



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 24, 2023 – 07:50 AM EDT

PDB ID : 5TXM  
Title : STRUCTURE OF HIV-1 REVERSE TRANSCRIPTASE (RT) TERNARY COMPLEX WITH A DOUBLE STRANDED DNA AND AN INCOMING DDATP  
Authors : Das, K.; Martinez, S.M.; Arnold, E.  
Deposited on : 2016-11-17  
Resolution : 2.70 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 : 2.35.1  
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

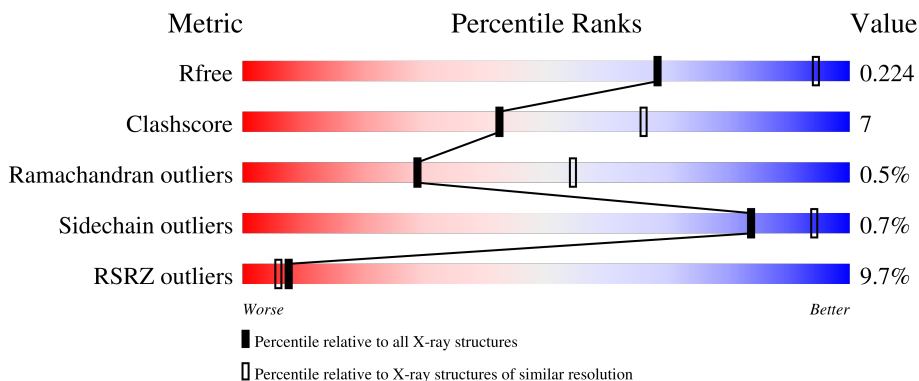
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*



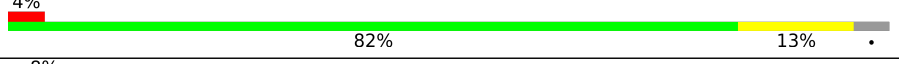
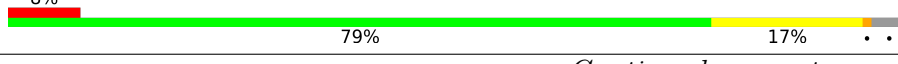
The reported resolution of this entry is 2.70 Å.

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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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	557	
1	C	557	
2	B	428	
2	D	428	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	E	27	<p>7% 48% 37% 11%</p>
3	T	27	<p>11% 70% 19% 11%</p>
4	F	21	<p>48% 33% 14% 5%</p>
4	P	21	<p>5% 57% 33% 5% 5%</p>
5	G	2	<p>50% 50%</p>
5	H	2	<p>100%</p>

## 2 Entry composition [i](#)

There are 10 unique types of molecules in this entry. The entry contains 17950 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 HIV-1 Reverse Transcriptase P66 subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	556	Total	C	N	O	S	0	0	0
			4516	2923	752	833	8			
1	C	556	Total	C	N	O	S	0	0	0
			4516	2923	752	833	8			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	initiating methionine	UNP P03366
A	0	VAL	-	expression tag	UNP P03366
A	258	CYS	GLN	engineered mutation	UNP P03366
A	280	SER	CYS	engineered mutation	UNP P03366
A	498	ASN	ASP	engineered mutation	UNP P03366
C	-1	MET	-	initiating methionine	UNP P03366
C	0	VAL	-	expression tag	UNP P03366
C	258	CYS	GLN	engineered mutation	UNP P03366
C	280	SER	CYS	engineered mutation	UNP P03366
C	498	ASN	ASP	engineered mutation	UNP P03366

- Molecule 2 is a protein called HIV-1 Reverse Transcriptase P51 subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	411	Total	C	N	O	S	0	0	0
			3392	2207	562	616	7			
2	D	412	Total	C	N	O	S	0	0	0
			3400	2212	563	619	6			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	280	SER	CYS	engineered mutation	UNP P03366
D	280	SER	CYS	engineered mutation	UNP P03366

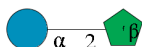
- Molecule 3 is a DNA chain called DNA (5'-D(\*TP\*GP\*GP\*TP\*CP\*GP\*GP\*CP\*GP\*CP\*CP\*GP\*AP\*AP\*CP\*AP\*GP\*GP\*GP\*AP\*CP\*TP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	T	24	Total	C	N	O	P	0	0	0
			494	233	97	141	23			
3	E	24	Total	C	N	O	P	0	0	0
			494	233	97	141	23			

- Molecule 4 is a DNA chain called DNA (5'-D(\*CP\*AP\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*(MRG)P\*CP\*GP\*CP\*CP\*G)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
4	P	20	Total	C	N	O	P	S	0	0	0
			407	195	72	120	19	1			
4	F	20	Total	C	N	O	P	S	0	0	0
			407	195	72	120	19	1			

- Molecule 5 is an oligosaccharide called beta-D-fructofuranose-(2-1)-alpha-D-glucopyranose.

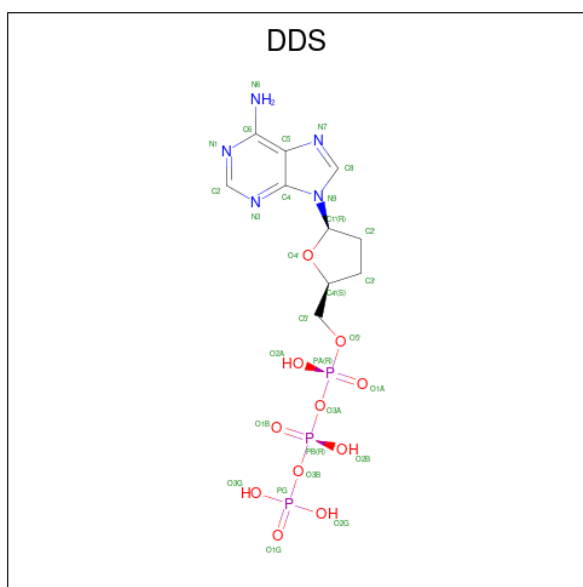


Mol	Chain	Residues	Atoms		ZeroOcc	AltConf	Trace	
5	G	2	Total	C	O	0	0	0
			23	12	11			
5	H	2	Total	C	O	0	0	0
			23	12	11			

- Molecule 6 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	2	Total	Mg	0	0
			2	2		
6	C	2	Total	Mg	0	0
			2	2		

- Molecule 7 is 2',3'-dideoxyadenosine triphosphate (three-letter code: DDS) (formula: C<sub>10</sub>H<sub>16</sub>N<sub>5</sub>O<sub>11</sub>P<sub>3</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	
7	A	1	Total	C	N	O	P	0	0
				29	10	5	11		
7	C	1	Total	C	N	O	P	0	0
				29	10	5	11		

- Molecule 8 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



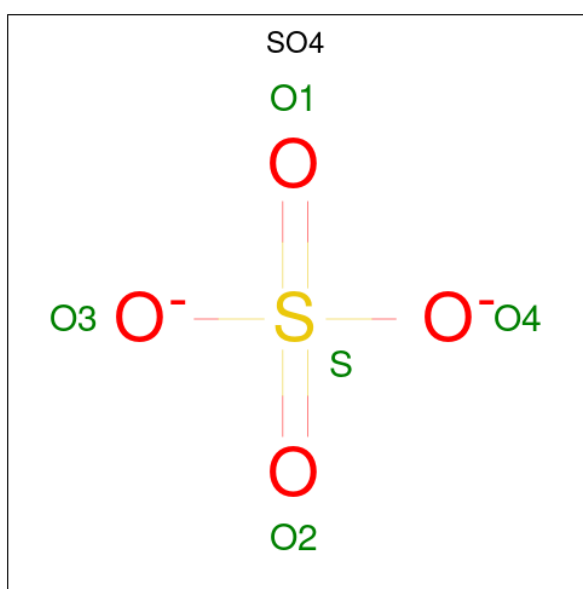
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf	
8	A	1	Total	C	O	0	0
				4	2		
8	B	1	Total	C	O	0	0
				4	2		

*Continued on next page...*

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
8	B	1	Total	C	O	0	0
			4	2	2		
8	T	1	Total	C	O	0	0
			4	2	2		
8	D	1	Total	C	O	0	0
			4	2	2		
8	D	1	Total	C	O	0	0
			4	2	2		

- Molecule 9 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
9	B	1	Total	O	S	0	0
			5	4	1		

- Molecule 10 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
10	A	52	Total	O	0	0
			52	52		
10	B	48	Total	O	0	0
			48	48		
10	T	2	Total	O	0	0
			2	2		
10	P	1	Total	O	0	0
			1	1		

Continued on next page...

*Continued from previous page...*

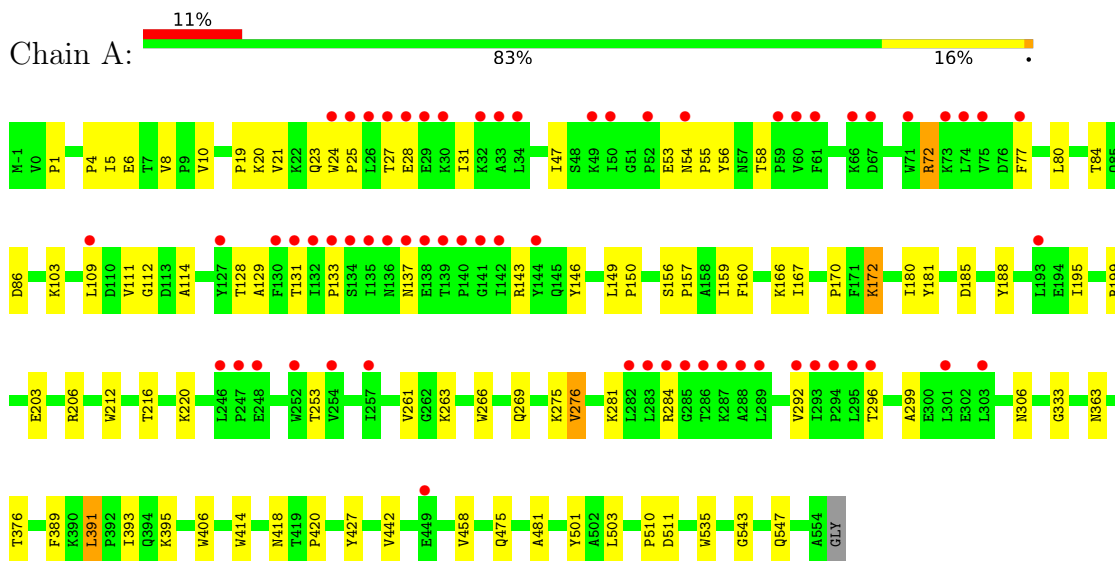
<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
10	C	50	Total	O	0	0
			50	50		
10	D	34	Total	O	0	0
			34	34		



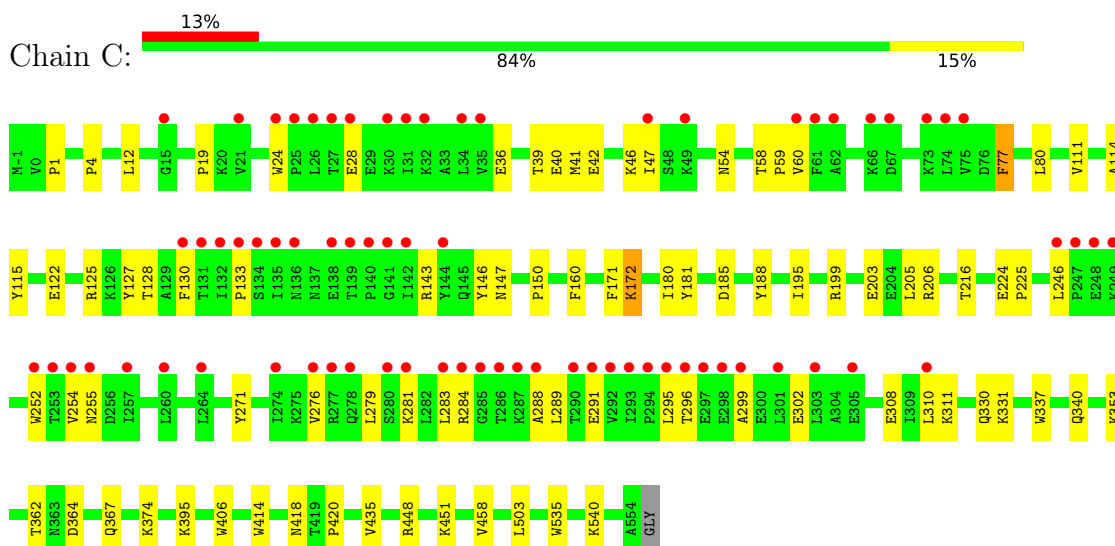
### 3 Residue-property plots

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.

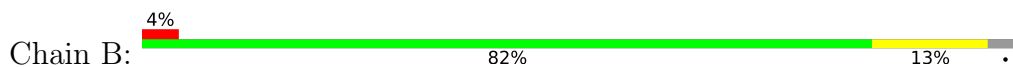
- Molecule 1: HIV-1 Reverse Transcriptase P66 subunit

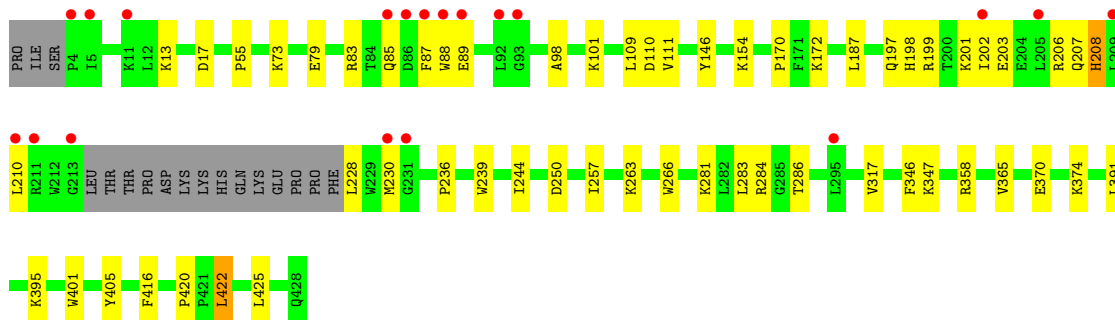


- Molecule 1: HIV-1 Reverse Transcriptase P66 subunit

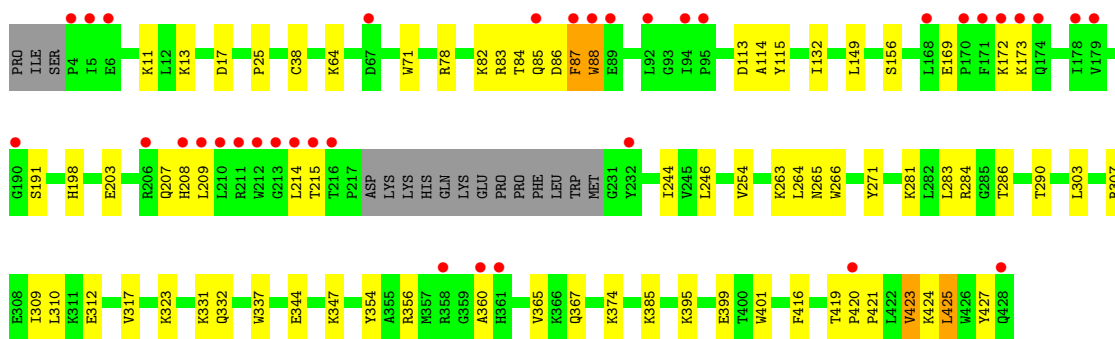
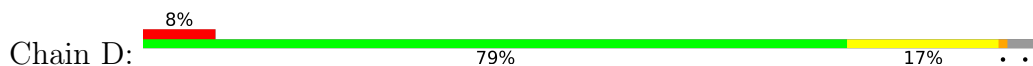


- Molecule 2: HIV-1 Reverse Transcriptase P51 subunit





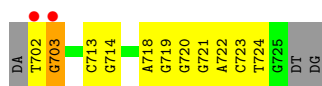
- Molecule 2: HIV-1 Reverse Transcriptase P51 subunit



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



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



- Molecule 4: DNA (5'-D(\*CP\*AP\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*(MRG)P\*CP\*GP\*CP\*CP\*G)-3')



- Molecule 4: DNA (5'-D(\*CP\*AP\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*(MRG)P\*CP\*GP\*CP\*CP\*G)-3')

Chain F:  48% 33% 14% 5%

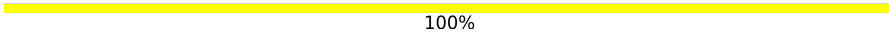


- Molecule 5: beta-D-fructofuranose-(2-1)-alpha-D-glucopyranose

Chain G:  50% 50%



- Molecule 5: beta-D-fructofuranose-(2-1)-alpha-D-glucopyranose

Chain H:  100%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	90.34Å 133.95Å 139.40Å 90.00° 97.71° 90.00°	Depositor
Resolution (Å)	46.05 – 2.70 46.05 – 2.60	Depositor EDS
% Data completeness (in resolution range)	97.4 (46.05-2.70) 95.3 (46.05-2.60)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.79 (at 2.61Å)	Xtrriage
Refinement program	PHENIX (1.10_2155: ???)	Depositor
R, $R_{free}$	0.184 , 0.223 0.185 , 0.224	Depositor DCC
$R_{free}$ test set	2864 reflections (2.98%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	61.9	Xtrriage
Anisotropy	0.160	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 70.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	17950	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	94.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.69% 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: MRG, SO4, MG, GLC, EDO, DDS, FRU

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	0.33	1/4634 (0.0%)	0.50	1/6297 (0.0%)
1	C	0.31	1/4634 (0.0%)	0.49	1/6297 (0.0%)
2	B	0.38	2/3488 (0.1%)	0.53	1/4736 (0.0%)
2	D	0.35	1/3497 (0.0%)	0.60	1/4751 (0.0%)
3	E	0.84	0/555	1.04	1/856 (0.1%)
3	T	0.59	0/555	0.84	0/856
4	F	0.72	0/424	1.11	3/649 (0.5%)
4	P	0.64	0/424	0.90	0/649
All	All	0.40	5/18211 (0.0%)	0.60	8/25091 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
2	D	0	2

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	391	LEU	C-N	9.23	1.51	1.34
1	A	172	LYS	CE-NZ	-6.33	1.33	1.49
1	C	172	LYS	CE-NZ	-6.29	1.33	1.49
2	D	172	LYS	CE-NZ	-6.26	1.33	1.49
2	B	172	LYS	CE-NZ	-6.21	1.33	1.49

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	F	804	DA	O4'-C1'-N9	9.75	114.83	108.00
3	E	703	DG	O4'-C1'-N9	-6.01	103.79	108.00
4	F	805	DG	OP1-P-O3'	5.79	117.93	105.20
2	D	172	LYS	CD-CE-NZ	5.42	124.15	111.70
4	F	805	DG	P-O3'-C3'	5.39	126.16	119.70

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	D	424	LYS	Peptide
2	D	88	TRP	Mainchain

## 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	4516	0	4572	59	0
1	C	4516	0	4572	55	0
2	B	3392	0	3423	41	0
2	D	3400	0	3433	51	0
3	E	494	0	269	13	0
3	T	494	0	269	5	0
4	F	407	0	229	17	0
4	P	407	0	229	6	0
5	G	23	0	21	0	0
5	H	23	0	21	0	0
6	A	2	0	0	0	0
6	C	2	0	0	0	0
7	A	29	0	12	4	0
7	C	29	0	12	2	0
8	A	4	0	6	1	0
8	B	8	0	12	1	0
8	D	8	0	12	1	0
8	T	4	0	6	0	0
9	B	5	0	0	0	0
10	A	52	0	0	0	0
10	B	48	0	0	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
10	C	50	0	0	0	0
10	D	34	0	0	0	0
10	P	1	0	0	0	0
10	T	2	0	0	0	0
All	All	17950	0	17098	240	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 7.

The worst 5 of 240 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)
2:D:419:THR:O	2:D:421:PRO:HD3	1.67	0.94
1:A:172:LYS:HE2	1:A:180:ILE:HB	1.49	0.92
2:B:109:LEU:HD23	2:B:230:MET:HB2	1.57	0.86
3:E:722:DA:H2''	3:E:723:DC:H5''	1.59	0.85
2:D:266:TRP:CZ3	2:D:425:LEU:HG	2.13	0.83

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	554/557 (100%)	535 (97%)	15 (3%)	4 (1%)	22	46
1	C	554/557 (100%)	534 (96%)	16 (3%)	4 (1%)	22	46
2	B	407/428 (95%)	396 (97%)	11 (3%)	0	100	100
2	D	408/428 (95%)	396 (97%)	11 (3%)	1 (0%)	47	73
All	All	1923/1970 (98%)	1861 (97%)	53 (3%)	9 (0%)	29	54

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	4	PRO
1	C	4	PRO
1	A	1	PRO
1	C	1	PRO
2	D	423	VAL

### 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	495/497 (100%)	491 (99%)	4 (1%)	81	93
1	C	495/497 (100%)	493 (100%)	2 (0%)	91	97
2	B	372/390 (95%)	369 (99%)	3 (1%)	81	93
2	D	374/390 (96%)	370 (99%)	4 (1%)	73	90
All	All	1736/1774 (98%)	1723 (99%)	13 (1%)	84	94

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	77	PHE
1	C	276	VAL
2	D	425	LEU
2	D	87	PHE
2	D	208	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	547	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](#)

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
4	MRG	F	817	4,3,1	21,28,29	2.45	8 (38%)	19,39,42	1.75	5 (26%)
4	MRG	P	817	4,3,1	21,28,29	2.48	8 (38%)	19,39,42	1.63	5 (26%)

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	MRG	F	817	4,3,1	-	2/8/26/27	0/3/3/3
4	MRG	P	817	4,3,1	-	2/8/26/27	0/3/3/3

The worst 5 of 16 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	P	817	MRG	C2-N2	7.12	1.49	1.34
4	F	817	MRG	C2-N2	7.06	1.49	1.34
4	F	817	MRG	C2-N1	4.47	1.43	1.36
4	P	817	MRG	C2-N1	4.26	1.43	1.36
4	P	817	MRG	C4-N3	3.39	1.45	1.37

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	F	817	MRG	C22-C21-N2	3.96	123.52	112.21
4	P	817	MRG	C22-C21-N2	3.39	121.91	112.21
4	F	817	MRG	C5-C6-N1	3.39	119.94	113.95
4	P	817	MRG	C5-C6-N1	3.27	119.73	113.95
4	F	817	MRG	O5'-C5'-C4'	3.11	119.58	108.99

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	P	817	MRG	C22-C21-N2-C2
4	P	817	MRG	N2-C21-C22-C23
4	F	817	MRG	C22-C21-N2-C2
4	F	817	MRG	N2-C21-C22-C23

There are no ring outliers.

2 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	F	817	MRG	4	0
4	P	817	MRG	1	0

## 5.5 Carbohydrates i

4 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
5	GLC	G	1	5	11,11,12	0.87	0	15,15,17	0.94	0
5	FRU	G	2	5	11,12,12	2.51	6 (54%)	10,18,18	5.28	3 (30%)
5	GLC	H	1	5	11,11,12	0.93	1 (9%)	15,15,17	1.31	2 (13%)
5	FRU	H	2	5	11,12,12	3.04	4 (36%)	10,18,18	5.25	5 (50%)

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
5	GLC	G	1	5	-	0/2/19/22	0/1/1/1
5	FRU	G	2	5	-	3/5/24/24	0/1/1/1

*Continued on next page...*

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	GLC	H	1	5	-	0/2/19/22	0/1/1/1
5	FRU	H	2	5	-	2/5/24/24	0/1/1/1

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	H	2	FRU	C4-C5	-7.62	1.33	1.53
5	G	2	FRU	O5-C2	5.51	1.51	1.43
5	H	2	FRU	C4-C3	4.01	1.69	1.52
5	H	2	FRU	O2-C2	3.44	1.46	1.40
5	G	2	FRU	O5-C5	-3.28	1.36	1.43

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	G	2	FRU	O2-C2-O5	-13.76	82.93	109.50
5	H	2	FRU	C5-C4-C3	-9.15	72.54	101.91
5	H	2	FRU	O5-C5-C6	-8.60	84.91	108.85
5	H	2	FRU	C6-C5-C4	-8.10	95.57	115.09
5	G	2	FRU	O5-C5-C4	-8.06	85.43	105.49

There are no chirality outliers.

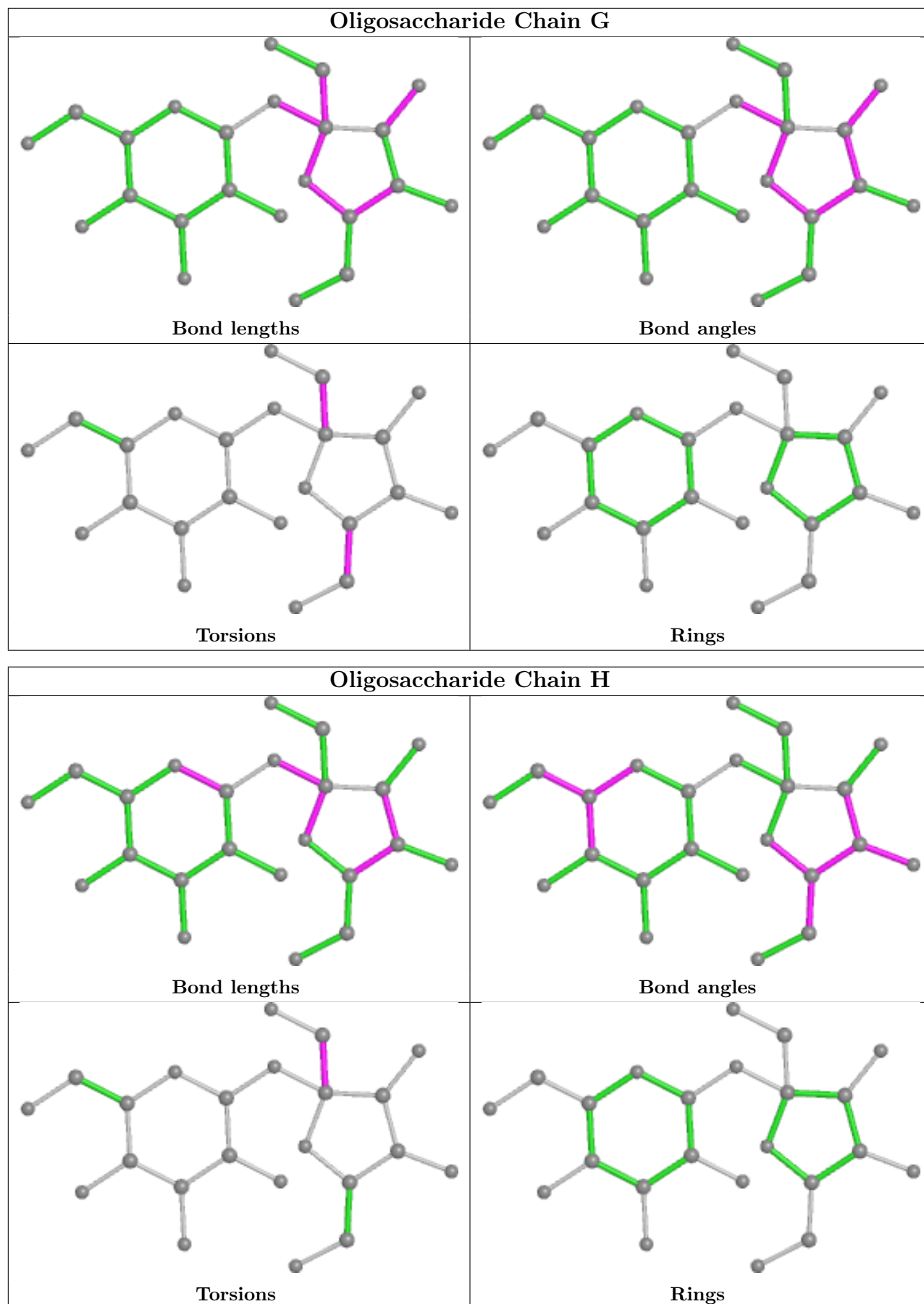
All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	H	2	FRU	O1-C1-C2-O2
5	G	2	FRU	O5-C5-C6-O6
5	G	2	FRU	O1-C1-C2-O2
5	G	2	FRU	O1-C1-C2-C3
5	H	2	FRU	O1-C1-C2-C3

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 13 ligands modelled in this entry, 4 are monoatomic - leaving 9 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
8	EDO	D	503	-	3,3,3	0.52	0	2,2,2	0.37	0
8	EDO	B	503	-	3,3,3	0.47	0	2,2,2	0.38	0
8	EDO	B	504	-	3,3,3	0.54	0	2,2,2	0.27	0
8	EDO	T	801	-	3,3,3	0.48	0	2,2,2	0.25	0
7	DDS	C	601	6	25,31,31	1.41	4 (16%)	26,48,48	1.57	4 (15%)
8	EDO	D	502	-	3,3,3	0.47	0	2,2,2	0.37	0
8	EDO	A	604	-	3,3,3	0.46	0	2,2,2	0.27	0
7	DDS	A	603	6	25,31,31	1.40	4 (16%)	26,48,48	1.52	6 (23%)
9	SO4	B	502	-	4,4,4	0.12	0	6,6,6	0.15	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. '2' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
8	EDO	D	503	-	-	1/1/1/1	-
8	EDO	B	503	-	-	1/1/1/1	-
8	EDO	B	504	-	-	0/1/1/1	-
8	EDO	T	801	-	-	0/1/1/1	-
7	DDS	C	601	6	-	5/18/31/31	0/3/3/3
8	EDO	D	502	-	-	1/1/1/1	-
8	EDO	A	604	-	-	0/1/1/1	-
7	DDS	A	603	6	-	5/18/31/31	0/3/3/3

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	A	603	DDS	C6-N6	3.59	1.47	1.34
7	C	601	DDS	C6-N6	3.58	1.47	1.34

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	C	601	DDS	C5'-C4'	-3.01	1.41	1.50
7	A	603	DDS	C5'-C4'	-2.87	1.41	1.50
7	C	601	DDS	O5'-C5'	-2.67	1.34	1.44

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	A	603	DDS	N3-C2-N1	-4.17	122.16	128.68
7	C	601	DDS	N3-C2-N1	-4.03	122.39	128.68
7	A	603	DDS	PB-O3B-PG	-3.44	121.02	132.83
7	C	601	DDS	C3'-C2'-C1'	3.36	106.66	102.78
7	C	601	DDS	PB-O3A-PA	-2.87	122.97	132.83

There are no chirality outliers.

5 of 13 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
7	A	603	DDS	C5'-O5'-PA-O1A
7	A	603	DDS	C3'-C4'-C5'-O5'
7	A	603	DDS	O4'-C4'-C5'-O5'
7	C	601	DDS	C3'-C4'-C5'-O5'
7	C	601	DDS	O4'-C4'-C5'-O5'

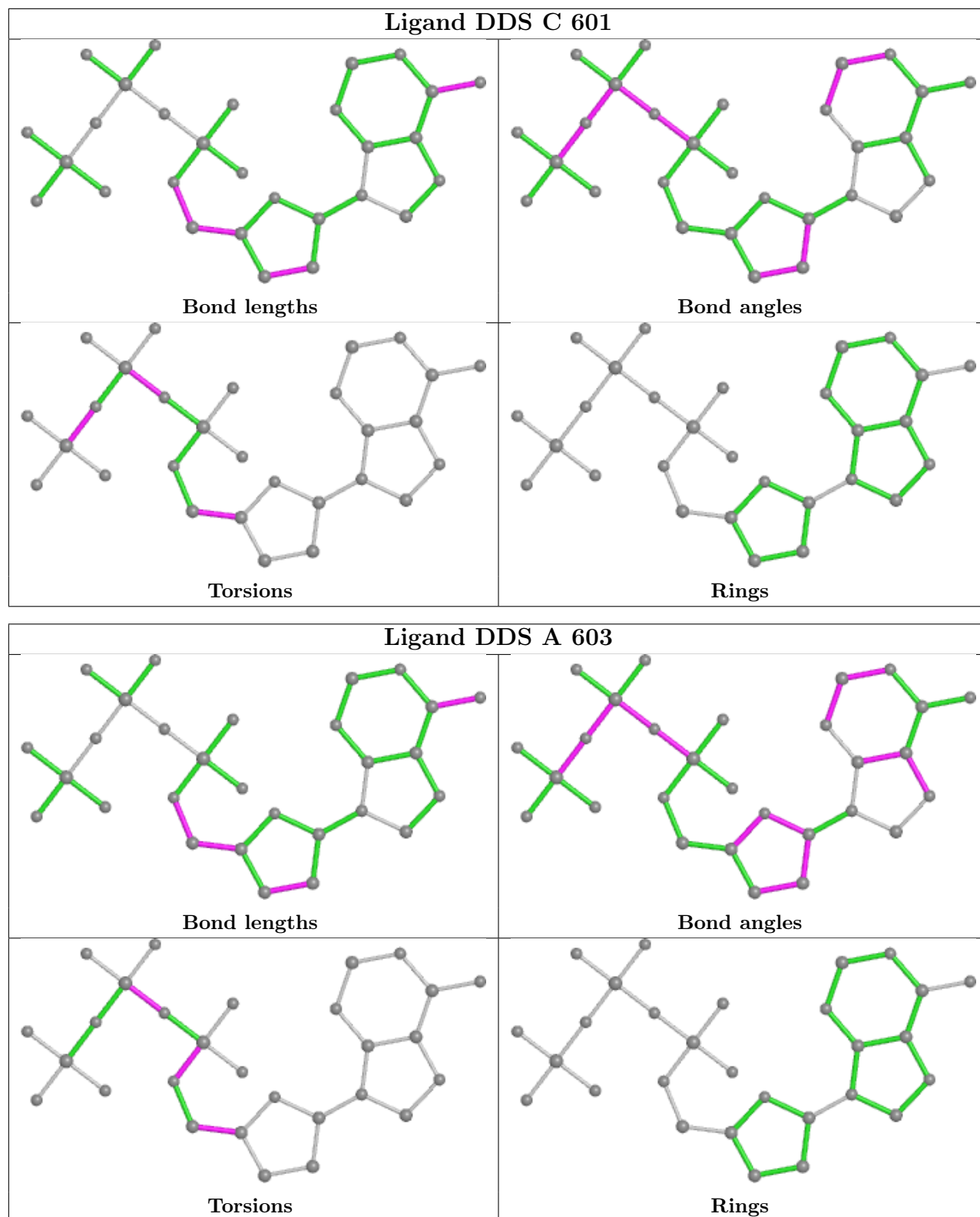
There are no ring outliers.

5 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
8	D	503	EDO	1	0
8	B	504	EDO	1	0
7	C	601	DDS	2	0
8	A	604	EDO	1	0
7	A	603	DDS	4	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 [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	556/557 (99%)	0.42	63 (11%) 5 4	31, 85, 182, 297	0
1	C	556/557 (99%)	0.53	72 (12%) 3 2	35, 92, 196, 259	0
2	B	411/428 (96%)	0.13	19 (4%) 32 31	34, 69, 132, 213	0
2	D	412/428 (96%)	0.25	36 (8%) 10 8	40, 83, 162, 226	0
3	E	24/27 (88%)	0.38	2 (8%) 11 9	80, 121, 199, 256	0
3	T	24/27 (88%)	0.14	3 (12%) 3 3	75, 115, 214, 240	0
4	F	19/21 (90%)	-0.29	0 100 100	64, 103, 209, 212	0
4	P	19/21 (90%)	-0.12	1 (5%) 26 25	69, 105, 182, 220	0
All	All	2021/2066 (97%)	0.34	196 (9%) 7 6	31, 83, 180, 297	0

The worst 5 of 196 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	293	ILE	10.3
1	C	142	ILE	9.6
2	D	214	LEU	8.9
1	C	292	VAL	8.6
2	D	4	PRO	8.4

### 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
4	MRG	P	817	26/27	0.90	0.14	95,122,147,149	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
4	MRG	F	817	26/27	0.91	0.13	111,128,154,182	0

### 6.3 Carbohydrates

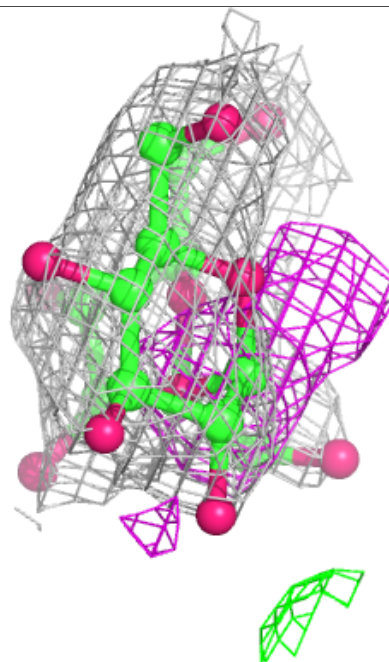
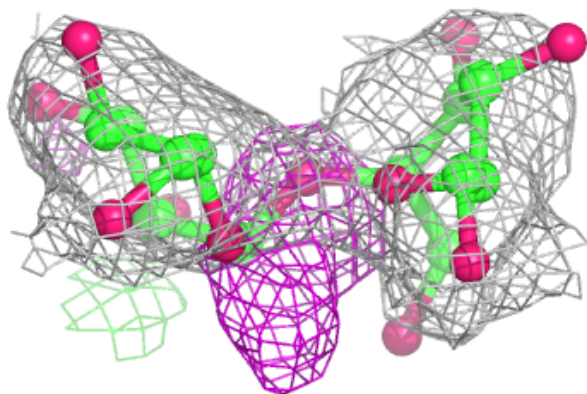
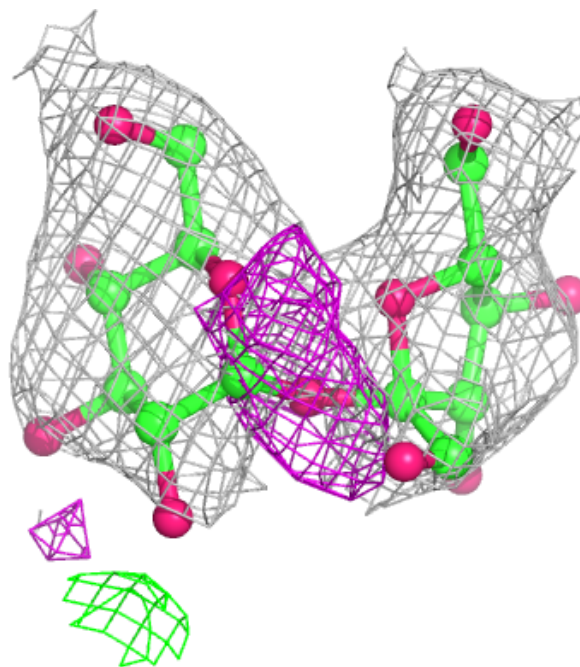
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

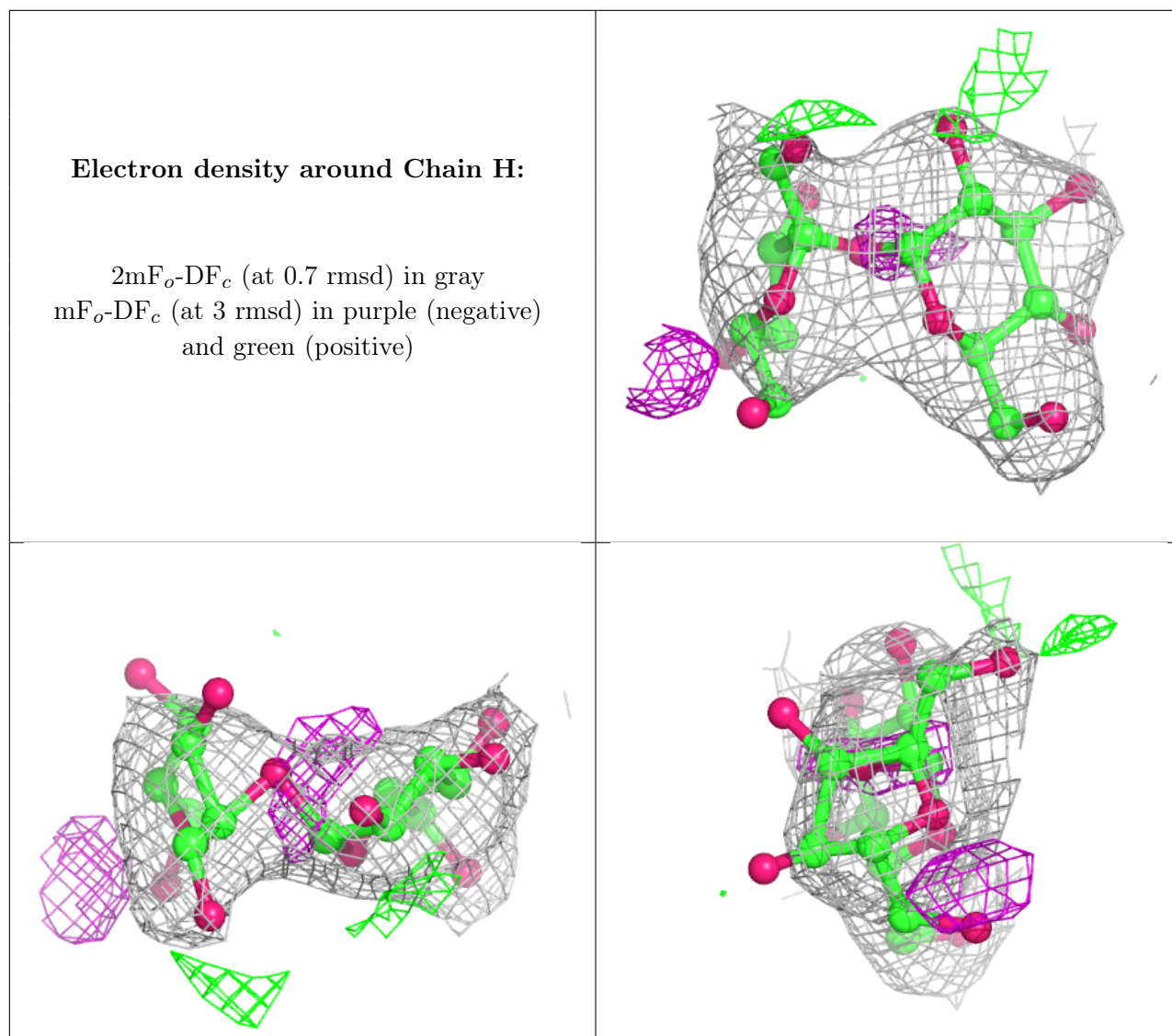
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
5	FRU	G	2	12/12	0.78	0.29	84,128,135,141	0
5	FRU	H	2	12/12	0.79	0.28	125,146,150,155	0
5	GLC	H	1	11/12	0.86	0.20	80,109,129,134	0
5	GLC	G	1	11/12	0.88	0.23	79,105,117,129	0

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](#)

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.

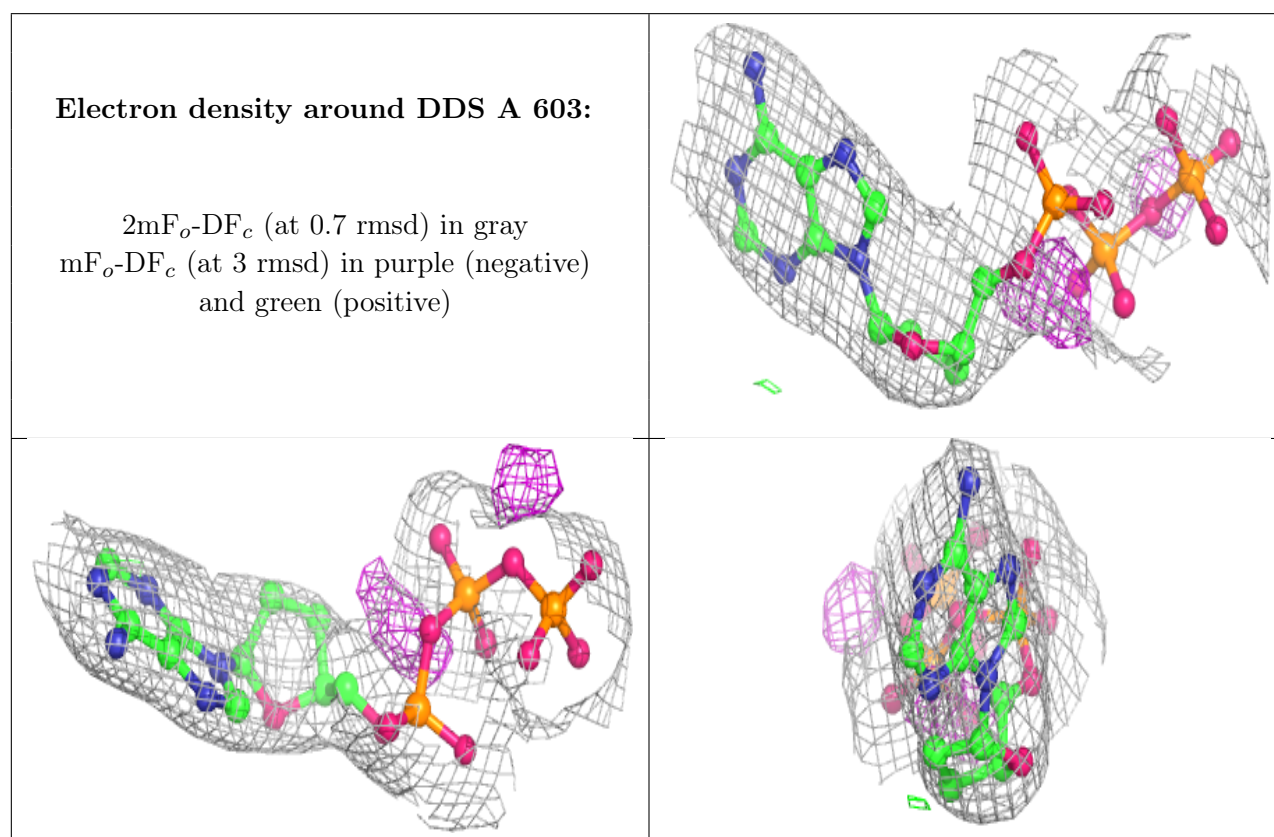
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
8	EDO	B	504	4/4	0.73	0.31	73,80,89,93	0
6	MG	A	602	1/1	0.79	0.16	117,117,117,117	0
8	EDO	D	502	4/4	0.80	0.27	91,93,93,96	0
9	SO4	B	502	5/5	0.82	0.40	150,152,154,161	0
6	MG	C	602	1/1	0.85	0.10	109,109,109,109	0
8	EDO	B	503	4/4	0.91	0.18	64,67,73,81	0
8	EDO	T	801	4/4	0.91	0.29	82,92,100,101	0

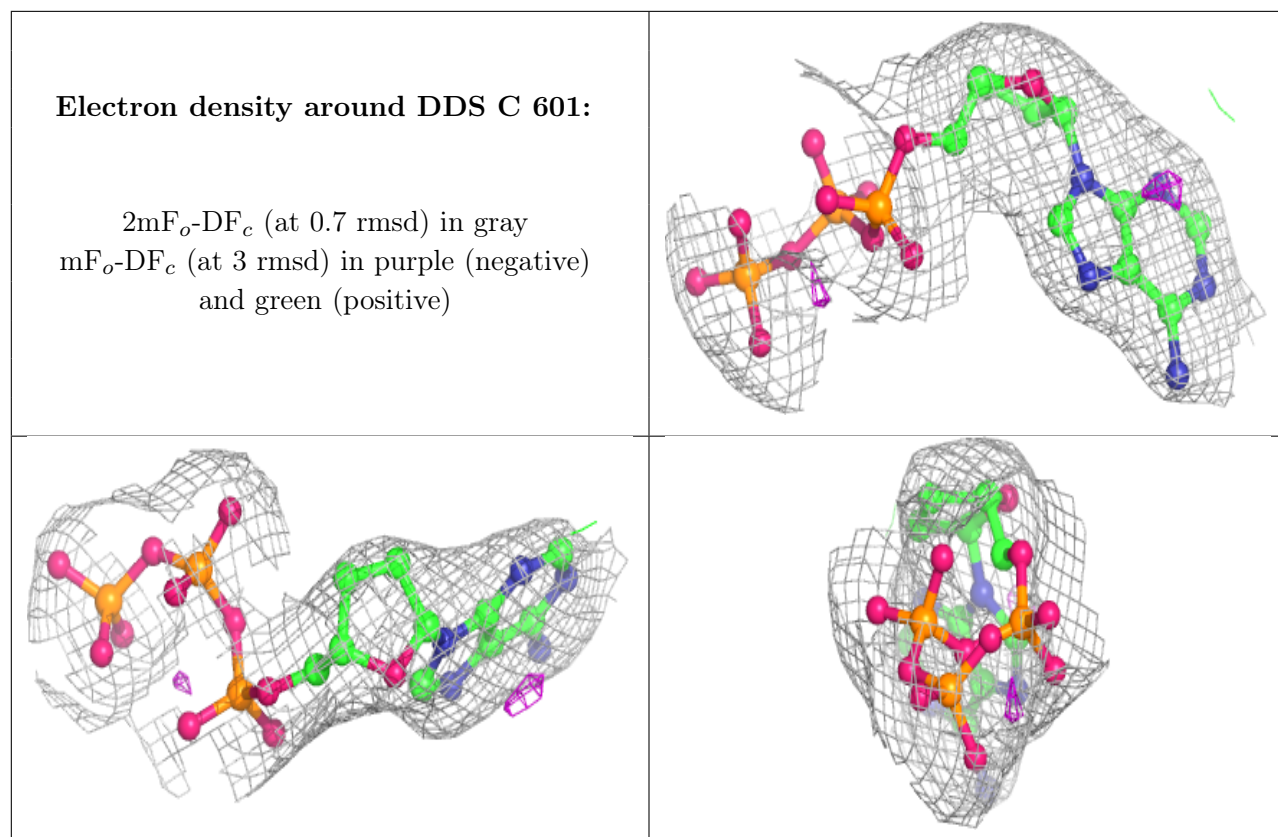
*Continued on next page...*

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
8	EDO	D	503	4/4	0.92	0.43	66,70,77,82	0
7	DDS	A	603	29/29	0.93	0.14	66,92,155,158	0
8	EDO	A	604	4/4	0.93	0.10	79,84,94,96	0
6	MG	A	601	1/1	0.94	0.14	159,159,159,159	0
7	DDS	C	601	29/29	0.96	0.14	66,81,131,138	0
6	MG	C	603	1/1	0.98	0.18	107,107,107,107	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.