Full wwPDB X-ray Structure Validation Report (i)

Sep 7, 2023 – 04:25 AM EDT

PDB ID	:	4F2S
Title	:	DNA Polymerase I Large Fragment complex 4
Authors	:	Wang, W.; Beese, L.S.
Deposited on	:	2012-05-08
Resolution	:	1.65 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) 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 (i)) were used in the production of this report:

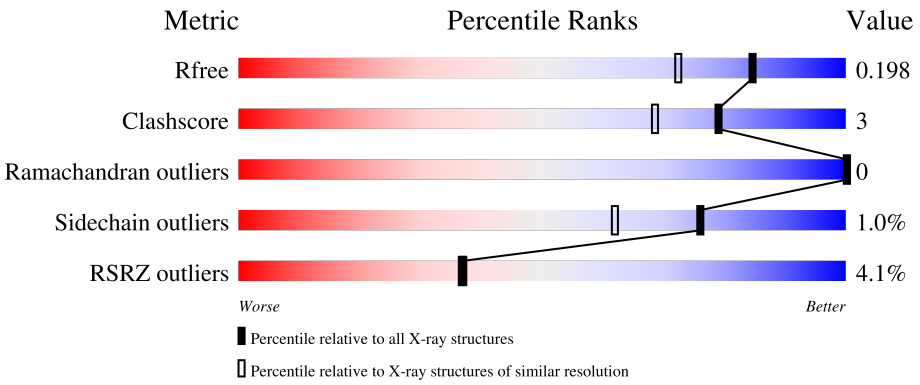
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.65 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	1827 (1.66-1.66)
Clashscore	141614	1931 (1.66-1.66)
Ramachandran outliers	138981	1891 (1.66-1.66)
Sidechain outliers	138945	1891 (1.66-1.66)
RSRZ outliers	127900	1791 (1.66-1.66)

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 >=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 <=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	5%; 92%; 5% •
1	D	592	3%; 92%; 6% •
2	B	9	56%; 33%; 11%
2	E	9	56%; 44%
3	C	13	31%; 46%; 8%; 15%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	13	8% 46% 46% 8%
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 21716 atoms, of which 10082 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	9402	2966	4743	805	869	19	0	5	0
1	D	583	9524	3004	4814	813	876	17	0	9	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

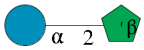
- Molecule 2 is a DNA chain called DNA (5'-D(*CP*CP*TP*GP*AP*CP*TP*CP*(DOC))-3').

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').

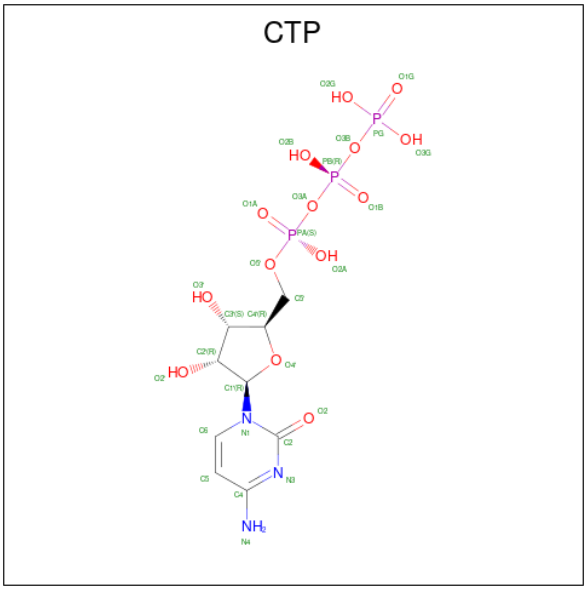
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
3	C	11	356	109	123	47	66	11	0	0	0
3	F	12	385	119	134	52	69	11	0	0	0

- 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
			Total	C	H	O			
4	H	2	45	12	22	11	0	0	0

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

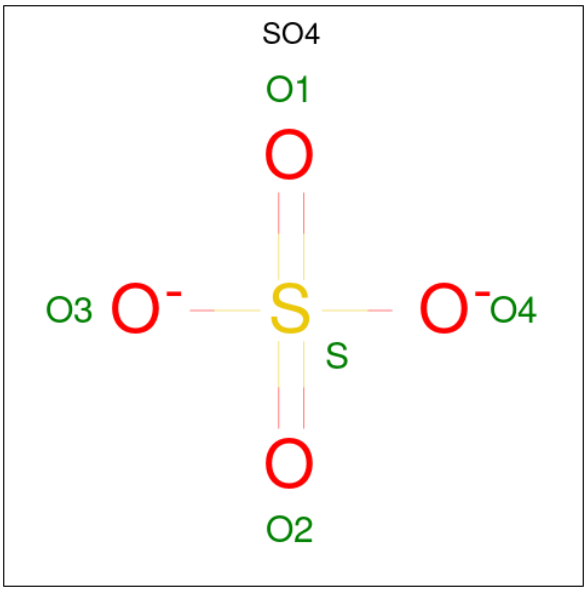


Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
			Total	C	H	N	O	P		
5	A	1	41	9	12	3	14	3	0	0
			Total	C	H	N	O	P		
5	D	1	41	9	12	3	14	3	0	0

- Molecule 6 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mn		
6	D	1	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

- Molecule 8 is water.

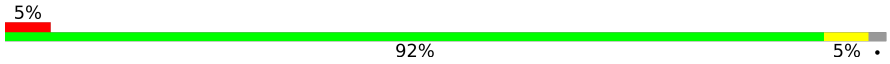
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	A	453	Total O 453 453	0	0
8	D	692	Total O 692 692	0	0
8	B	27	Total O 27 27	0	0
8	C	44	Total O 44 44	0	0
8	E	39	Total O 39 39	0	0
8	F	68	Total O 68 68	0	0

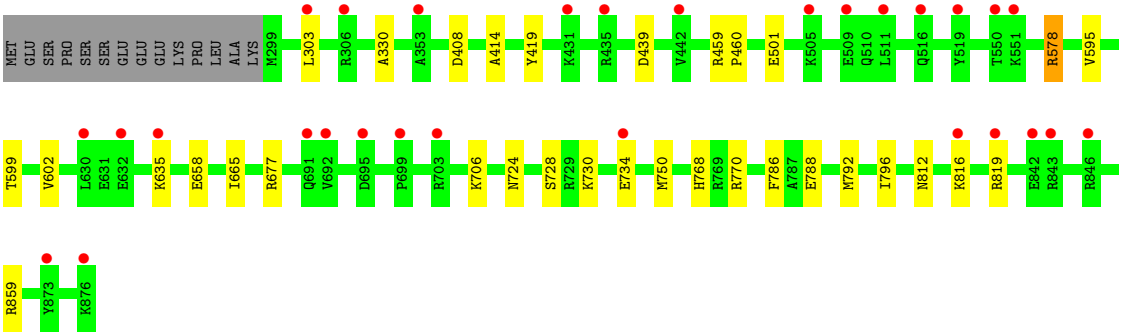
3 Residue-property plots (i)

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

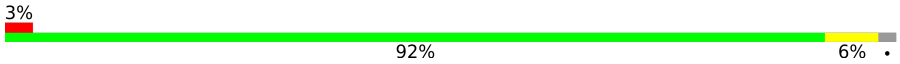
Chain A:

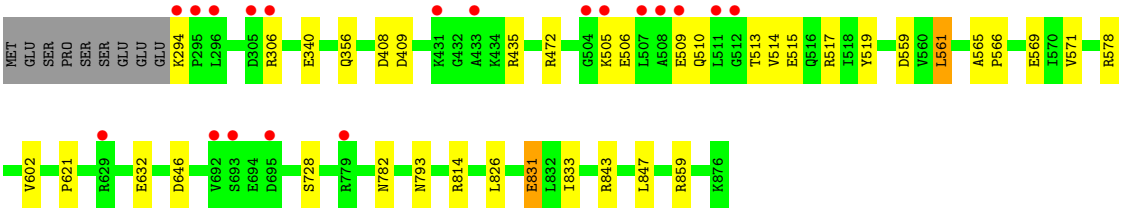




• Molecule 1: DNA polymerase

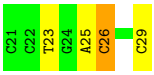
Chain D:





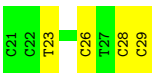
• Molecule 2: DNA (5'-D(*CP*CP*TP*GP*AP*CP*TP*CP*(DOC))-3')

Chain B:



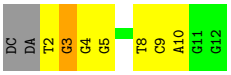
• Molecule 2: DNA (5'-D(*CP*CP*TP*GP*AP*CP*TP*CP*(DOC))-3')

Chain E:



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

Chain C:



• Molecule 3: DNA (5'-D(*CP*AP*TP*GP*GP*GP*AP*GP*TP*CP*AP*GP*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:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	93.18Å 108.59Å 150.99Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	34.19 – 1.65 70.71 – 1.65	Depositor EDS
% Data completeness (in resolution range)	95.8 (34.19-1.65) 95.8 (70.71-1.65)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.90 (at 1.65Å)	Xtriage
Refinement program	PHENIX dev_1026	Depositor
R, R_{free}	0.174 , 0.199 0.173 , 0.198	Depositor DCC
R_{free} test set	7732 reflections (4.40%)	wwPDB-VP
Wilson B-factor (Å^2)	19.6	Xtriage
Anisotropy	0.211	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.42 , 53.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	21716	wwPDB-VP
Average B, all atoms (Å^2)	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.24% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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: CTP, GLC, DOC, FRU, MN, SO4

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	0.46	0/4758	0.59	1/6429 (0.0%)
1	D	0.59	1/4827 (0.0%)	0.72	8/6525 (0.1%)
2	B	0.99	1/173 (0.6%)	1.76	6/264 (2.3%)
2	E	1.00	0/173	1.87	3/264 (1.1%)
3	C	0.98	0/262	1.62	5/404 (1.2%)
3	F	1.15	0/283	1.65	4/437 (0.9%)
All	All	0.59	2/10476 (0.0%)	0.82	27/14323 (0.2%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	831	GLU	CB-CG	-5.84	1.41	1.52
2	B	25	DA	C3'-O3'	-5.30	1.37	1.44

All (27) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	23	DT	O4'-C1'-N1	-12.13	99.51	108.00
3	C	3	DG	O5'-P-OP1	-9.04	97.56	105.70
1	D	578	ARG	NE-CZ-NH2	-8.19	116.20	120.30
1	D	843	ARG	NE-CZ-NH2	-8.03	116.29	120.30
3	F	8	DT	O4'-C1'-N1	-8.01	102.40	108.00
3	F	5	DG	O4'-C4'-C3'	-7.84	101.30	106.00
1	D	843	ARG	NE-CZ-NH1	7.07	123.84	120.30
1	D	859	ARG	NE-CZ-NH1	7.06	123.83	120.30
1	A	578	ARG	NE-CZ-NH2	-7.02	116.79	120.30
2	B	26	DC	C6-N1-C2	6.91	123.06	120.30
2	B	23	DT	N3-C4-O4	6.58	123.85	119.90
3	C	8	DT	O4'-C1'-N1	-6.56	103.41	108.00
2	B	23	DT	O4'-C1'-N1	-6.44	103.49	108.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	26	DC	O4'-C1'-N1	6.42	112.50	108.00
1	D	859	ARG	NE-CZ-NH2	-6.09	117.25	120.30
2	E	28	DC	C2-N3-C4	5.93	122.87	119.90
3	C	2	DT	O4'-C1'-N1	5.90	112.13	108.00
3	C	4	DG	N3-C4-C5	5.81	131.50	128.60
1	D	578	ARG	NE-CZ-NH1	5.63	123.12	120.30
3	C	5	DG	N3-C4-N9	-5.45	122.73	126.00
1	D	409	ASP	CB-CG-OD2	5.41	123.17	118.30
2	B	26	DC	O4'-C1'-C2'	5.27	110.11	105.90
2	B	23	DT	C5-C4-O4	-5.23	121.24	124.90
2	B	26	DC	N3-C4-C5	5.16	123.96	121.90
1	D	578	ARG	CB-CG-CD	-5.14	98.23	111.60
3	F	6	DA	O4'-C1'-C2'	5.14	110.01	105.90
3	F	6	DA	C5-N7-C8	-5.12	101.34	103.90

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	4659	4743	4731	25	0
1	D	4710	4814	4795	26	0
2	B	174	100	103	1	0
2	E	174	100	103	1	0
3	C	233	123	124	3	0
3	F	251	134	136	2	0
4	G	23	22	21	2	0
4	H	23	22	21	1	0
5	A	29	12	11	2	0
5	D	29	12	12	0	0
6	D	1	0	0	0	0
7	D	5	0	0	0	0
8	A	453	0	0	11	0
8	B	27	0	0	0	0
8	C	44	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
8	D	692	0	0	15	0
8	E	39	0	0	0	0
8	F	68	0	0	1	0
All	All	11634	10082	10057	56	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 (56) 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:NH2	8:D:1607:HOH:O	1.89	1.02
1:A:599:THR:O	8:A:1446:HOH:O	1.92	0.86
1:A:812:ASN:HB3	8:A:1242:HOH:O	1.79	0.80
1:D:340:GLU:OE2	8:D:1453:HOH:O	2.03	0.75
1:D:632:GLU:OE1	8:D:1652:HOH:O	2.05	0.75
1:A:724:ASN:ND2	3:C:3:DG:O6	2.24	0.71
1:A:501:GLU:OE2	8:A:1330:HOH:O	2.07	0.71
1:A:658:GLU:OE2	8:A:1105:HOH:O	2.10	0.69
1:D:472:ARG:NE	8:D:1410:HOH:O	2.26	0.68
1:D:559:ASP:OD1	8:D:1352:HOH:O	2.12	0.67
1:D:569:GLU:OE1	8:D:1617:HOH:O	2.14	0.66
1:A:724:ASN:ND2	3:C:3:DG:C6	2.67	0.63
1:A:439:ASP:HB3	8:A:1166:HOH:O	2.00	0.61
1:D:602[B]:VAL:HG21	1:D:621:PRO:HG3	1.83	0.61
1:D:569:GLU:CD	8:D:1617:HOH:O	2.39	0.60
1:D:435:ARG:NH2	8:D:1298:HOH:O	2.37	0.57
1:A:330:ALA:O	8:A:1036:HOH:O	2.18	0.56
3:F:9:DC:H2"	3:F:10:DA:C8	2.43	0.53
1:A:788:GLU:O	1:A:792:MET:HG3	2.09	0.52
1:D:506:GLU:O	1:D:509:GLU:HB3	2.10	0.52
1:D:509:GLU:O	1:D:513:THR:HG23	2.10	0.52
1:D:472:ARG:CZ	8:D:1410:HOH:O	2.56	0.51
1:D:408:ASP:HB2	4:H:2:FRU:H11	1.92	0.50
1:A:408:ASP:HB2	4:G:2:FRU:H11	1.94	0.49
1:D:510:GLN:OE1	8:D:1499:HOH:O	2.20	0.48
1:A:770:ARG:HD2	8:A:1366:HOH:O	2.14	0.48
1:A:750[B]:MET:SD	1:A:792:MET:HB3	2.54	0.47
1:D:294:LYS:NZ	8:D:1354:HOH:O	2.46	0.47
1:D:306:ARG:NE	8:D:1485:HOH:O	2.45	0.47
1:D:831:GLU:CD	8:D:1660:HOH:O	2.53	0.47

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:595:VAL:HG22	1:A:602:VAL:HG13	1.97	0.46
1:A:706:LYS:NZ	5:A:901:CTP:O1A	2.47	0.46
1:D:814:ARG:NH2	1:D:847[A]:LEU:HD13	2.32	0.45
1:A:414:ALA:HB1	1:A:419:TYR:HB3	1.99	0.45
1:D:646:ASP:OD1	1:D:646:ASP:N	2.34	0.45
1:D:728:SER:HB3	8:D:1220:HOH:O	2.15	0.45
1:A:635:LYS:HE2	8:A:1237:HOH:O	2.17	0.44
1:A:816:LYS:O	1:A:819:ARG:HD2	2.18	0.44
1:A:665:ILE:HD12	1:A:796:ILE:HD13	1.99	0.44
1:D:510:GLN:O	1:D:514:VAL:HG23	2.18	0.44
1:D:565:ALA:N	1:D:566:PRO:CD	2.82	0.43
3:F:1:DA:C2	8:F:166:HOH:O	2.71	0.43
3:C:9:DC:H2"	3:C:10:DA:C8	2.53	0.43
1:A:578:ARG:HD3	2:B:26:DC:OP1	2.19	0.42
1:D:561:LEU:O	1:D:571:VAL:HG11	2.18	0.42
1:A:595:VAL:HG22	1:A:602:VAL:CG1	2.49	0.42
1:A:859:ARG:NH2	8:A:1178:HOH:O	2.52	0.42
8:A:1009:HOH:O	4:G:2:FRU:H62	2.20	0.42
1:D:826:LEU:HB2	1:D:833:ILE:HD13	2.02	0.42
1:A:459:ARG:HB2	1:A:460:PRO:HD3	2.02	0.41
8:D:1660:HOH:O	2:E:29:DOC:H5'	2.20	0.41
1:A:730:LYS:O	1:A:734:GLU:HG2	2.20	0.41
1:A:768:HIS:HD2	8:A:1319:HOH:O	2.04	0.41
1:A:706:LYS:CE	5:A:901:CTP:O1A	2.68	0.41
1:D:515:GLU:HG2	1:D:519:TYR:CZ	2.56	0.41
1:D:814:ARG:CZ	1:D:847[A]:LEU:HD13	2.50	0.41

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	581/592 (98%)	568 (98%)	13 (2%)	0	100	100
1	D	590/592 (100%)	577 (98%)	13 (2%)	0	100	100
All	All	1171/1184 (99%)	1145 (98%)	26 (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	499/507 (98%)	495 (99%)	4 (1%)	81	70
1	D	506/507 (100%)	500 (99%)	6 (1%)	71	53
All	All	1005/1014 (99%)	995 (99%)	10 (1%)	76	62

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	303	LEU
1	A	677	ARG
1	A	728	SER
1	A	786	PHE
1	D	356	GLN
1	D	505	LYS
1	D	517	ARG
1	D	561	LEU
1	D	782	ASN
1	D	793	ASN

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

Mol	Chain	Res	Type
1	D	768	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	B	29	3,2	16,19,20	0.68	0	20,26,29	0.99	1 (5%)
2	DOC	E	29	3,2	16,19,20	0.62	0	20,26,29	0.80	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	B	29	3,2	-	0/7/18/19	0/2/2/2
2	DOC	E	29	3,2	-	0/7/18/19	0/2/2/2

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	29	DOC	O4'-C1'-N1	2.02	111.47	107.86

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)

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	2.07	3 (20%)
4	FRU	G	2	4	11,12,12	0.70	1 (9%)	10,18,18	1.11	1 (10%)
4	GLC	H	1	4	11,11,12	0.57	0	15,15,17	0.95	1 (6%)
4	FRU	H	2	4	11,12,12	0.84	1 (9%)	10,18,18	1.22	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 (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	G	2	FRU	O2-C2	2.14	1.44	1.40
4	H	2	FRU	O2-C2	2.14	1.44	1.40

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	G	1	GLC	O3-C3-C2	4.64	118.88	109.99

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	G	1	GLC	O3-C3-C4	4.04	119.70	110.35
4	G	1	GLC	C2-C3-C4	3.67	117.24	110.89
4	H	1	GLC	O5-C1-C2	-2.08	107.56	110.77
4	G	2	FRU	O6-C6-C5	-2.07	104.20	111.29

There are no chirality outliers.

There are no torsion outliers.

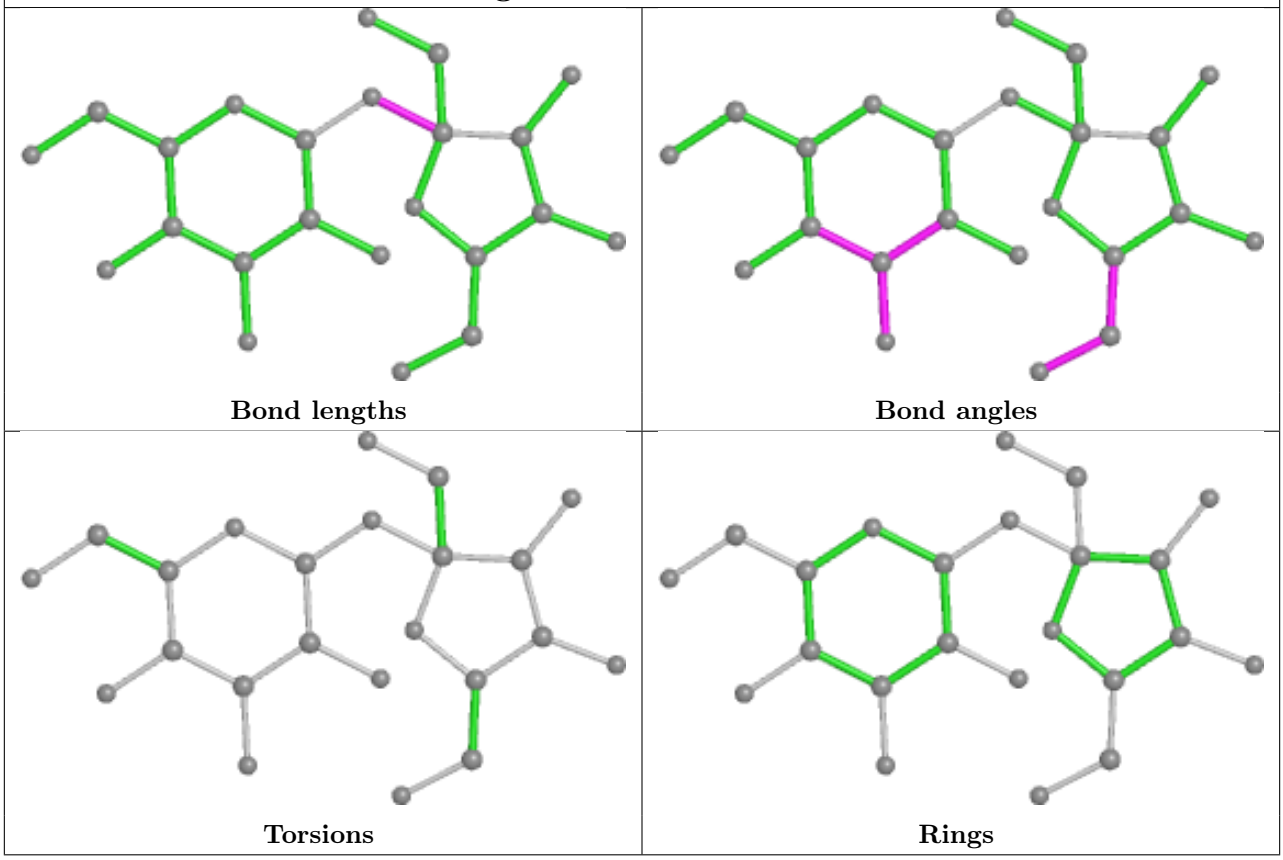
There are no ring outliers.

2 monomers are involved in 3 short contacts:

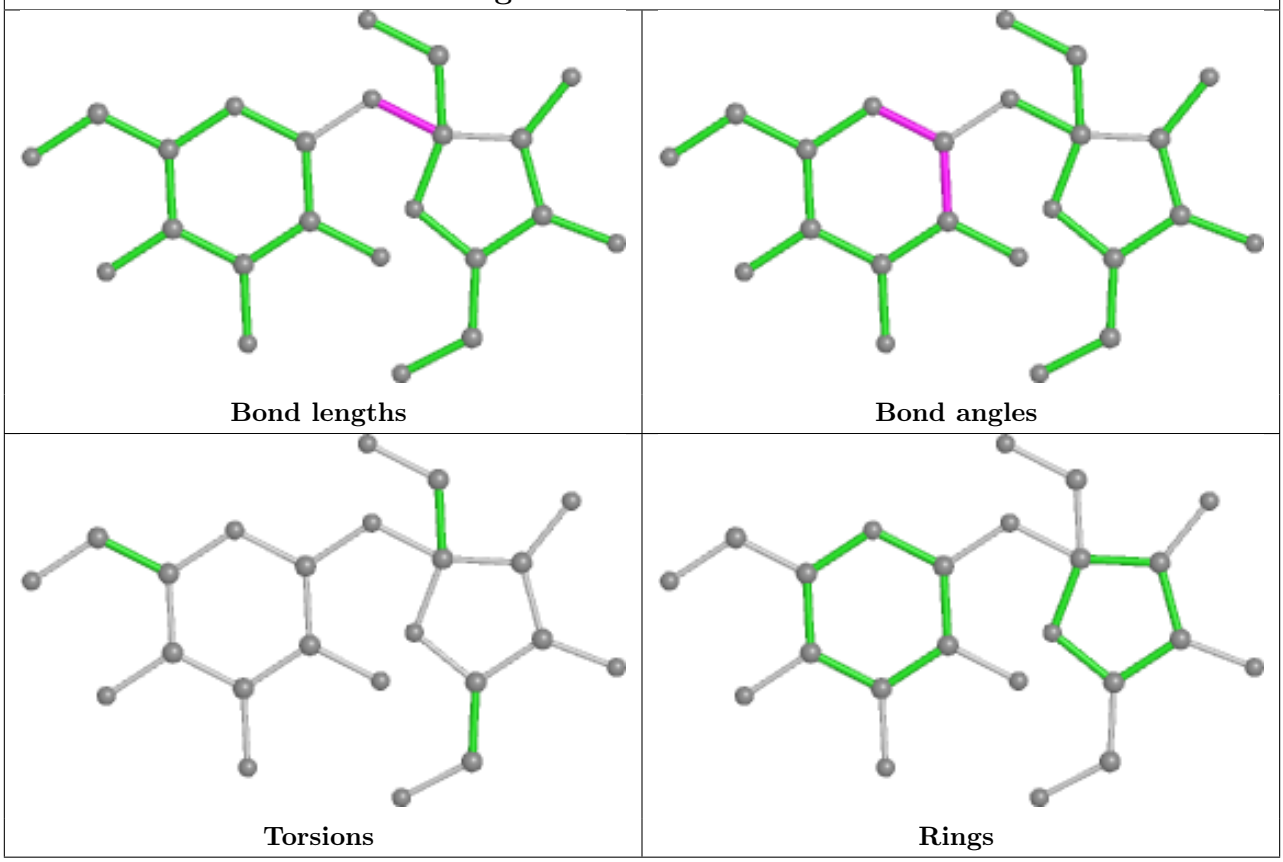
Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	H	2	FRU	1	0
4	G	2	FRU	2	0

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

Oligosaccharide Chain G



Oligosaccharide Chain H



5.6 Ligand geometry (i)

Of 4 ligands modelled in this entry, 1 is monoatomic - leaving 3 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	A	901	-	26,30,30	2.12	5 (19%)	39,47,47	1.82	8 (20%)
5	CTP	D	901	6	26,30,30	1.72	3 (11%)	39,47,47	1.94	13 (33%)
7	SO4	D	903	-	4,4,4	0.19	0	6,6,6	0.74	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	A	901	-	-	3/22/38/38	0/2/2/2
5	CTP	D	901	6	-	8/22/38/38	0/2/2/2

All (8) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	901	CTP	O2-C2	7.51	1.37	1.23
5	D	901	CTP	O2-C2	6.73	1.36	1.23
5	A	901	CTP	C4-N4	3.73	1.42	1.33
5	A	901	CTP	C2-N3	3.61	1.43	1.36
5	D	901	CTP	C4-N4	2.75	1.40	1.33
5	A	901	CTP	O2'-C2'	-2.41	1.37	1.43
5	A	901	CTP	O3'-C3'	-2.26	1.37	1.43
5	D	901	CTP	O2'-C2'	-2.18	1.37	1.43

All (21) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	A	901	CTP	O4'-C1'-N1	6.98	124.31	108.36
5	D	901	CTP	O4'-C1'-N1	5.47	120.87	108.36
5	D	901	CTP	O3G-PG-O3B	4.16	118.58	104.64
5	A	901	CTP	O3G-PG-O3B	3.49	116.33	104.64
5	D	901	CTP	C3'-C2'-C1'	-3.48	94.82	101.43
5	D	901	CTP	C4'-O4'-C1'	-3.45	101.85	109.47
5	A	901	CTP	C2'-C1'-N1	3.38	122.80	113.22
5	D	901	CTP	C6-N1-C2	2.97	125.66	120.49
5	D	901	CTP	C1'-N1-C6	-2.96	114.38	120.84
5	D	901	CTP	C5-C6-N1	-2.65	117.38	121.81
5	A	901	CTP	O2G-PG-O3B	2.63	113.45	104.64
5	D	901	CTP	C2'-C1'-N1	2.54	120.42	113.22
5	A	901	CTP	C4'-O4'-C1'	-2.50	103.96	109.47
5	A	901	CTP	C3'-C2'-C1'	-2.43	96.80	101.43
5	D	901	CTP	O4'-C1'-C2'	2.23	111.50	106.64
5	D	901	CTP	PB-O3B-PG	-2.20	125.28	132.83
5	D	901	CTP	O2A-PA-O1A	-2.19	101.42	112.24
5	A	901	CTP	O3G-PG-O1G	-2.18	102.16	110.68
5	D	901	CTP	O5'-PA-O1A	2.16	117.49	109.07
5	D	901	CTP	O2G-PG-O3B	2.15	111.86	104.64
5	A	901	CTP	O2A-PA-O1A	-2.15	101.61	112.24

There are no chirality outliers.

All (11) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	901	CTP	C5'-O5'-PA-O1A
5	A	901	CTP	C5'-O5'-PA-O2A
5	D	901	CTP	C5'-O5'-PA-O1A
5	D	901	CTP	C5'-O5'-PA-O2A
5	A	901	CTP	C5'-O5'-PA-O3A
5	D	901	CTP	C5'-O5'-PA-O3A
5	D	901	CTP	PG-O3B-PB-O2B
5	D	901	CTP	PB-O3A-PA-O1A
5	D	901	CTP	PB-O3A-PA-O2A
5	D	901	CTP	PG-O3B-PB-O1B
5	D	901	CTP	C2'-C1'-N1-C2

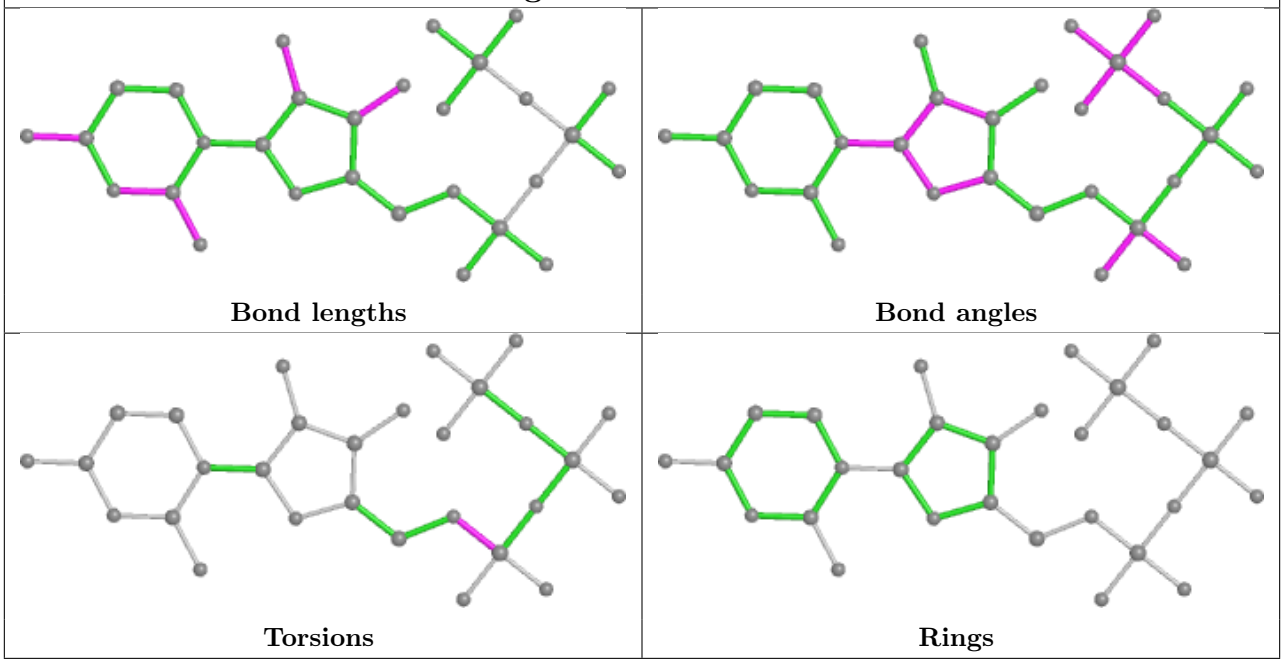
There are no ring outliers.

1 monomer is involved in 2 short contacts:

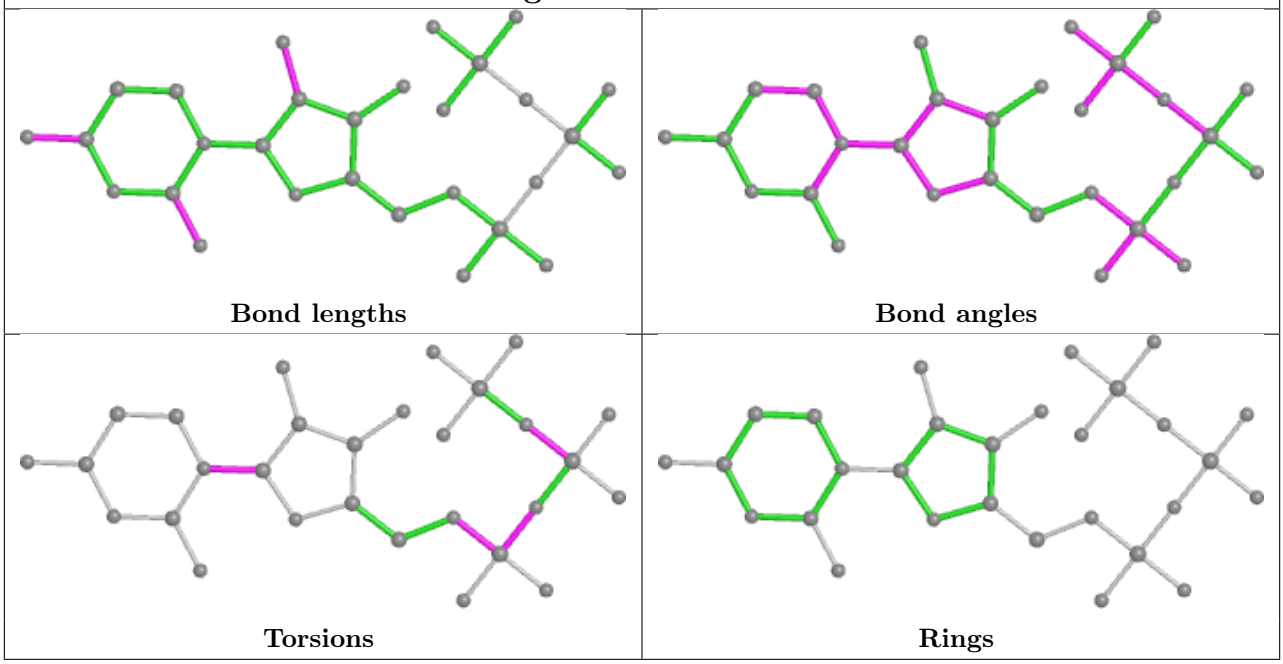
Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	901	CTP	2	0

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

Ligand CTP A 901



Ligand CTP D 901



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, 95th 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^2)	Q<0.9
1	A	578/592 (97%)	0.18	29 (5%)	28	27	15, 35, 55, 71	0
1	D	583/592 (98%)	0.08	19 (3%)	46	47	9, 23, 42, 61	0
2	B	8/9 (88%)	-0.50	0	100	100	21, 27, 42, 49	0
2	E	8/9 (88%)	-0.39	0	100	100	16, 22, 39, 49	0
3	C	11/13 (84%)	-0.37	0	100	100	18, 25, 43, 59	0
3	F	12/13 (92%)	0.11	1 (8%)	11	11	13, 25, 49, 60	0
All	All	1200/1228 (97%)	0.12	49 (4%)	37	37	9, 29, 51, 71	0

All (49) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	509	GLU	6.9
1	D	433	ALA	5.7
1	A	505	LYS	4.7
1	A	630	LEU	4.4
1	A	703	ARG	4.2
1	A	819	ARG	4.2
1	A	303	LEU	3.9
1	D	505	LYS	3.8
1	A	551	LYS	3.5
1	A	519	TYR	3.3
3	F	1	DA	3.3
1	A	353	ALA	3.2
1	D	508	ALA	3.2
1	D	692	VAL	3.2
1	A	876	LYS	3.1
1	A	692	VAL	3.1
1	A	442	VAL	3.1
1	D	431	LYS	3.1
1	D	295	PRO	3.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	D	296	LEU	3.0
1	D	294	LYS	2.9
1	D	306	ARG	2.9
1	A	699	PRO	2.9
1	D	305	ASP	2.8
1	D	511	LEU	2.8
1	A	509	GLU	2.8
1	A	695	ASP	2.7
1	D	779	ARG	2.7
1	D	507	LEU	2.6
1	A	511	LEU	2.6
1	A	691	GLN	2.6
1	D	512	GLY	2.6
1	A	846	ARG	2.5
1	D	695	ASP	2.5
1	A	516	GLN	2.5
1	A	734	GLU	2.4
1	D	504	GLY	2.4
1	D	693[A]	SER	2.4
1	A	632	GLU	2.3
1	A	816	LYS	2.3
1	A	550	THR	2.3
1	A	306	ARG	2.2
1	A	843	ARG	2.2
1	D	629	ARG	2.2
1	A	435	ARG	2.1
1	A	635	LYS	2.1
1	A	842	GLU	2.1
1	A	873	TYR	2.0
1	A	431	LYS	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, 95th 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^2)	Q<0.9
2	DOC	B	29	18/19	0.97	0.08	19,26,34,36	0
2	DOC	E	29	18/19	0.99	0.10	14,18,23,24	0

6.3 Carbohydrates (i)

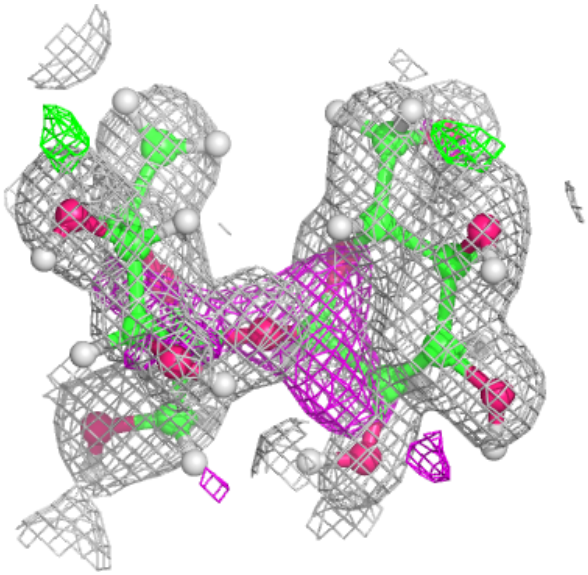
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists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

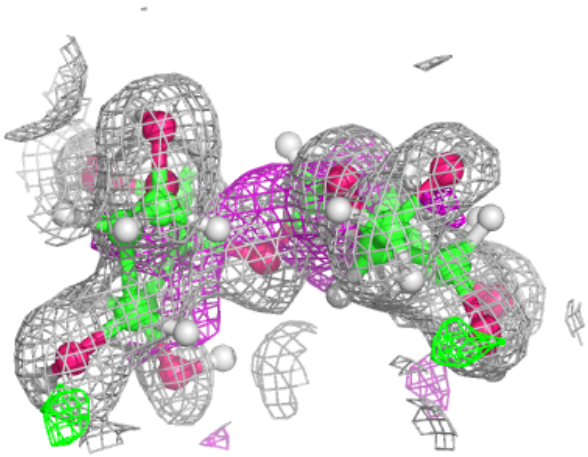
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	GLC	G	1	11/12	0.76	0.19	27,36,46,48	0
4	FRU	G	2	12/12	0.83	0.16	25,37,46,46	0
4	FRU	H	2	12/12	0.90	0.11	16,24,30,33	0
4	GLC	H	1	11/12	0.95	0.07	20,25,31,32	0

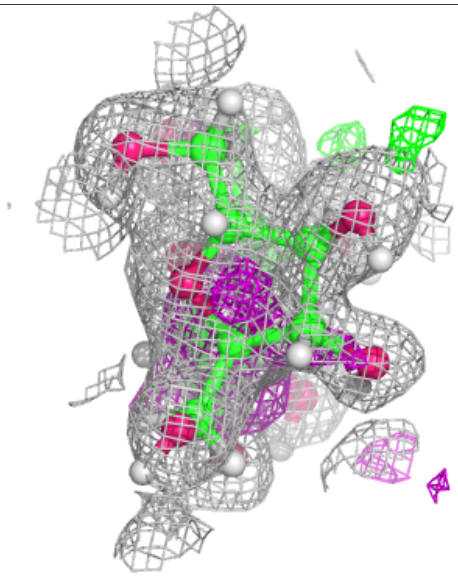
The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around Chain G:

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

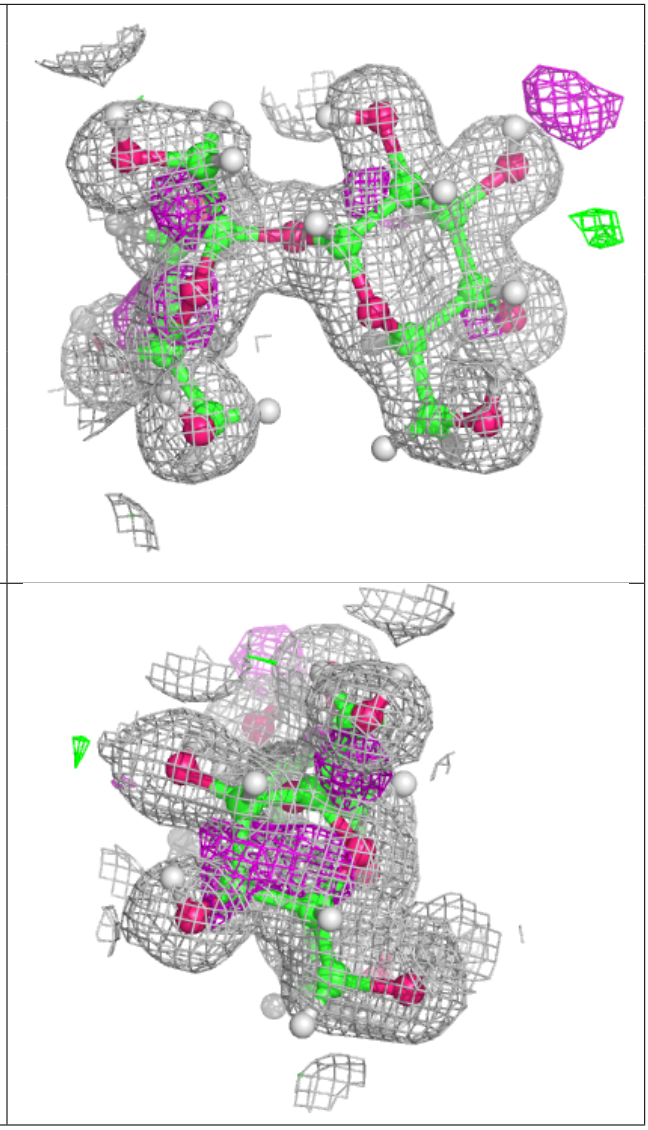






Electron density around Chain H:

$2mF_o$-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, 95th 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.

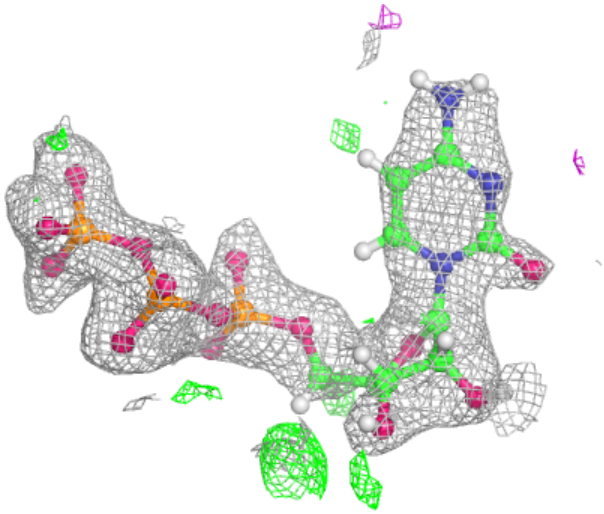
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
5	CTP	A	901	29/29	0.89	0.14	27,41,51,57	41
5	CTP	D	901	29/29	0.91	0.10	19,27,37,38	0
7	SO4	D	903	5/5	0.97	0.09	29,36,43,49	0
6	MN	D	902	1/1	0.99	0.06	29,29,29,29	0

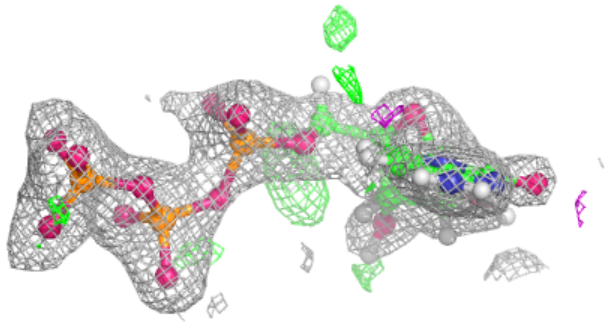
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers

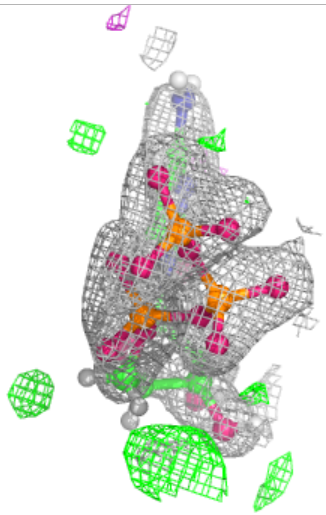
as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around CTP A 901:

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

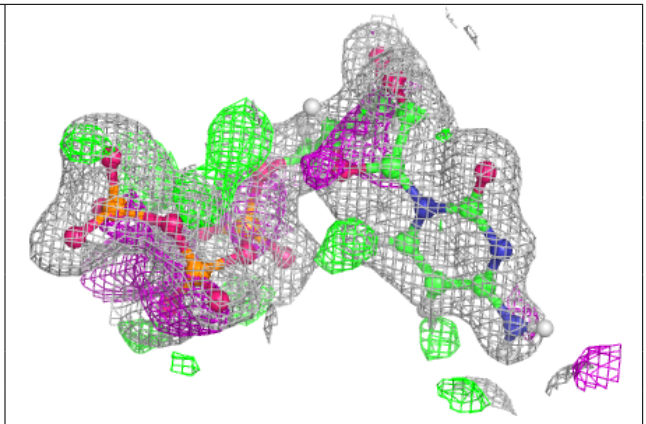


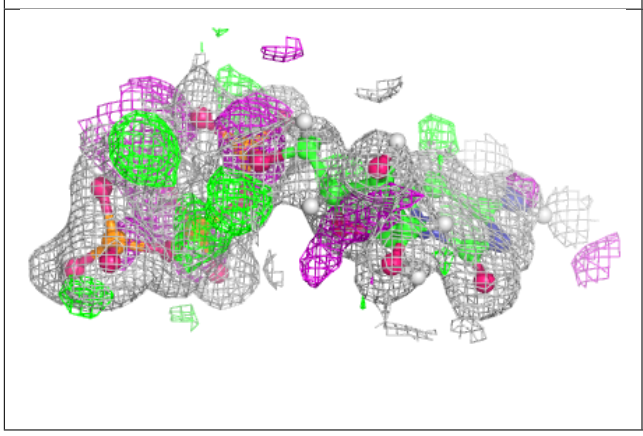


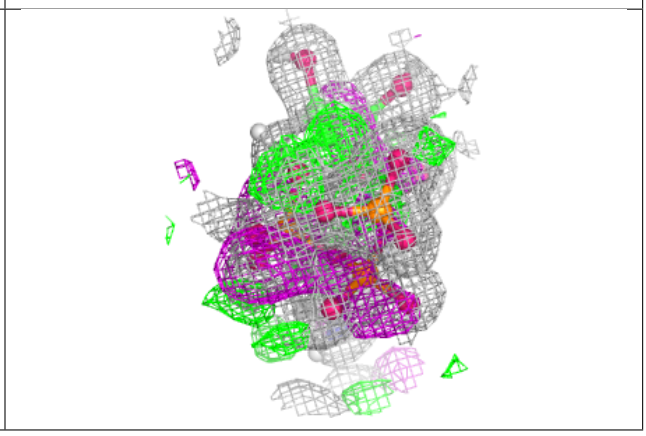


Electron density around CTP D 901:

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







6.5 Other polymers (i)

There are no such residues in this entry.