



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 27, 2023 – 05:52 AM EDT

PDB ID : 3HT3  
Title : Crystal structure of fragment DNA polymerase I from *Bacillus stearothermophilus* V713P mutant bound to G:dCTP  
Authors : Wu, E.Y.; Beese, L.S.  
Deposited on : 2009-06-11  
Resolution : 1.70 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 : 2.35  
buster-report : 1.1.7 (2018)  
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) : 2.35

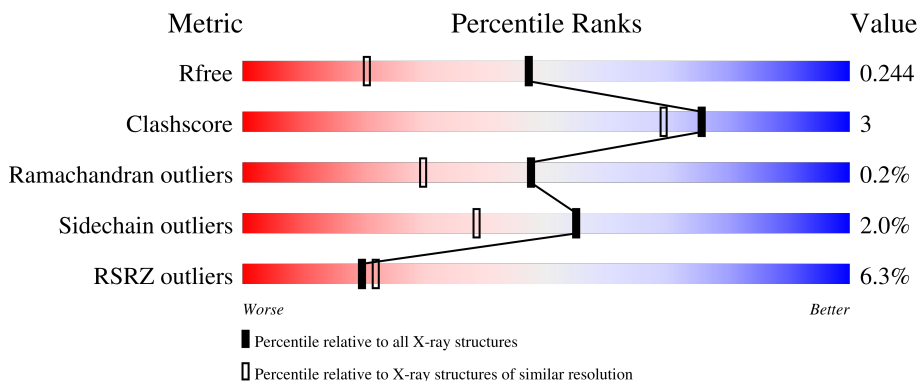
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.70 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.



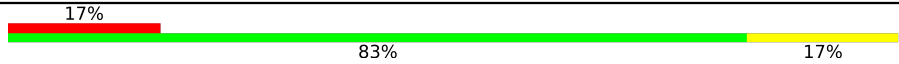
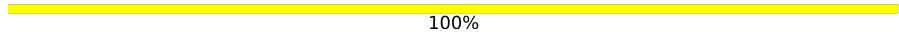
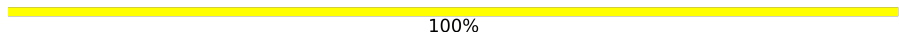

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	4298 (1.70-1.70)
Clashscore	141614	4695 (1.70-1.70)
Ramachandran outliers	138981	4610 (1.70-1.70)
Sidechain outliers	138945	4610 (1.70-1.70)
RSRZ outliers	127900	4222 (1.70-1.70)

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	579	
1	D	579	
2	B	9	
2	E	9	
3	C	12	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	F	12	 17% 83% 17%
4	G	2	 100%
4	H	2	 100%
4	I	2	 50% 50%

## 2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 11370 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 DNA polymerase I, large fragment.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	579	4678	2975	811	875	17	0	5	0
1	D	579	4678	2978	810	872	18	0	6	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	598	ALA	ASP	engineered mutation	PDB 3HT3
A	713	PRO	VAL	engineered mutation	PDB 3HT3
D	598	ALA	ASP	engineered mutation	PDB 3HT3
D	713	PRO	VAL	engineered mutation	PDB 3HT3

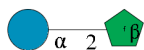
- Molecule 2 is a DNA chain called 5'-D(\*CP\*GP\*AP\*TP\*CP\*AP\*CP\*GP\*(DOC))-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	9	178	86	34	50	8	0	0	0
2	E	9	178	86	34	50	8	0	0	0

- Molecule 3 is a DNA chain called 5'-D(\*AP\*CP\*GP\*GP\*CP\*GP\*TP\*GP\*AP\*TP\*CP\*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	12	246	117	48	70	11	0	0	0
3	F	12	246	117	48	70	11	0	0	0

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

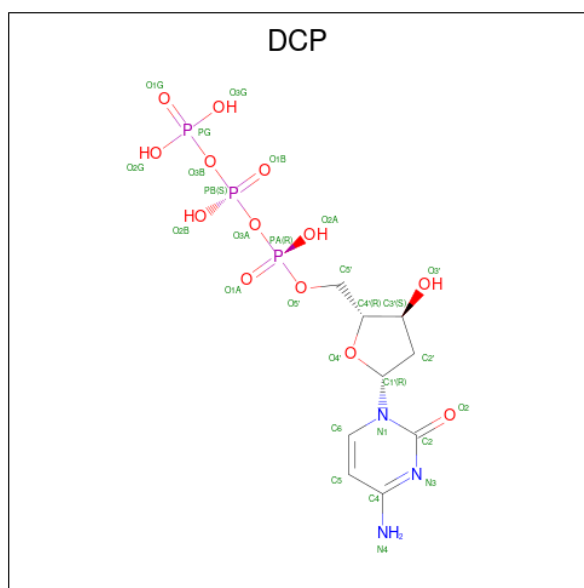


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
4	G	2	Total	C	O	0	0	0
			23	12	11			
4	H	2	Total	C	O	0	0	0
			23	12	11			
4	I	2	Total	C	O	0	0	0
			23	12	11			

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total	Mg	0	0
			1	1		
5	D	1	Total	Mg	0	0
			1	1		

- Molecule 6 is 2'-DEOXYCYTIDINE-5'-TRIPHOSPHATE (three-letter code: DCP) (formula: C<sub>9</sub>H<sub>16</sub>N<sub>3</sub>O<sub>13</sub>P<sub>3</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
6	A	1	Total	C	N	O	P	0	0
			28	9	3	13	3		
6	A	1	Total	C	N	O	P	0	0
			28	9	3	13	3		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
6	D	1	28	9	3	13	3	0	0

- Molecule 7 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	S		
7	D	1	5	4	1	0	0
7	D	1	5	4	1	0	0
7	D	1	5	4	1	0	0

- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
8	A	365	365	365	0	0
8	D	521	521	521	0	0
8	B	21	21	21	0	0
8	C	32	32	32	0	0
8	E	16	16	16	0	0

Continued on next page...

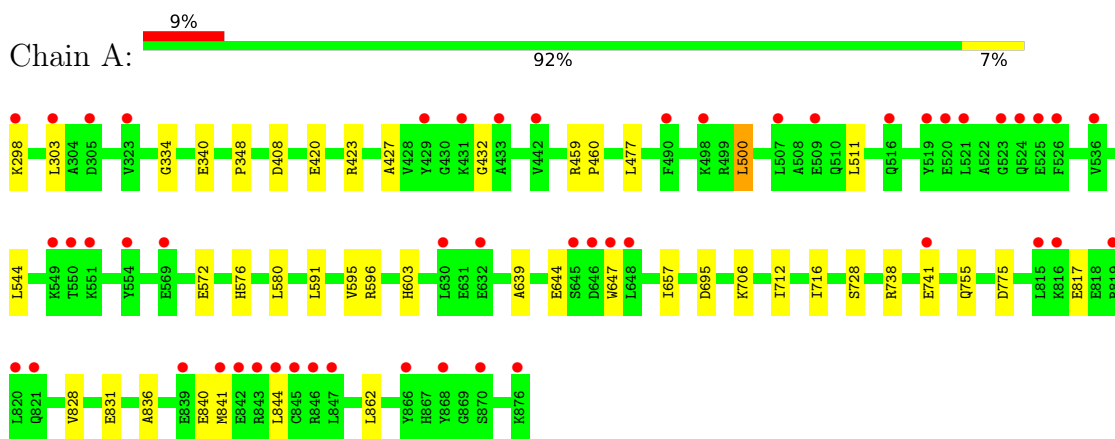
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
8	F	41	Total	O	0	0
			41	41		

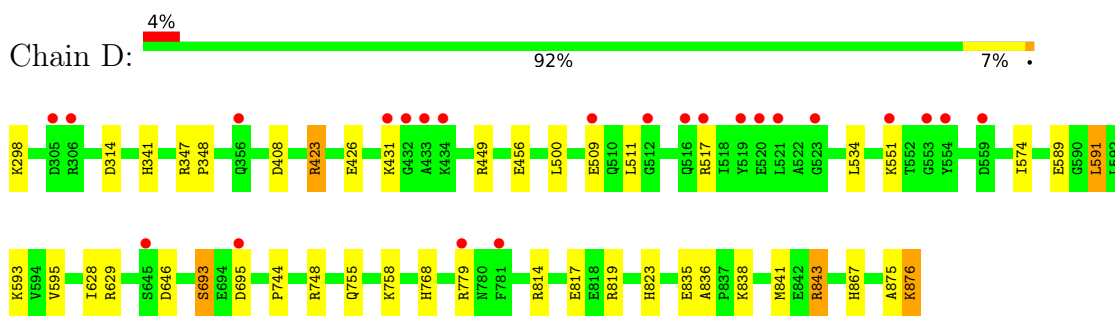
### 3 Residue-property plots [i](#)

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.

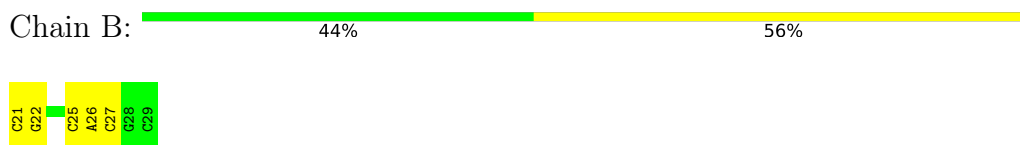
- Molecule 1: DNA polymerase I, large fragment



- Molecule 1: DNA polymerase I, large fragment



- Molecule 2: 5'-D(\*CP\*GP\*AP\*TP\*CP\*AP\*CP\*GP\*(DOC))-3'



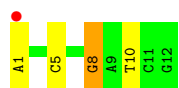
- Molecule 2: 5'-D(\*CP\*GP\*AP\*TP\*CP\*AP\*CP\*GP\*(DOC))-3'






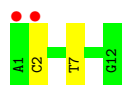
- Molecule 3: 5'-D(\*AP\*CP\*GP\*GP\*CP\*GP\*TP\*GP\*AP\*TP\*CP\*G)-3'

Chain C: 



- Molecule 3: 5'-D(\*AP\*CP\*GP\*GP\*CP\*GP\*TP\*GP\*AP\*TP\*CP\*G)-3'

Chain F: 



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

Chain G: 



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

Chain H: 



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

Chain I: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	93.28Å 108.79Å 152.09Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.98 – 1.70 46.99 – 1.70	Depositor EDS
% Data completeness (in resolution range)	100.0 (46.98-1.70) 99.8 (46.99-1.70)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.57 (at 1.70Å)	Xtrriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.210 , 0.247 0.208 , 0.244	Depositor DCC
$R_{free}$ test set	7528 reflections (4.44%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	17.9	Xtrriage
Anisotropy	0.056	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 48.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	11370	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	23.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.19% 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: DCP, SO4, GLC, DOC, FRU, MG

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	0.45	0/4775	0.57	0/6452
1	D	0.52	0/4781	0.65	2/6459 (0.0%)
2	B	0.80	0/179	1.51	5/274 (1.8%)
2	E	0.82	0/179	1.61	2/274 (0.7%)
3	C	0.89	0/276	1.55	4/425 (0.9%)
3	F	0.88	0/276	1.46	3/425 (0.7%)
All	All	0.53	0/10466	0.75	16/14309 (0.1%)

There are no bond length outliers.

All (16) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	25	DC	O4'-C4'-C3'	-10.07	99.96	106.00
1	D	843	ARG	NE-CZ-NH2	-10.01	115.30	120.30
2	B	26	DA	O4'-C1'-N9	8.76	114.13	108.00
1	D	843	ARG	NE-CZ-NH1	8.34	124.47	120.30
3	C	1	DA	O4'-C4'-C3'	-8.11	101.14	106.00
3	F	2	DC	O4'-C1'-N1	7.43	113.20	108.00
3	C	8	DG	O4'-C1'-N9	6.88	112.82	108.00
2	B	27	DC	C4'-C3'-C2'	-6.76	97.01	103.10
2	B	25	DC	O4'-C1'-N1	6.30	112.41	108.00
3	C	10	DT	C1'-O4'-C4'	-6.13	103.97	110.10
3	F	7	DT	N3-C4-O4	5.80	123.38	119.90
2	B	27	DC	O4'-C4'-C3'	-5.76	102.19	104.50
2	B	27	DC	P-O3'-C3'	5.54	126.34	119.70
3	C	5	DC	C2-N3-C4	5.29	122.55	119.90
2	E	25	DC	C4'-C3'-C2'	-5.06	98.55	103.10
3	F	7	DT	C6-C5-C7	-5.02	119.89	122.90

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	4678	0	4736	26	0
1	D	4678	0	4751	36	0
2	B	178	0	102	1	0
2	E	178	0	102	2	0
3	C	246	0	136	1	0
3	F	246	0	136	0	0
4	G	23	0	21	1	0
4	H	23	0	21	2	0
4	I	23	0	21	0	0
5	A	1	0	0	0	0
5	D	1	0	0	0	0
6	A	56	0	24	1	0
6	D	28	0	12	0	0
7	D	15	0	0	0	0
8	A	365	0	0	7	0
8	B	21	0	0	0	0
8	C	32	0	0	1	0
8	D	521	0	0	9	0
8	E	16	0	0	0	0
8	F	41	0	0	0	0
All	All	11370	0	10062	64	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 3.

All (64) 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:758:LYS:HE2	8:D:972:HOH:O	1.56	1.04
1:D:423:ARG:HG2	1:D:423:ARG:HH11	1.31	0.96
1:D:876:LYS:HE3	8:D:14:HOH:O	1.74	0.86
1:D:456:GLU:HG2	8:D:295:HOH:O	1.74	0.86

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:843:ARG:HD2	8:D:180:HOH:O	1.75	0.86
1:D:819:ARG:HH11	1:D:819:ARG:HG3	1.40	0.85
1:D:408:ASP:HB2	4:H:2:FRU:H11	1.65	0.78
1:D:693:SER:HB2	1:D:695:ASP:OD1	1.92	0.70
1:A:657:ILE:HD13	1:A:862:LEU:HD22	1.78	0.66
1:D:867:HIS:CG	1:D:876:LYS:HA	2.32	0.65
1:D:875:ALA:O	1:D:876:LYS:HG2	1.96	0.65
1:A:423:ARG:HH11	1:A:427:ALA:HB1	1.64	0.63
1:D:534:LEU:HD11	1:D:574:ILE:HD13	1.80	0.63
1:A:596:ARG:HD3	1:A:603:HIS:CD2	2.34	0.62
1:D:298:LYS:HB3	1:D:449:ARG:NH2	2.15	0.62
1:D:875:ALA:C	1:D:876:LYS:HG2	2.22	0.60
1:A:695:ASP:HB3	8:A:996:HOH:O	2.02	0.59
1:D:819:ARG:HG3	1:D:819:ARG:NH1	2.13	0.58
1:A:459:ARG:HG3	8:A:1107:HOH:O	2.04	0.57
1:A:596:ARG:HD3	1:A:603:HIS:HD2	1.70	0.57
1:A:840:GLU:HB3	8:A:1052:HOH:O	2.04	0.57
1:D:426:GLU:OE2	1:D:431:LYS:HD3	2.05	0.56
1:D:744:PRO:HB2	1:D:748[B]:ARG:HH12	1.73	0.54
1:D:629:ARG:HH22	2:E:29:DOC:H5	1.73	0.54
1:D:423:ARG:HG2	1:D:423:ARG:NH1	2.08	0.53
1:D:456:GLU:CG	8:D:295:HOH:O	2.44	0.53
1:A:591:LEU:O	1:A:595:VAL:HG23	2.10	0.52
1:D:836:ALA:HB3	1:D:841:MET:CE	2.40	0.52
1:D:591:LEU:O	1:D:595:VAL:HG23	2.11	0.51
1:A:408:ASP:HB2	4:G:2:FRU:H11	1.92	0.51
1:A:596:ARG:NH2	8:A:121:HOH:O	2.43	0.51
1:D:744:PRO:HB2	1:D:748[B]:ARG:NH1	2.26	0.50
1:D:408:ASP:HB2	4:H:2:FRU:C1	2.39	0.50
1:D:779:ARG:HD3	8:D:1115:HOH:O	2.12	0.48
1:A:706:LYS:HE3	6:A:201:DCP:O3B	2.14	0.48
1:D:646:ASP:HB3	1:D:838:LYS:HE3	1.96	0.48
1:A:738:ARG:HA	1:A:741:GLU:HG2	1.96	0.48
1:A:459:ARG:NH2	1:D:314:ASP:O	2.47	0.48
1:D:589:GLU:O	1:D:593:LYS:HG3	2.13	0.47
2:B:21:DC:H2''	2:B:22:DG:C8	2.49	0.47
2:E:21:DC:H2'	2:E:22:DG:C8	2.50	0.46
1:D:814:ARG:HA	1:D:817:GLU:HG2	1.98	0.46
1:A:459:ARG:HB3	1:A:460:PRO:HD3	1.98	0.45
1:A:334:GLY:HA2	1:A:348:PRO:HD3	2.00	0.44
1:A:828:VAL:HB	1:A:831:GLU:CG	2.48	0.44

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:341:HIS:HE1	8:D:930:HOH:O	1.99	0.44
1:D:768:HIS:HD2	8:D:1152:HOH:O	2.01	0.43
1:A:755[A]:GLN:HG2	8:A:1053:HOH:O	2.17	0.43
1:D:423:ARG:NH1	1:D:423:ARG:CG	2.79	0.43
1:D:823:HIS:HE1	1:D:835:GLU:OE2	2.02	0.43
1:D:867:HIS:CB	1:D:876:LYS:HA	2.48	0.43
1:D:755:GLN:HG2	8:D:126:HOH:O	2.19	0.42
1:A:738:ARG:HA	1:A:741:GLU:CG	2.50	0.42
1:A:500:LEU:HD13	1:A:639:ALA:CB	2.50	0.42
1:A:817:GLU:HG2	8:A:1143:HOH:O	2.19	0.42
1:A:576:HIS:CE1	1:A:580:LEU:HD11	2.55	0.42
1:A:775:ASP:HB2	8:A:67:HOH:O	2.20	0.41
1:A:712:ILE:HA	1:A:716:ILE:HG22	2.02	0.41
1:A:836:ALA:HB3	1:A:841:MET:CE	2.50	0.40
1:D:534:LEU:HD11	1:D:574:ILE:CD1	2.50	0.40
1:A:644:GLU:HB2	1:A:647:TRP:CD1	2.57	0.40
1:D:347:ARG:HA	1:D:348:PRO:HD3	1.98	0.40
3:C:8:DG:H4'	8:C:587:HOH:O	2.22	0.40

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	582/579 (100%)	569 (98%)	12 (2%)	1 (0%)	47	30
1	D	583/579 (101%)	572 (98%)	10 (2%)	1 (0%)	47	30
All	All	1165/1158 (101%)	1141 (98%)	22 (2%)	2 (0%)	47	30

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	432	GLY
1	D	628	ILE

### 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	500/495 (101%)	488 (98%)	12 (2%)	49 31
1	D	501/495 (101%)	492 (98%)	9 (2%)	59 43
All	All	1001/990 (101%)	980 (98%)	21 (2%)	55 36

All (21) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	298	LYS
1	A	303	LEU
1	A	340	GLU
1	A	420[A]	GLU
1	A	420[B]	GLU
1	A	477	LEU
1	A	500	LEU
1	A	511	LEU
1	A	544	LEU
1	A	572	GLU
1	A	728	SER
1	A	844	LEU
1	D	423	ARG
1	D	500	LEU
1	D	509	GLU
1	D	511	LEU
1	D	517	ARG
1	D	551	LYS
1	D	591	LEU
1	D	693	SER
1	D	876	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (8) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	579	GLN
1	A	867	HIS
1	D	341	HIS
1	D	405	GLN
1	D	502	GLN
1	D	755	GLN
1	D	768	HIS
1	D	823	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](#)

2 non-standard protein/DNA/RNA residues 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	DOC	E	29	2,3	16,19,20	0.87	0	20,26,29	0.98	0
2	DOC	B	29	2,3	16,19,20	0.87	0	20,26,29	1.02	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
2	DOC	E	29	2,3	-	0/7/18/19	0/2/2/2
2	DOC	B	29	2,3	-	0/7/18/19	0/2/2/2

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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	E	29	DOC	1	0

## 5.5 Carbohydrates i

6 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
4	GLC	G	1	4	11,11,12	0.58	0	15,15,17	1.23	1 (6%)
4	FRU	G	2	4	11,12,12	0.56	0	10,18,18	0.78	0
4	GLC	H	1	4	11,11,12	0.46	0	15,15,17	0.96	1 (6%)
4	FRU	H	2	4	11,12,12	0.61	0	10,18,18	0.71	0
4	GLC	I	1	4	11,11,12	0.53	0	15,15,17	0.91	1 (6%)
4	FRU	I	2	4	11,12,12	0.57	0	10,18,18	0.78	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	GLC	G	1	4	-	0/2/19/22	0/1/1/1
4	FRU	G	2	4	-	0/5/24/24	0/1/1/1
4	GLC	H	1	4	-	0/2/19/22	0/1/1/1
4	FRU	H	2	4	-	0/5/24/24	0/1/1/1
4	GLC	I	1	4	-	0/2/19/22	0/1/1/1
4	FRU	I	2	4	-	0/5/24/24	0/1/1/1

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	G	1	GLC	C1-O5-C5	4.31	118.03	112.19
4	H	1	GLC	C1-O5-C5	2.98	116.23	112.19
4	I	1	GLC	C1-O5-C5	2.42	115.47	112.19

There are no chirality outliers.

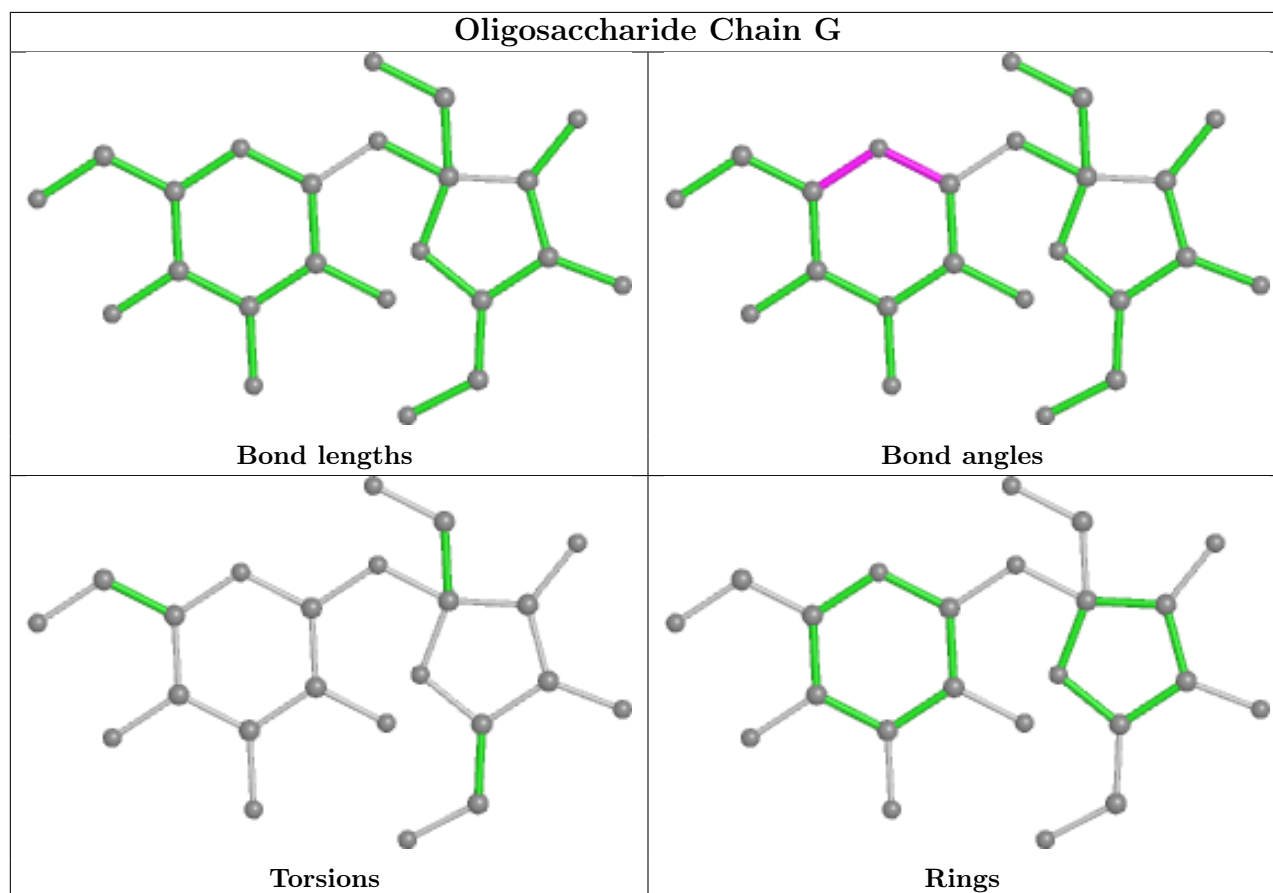
There are no torsion outliers.

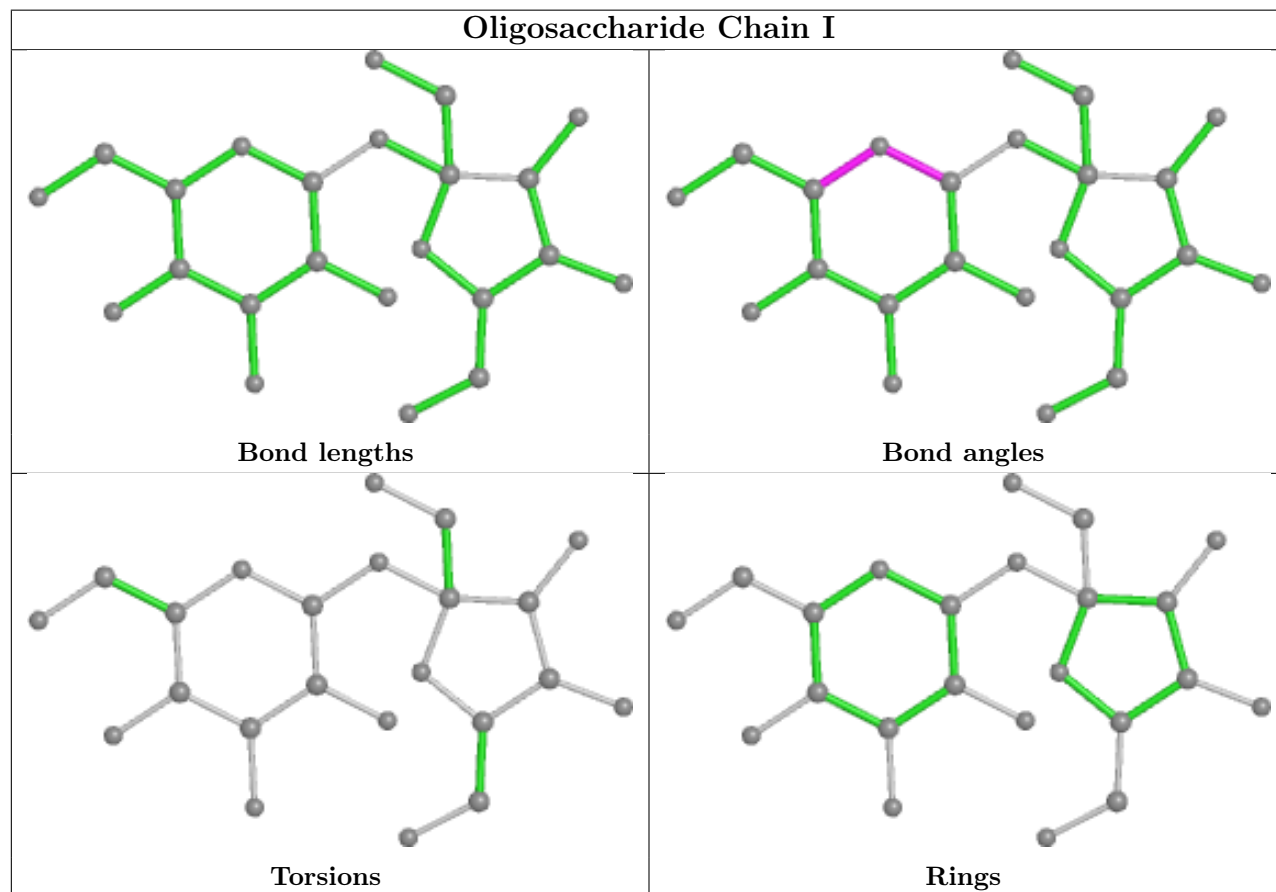
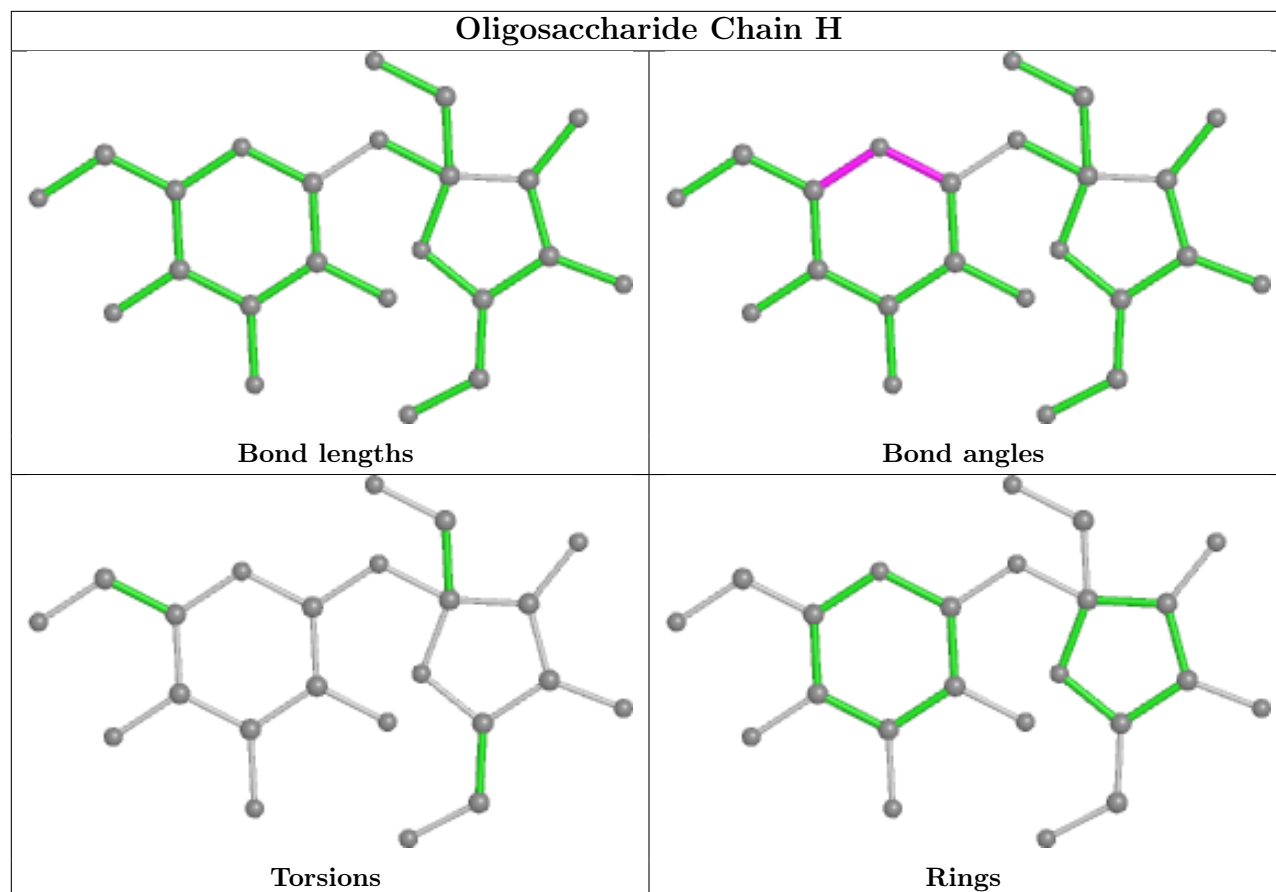
There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	H	2	FRU	2	0
4	G	2	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

Of 8 ligands modelled in this entry, 2 are monoatomic - leaving 6 for Mogul analysis.

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
6	DCP	D	202	5	25,29,29	0.75	0	37,45,45	1.18	4 (10%)
6	DCP	A	201	5	25,29,29	0.81	0	37,45,45	1.12	1 (2%)
7	SO4	D	2	-	4,4,4	0.14	0	6,6,6	0.07	0
7	SO4	D	3	-	4,4,4	0.16	0	6,6,6	0.12	0
7	SO4	D	1	-	4,4,4	0.20	0	6,6,6	0.41	0
6	DCP	A	203	-	25,29,29	0.79	1 (4%)	37,45,45	1.18	2 (5%)

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
6	DCP	D	202	5	-	3/22/34/34	0/2/2/2
6	DCP	A	201	5	-	2/22/34/34	0/2/2/2
6	DCP	A	203	-	-	10/22/34/34	0/2/2/2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	A	203	DCP	C6-C5	2.05	1.39	1.35

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	A	203	DCP	PB-O3B-PG	-3.71	120.09	132.83
6	D	202	DCP	C2'-C1'-N1	-3.42	105.89	113.77
6	A	201	DCP	PB-O3B-PG	-2.62	123.83	132.83
6	D	202	DCP	PB-O3B-PG	-2.45	124.40	132.83

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	D	202	DCP	O2A-PA-O1A	2.18	123.00	112.24
6	D	202	DCP	O3G-PG-O2G	2.12	115.75	107.64
6	A	203	DCP	PB-O3A-PA	-2.10	125.61	132.83

There are no chirality outliers.

All (15) torsion outliers are listed below:

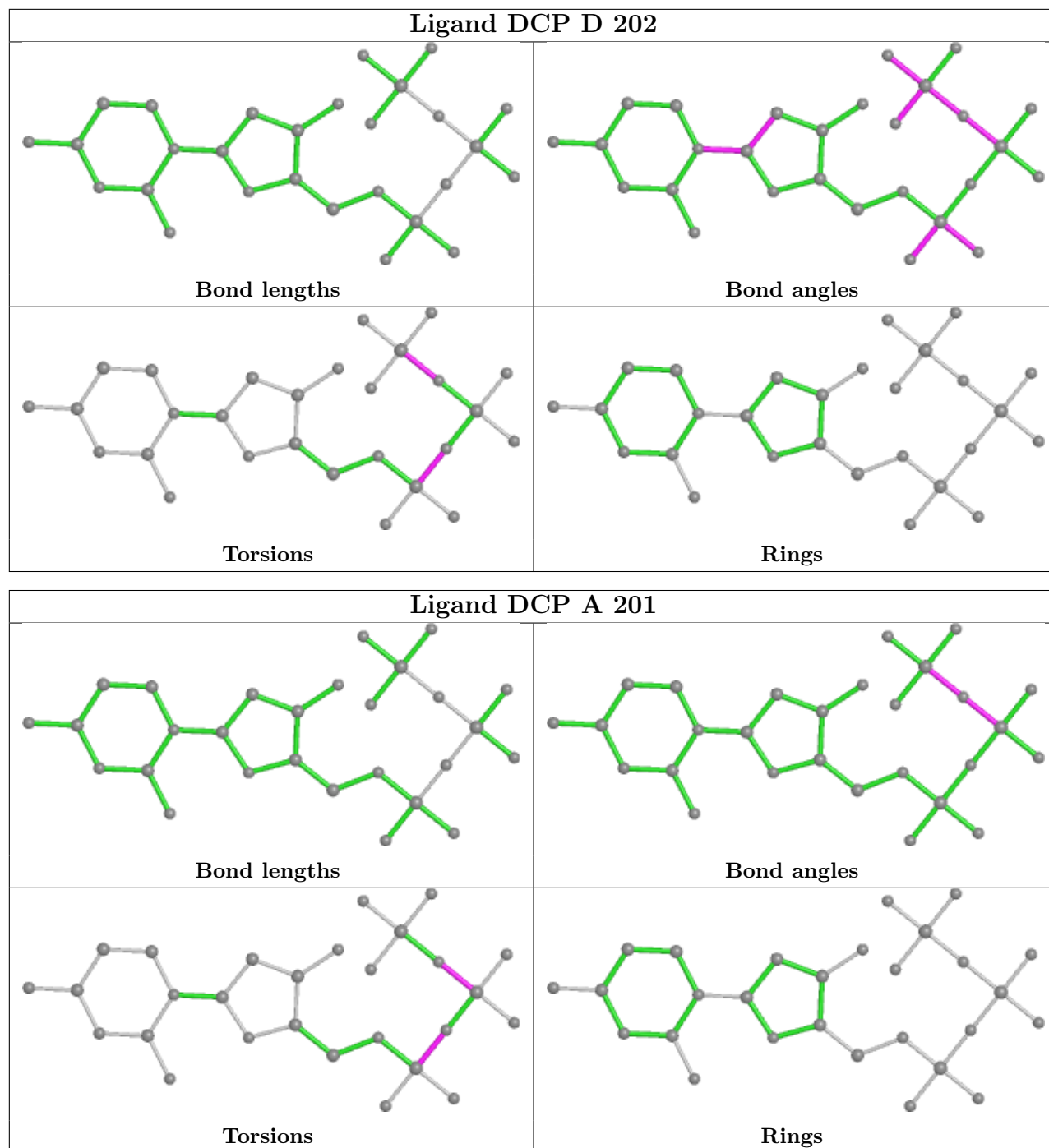
Mol	Chain	Res	Type	Atoms
6	A	203	DCP	C3'-C4'-C5'-O5'
6	A	203	DCP	O4'-C4'-C5'-O5'
6	A	203	DCP	PB-O3B-PG-O2G
6	A	203	DCP	PB-O3B-PG-O3G
6	A	203	DCP	C5'-O5'-PA-O3A
6	A	203	DCP	PG-O3B-PB-O2B
6	D	202	DCP	PB-O3A-PA-O2A
6	D	202	DCP	PB-O3B-PG-O1G
6	A	203	DCP	PA-O3A-PB-O2B
6	A	201	DCP	PB-O3A-PA-O1A
6	A	201	DCP	PG-O3B-PB-O1B
6	A	203	DCP	PA-O3A-PB-O1B
6	A	203	DCP	PG-O3B-PB-O1B
6	D	202	DCP	PB-O3A-PA-O1A
6	A	203	DCP	C5'-O5'-PA-O1A

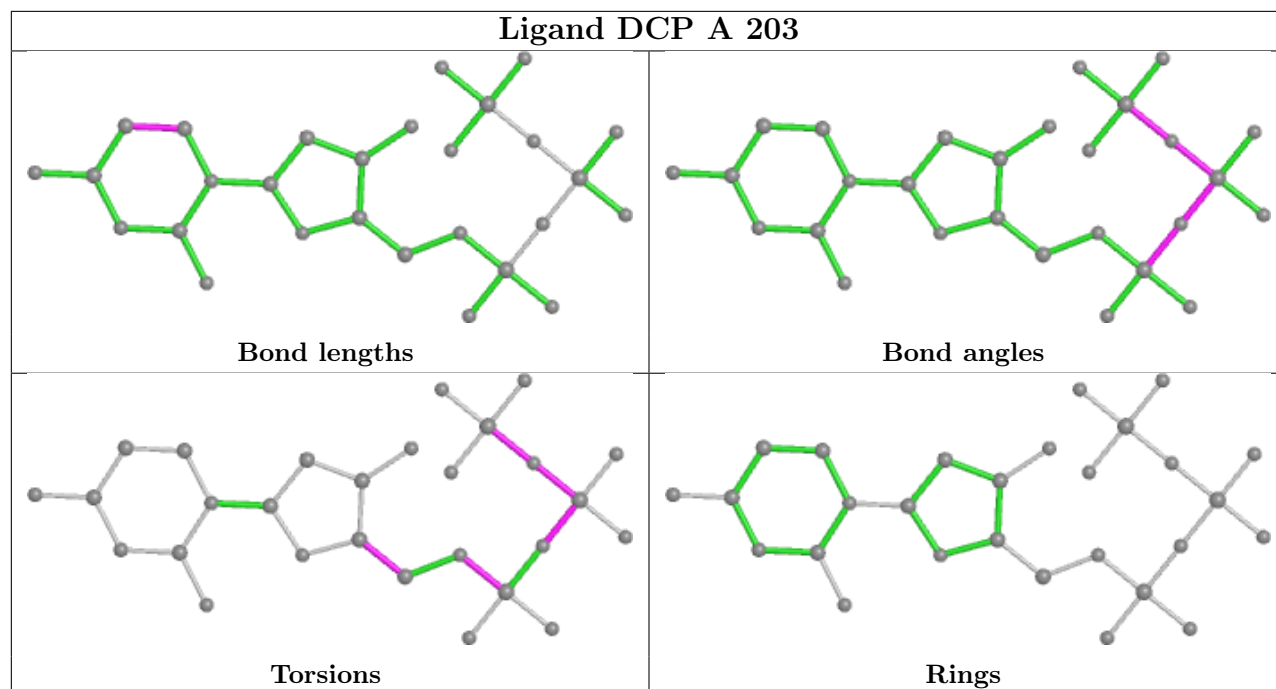
There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	A	201	DCP	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 all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less than 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.





## 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

### 6.1 Protein, DNA and RNA chains

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	579/579 (100%)	0.66	50 (8%) 10 12	10, 26, 42, 50	0
1	D	579/579 (100%)	0.33	23 (3%) 38 42	6, 16, 38, 45	0
2	B	8/9 (88%)	-0.05	0 100 100	14, 25, 45, 49	0
2	E	8/9 (88%)	0.15	0 100 100	11, 28, 50, 57	0
3	C	12/12 (100%)	0.32	1 (8%) 11 13	11, 26, 43, 61	0
3	F	12/12 (100%)	0.57	2 (16%) 1 1	8, 20, 47, 67	0
All	All	1198/1200 (99%)	0.49	76 (6%) 20 22	6, 21, 41, 67	0

All (76) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	F	1	DA	5.2
1	A	551	LYS	5.2
1	A	550	THR	5.0
1	A	819	ARG	4.9
1	A	820	LEU	4.8
1	D	433	ALA	4.7
1	A	844	LEU	4.5
1	A	846	ARG	4.4
1	A	298	LYS	4.1
1	A	519	TYR	3.9
3	C	1	DA	3.8
1	A	523	GLY	3.8
1	A	647	TRP	3.7
1	A	847	LEU	3.3
1	A	442	VAL	3.3
1	D	509	GLU	3.3
1	D	434	LYS	3.3
1	D	521	LEU	3.3
1	D	512	GLY	3.2

*Continued on next page...*



*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
1	A	821	GLN	3.2
1	A	305	ASP	3.2
1	A	516	GLN	3.2
1	A	509	GLU	3.1
1	A	526	PHE	3.1
1	A	842	GLU	3.1
1	D	551	LYS	3.1
1	A	646	ASP	3.1
1	A	630	LEU	3.0
1	D	432	GLY	3.0
1	A	645	SER	2.9
1	A	569	GLU	2.8
1	A	741	GLU	2.8
1	A	429	TYR	2.7
1	D	520	GLU	2.7
1	D	356	GLN	2.7
1	A	525	GLU	2.6
1	A	868	TYR	2.6
1	A	632	GLU	2.6
1	D	519	TYR	2.6
1	D	781	PHE	2.6
1	A	433	ALA	2.6
1	A	816	LYS	2.5
1	A	524	GLN	2.5
1	A	520	GLU	2.5
1	D	517	ARG	2.5
1	A	303	LEU	2.5
1	D	523	GLY	2.5
1	A	815	LEU	2.5
1	A	554	TYR	2.4
1	D	554	TYR	2.4
1	D	305	ASP	2.4
1	A	843	ARG	2.4
1	A	876	LYS	2.4
1	D	553	GLY	2.4
1	A	507	LEU	2.4
3	F	2	DC	2.4
1	A	521	LEU	2.3
1	A	845	CYS	2.3
1	A	431	LYS	2.3
1	D	695	ASP	2.3
1	A	549	LYS	2.3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	841	MET	2.3
1	D	779	ARG	2.2
1	A	870	SER	2.2
1	A	648	LEU	2.2
1	A	498	LYS	2.2
1	D	431	LYS	2.2
1	A	536	VAL	2.2
1	A	839	GLU	2.2
1	A	323	VAL	2.1
1	D	516	GLN	2.1
1	D	645	SER	2.1
1	A	866	TYR	2.1
1	A	490	PHE	2.1
1	D	306	ARG	2.0
1	D	559	ASP	2.0

## 6.2 Non-standard residues in protein, DNA, RNA chains [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.

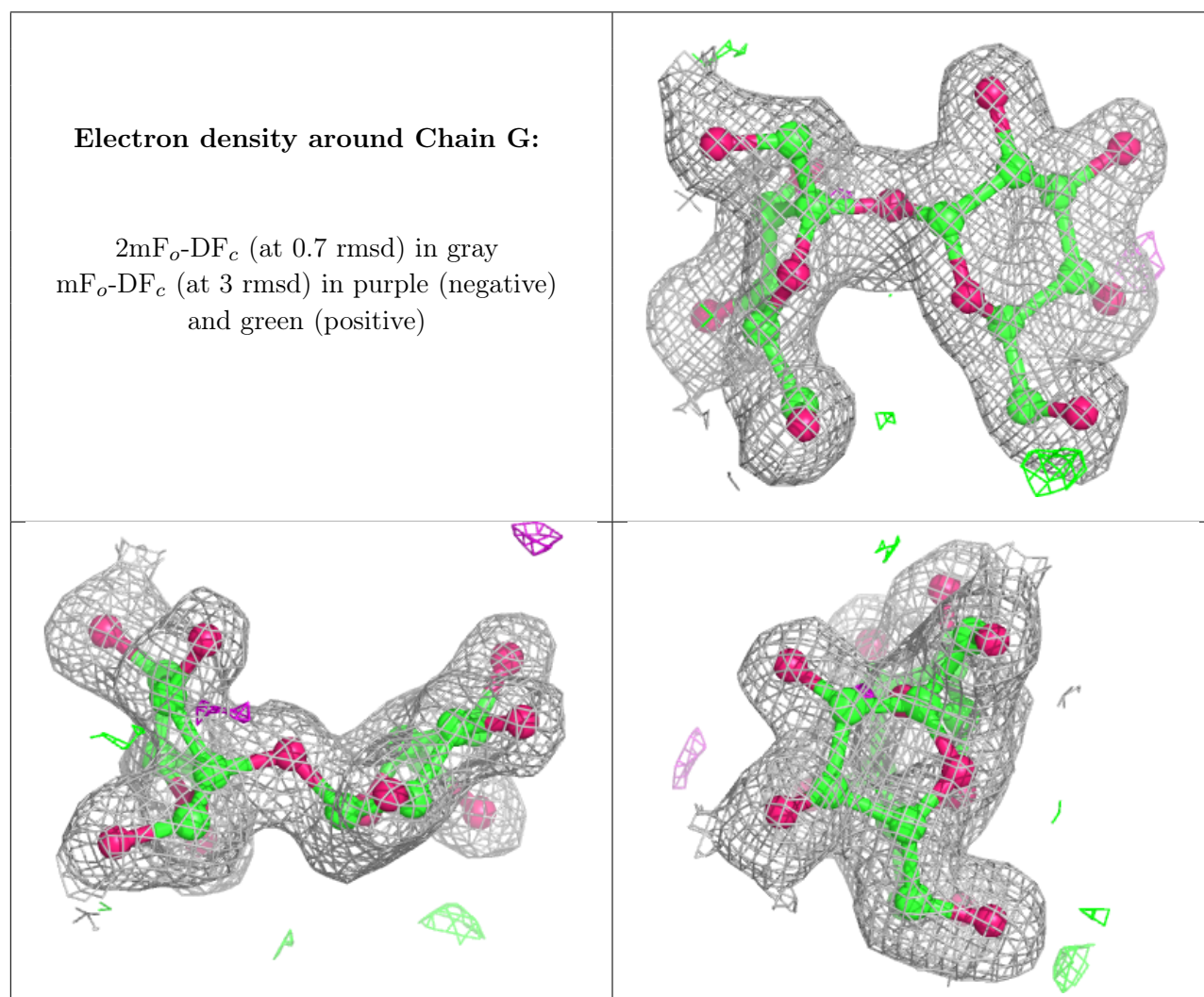
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
2	DOC	B	29	18/19	0.97	0.07	13,14,18,20	0
2	DOC	E	29	18/19	0.98	0.09	8,10,16,18	0

## 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.

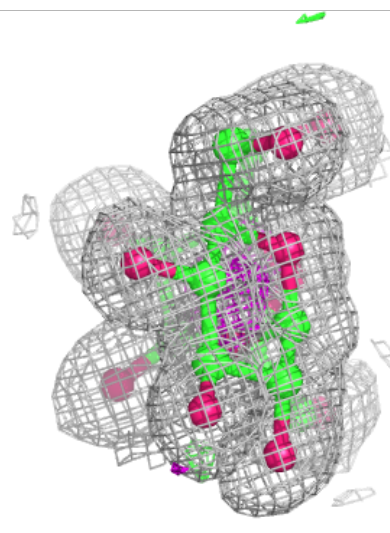
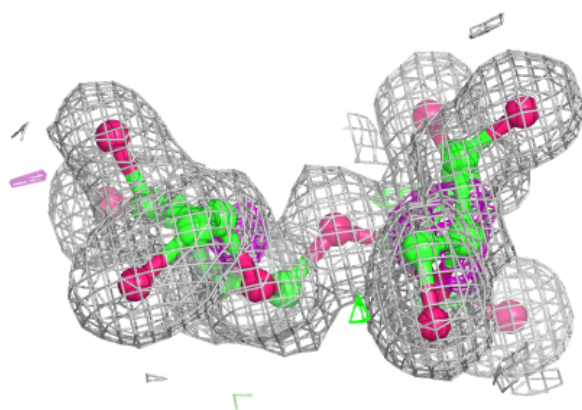
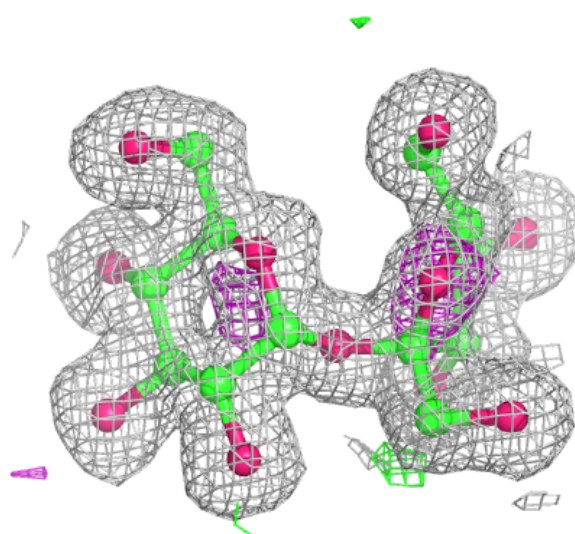
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
4	FRU	G	2	12/12	0.79	0.13	23,29,30,31	0
4	FRU	H	2	12/12	0.91	0.11	11,14,15,15	0
4	GLC	I	1	11/12	0.91	0.11	15,16,17,18	0
4	GLC	G	1	11/12	0.92	0.11	28,30,32,32	0
4	GLC	H	1	11/12	0.93	0.09	15,16,19,20	0
4	FRU	I	2	12/12	0.93	0.08	15,17,19,20	0

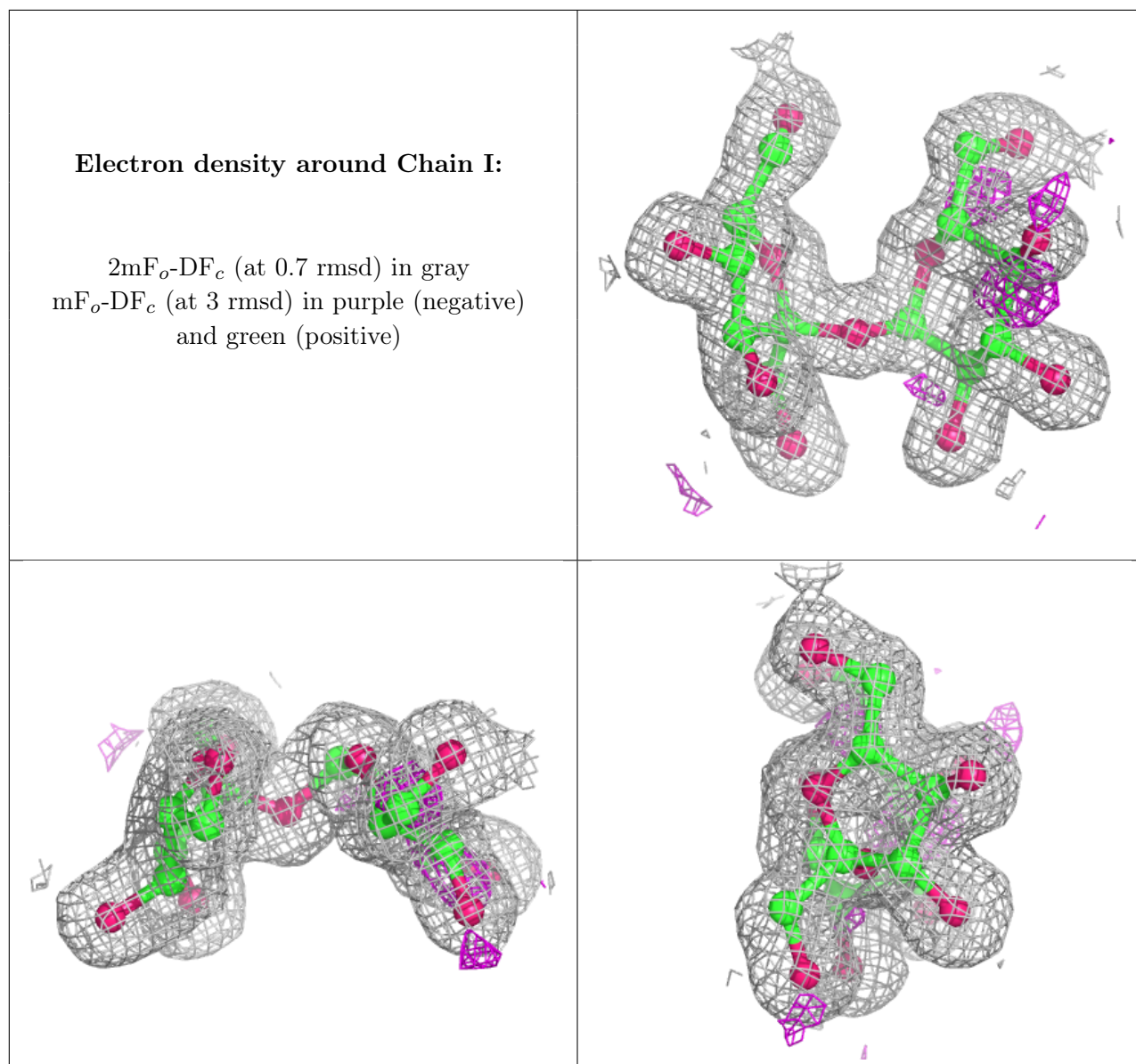
The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.



**Electron density around Chain H:**

$2mF_o-DF_c$  (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
5	MG	A	200	1/1	0.70	0.28	39,39,39,39	0
6	DCP	A	203	28/28	0.77	0.32	23,38,51,51	8
7	SO4	D	3	5/5	0.89	0.20	38,39,40,40	0
6	DCP	A	201	28/28	0.91	0.13	25,28,38,39	0
7	SO4	D	2	5/5	0.95	0.24	64,64,64,64	0

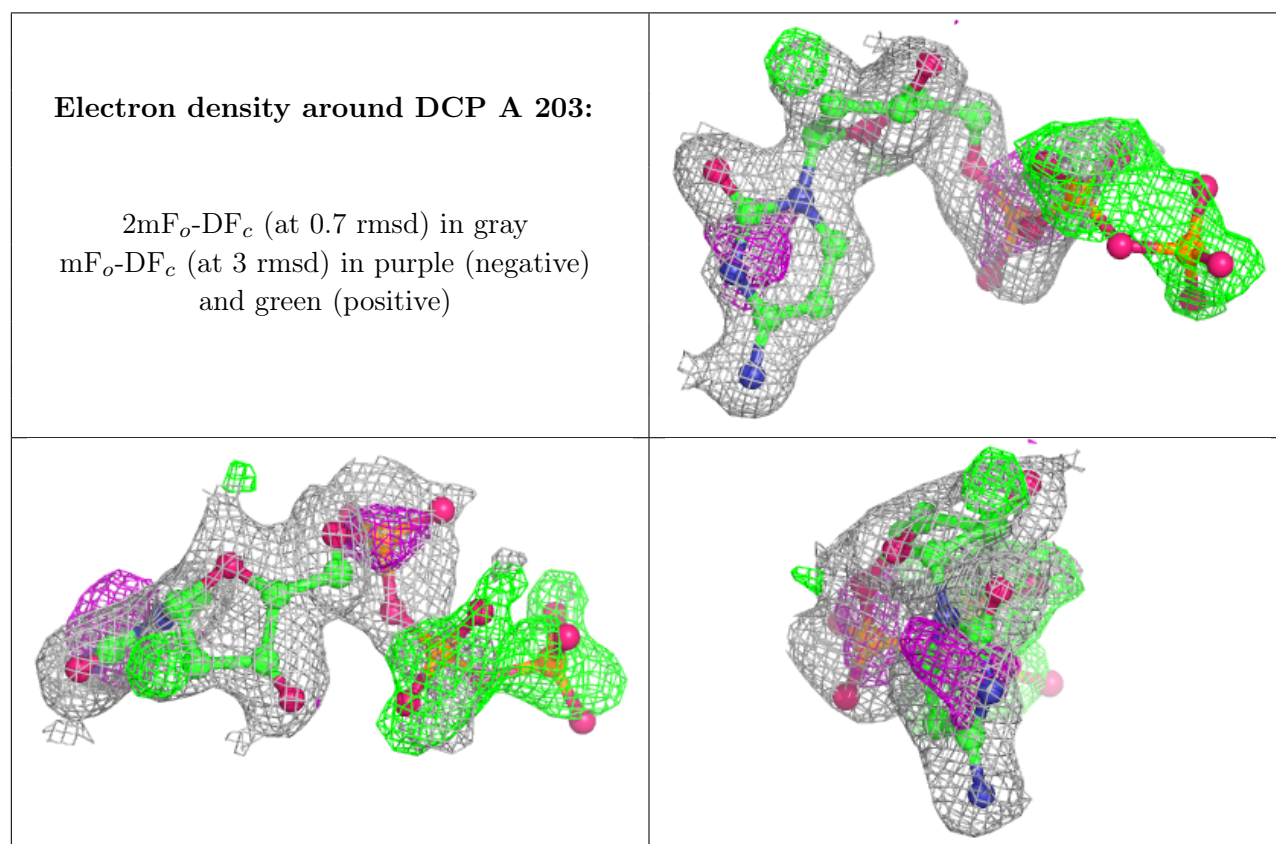
*Continued on next page...*

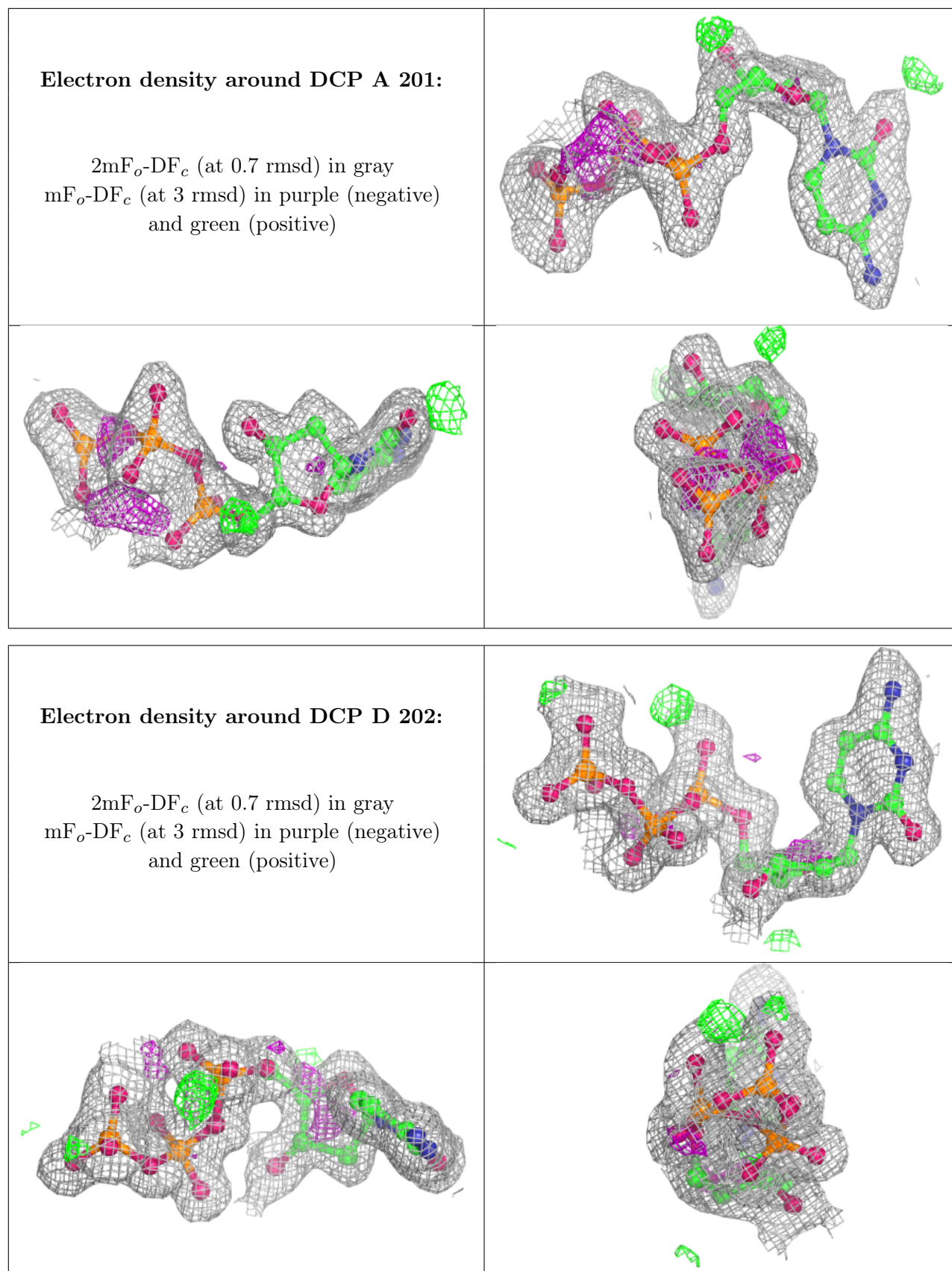


*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
5	MG	D	200	1/1	0.96	0.06	17,17,17,17	0
6	DCP	D	202	28/28	0.96	0.09	16,18,21,24	0
7	SO4	D	1	5/5	0.97	0.13	26,28,31,32	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.





## 6.5 Other polymers [i](#)

There are no such residues in this entry.