



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 21, 2020 – 04:34 PM BST

PDB ID : 4F2R  
Title : DNA Polymerase I Large Fragment complex 3  
Authors : Wang, W.; Beese, L.S.  
Deposited on : 2012-05-08  
Resolution : 1.63 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.13.1  
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) : 2.13.1

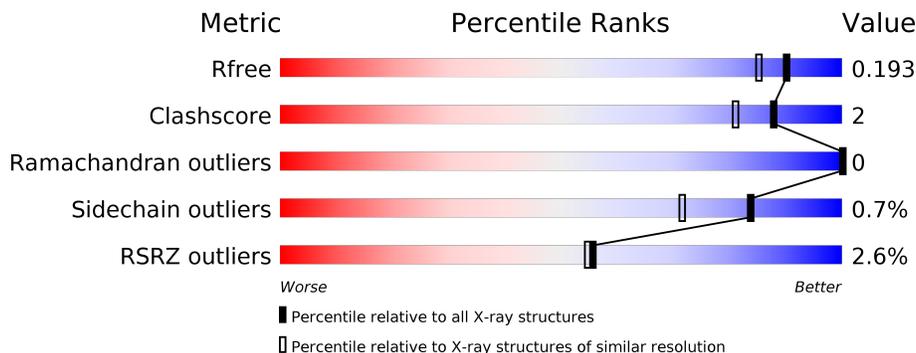
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.63 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	3122 (1.66-1.62)
Clashscore	141614	3268 (1.66-1.62)
Ramachandran outliers	138981	3215 (1.66-1.62)
Sidechain outliers	138945	3215 (1.66-1.62)
RSRZ outliers	127900	3079 (1.66-1.62)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	592	 3% 94% 2% 1% 1%
1	D	592	 2% 93% 5% 1% 1%
2	B	9	 56% 33% 11%
2	E	9	 56% 33% 11%
3	C	13	 46% 31% 8% 15%
3	F	13	 8% 38% 38% 15% 8%

*Continued on next page...*

*Continued from previous page...*

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

## 2 Entry composition i

There are 8 unique types of molecules in this entry. The entry contains 21703 atoms, of which 10119 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.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	578	9418	2971	4754	805	869	19	0	6	0
1	D	583	9570	3016	4840	815	882	17	11	15	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	598	ALA	ASP	engineered mutation	UNP Q5KWC1
A	710	TYR	PHE	engineered mutation	UNP Q5KWC1
A	823	HIS	ARG	SEE REMARK 999	UNP Q5KWC1
D	598	ALA	ASP	engineered mutation	UNP Q5KWC1
D	710	TYR	PHE	engineered mutation	UNP Q5KWC1
D	823	HIS	ARG	SEE REMARK 999	UNP Q5KWC1

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
2	B	9	274	85	100	29	52	8	0	0	0
2	E	9	274	85	100	29	52	8	0	0	0

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

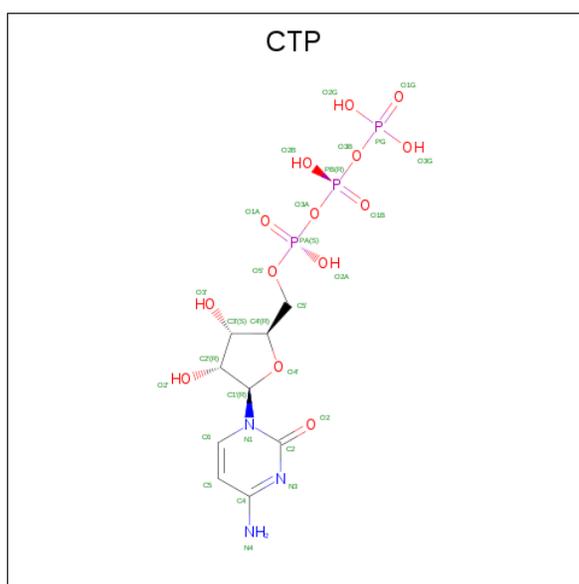
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
3	C	11	356	109	123	47	66	11	0	0	0
3	F	12	385	119	134	52	69	11	0	0	0

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



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	H	O			
4	G	2	45	12	22	11	0	0	0
4	H	2	45	12	22	11	0	0	0

- Molecule 5 is CYTIDINE-5'-TRIPHOSPHATE (three-letter code: CTP) (formula:  $C_9H_{16}N_3O_{14}P_3$ ).

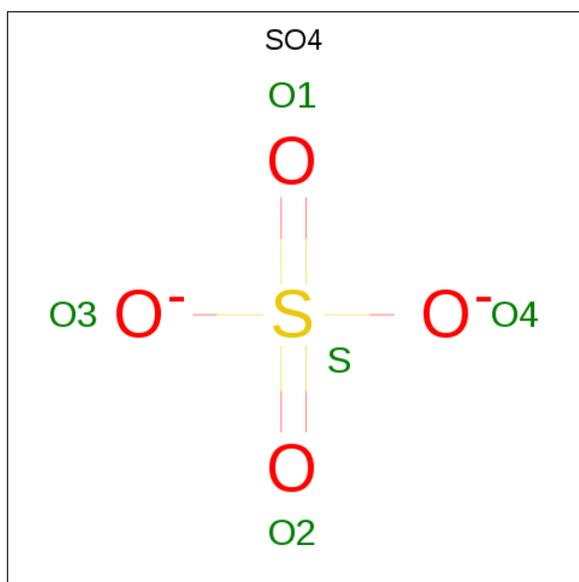


Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
			Total	C	H	N	O	P		
5	A	1	41	9	12	3	14	3	0	0
5	D	1	41	9	12	3	14	3	0	0

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

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	D	1	Total Mg 1 1	0	0

- Molecule 7 is SULFATE ION (three-letter code: SO4) (formula:  $O_4S$ ).



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

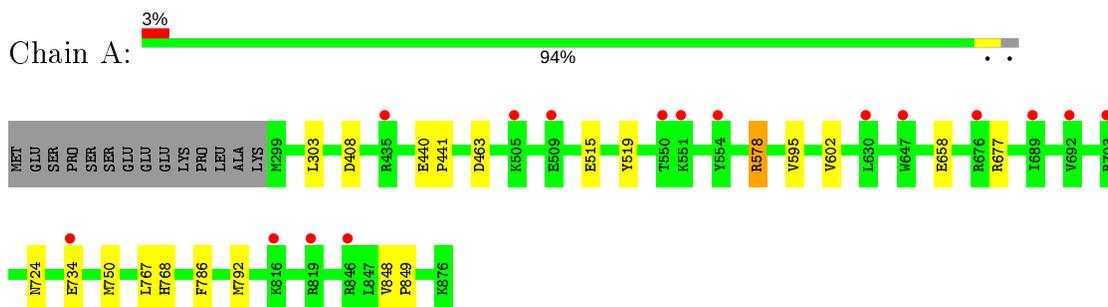
- Molecule 8 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	A	435	Total O 435 435	0	0
8	D	643	Total O 643 643	0	0
8	B	29	Total O 29 29	0	0
8	C	45	Total O 45 45	0	0
8	E	34	Total O 34 34	0	0
8	F	52	Total O 52 52	0	0

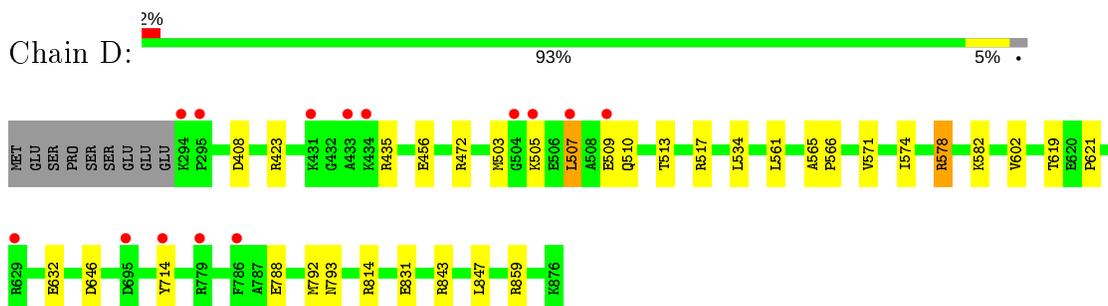
### 3 Residue-property plots [i](#)

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: DNA polymerase



- Molecule 1: DNA polymerase



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



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



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



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



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



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



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	93.22Å 108.48Å 150.95Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	34.98 – 1.63 88.09 – 1.63	Depositor EDS
% Data completeness (in resolution range)	96.8 (34.98-1.63) 96.8 (88.09-1.63)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.71 (at 1.63Å)	Xtrriage
Refinement program	PHENIX dev_1026	Depositor
R, $R_{free}$	0.170 , 0.192 0.171 , 0.193	Depositor DCC
$R_{free}$ test set	8311 reflections (4.51%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	19.2	Xtrriage
Anisotropy	0.080	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.42 , 52.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	21703	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	33.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.16% 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: MG, CTP, DOC, GLC, SO4, 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	0.45	0/4766	0.57	1/6440 (0.0%)
1	D	0.58	1/4871 (0.0%)	0.71	6/6586 (0.1%)
2	B	0.99	1/173 (0.6%)	1.56	2/264 (0.8%)
2	E	1.06	0/173	1.86	3/264 (1.1%)
3	C	0.98	0/262	1.64	4/404 (1.0%)
3	F	1.07	0/283	1.66	5/437 (1.1%)
All	All	0.58	2/10528 (0.0%)	0.80	21/14395 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	831	GLU	CB-CG	-6.83	1.39	1.52
2	B	25	DA	C3'-O3'	-5.09	1.37	1.44

All (21) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	23	DT	O4'-C1'-N1	-10.95	100.33	108.00
3	F	4	DG	O4'-C1'-N9	10.73	115.51	108.00
1	D	578	ARG	NE-CZ-NH2	-9.88	115.36	120.30
2	B	23	DT	O4'-C1'-N1	-8.65	101.95	108.00
3	C	8	DT	O4'-C1'-N1	-8.49	102.05	108.00
3	C	3	DG	O5'-P-OP1	-8.48	98.07	105.70
1	D	843	ARG	NE-CZ-NH2	-7.60	116.50	120.30
3	F	8	DT	O4'-C1'-N1	-7.45	102.78	108.00
1	D	578	ARG	NE-CZ-NH1	7.34	123.97	120.30
1	A	578	ARG	NE-CZ-NH2	-6.88	116.86	120.30
1	D	859	ARG	NE-CZ-NH2	-6.45	117.08	120.30
3	F	5	DG	O4'-C4'-C3'	-6.28	101.99	104.50
1	D	859	ARG	NE-CZ-NH1	6.23	123.42	120.30

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	22	DC	O4'-C1'-N1	-5.74	103.98	108.00
1	D	843	ARG	NE-CZ-NH1	5.72	123.16	120.30
2	E	26	DC	O4'-C1'-N1	5.47	111.83	108.00
2	B	26	DC	O4'-C1'-N1	5.37	111.76	108.00
3	F	3	DG	OP1-P-OP2	5.31	127.57	119.60
3	F	4	DG	O4'-C4'-C3'	5.30	109.18	106.00
3	C	5	DG	N3-C4-N9	-5.20	122.88	126.00
3	C	11	DG	N3-C2-N2	-5.09	116.34	119.90

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	4664	4754	4742	13	0
1	D	4730	4840	4805	23	0
2	B	174	100	103	1	0
2	E	174	100	103	2	0
3	C	233	123	124	3	0
3	F	251	134	136	6	0
4	G	23	22	20	1	0
4	H	23	22	21	1	0
5	A	29	12	12	0	0
5	D	29	12	12	0	0
6	D	1	0	0	0	0
7	D	15	0	0	0	0
8	A	435	0	0	3	1
8	B	29	0	0	0	0
8	C	45	0	0	1	0
8	D	643	0	0	4	2
8	E	34	0	0	0	0
8	F	52	0	0	2	0
All	All	11584	10119	10078	40	2

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 2.

All (40) 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:472:ARG:NE	8:D:1481:HOH:O	2.22	0.72
1:D:435:ARG:NH2	8:D:1360:HOH:O	2.24	0.70
1:A:463:ASP:OD2	8:A:1356:HOH:O	2.13	0.67
1:A:734:GLU:OE2	8:A:1422:HOH:O	2.13	0.64
1:D:423:ARG:NH2	8:D:1315:HOH:O	2.31	0.64
1:D:714:TYR:HD2	3:F:4:DG:O4'	1.81	0.63
3:F:12:DG:O4'	8:F:143:HOH:O	2.16	0.61
1:D:602[B]:VAL:HG21	1:D:621:PRO:HG3	1.82	0.60
1:A:595:VAL:HG22	1:A:602:VAL:HG13	1.86	0.57
1:A:724:ASN:ND2	3:C:3:DG:C6	2.72	0.57
1:A:408:ASP:HB2	4:G:2:FRU:H11	1.87	0.56
1:A:658:GLU:OE2	8:A:1116:HOH:O	2.19	0.54
1:D:510:GLN:NE2	1:D:632:GLU:OE2	2.41	0.54
1:D:509:GLU:O	1:D:513:THR:HG23	2.09	0.53
1:D:503:MET:O	1:D:507:LEU:HB2	2.09	0.52
1:D:408:ASP:HB2	4:H:2:FRU:H11	1.92	0.52
3:C:12:DG:H1'	8:C:129:HOH:O	2.11	0.51
1:D:561:LEU:HB3	1:D:571:VAL:HG13	1.92	0.51
1:A:724:ASN:ND2	3:C:3:DG:O6	2.44	0.50
3:F:9:DC:H2''	3:F:10:DA:C8	2.47	0.50
1:A:750[B]:MET:SD	1:A:792:MET:HG2	2.52	0.49
1:D:814:ARG:NH2	1:D:847[A]:LEU:HD13	2.29	0.48
1:A:578:ARG:HD3	2:B:26:DC:OP1	2.15	0.47
1:D:534:LEU:HD11	1:D:574:ILE:HD13	1.97	0.47
1:D:814:ARG:CZ	1:D:847[A]:LEU:CD1	2.93	0.46
1:A:440:GLU:HB3	1:A:441:PRO:HD3	1.99	0.44
1:D:565:ALA:N	1:D:566:PRO:CD	2.81	0.44
3:F:12:DG:C4'	8:F:143:HOH:O	2.65	0.44
1:D:619[B]:THR:HG23	3:F:8:DT:OP1	2.18	0.44
1:A:515:GLU:HG2	1:A:519:TYR:CE2	2.53	0.44
1:D:788:GLU:O	1:D:792:MET:HG3	2.18	0.43
1:D:602[B]:VAL:HG21	1:D:621:PRO:CG	2.49	0.42
1:D:561:LEU:O	1:D:571:VAL:HG11	2.20	0.42
1:A:848:VAL:HB	1:A:849:PRO:HD3	2.02	0.42
1:D:646:ASP:N	1:D:646:ASP:OD1	2.37	0.42
1:D:456:GLU:HG3	8:D:1240:HOH:O	2.20	0.41
1:D:578:ARG:HD3	2:E:26:DC:OP1	2.20	0.41
1:D:582:LYS:HE3	2:E:26:DC:H1'	2.01	0.41

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:767:LEU:O	1:A:768:HIS:HB2	2.21	0.41
1:D:619[B]:THR:HG23	3:F:8:DT:P	2.62	0.40

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
8:D:1226:HOH:O	8:D:1393:HOH:O[4_445]	2.16	0.04
8:A:1345:HOH:O	8:D:1576:HOH:O[2_755]	2.19	0.01

## 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	582/592 (98%)	566 (97%)	16 (3%)	0	100	100
1	D	596/592 (101%)	583 (98%)	13 (2%)	0	100	100
All	All	1178/1184 (100%)	1149 (98%)	29 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	500/507 (99%)	497 (99%)	3 (1%)	86	75

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	512/507 (101%)	508 (99%)	4 (1%)	81	68
All	All	1012/1014 (100%)	1005 (99%)	7 (1%)	84	71

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	303	LEU
1	A	677	ARG
1	A	786	PHE
1	D	505	LYS
1	D	507	LEU
1	D	517	ARG
1	D	793	ASN

Some 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](#)

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	3,2	14,19,20	0.93	0	13,26,29	1.72	1 (7%)
2	DOC	B	29	3,2	14,19,20	0.87	0	13,26,29	1.31	1 (7%)

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	3,2	-	0/4/18/19	0/2/2/2
2	DOC	B	29	3,2	-	0/4/18/19	0/2/2/2

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	29	DOC	C2-N3-C4	5.56	121.97	116.34
2	B	29	DOC	C2-N3-C4	3.90	120.30	116.34

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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
4	GLC	G	1	4	11,11,12	0.58	0	15,15,17	0.97	2 (13%)
4	FRU	G	2	4	11,12,12	0.60	0	10,18,18	2.26	3 (30%)
4	GLC	H	1	4	11,11,12	0.46	0	15,15,17	1.13	2 (13%)
4	FRU	H	2	4	11,12,12	0.78	1 (9%)	10,18,18	1.05	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

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	H	2	FRU	O2-C2	2.06	1.44	1.40

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	G	2	FRU	O4-C4-C3	4.97	127.05	112.15
4	G	2	FRU	O4-C4-C5	3.98	122.55	111.05
4	H	1	GLC	O5-C1-C2	-2.64	106.69	110.77
4	G	1	GLC	O5-C1-C2	-2.31	107.20	110.77
4	H	1	GLC	C1-O5-C5	2.14	115.08	112.19
4	G	1	GLC	C1-O5-C5	2.02	114.93	112.19
4	G	2	FRU	O6-C6-C5	-2.01	104.39	111.29

There are no chirality outliers.

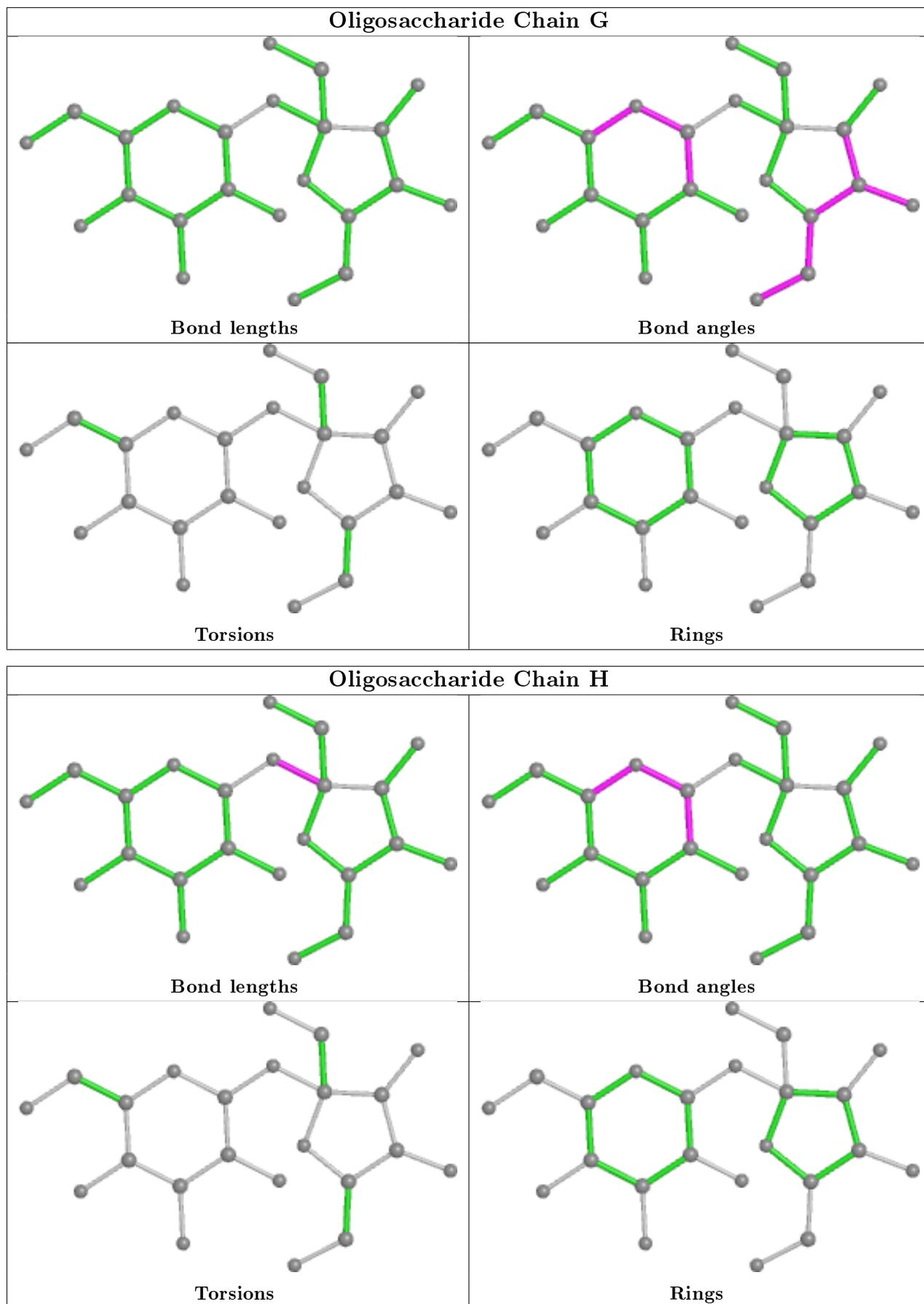
There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	H	2	FRU	1	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

Of 6 ligands modelled in this entry, 1 is monoatomic - leaving 5 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
5	CTP	D	901	6	23,30,30	1.62	4 (17%)	30,47,47	1.60	5 (16%)
7	SO4	D	903	-	4,4,4	0.17	0	6,6,6	0.59	0
5	CTP	A	901	-	23,30,30	1.75	6 (26%)	30,47,47	1.33	5 (16%)
7	SO4	D	906	-	4,4,4	0.15	0	6,6,6	0.10	0
7	SO4	D	905	-	4,4,4	0.17	0	6,6,6	0.19	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
5	CTP	D	901	6	-	6/20/38/38	0/2/2/2
5	CTP	A	901	-	-	3/20/38/38	0/2/2/2

All (10) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	D	901	CTP	C2'-C1'	-4.58	1.46	1.53
5	A	901	CTP	C6-N1	4.06	1.40	1.35
5	A	901	CTP	C2'-C1'	-3.50	1.48	1.53
5	D	901	CTP	C4-N4	2.89	1.43	1.35
5	A	901	CTP	C4-N4	2.60	1.42	1.35
5	D	901	CTP	C6-N1	2.52	1.38	1.35
5	D	901	CTP	O2'-C2'	-2.50	1.37	1.43
5	A	901	CTP	O2'-C2'	-2.46	1.37	1.43
5	A	901	CTP	C2-N3	2.41	1.42	1.38
5	A	901	CTP	O3'-C3'	-2.06	1.38	1.43

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	D	901	CTP	O3G-PG-O3B	4.09	118.37	104.64
5	A	901	CTP	O3G-PG-O3B	3.37	115.94	104.64
5	D	901	CTP	PB-O3A-PA	-3.29	121.55	132.83
5	D	901	CTP	C2-N3-C4	3.27	119.65	116.34
5	A	901	CTP	C2-N3-C4	3.06	119.44	116.34
5	A	901	CTP	O2G-PG-O3B	2.82	114.09	104.64
5	D	901	CTP	O2G-PG-O3B	2.51	113.06	104.64
5	A	901	CTP	PB-O3B-PG	-2.15	125.45	132.83
5	A	901	CTP	O2A-PA-O1A	-2.13	101.71	112.24
5	D	901	CTP	O5'-PA-O1A	2.13	117.39	109.07

There are no chirality outliers.

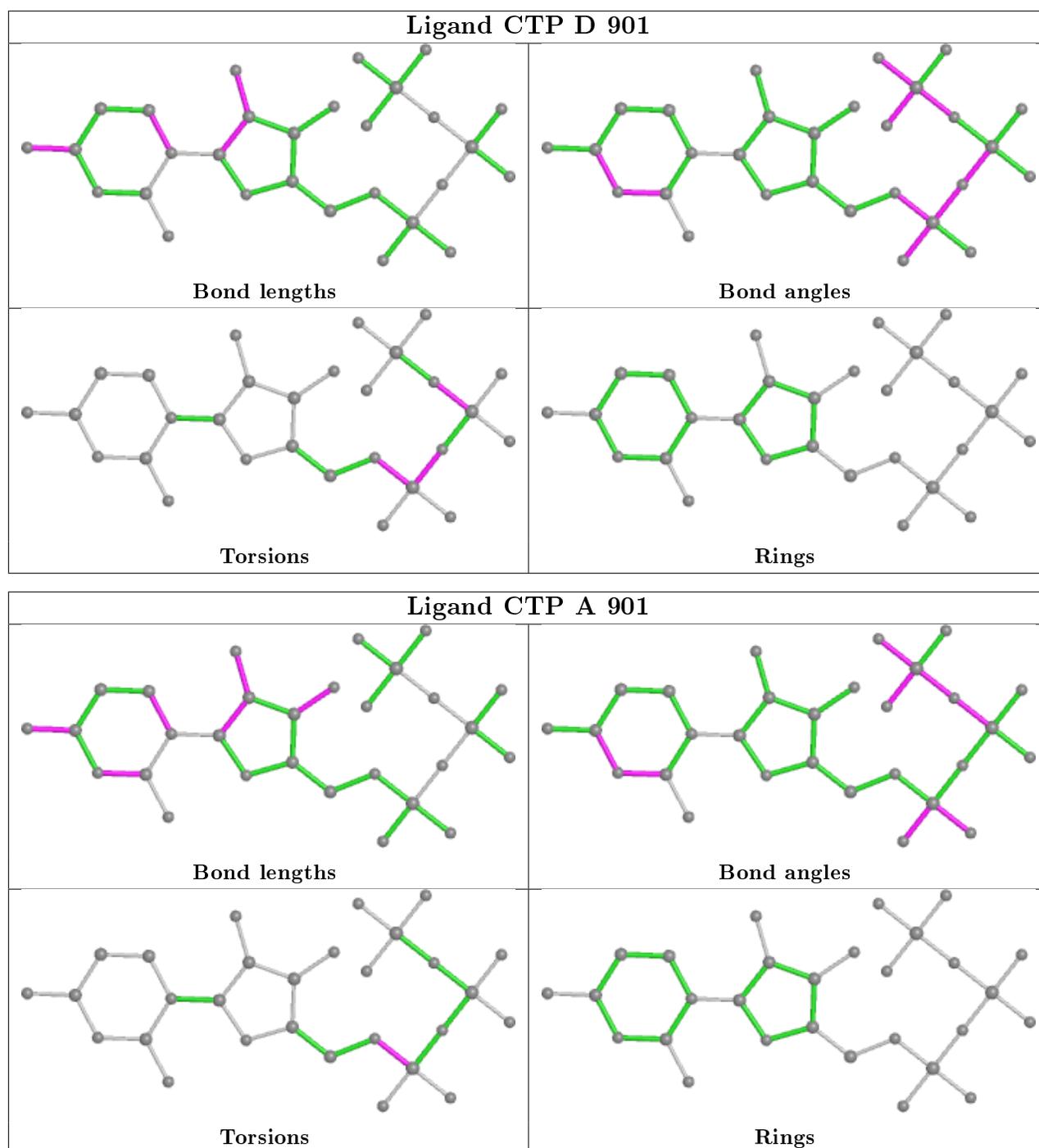
All (9) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	D	901	CTP	C5'-O5'-PA-O1A
5	A	901	CTP	C5'-O5'-PA-O1A
5	A	901	CTP	C5'-O5'-PA-O2A
5	D	901	CTP	C5'-O5'-PA-O3A
5	A	901	CTP	C5'-O5'-PA-O3A
5	D	901	CTP	PG-O3B-PB-O2B
5	D	901	CTP	C5'-O5'-PA-O2A
5	D	901	CTP	PB-O3A-PA-O1A
5	D	901	CTP	PB-O3A-PA-O2A

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 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	578/592 (97%)	-0.02	16 (2%) 53 51	15, 34, 54, 69	0
1	D	583/592 (98%)	-0.01	14 (2%) 59 58	9, 23, 43, 62	0
2	B	8/9 (88%)	-0.51	0 100 100	20, 28, 44, 50	0
2	E	8/9 (88%)	-0.41	0 100 100	17, 23, 40, 51	0
3	C	11/13 (84%)	-0.38	0 100 100	17, 23, 44, 62	0
3	F	12/13 (92%)	0.16	1 (8%) 11 9	14, 27, 53, 64	0
All	All	1200/1228 (97%)	-0.03	31 (2%) 56 55	9, 29, 51, 69	0

All (31) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	819	ARG	5.1
1	D	433	ALA	4.5
1	D	431	LYS	4.2
1	D	509	GLU	3.7
1	D	505	LYS	3.5
1	D	507	LEU	3.3
1	A	734	GLU	3.2
1	A	692	VAL	3.2
1	D	295	PRO	3.1
1	A	630	LEU	3.0
1	A	551	LYS	3.0
3	F	1	DA	3.0
1	A	550	THR	2.9
1	A	505	LYS	2.7
1	A	435	ARG	2.6
1	D	779	ARG	2.6
1	A	647	TRP	2.5
1	D	434	LYS	2.4
1	D	786	PHE	2.4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	D	714	TYR	2.3
1	A	703	ARG	2.3
1	A	846	ARG	2.2
1	D	504	GLY	2.2
1	D	695	ASP	2.2
1	D	629	ARG	2.2
1	A	689	ILE	2.2
1	A	676	ARG	2.1
1	A	554	TYR	2.1
1	A	816	LYS	2.1
1	D	294	LYS	2.1
1	A	509	GLU	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	19,23,31,32	0
2	DOC	E	29	18/19	0.98	0.10	14,19,23,27	0

## 6.3 Carbohydrates [i](#)

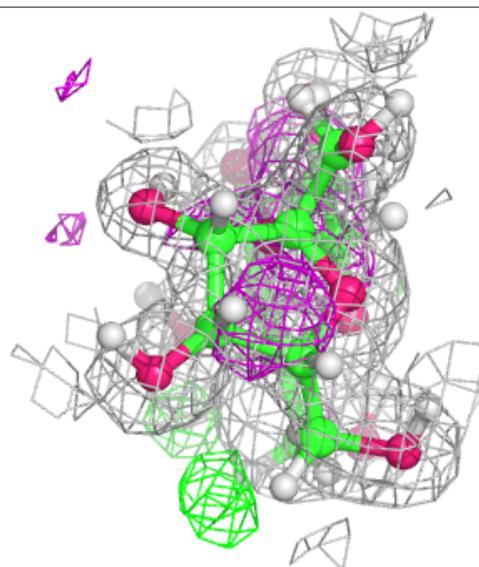
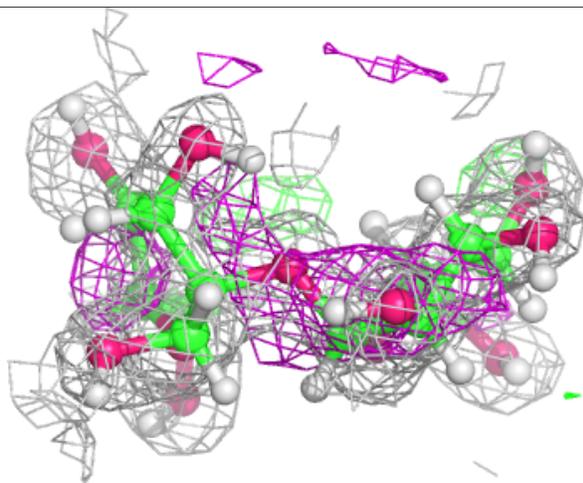
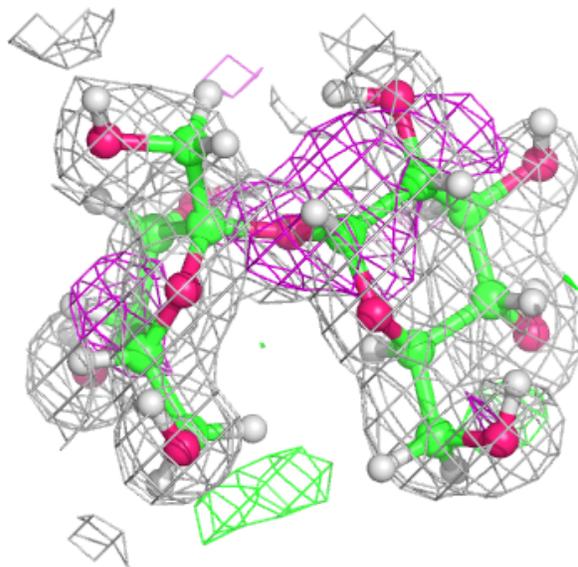
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q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

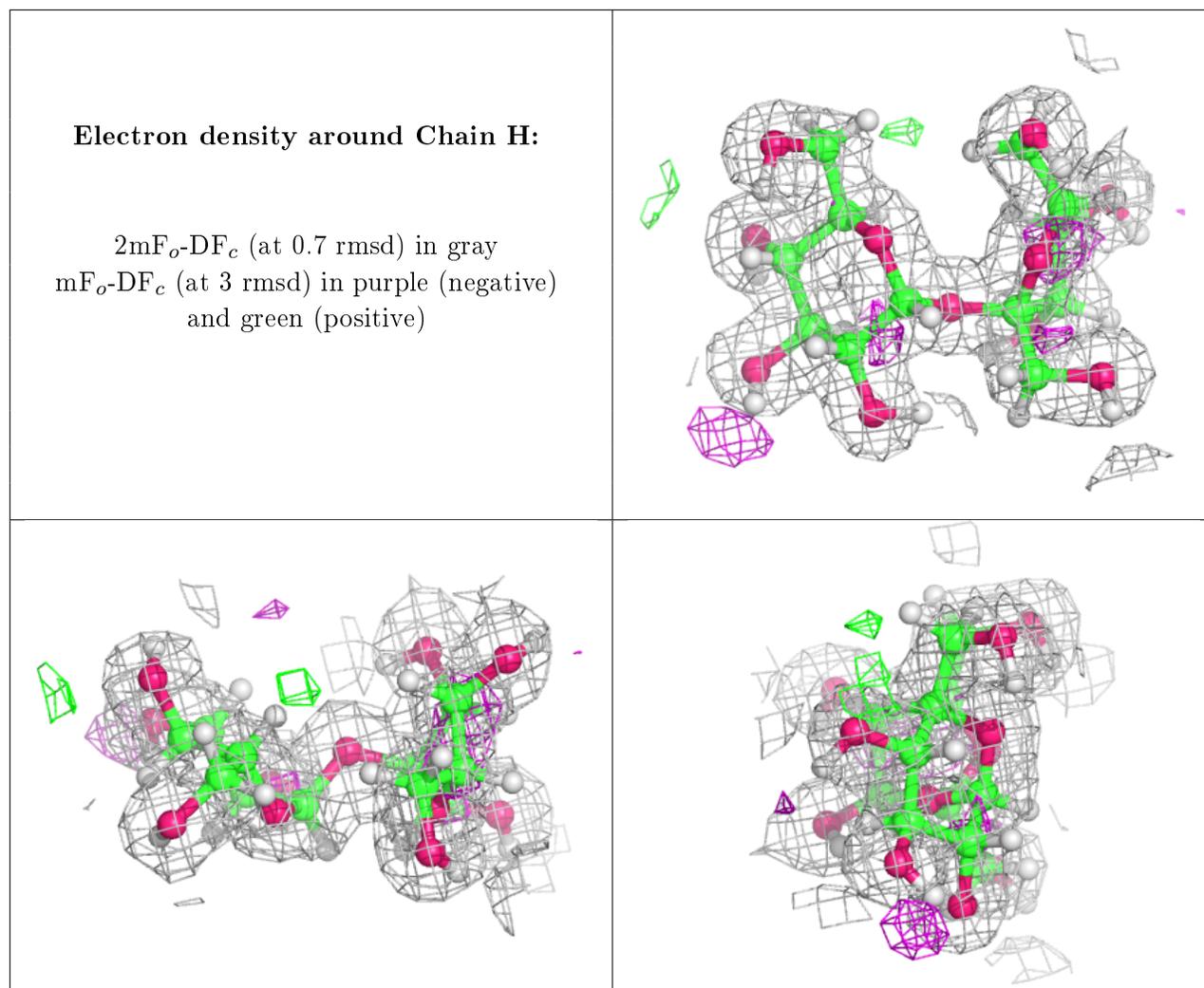
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
4	GLC	G	1	11/12	0.83	0.22	29,36,43,46	0
4	FRU	G	2	12/12	0.89	0.18	26,36,45,48	0
4	FRU	H	2	12/12	0.93	0.09	17,22,27,28	0
4	GLC	H	1	11/12	0.96	0.08	19,26,32,33	0

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)





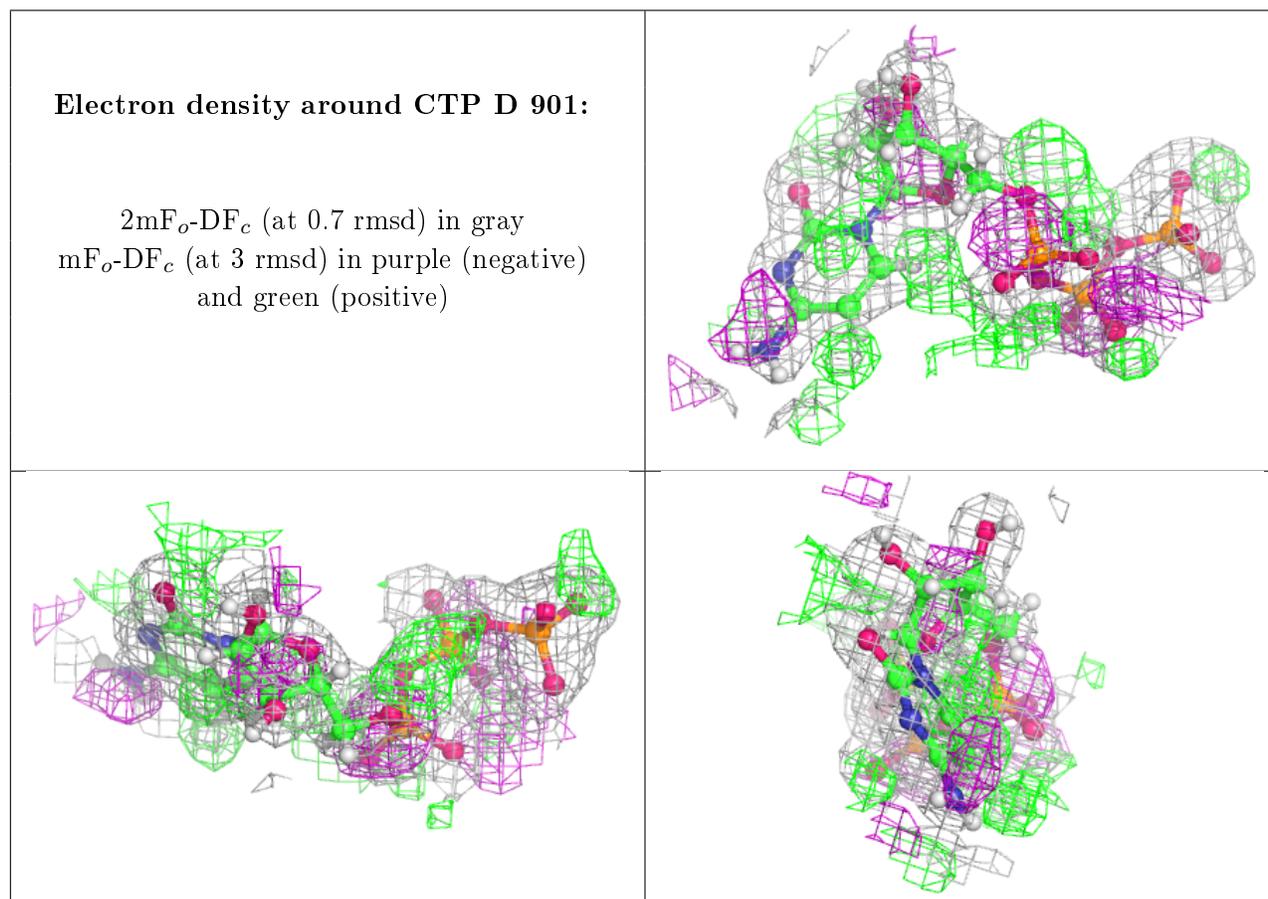
## 6.4 Ligands [\(i\)](#)

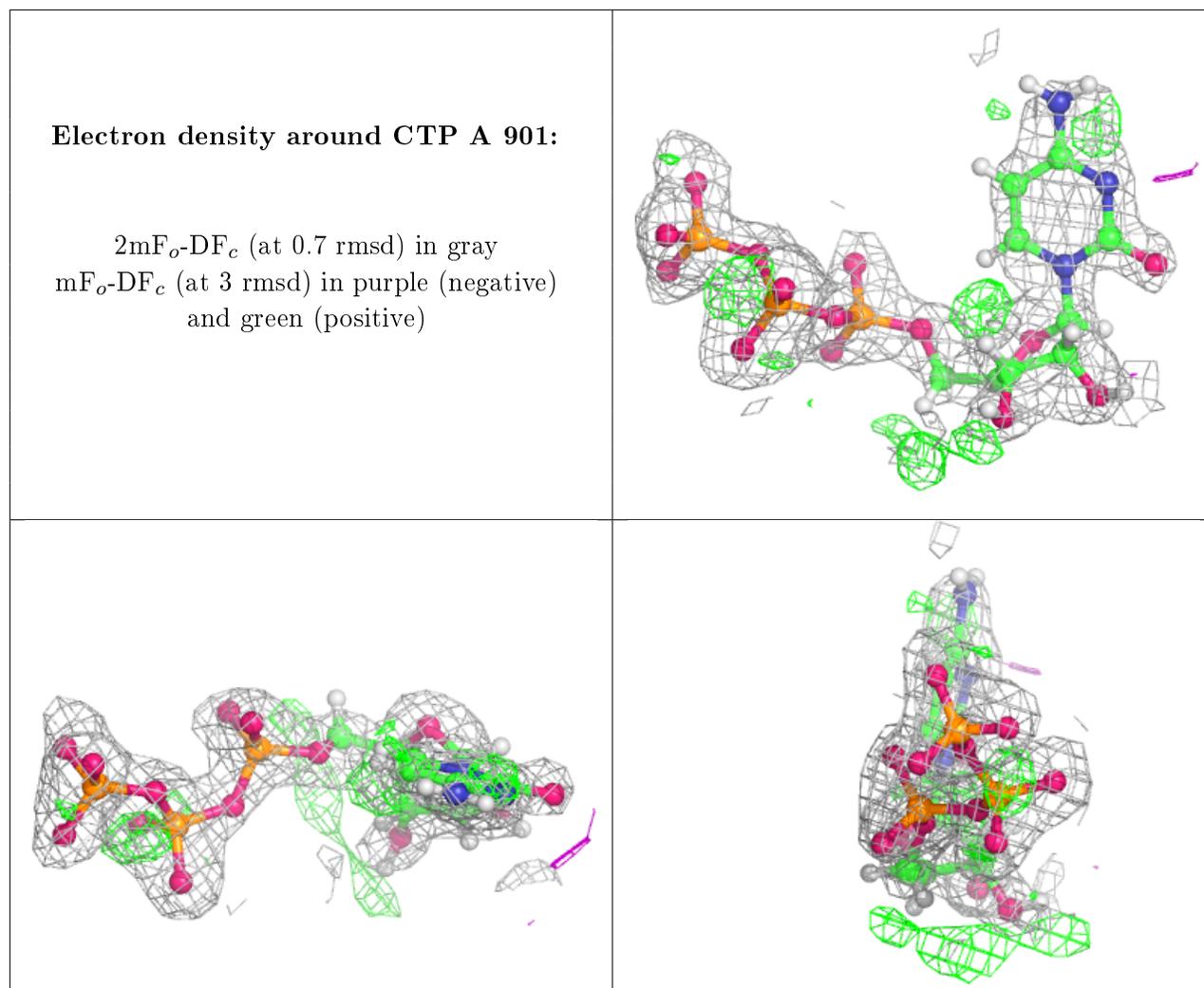
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 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
6	MG	D	902	1/1	0.69	0.23	38,38,38,38	0
5	CTP	D	901	29/29	0.86	0.13	22,33,45,49	0
7	SO4	D	906	5/5	0.89	0.13	30,37,44,47	0
5	CTP	A	901	29/29	0.90	0.13	21,31,42,45	41
7	SO4	D	905	5/5	0.92	0.12	71,71,75,76	0
7	SO4	D	903	5/5	0.97	0.09	30,38,45,54	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.