



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 31, 2023 – 08:21 PM JST

PDB ID : 5H06  
Title : Crystal structure of AmyP in complex with maltose  
Authors : He, C.; Liu, Y.  
Deposited on : 2016-10-03  
Resolution : 1.95 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](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 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 : 4.02b-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

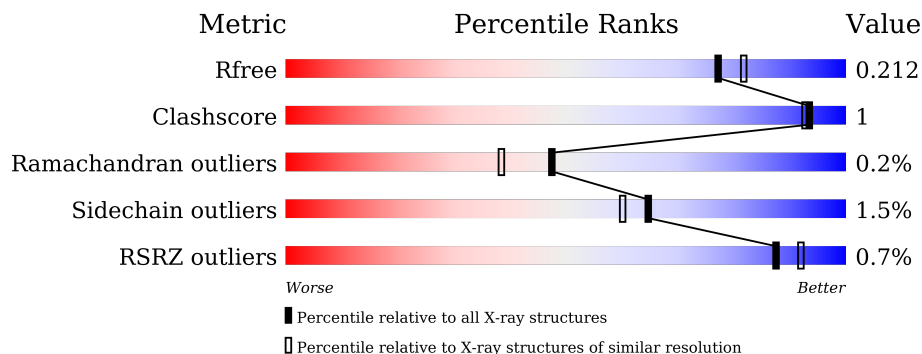
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.95 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.



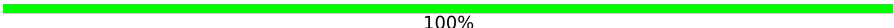

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	2580 (1.96-1.96)
Clashscore	141614	2705 (1.96-1.96)
Ramachandran outliers	138981	2678 (1.96-1.96)
Sidechain outliers	138945	2678 (1.96-1.96)
RSRZ outliers	127900	2539 (1.96-1.96)

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  $\geq 3$ , 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	640	 75% 22%
1	B	640	 74% 22%
1	C	640	 74% 22%
1	D	640	 73% 22%
2	E	2	 50% 50%
2	F	2	 100%

*Continued on next page...*

*Continued from previous page...*

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

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 17827 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 AmyP.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	497	3880	2444	652	771	13	0	1	0
1	B	499	3898	2455	656	774	13	0	0	0
1	C	497	3887	2448	652	774	13	0	2	0
1	D	497	3885	2447	654	771	13	0	1	0

- Molecule 2 is an oligosaccharide called alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose.



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

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Ca		
3	A	2	2	2	0	0
3	B	2	2	2	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	C	2	Total 2	Ca 2	0	0
3	D	2	Total 2	Ca 2	0	0

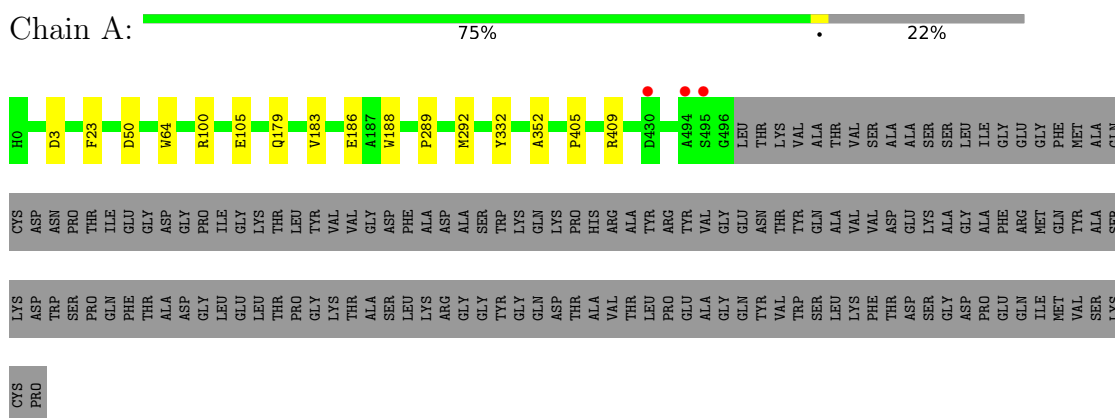
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	570	Total 570	O 570	0	0
4	B	582	Total 582	O 582	0	0
4	C	506	Total 506	O 506	0	0
4	D	519	Total 519	O 519	0	0

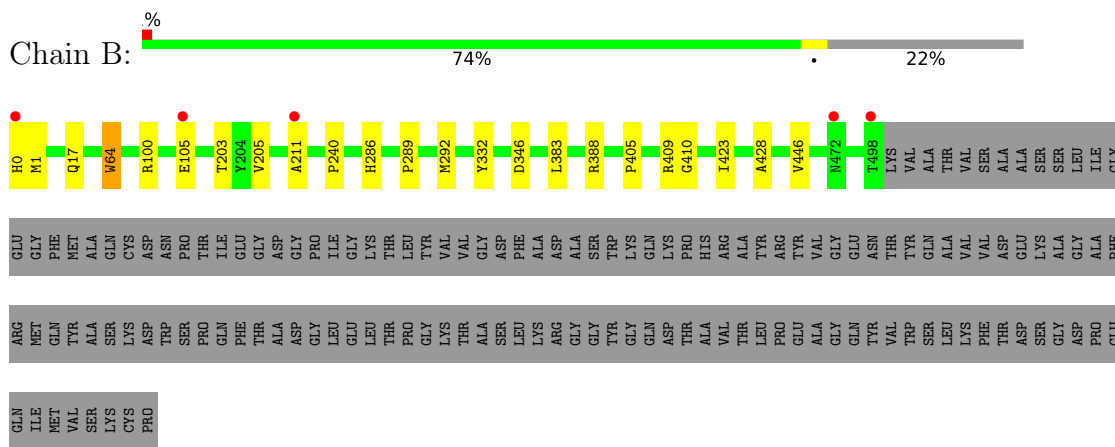
### 3 Residue-property plots [i](#)

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.

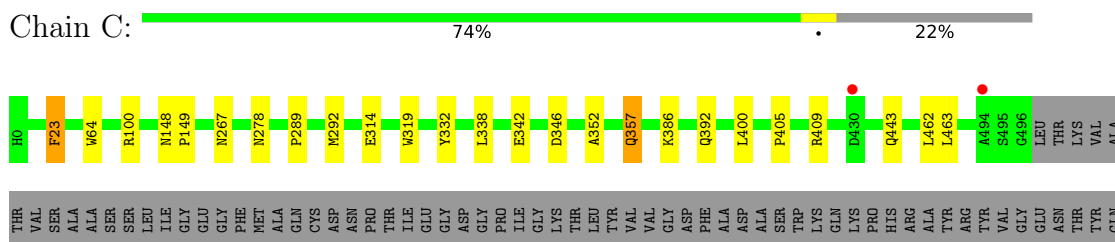
- Molecule 1: AmyP



- Molecule 1: AmyP



- Molecule 1: AmyP



ALA VAL VAL VAL ASP GLU LYS ALA GLY ALA ALA PHE ARG MET MET GLN TYR ALA SER LYS ASP TRP SER PRO PHE GLN THR ALA ASP GLY LEU LEU LEU THR PRO GLY THR ALA SER LEU LEU ARG GLY TYR GLY GLN ASP THR ALA VAL THR LEU PRO GLU ALA GLN TYR VAL TRP VAL TRP SER

LEU LYS THR ASP SER GLY ASP PRO GLU GLN ILE MET VAL SER LYS CYS PRO

- Molecule 1: AmyP

Chain D: 73% 22%

HO D12 W64 R100 K171 Q179 Y181 Q182 V183 W188 T203 E210 N213 A220 L242 R251 R281 M292 W319 K323 A352 K386 Y393 P405 R409 G410 E411 D430 Q443 D461 L487 A494 S495

G496 LEU THR LYS VAL THR VAL SER ALA ALA ASP GLU LYS SER SER LEU ILE PHE GLU MET GLN MET ALA SER GLN CYS ASP ASN PRO THR ILE GLU THR ALA ASP GLY PRO ILE GLY LYS THR LEU LEU THR PRO GLY VAL THR ALA ASP PHE ALA ASP THR ALA SER TRP LYS GLN ASP LYS PRO THR ALA HIS VAL ARG TYR ARG TYR VAL

GLY LEU THR LYS VAL THR VAL SER ALA ALA ASP GLU LYS SER SER LEU ILE PHE GLU MET GLN MET ALA SER GLN CYS ASP ASN PRO THR ILE GLU THR ALA ASP GLY PRO ILE GLY LYS THR LEU LEU THR PRO GLY VAL THR ALA ASP PHE ALA ASP THR ALA SER TRP LYS GLN ASP LYS PRO THR ALA HIS VAL ARG TYR ARG TYR VAL

GLY GLN TYR VAL TRP SER LEU LYS PHE THR ASP SER SER GLY ASP PRO GLU ILE MET VAL SER GLN LYS CYS PRO

- Molecule 2: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose

Chain E: 50% 50%

GLC1  
GLC2

- Molecule 2: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose

Chain F: 100%

GLC1  
GLC2

- Molecule 2: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose

Chain G: 100%

GLC1  
GLC2

- Molecule 2: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose

Chain H: 50% 50%

GLC1  
GLC2

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	79.87Å 130.53Å 217.53Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.94 – 1.95 39.94 – 1.95	Depositor EDS
% Data completeness (in resolution range)	87.3 (39.94-1.95) 87.5 (39.94-1.95)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.73 (at 1.95Å)	Xtrriage
Refinement program	PHENIX 1.10.1_2155: ???	Depositor
R, $R_{free}$	0.164 , 0.212 0.164 , 0.212	Depositor DCC
$R_{free}$ test set	7320 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	13.3	Xtrriage
Anisotropy	0.233	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 43.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	17827	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	14.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.65% 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, CA

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	0.36	0/3973	0.54	0/5403
1	B	0.36	0/3989	0.54	0/5425
1	C	0.36	0/3983	0.54	0/5417
1	D	0.36	0/3979	0.53	0/5411
All	All	0.36	0/15924	0.54	0/21656

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	3880	0	3670	7	0
1	B	3898	0	3691	10	0
1	C	3887	0	3679	13	0
1	D	3885	0	3675	13	0
2	E	23	0	21	0	0
2	F	23	0	21	0	0
2	G	23	0	21	0	0
2	H	23	0	21	0	0
3	A	2	0	0	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	2	0	0	0	0
3	C	2	0	0	0	0
3	D	2	0	0	0	0
4	A	570	0	0	4	0
4	B	582	0	0	2	0
4	C	506	0	0	5	0
4	D	519	0	0	4	0
All	All	17827	0	14799	43	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 1.

The worst 5 of 43 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:278:ASN:HB3	4:C:1093:HOH:O	1.90	0.71
1:D:281:ARG:NH2	4:D:803:HOH:O	2.32	0.61
1:B:383:LEU:O	1:B:388:ARG:NH1	2.39	0.54
1:D:405:PRO:HB3	1:D:409:ARG:CZ	2.39	0.52
1:C:267:ASN:HB2	4:C:1225:HOH:O	2.10	0.51

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	496/640 (78%)	482 (97%)	13 (3%)	1 (0%)	47 38
1	B	497/640 (78%)	483 (97%)	14 (3%)	0	100 100
1	C	497/640 (78%)	484 (97%)	11 (2%)	2 (0%)	34 22
1	D	496/640 (78%)	483 (97%)	12 (2%)	1 (0%)	47 38

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1986/2560 (78%)	1932 (97%)	50 (2%)	4 (0%)	47 38

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	352	ALA
1	D	352	ALA
1	C	23	PHE
1	A	352	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	407/521 (78%)	402 (99%)	5 (1%)	71 68
1	B	410/521 (79%)	404 (98%)	6 (2%)	65 60
1	C	409/521 (78%)	403 (98%)	6 (2%)	65 60
1	D	408/521 (78%)	400 (98%)	8 (2%)	55 48
All	All	1634/2084 (78%)	1609 (98%)	25 (2%)	65 60

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	346	ASP
1	D	0	HIS
1	D	443	GLN
1	C	443	GLN
1	D	64	TRP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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	12,12,12	0.55	0	17,17,17	0.81	0
2	GLC	E	2	2	11,11,12	0.44	0	15,15,17	1.04	1 (6%)
2	GLC	F	1	2	12,12,12	0.55	0	17,17,17	1.01	1 (5%)
2	GLC	F	2	2	11,11,12	0.63	0	15,15,17	1.07	1 (6%)
2	GLC	G	1	2	12,12,12	0.49	0	17,17,17	0.86	0
2	GLC	G	2	2	11,11,12	0.48	0	15,15,17	0.68	0
2	GLC	H	1	2	12,12,12	0.44	0	17,17,17	0.73	0
2	GLC	H	2	2	11,11,12	0.60	0	15,15,17	1.33	1 (6%)

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

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GLC	H	1	2	-	0/2/22/22	0/1/1/1
2	GLC	H	2	2	-	0/2/19/22	0/1/1/1

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	H	2	GLC	C1-O5-C5	4.43	118.20	112.19
2	E	2	GLC	C1-O5-C5	3.04	116.31	112.19
2	F	2	GLC	C1-O5-C5	2.79	115.97	112.19
2	F	1	GLC	C6-C5-C4	-2.11	108.07	113.00

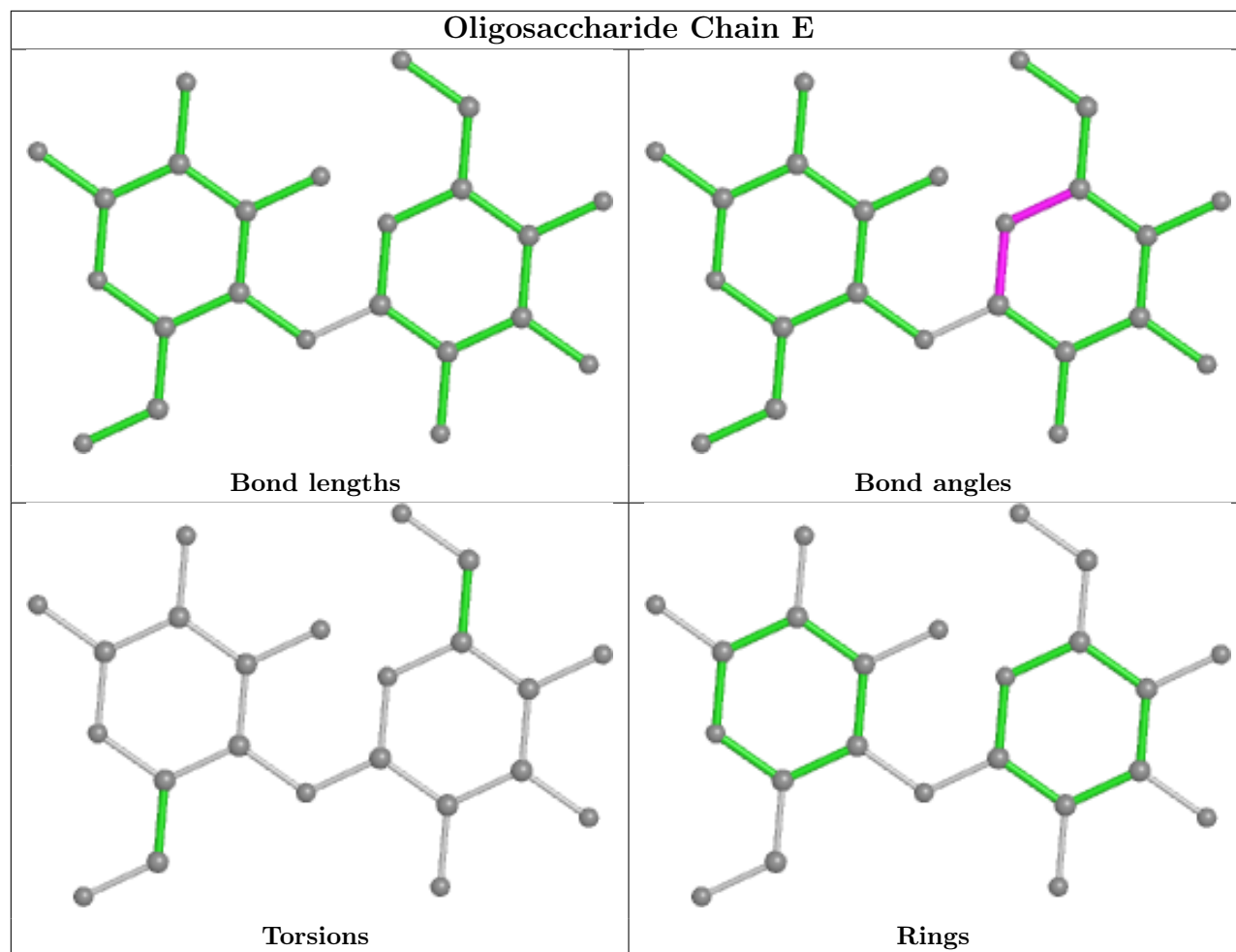
There are no chirality outliers.

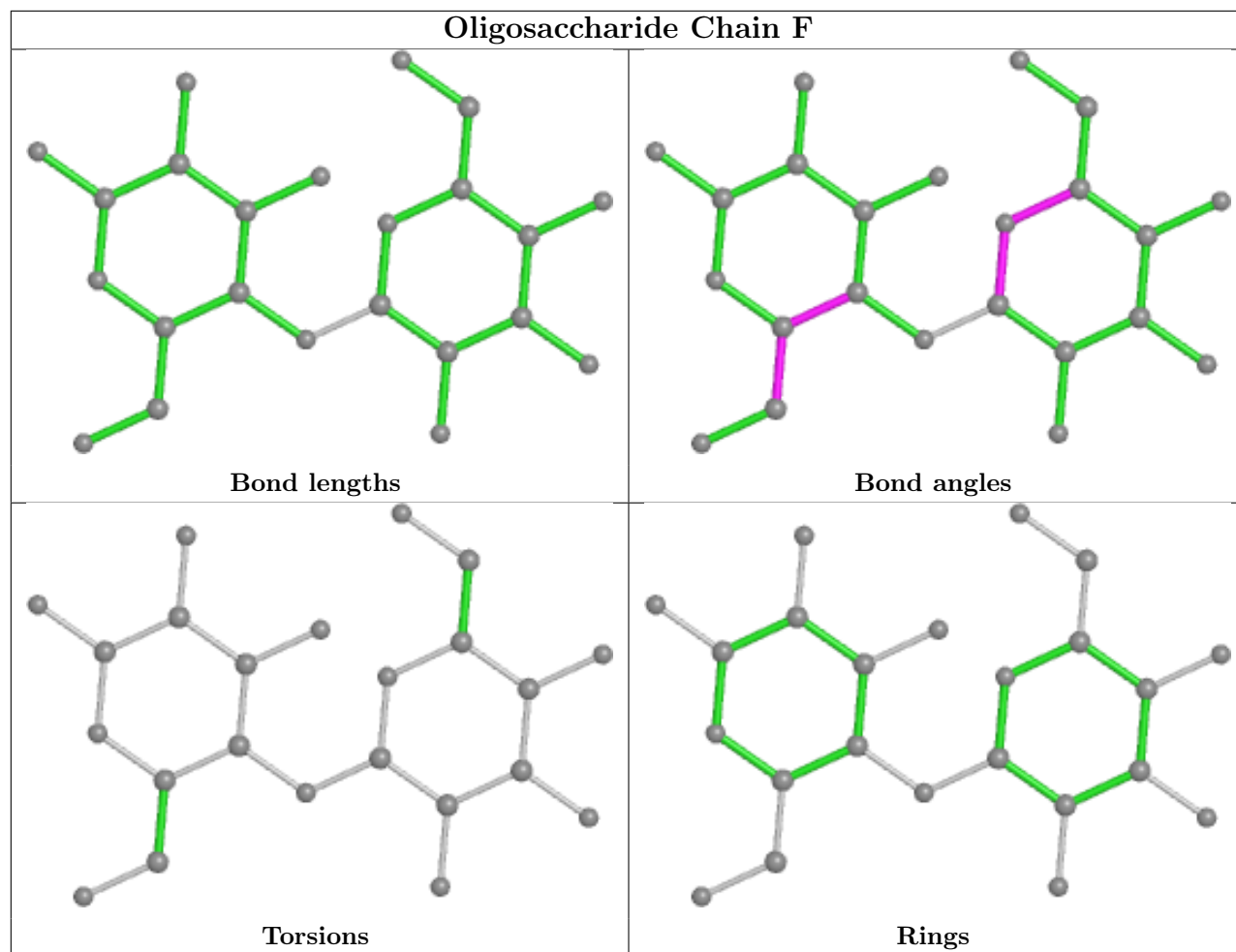
There are no torsion outliers.

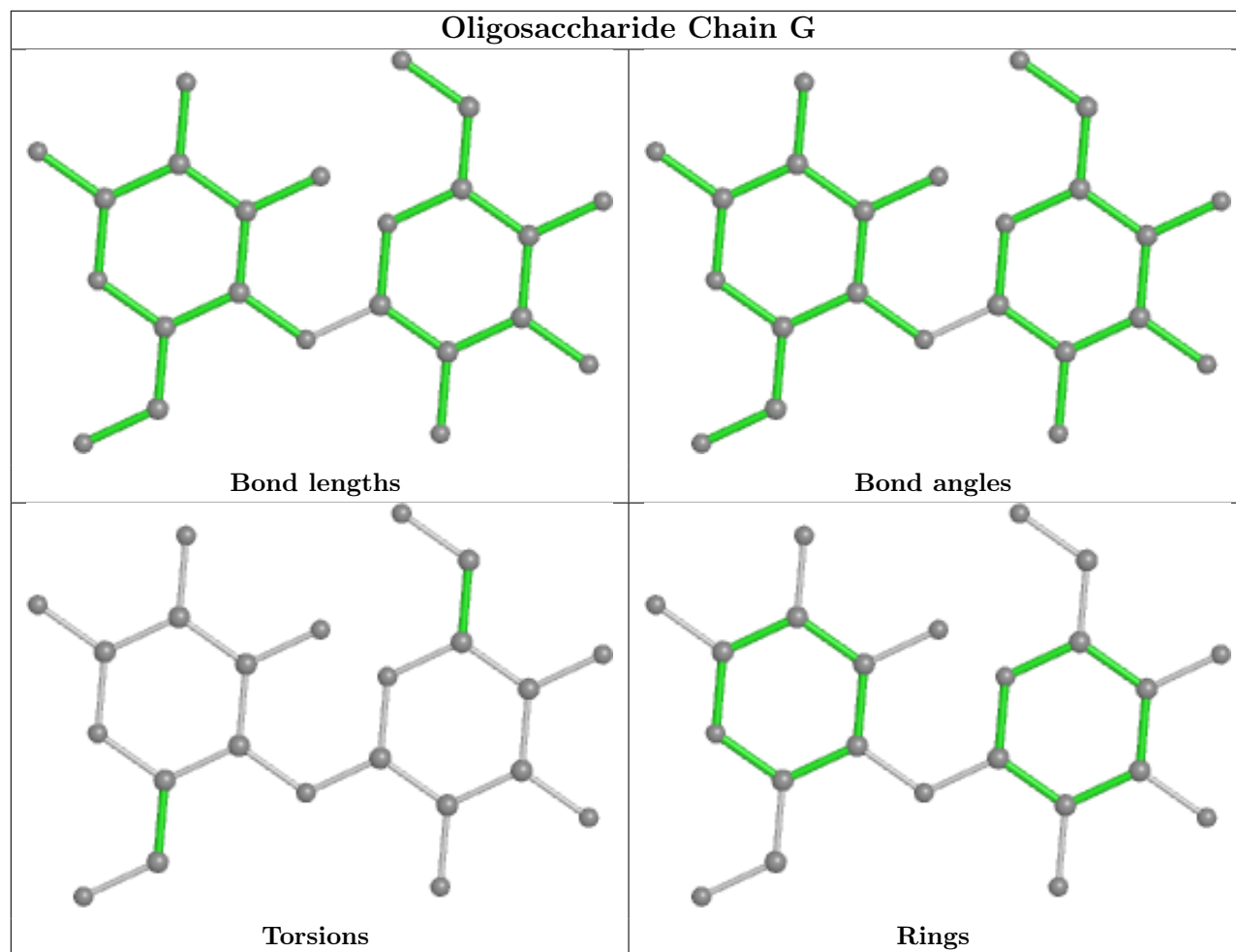
There are no ring outliers.

No monomer is involved in short contacts.

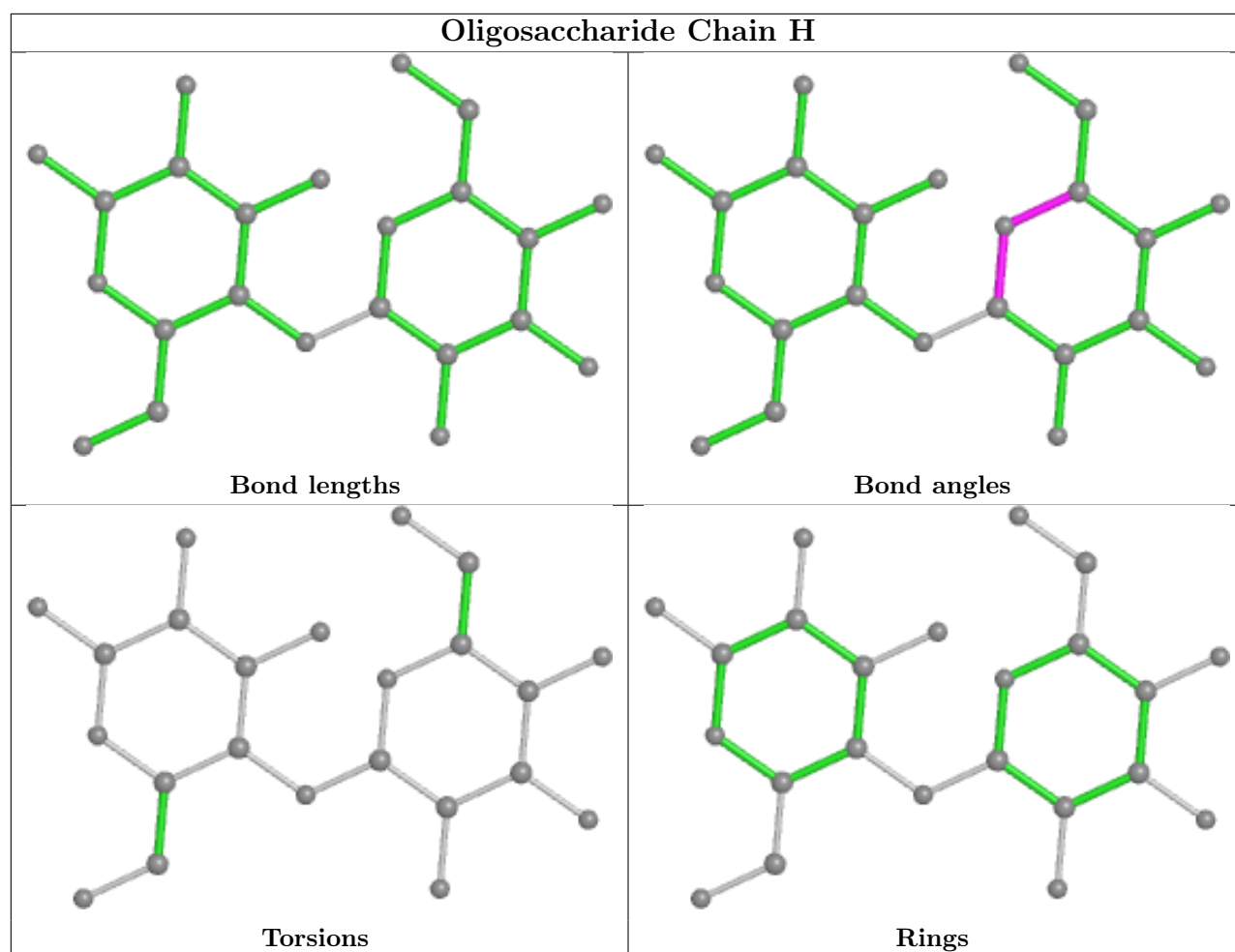
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.











## 5.6 Ligand geometry [i](#)

Of 8 ligands modelled in this entry, 8 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

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

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	497/640 (77%)	-0.28	3 (0%) 89 93	6, 11, 25, 49	0
1	B	499/640 (77%)	-0.25	5 (1%) 82 87	6, 12, 26, 57	0
1	C	497/640 (77%)	-0.21	2 (0%) 92 95	6, 12, 29, 46	1 (0%)
1	D	497/640 (77%)	-0.19	4 (0%) 86 90	6, 13, 28, 53	1 (0%)
All	All	1990/2560 (77%)	-0.23	14 (0%) 87 92	6, 12, 28, 57	2 (0%)

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	498	THR	6.2
1	A	495	SER	3.3
1	D	495	SER	3.0
1	C	430	ASP	2.7
1	A	430	ASP	2.7

### 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.

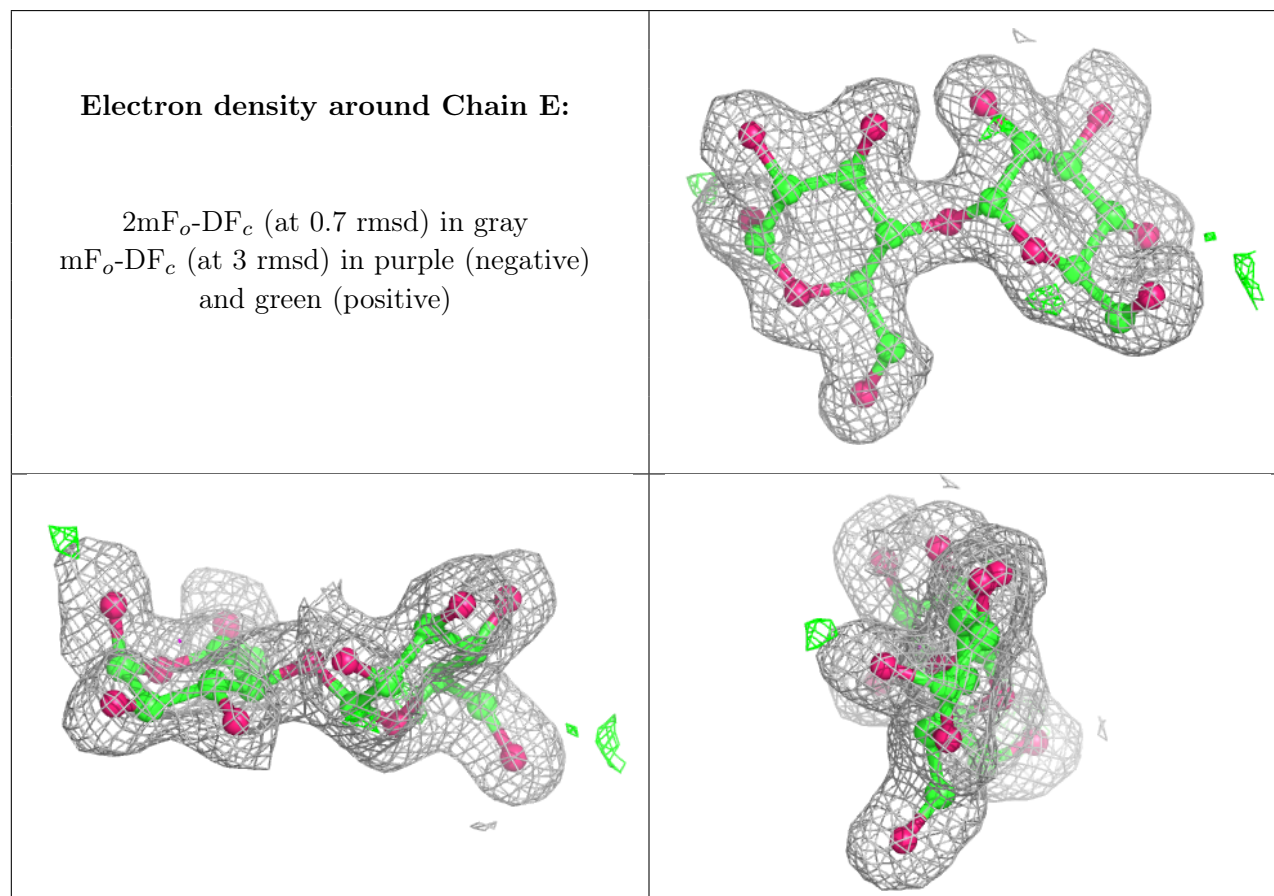
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
2	GLC	F	1	12/12	0.94	0.11	9,17,20,21	0
2	GLC	E	1	12/12	0.95	0.10	11,15,21,22	0

*Continued on next page...*

Continued from previous page...

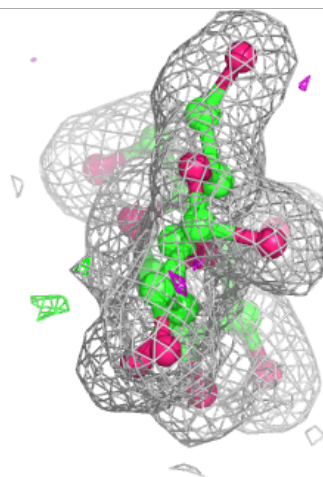
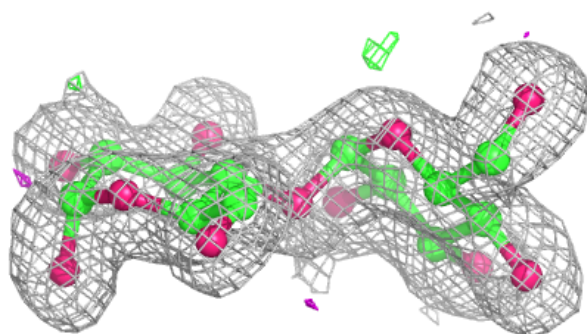
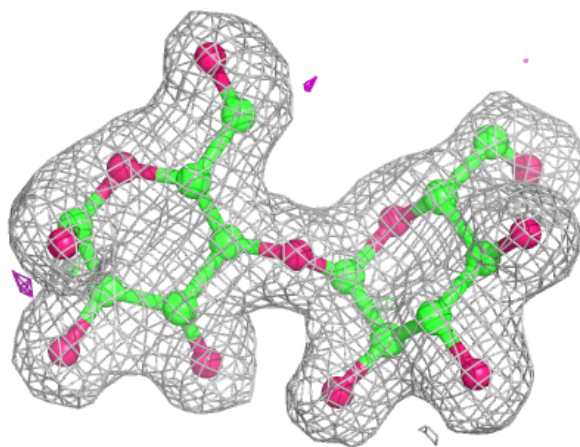
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
2	GLC	G	1	12/12	0.95	0.10	9,13,18,24	0
2	GLC	E	2	11/12	0.96	0.10	9,14,16,17	0
2	GLC	H	1	12/12	0.96	0.09	10,14,22,25	0
2	GLC	H	2	11/12	0.97	0.10	8,10,13,13	0
2	GLC	F	2	11/12	0.98	0.08	11,14,16,19	0
2	GLC	G	2	11/12	0.98	0.08	9,12,15,15	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.



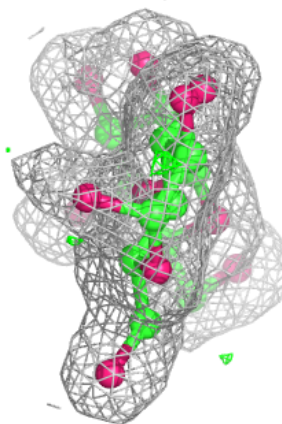
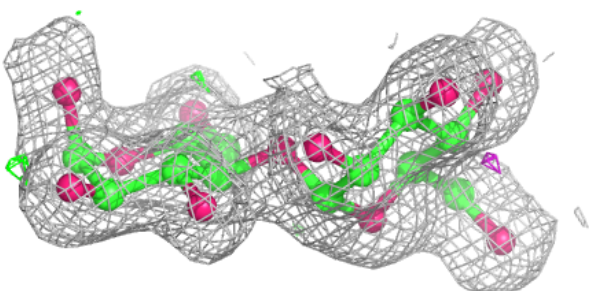
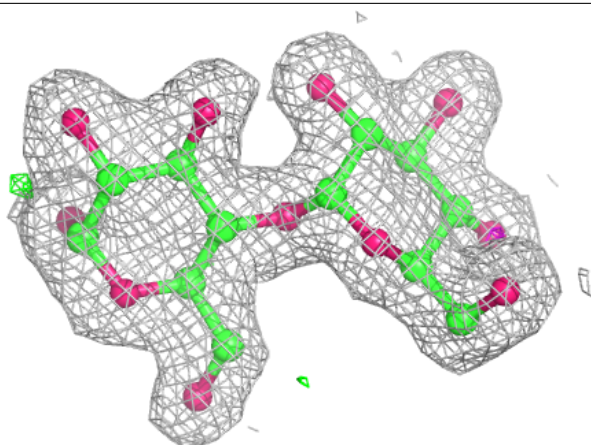
**Electron density around Chain F:**

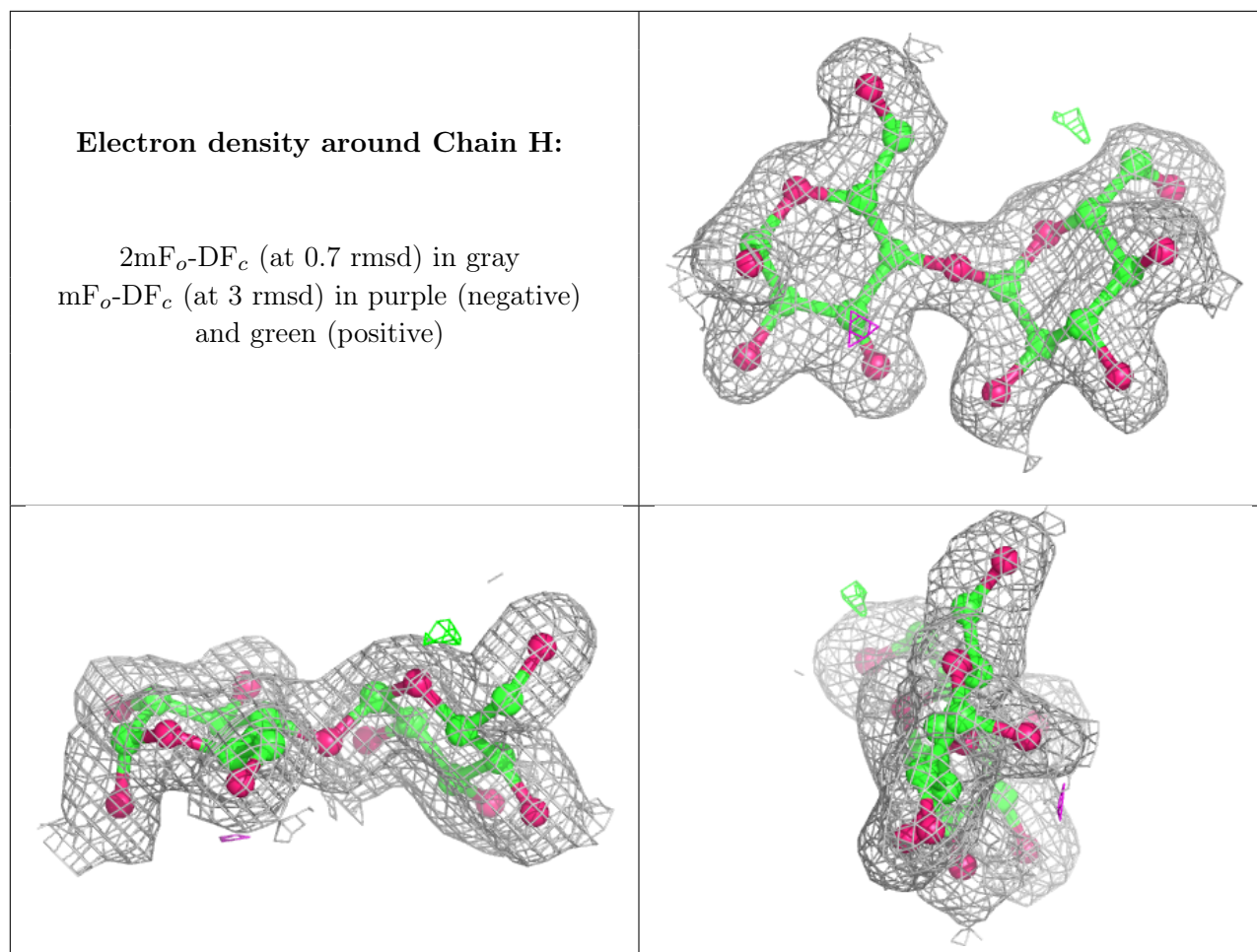
$2mF_o-DF_c$  (at 0.7 rmsd) in gray  
 $mF_o-DF_c$  (at 3 rmsd) in purple (negative)  
and green (positive)



**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](#)

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
3	CA	C	702	1/1	0.96	0.21	32,32,32,32	0
3	CA	B	702	1/1	0.98	0.19	24,24,24,24	0
3	CA	A	702	1/1	0.98	0.15	28,28,28,28	0
3	CA	A	703	1/1	0.99	0.04	10,10,10,10	0
3	CA	C	703	1/1	0.99	0.04	12,12,12,12	0
3	CA	D	702	1/1	0.99	0.22	27,27,27,27	0
3	CA	B	703	1/1	1.00	0.03	9,9,9,9	0
3	CA	D	703	1/1	1.00	0.03	10,10,10,10	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.