 0.255 , 0.302	Depositor DCC
$R_{free}$ test set	2106 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	63.6	Xtriage
Anisotropy	0.002	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 50.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.024 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	7919	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.61% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: GLC

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	0.75	0/1839	0.83	5/2520 (0.2%)
1	B	0.85	1/1899 (0.1%)	0.89	4/2601 (0.2%)
1	C	0.91	0/1933	1.01	11/2650 (0.4%)
1	D	0.86	2/1939 (0.1%)	0.90	4/2662 (0.2%)
All	All	0.85	3/7610 (0.0%)	0.91	24/10433 (0.2%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	85	TYR	CD2-CE2	5.44	1.47	1.39
1	B	252	TRP	CB-CG	5.43	1.60	1.50
1	D	162	TRP	CB-CG	-5.24	1.40	1.50

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	75	ASP	CB-CG-OD2	10.32	127.58	118.30
1	B	84	ASP	CB-CG-OD2	7.61	125.15	118.30
1	D	59	ASP	CB-CG-OD2	7.30	124.87	118.30
1	C	84	ASP	CB-CG-OD2	6.96	124.57	118.30
1	C	259	ASP	CB-CG-OD2	6.84	124.46	118.30

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	1794	0	1647	93	0
1	B	1853	0	1765	121	0
1	C	1883	0	1801	100	0
1	D	1887	0	1805	74	0
2	E	23	0	21	5	0
2	F	23	0	21	2	0
2	G	23	0	21	0	0
2	H	23	0	21	0	0
3	A	63	0	0	24	0
3	B	113	0	0	38	0
3	C	138	0	0	29	0
3	D	96	0	0	18	0
All	All	7919	0	7102	383	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 26.

The worst 5 of 383 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:122:SER:HB3	3:C:580:HOH:O	1.29	1.28
1:D:44:THR:CG2	1:D:46:MET:HB2	1.75	1.16
1:C:15:ARG:HG3	1:C:15:ARG:HH11	1.06	1.11
1:C:28:THR:HG22	1:C:30:VAL:H	1.18	1.07
1:B:249:SER:HA	3:B:513:HOH:O	1.53	1.06

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	228/287 (79%)	181 (79%)	38 (17%)	9 (4%)	3 4
1	B	235/287 (82%)	217 (92%)	14 (6%)	4 (2%)	9 18
1	C	238/287 (83%)	222 (93%)	12 (5%)	4 (2%)	9 18
1	D	241/287 (84%)	220 (91%)	19 (8%)	2 (1%)	19 39
All	All	942/1148 (82%)	840 (89%)	83 (9%)	19 (2%)	7 14

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	63	LEU
1	A	194	GLY
1	B	119	ARG
1	B	218	LEU
1	A	182	ALA

### 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	172/245 (70%)	131 (76%)	41 (24%)	0 1
1	B	185/245 (76%)	146 (79%)	39 (21%)	1 2
1	C	192/245 (78%)	168 (88%)	24 (12%)	4 8
1	D	190/245 (78%)	155 (82%)	35 (18%)	1 2
All	All	739/980 (75%)	600 (81%)	139 (19%)	1 2

5 of 139 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	124	LEU
1	B	255	ARG
1	D	186	THR
1	B	134	SER
1	B	200	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	149	GLN
1	B	183	HIS
1	D	39	GLN
1	B	87	GLN
1	B	111	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](#)

8 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	GLC	E	1	2	11,11,12	0.77	0	15,15,17	4.19	9 (60%)
2	GLC	E	2	2	12,12,12	1.00	0	17,17,17	3.10	10 (58%)
2	GLC	F	1	2	11,11,12	1.28	2 (18%)	15,15,17	4.63	12 (80%)
2	GLC	F	2	2	12,12,12	0.72	0	17,17,17	3.12	11 (64%)
2	GLC	G	1	2	11,11,12	1.57	2 (18%)	15,15,17	4.25	10 (66%)
2	GLC	G	2	2	12,12,12	0.78	0	17,17,17	3.14	12 (70%)
2	GLC	H	1	2	11,11,12	1.13	0	15,15,17	4.02	7 (46%)
2	GLC	H	2	2	12,12,12	0.94	1 (8%)	17,17,17	3.10	10 (58%)

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	GLC	E	1	2	-	2/2/19/22	0/1/1/1
2	GLC	E	2	2	-	1/2/22/22	0/1/1/1
2	GLC	F	1	2	-	1/2/19/22	0/1/1/1
2	GLC	F	2	2	-	0/2/22/22	0/1/1/1
2	GLC	G	1	2	-	1/2/19/22	0/1/1/1
2	GLC	G	2	2	-	2/2/22/22	0/1/1/1
2	GLC	H	1	2	-	0/2/19/22	0/1/1/1
2	GLC	H	2	2	-	0/2/22/22	0/1/1/1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	G	1	GLC	C2-C3	-3.55	1.47	1.52
2	F	1	GLC	O5-C1	-2.31	1.40	1.43
2	F	1	GLC	C2-C3	-2.11	1.49	1.52
2	G	1	GLC	C4-C3	-2.04	1.47	1.52
2	H	2	GLC	C3-C2	-2.02	1.47	1.52

The worst 5 of 81 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	F	1	GLC	C1-O5-C5	11.16	127.32	112.19
2	H	1	GLC	C1-O5-C5	10.14	125.94	112.19
2	E	1	GLC	C1-O5-C5	9.95	125.67	112.19
2	F	1	GLC	C1-C2-C3	9.09	120.84	109.67
2	G	1	GLC	C1-O5-C5	9.06	124.47	112.19

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

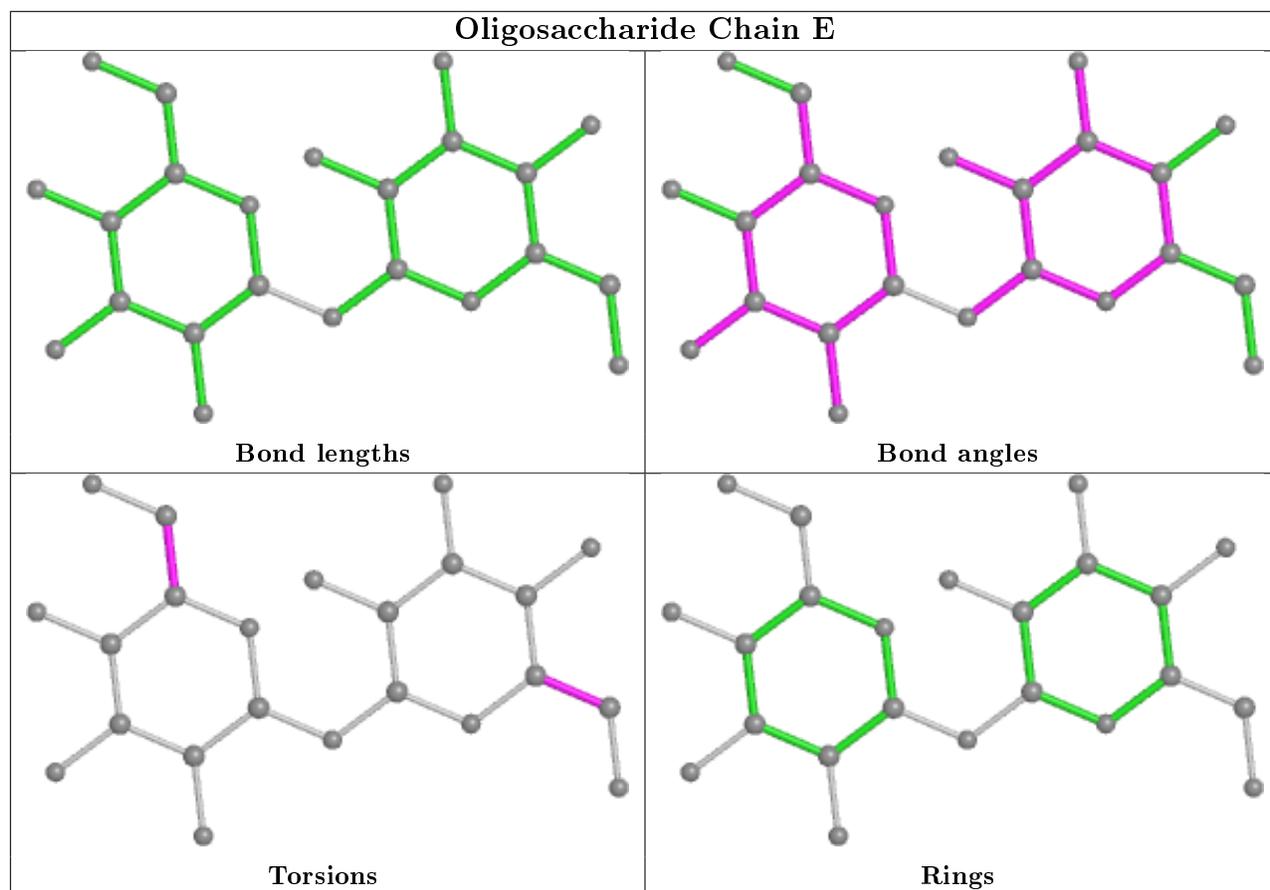
Mol	Chain	Res	Type	Atoms
2	E	1	GLC	O5-C5-C6-O6
2	E	1	GLC	C4-C5-C6-O6
2	E	2	GLC	O5-C5-C6-O6
2	G	2	GLC	C4-C5-C6-O6
2	G	2	GLC	O5-C5-C6-O6

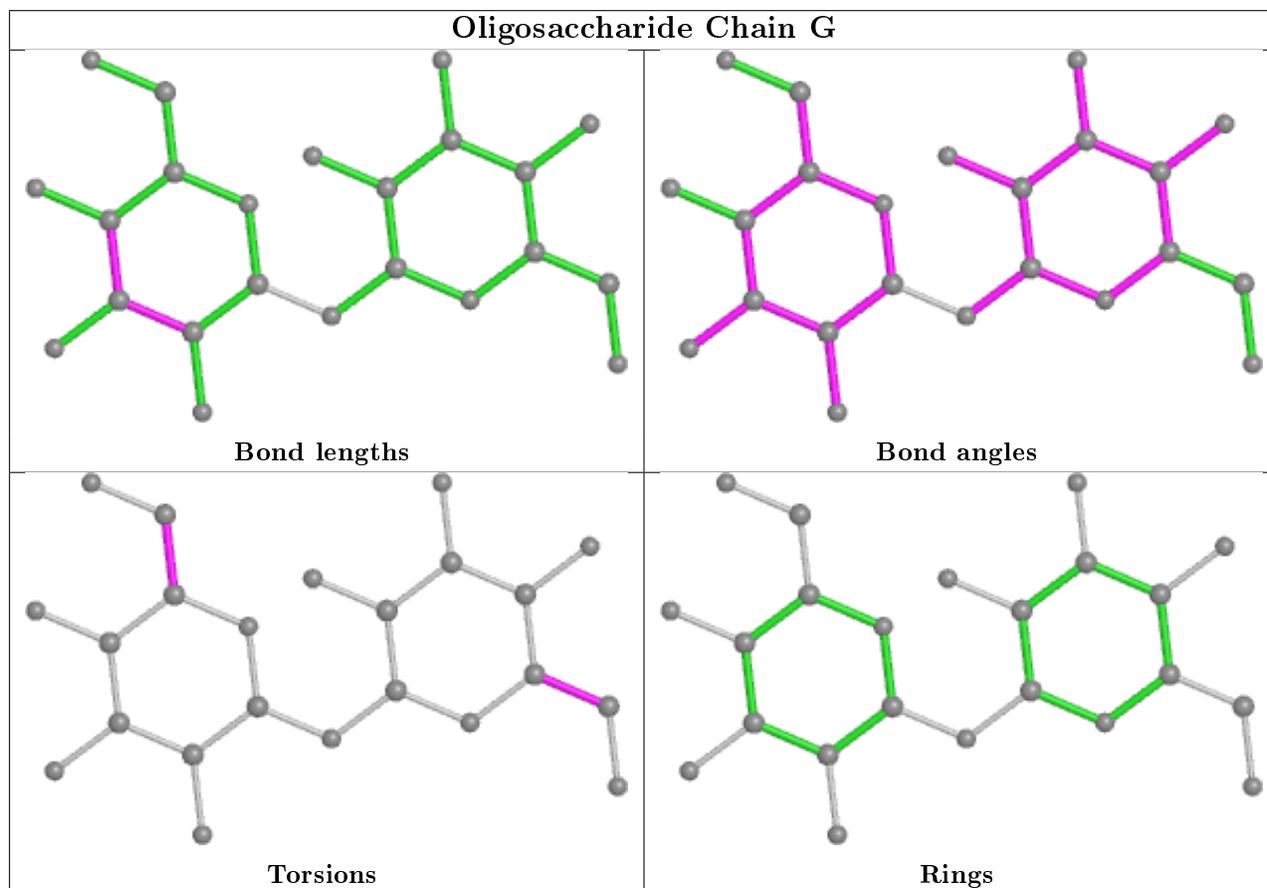
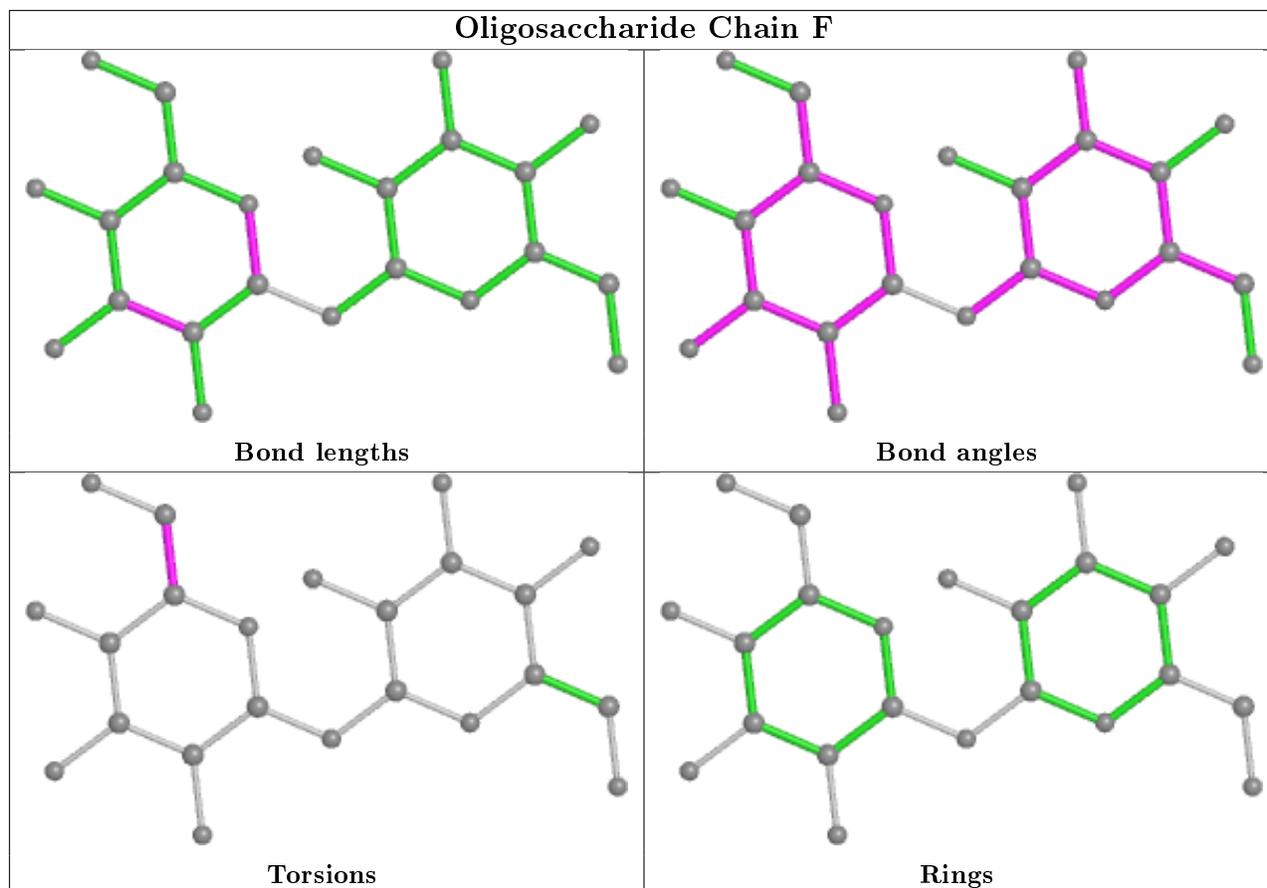
There are no ring outliers.

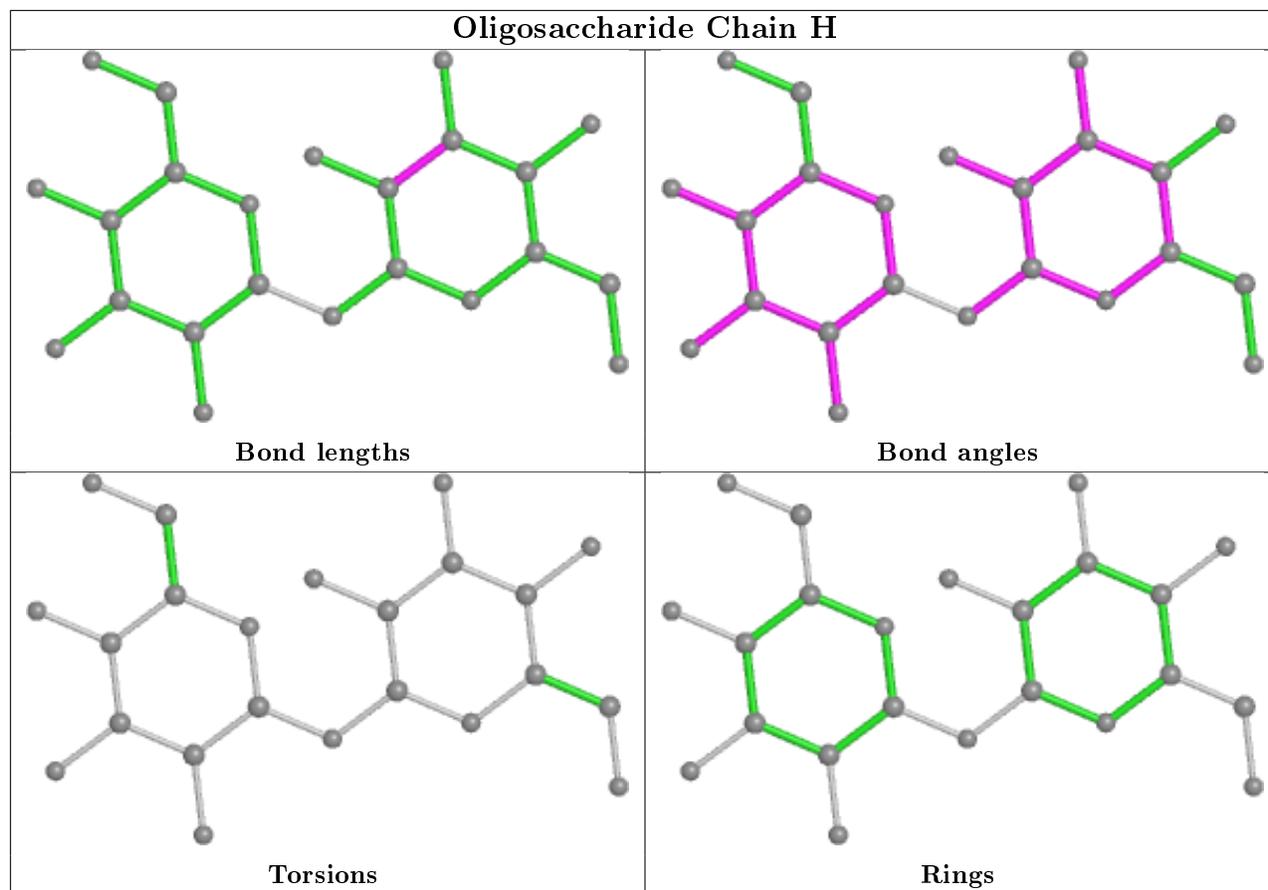
2 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	E	2	GLC	5	0
2	F	2	GLC	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](#)

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	239/287 (83%)	0.21	20 (8%) 11 7	20, 54, 93, 97	0
1	B	243/287 (84%)	-0.18	6 (2%) 57 51	19, 40, 76, 89	0
1	C	244/287 (85%)	-0.19	4 (1%) 72 68	12, 35, 62, 81	0
1	D	247/287 (86%)	-0.18	6 (2%) 59 53	17, 34, 58, 70	1 (0%)
All	All	973/1148 (84%)	-0.09	36 (3%) 41 34	12, 38, 79, 97	1 (0%)

The worst 5 of 36 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	264	GLY	5.4
1	D	3	ASP	4.5
1	A	265	LEU	4.4
1	B	249	SER	4.3
1	B	171	ASP	4.1

### 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, 95<sup>th</sup> 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.

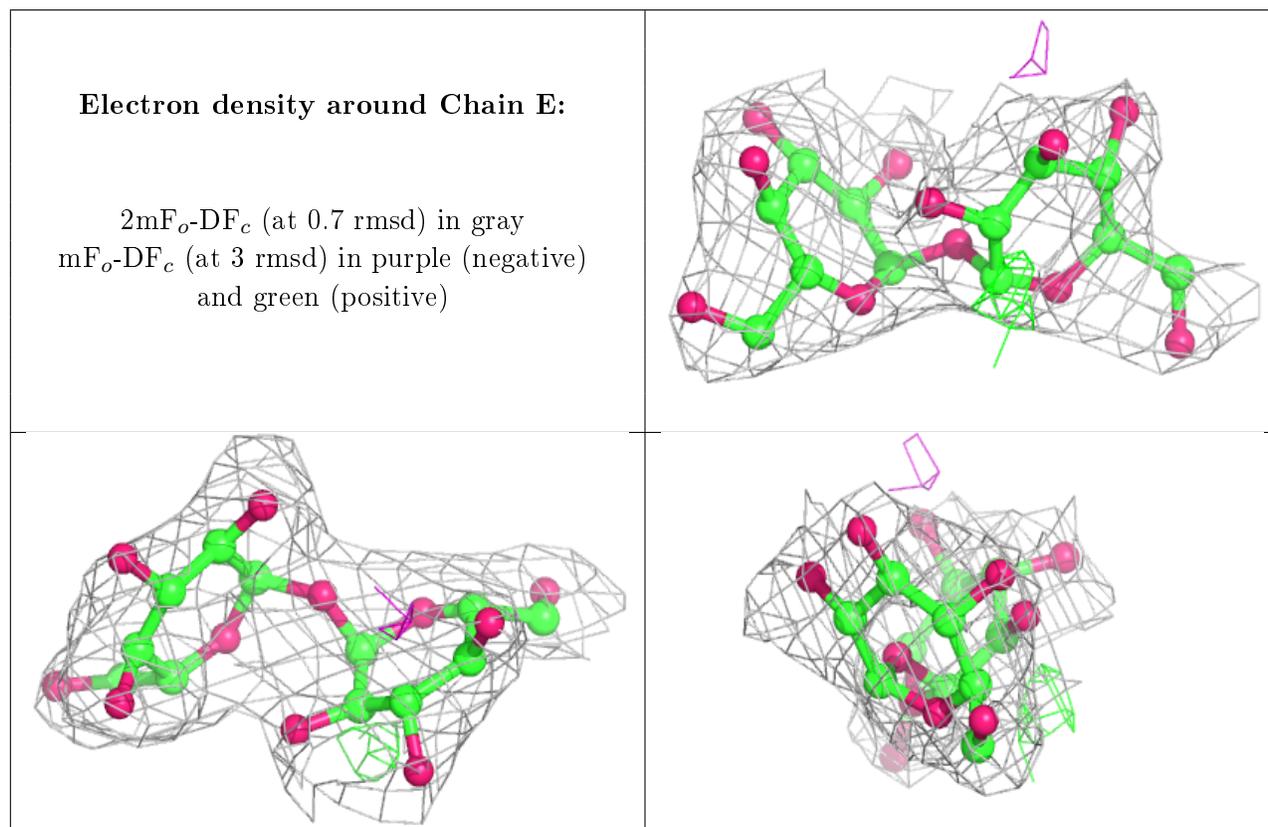
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
2	GLC	E	2	12/12	0.91	0.13	33,39,42,42	0
2	GLC	E	1	11/12	0.93	0.14	28,30,33,36	0

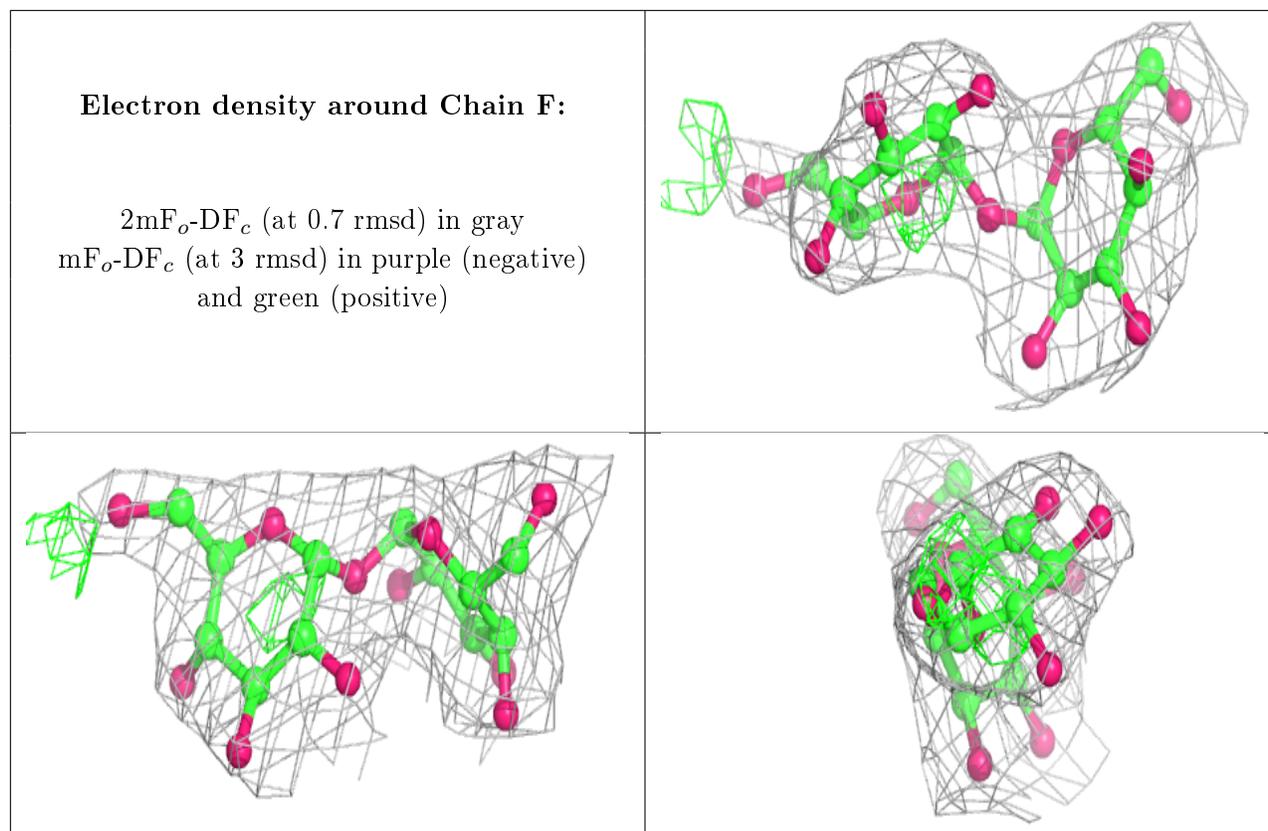
*Continued on next page...*

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
2	GLC	H	1	11/12	0.93	0.20	24,26,31,31	0
2	GLC	H	2	12/12	0.94	0.19	25,29,32,37	0
2	GLC	G	2	12/12	0.95	0.17	26,34,36,38	0
2	GLC	G	1	11/12	0.95	0.19	24,26,31,35	0
2	GLC	F	1	11/12	0.95	0.20	27,31,42,49	0
2	GLC	F	2	12/12	0.96	0.18	23,28,32,32	0

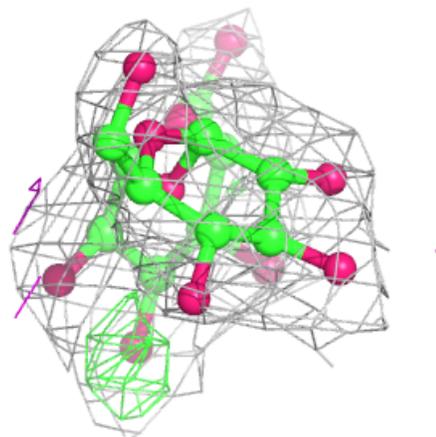
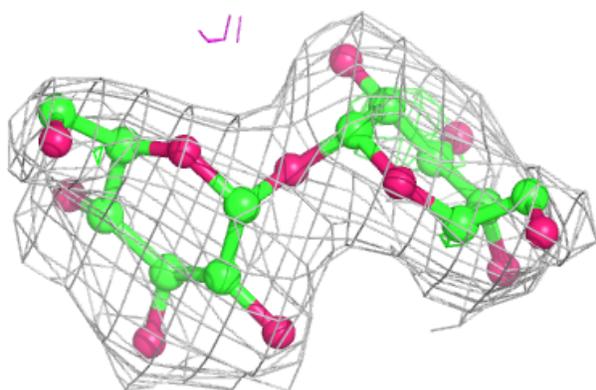
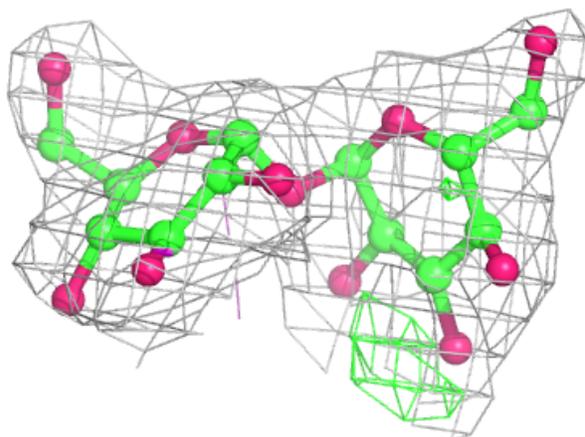
The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.

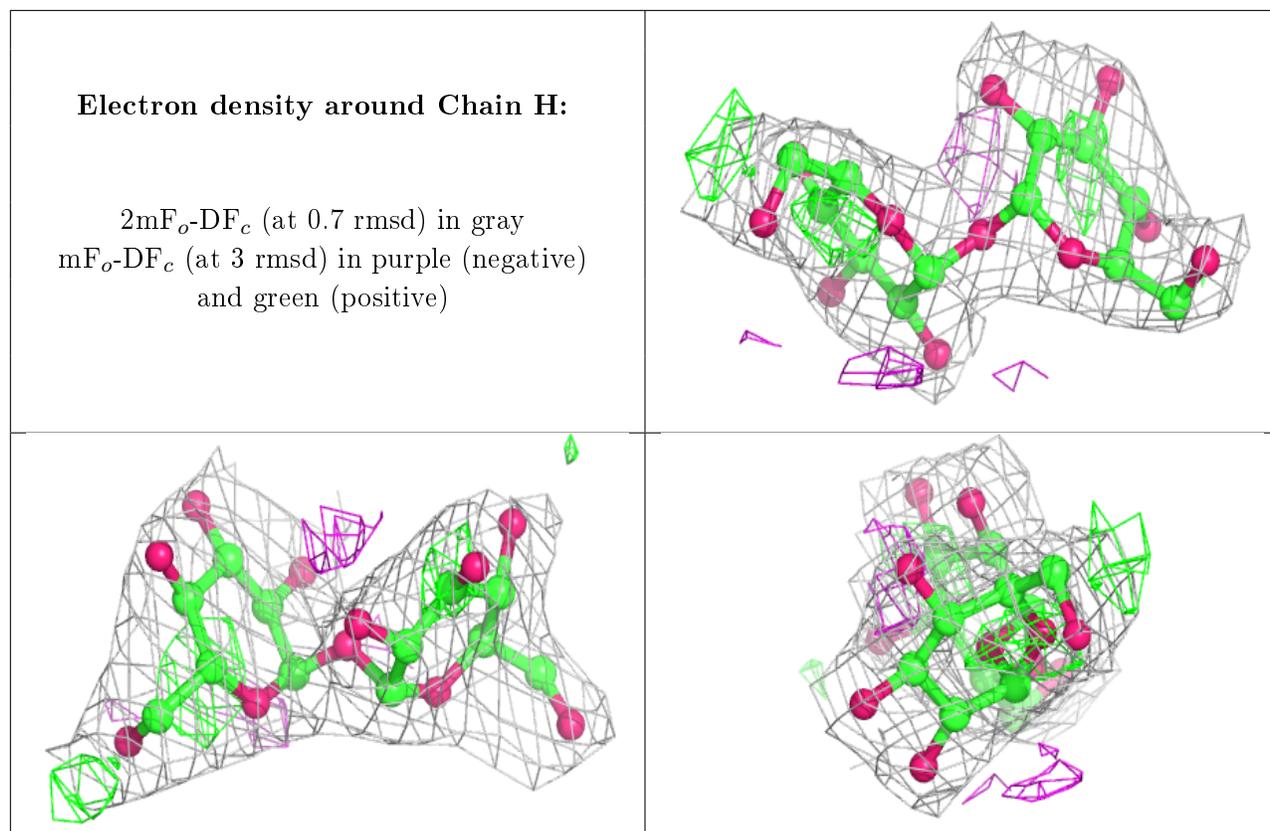




**Electron density around Chain G:**

$2mF_o-DF_c$  (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.