Full wwPDB X-ray Structure Validation Report ⓘ

Aug 27, 2023 – 07:12 PM EDT

PDB ID	:	3HP6
Title	:	Crystal structure of fragment DNA polymerase I from Bacillus stearothermophilus F710Y mutant bound to G:T mismatch
Authors	:	Wu, E.Y.; Beese, L.S.
Deposited on	:	2009-06-03
Resolution	:	1.81 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 ⓘ symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references ⓘ) were used in the production of this report:

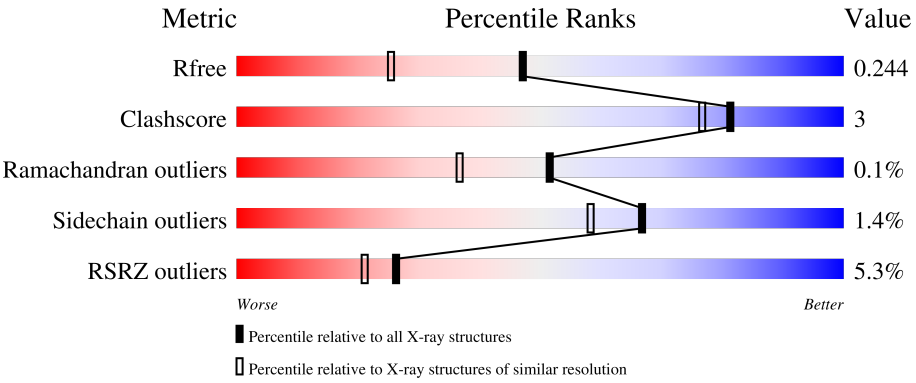
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber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

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 1.81 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.



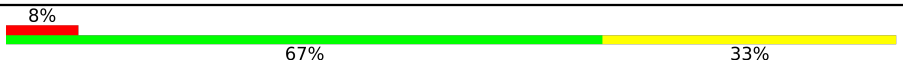
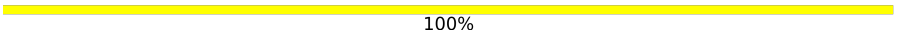
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	7484 (1.84-1.80)
Clashscore	141614	8401 (1.84-1.80)
Ramachandran outliers	138981	8290 (1.84-1.80)
Sidechain outliers	138945	8290 (1.84-1.80)
RSRZ outliers	127900	7371 (1.84-1.80)

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	580	8% 91% 9%
1	D	580	3% 95% 5%
2	B	9	33% 67%
2	E	9	33% 67%
3	C	12	8% 58% 42%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	12	
4	G	2	
4	H	2	

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 11210 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called DNA polymerase I, large fragment.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	580	Total	C	N	O	S	0	3	0
			4673	2972	812	870	19			
1	D	580	Total	C	N	O	S	0	8	0
			4701	2995	813	875	18			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	598	ALA	ASP	engineered mutation	PDB 3HP6
A	710	TYR	PHE	engineered mutation	PDB 3HP6
D	598	ALA	ASP	engineered mutation	PDB 3HP6
D	710	TYR	PHE	engineered mutation	PDB 3HP6

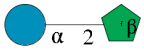
- Molecule 2 is a DNA chain called 5'-D(*CP*GP*AP*TP*CP*AP*CP*GP*(DDG))-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	9	Total	C	N	O	P	0	0	0
			181	87	36	50	8			
2	E	9	Total	C	N	O	P	0	0	0
			181	87	36	50	8			

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

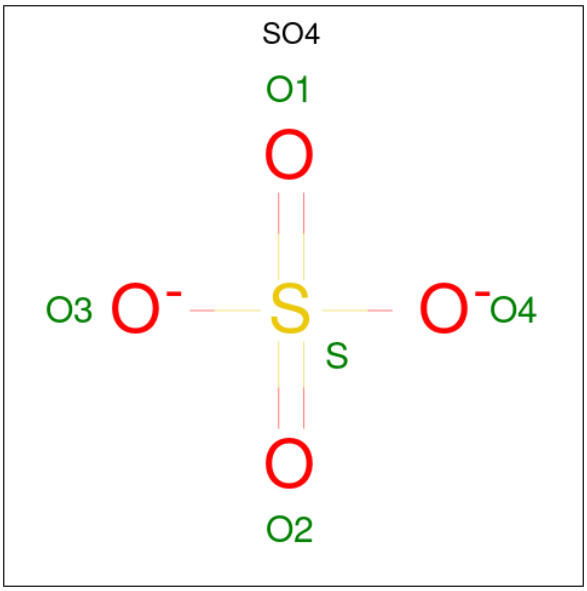
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	12	Total	C	N	O	P	0	0	0
			243	116	46	70	11			
3	F	12	Total	C	N	O	P	0	0	0
			243	116	46	70	11			

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



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

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O_4S).

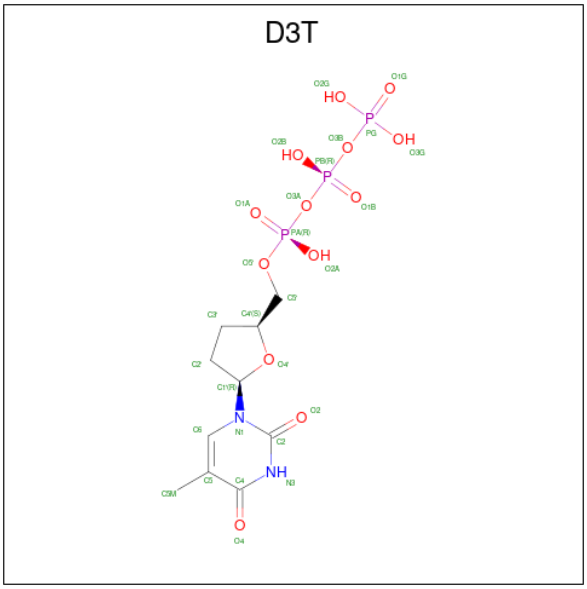


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total O S 5 4 1	0	0
5	D	1	Total O S 5 4 1	0	0
5	D	1	Total O S 5 4 1	0	0

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

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

- Molecule 7 is 2',3'-DIDEOXY-THYMIDINE-5'-TRIPHOSPHATE (three-letter code: D3T) (formula: $C_{10}H_{17}N_2O_{13}P_3$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
7	A	1	Total	C	N	O	P	0	0
			28	10	2	13	3		
7	D	1	Total	C	N	O	P	0	0
			28	10	2	13	3		

- Molecule 8 is water.

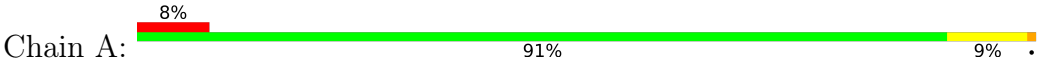
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	269	Total	O	0	0
			269	269		
8	D	476	Total	O	0	0
			476	476		
8	B	27	Total	O	0	0
			27	27		
8	C	33	Total	O	0	0
			33	33		
8	E	26	Total	O	0	0
			26	26		
8	F	38	Total	O	0	0
			38	38		

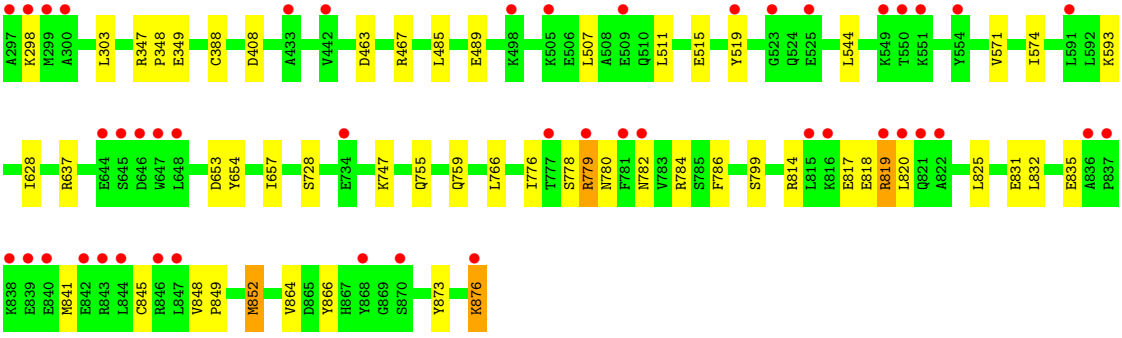
3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: DNA polymerase I, large fragment

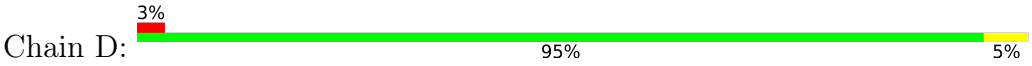
Chain A:

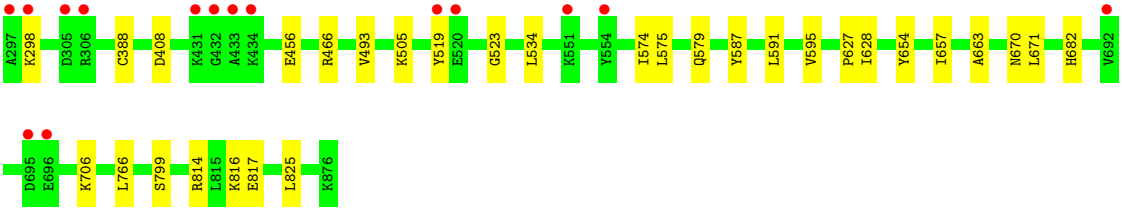




• Molecule 1: DNA polymerase I, large fragment

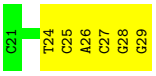
Chain D:





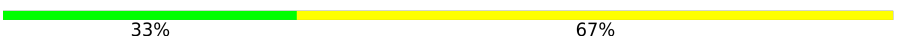
• Molecule 2: 5'-D(*CP*GP*AP*TP*CP*AP*CP*GP*(DDG))-3'

Chain B:



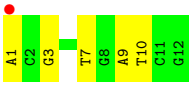
• Molecule 2: 5'-D(*CP*GP*AP*TP*CP*AP*CP*GP*(DDG))-3'

Chain E:



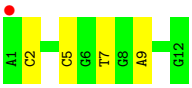
• Molecule 3: 5'-D(*AP*CP*GP*CP*CP*GP*TP*GP*AP*TP*CP*G)-3'

Chain C:



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

Chain F:

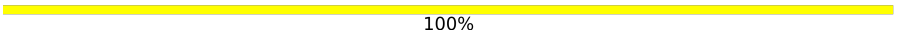


• Molecule 4: beta-D-fructofuranose-(2-1)-alpha-D-glucopyranose

Chain G:

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

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	93.17Å 108.71Å 152.21Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	42.84 – 1.81 42.82 – 1.81	Depositor EDS
% Data completeness (in resolution range)	100.0 (42.84-1.81) 98.9 (42.82-1.81)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.66 (at 1.81Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.211 , 0.247 0.209 , 0.244	Depositor DCC
R_{free} test set	6014 reflections (4.31%)	wwPDB-VP
Wilson B-factor (Å^2)	25.2	Xtriage
Anisotropy	0.037	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 49.0	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	0.95	EDS
Total number of atoms	11210	wwPDB-VP
Average B, all atoms (Å^2)	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.57% 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: D3T, CME, FRU, SO4, GLC, DDG, MG

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\| > 5$	RMSZ	$\#\|Z\| > 5$
1	A	0.55	0/4755	0.62	0/6422
1	D	0.65	0/4796	0.67	0/6480
2	B	1.07	0/179	1.97	9/274 (3.3%)
2	E	1.16	0/179	1.88	8/274 (2.9%)
3	C	1.25	0/272	1.74	5/418 (1.2%)
3	F	1.22	0/272	1.73	6/418 (1.4%)
All	All	0.67	0/10453	0.83	28/14286 (0.2%)

There are no bond length outliers.

All (28) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	26	DA	O4'-C1'-N9	14.39	118.07	108.00
2	B	28	DG	O4'-C4'-C3'	-7.43	101.53	104.50
3	F	7	DT	O4'-C1'-N1	-7.37	102.84	108.00
3	F	5	DC	O4'-C4'-C3'	-7.19	101.62	104.50
3	F	9	DA	O5'-P-OP2	-7.18	99.24	105.70
2	B	27	DC	O4'-C4'-C3'	-7.18	101.63	104.50
3	C	10	DT	C1'-O4'-C4'	-7.17	102.93	110.10
3	C	9	DA	O5'-P-OP2	-7.11	99.31	105.70
2	B	27	DC	O4'-C1'-N1	7.04	112.93	108.00
2	E	25	DC	O4'-C4'-C3'	-6.84	101.76	104.50
2	E	27	DC	O4'-C1'-N1	6.83	112.78	108.00
2	B	24	DT	C4-C5-C7	6.49	122.89	119.00
2	B	24	DT	C6-C5-C7	-6.33	119.10	122.90
2	B	25	DC	O4'-C1'-N1	6.08	112.25	108.00
2	B	27	DC	C4'-C3'-C2'	-5.90	97.79	103.10
3	C	7	DT	C4-C5-C7	5.86	122.51	119.00
3	C	1	DA	O4'-C1'-N9	5.82	112.07	108.00
2	E	25	DC	C4'-C3'-C2'	-5.79	97.89	103.10

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	3	DG	C2-N3-C4	5.67	114.73	111.90
3	F	2	DC	O4'-C1'-N1	5.59	111.91	108.00
2	B	27	DC	P-O3'-C3'	5.54	126.35	119.70
2	E	25	DC	O4'-C1'-N1	5.48	111.84	108.00
3	F	7	DT	C6-C5-C7	-5.31	119.72	122.90
2	E	25	DC	O5'-P-OP2	-5.15	101.07	105.70
2	E	25	DC	C1'-O4'-C4'	-5.12	104.98	110.10
2	E	24	DT	C4-C5-C7	5.07	122.04	119.00
3	F	7	DT	C4-C5-C7	5.07	122.04	119.00
2	E	27	DC	N1-C2-O2	-5.00	115.90	118.90

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	4673	0	4744	30	0
1	D	4701	0	4781	21	0
2	B	181	0	102	0	0
2	E	181	0	102	1	0
3	C	243	0	136	0	0
3	F	243	0	136	0	0
4	G	23	0	21	1	0
4	H	23	0	21	1	0
5	A	5	0	0	0	0
5	D	10	0	0	0	0
6	A	1	0	0	0	0
6	D	1	0	0	0	0
7	A	28	0	13	0	0
7	D	28	0	13	0	0
8	A	269	0	0	1	0
8	B	27	0	0	0	0
8	C	33	0	0	0	0
8	D	476	0	0	9	0
8	E	26	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
8	F	38	0	0	0	0
All	All	11210	0	10069	52	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 (52) 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:816:LYS:HG2	8:D:1037:HOH:O	1.56	1.03
1:A:818:GLU:C	1:A:819:ARG:HG2	1.97	0.84
1:A:832:LEU:HD12	1:A:852:MET:CE	2.12	0.79
1:D:466[B]:ARG:NH2	8:D:887:HOH:O	2.17	0.76
1:A:832:LEU:HD12	1:A:852:MET:HE1	1.68	0.73
1:D:816:LYS:HE3	8:D:1037:HOH:O	1.90	0.71
1:D:456:GLU:HG2	8:D:932:HOH:O	1.92	0.70
1:D:766:LEU:HD12	1:D:799:SER:HB3	1.76	0.68
1:D:534:LEU:HD11	1:D:574:ILE:HD13	1.76	0.66
1:A:778:SER:O	1:A:784:ARG:HD3	1.96	0.66
1:A:463:ASP:O	1:A:467[A]:ARG:HG3	1.97	0.64
1:D:408:ASP:HB2	4:H:2:FRU:H11	1.81	0.60
1:A:779:ARG:NE	1:A:779:ARG:H	2.01	0.58
1:A:832:LEU:HD12	1:A:852:MET:HE2	1.86	0.57
1:A:845[A]:CYS:SG	1:A:866:TYR:HE2	2.28	0.57
1:A:873:TYR:O	1:A:876:LYS:HD2	2.05	0.57
1:A:845[A]:CYS:HG	1:A:866:TYR:HE2	1.51	0.57
1:D:505:LYS:HE2	8:D:938:HOH:O	2.04	0.56
1:D:591:LEU:O	1:D:595:VAL:HG23	2.05	0.56
1:D:814:ARG:HA	1:D:817:GLU:HG2	1.88	0.56
1:D:816:LYS:HE2	8:D:1027:HOH:O	2.06	0.56
1:A:408:ASP:HB2	4:G:2:FRU:H11	1.88	0.55
1:A:848:VAL:HB	1:A:849:PRO:HD3	1.89	0.54
1:A:852:MET:HB3	1:A:864:VAL:HG21	1.88	0.54
1:A:779:ARG:HD2	1:A:780:ASN:N	2.22	0.54
1:A:814:ARG:HA	1:A:817:GLU:HB3	1.90	0.54
1:A:515:GLU:HG2	1:A:519:TYR:CE2	2.45	0.52
1:A:593:LYS:NZ	8:A:1057:HOH:O	2.44	0.51
1:D:575:LEU:O	1:D:579:GLN:HG3	2.11	0.51
1:D:682:HIS:CE1	1:D:706:LYS:HG3	2.48	0.48
1:A:654:TYR:HB3	1:A:657:ILE:HB	1.96	0.48
1:A:653:ASP:CB	1:A:831:GLU:HG2	2.44	0.47

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:628:ILE:HD12	1:A:637:ARG:NH1	2.31	0.46
1:A:776:ILE:O	1:A:784:ARG:HG3	2.16	0.46
1:A:782:ASN:O	1:A:786:PHE:HD2	1.99	0.45
1:D:534:LEU:HD11	1:D:574:ILE:CD1	2.42	0.45
1:D:670:ASN:HB3	8:D:1180:HOH:O	2.16	0.45
1:A:485:LEU:O	1:A:489:GLU:HG3	2.18	0.44
1:A:825:LEU:HD11	1:A:835:GLU:HB3	2.00	0.44
1:D:654:TYR:HB3	1:D:657:ILE:HB	2.00	0.44
1:D:587[B]:TYR:CE1	1:D:627:PRO:HD3	2.52	0.43
1:D:816:LYS:CG	8:D:1037:HOH:O	2.38	0.43
1:A:467[A]:ARG:HD3	8:D:1084:HOH:O	2.18	0.43
1:D:663:ALA:HB2	1:D:671:LEU:HG	2.01	0.42
1:A:571:VAL:HA	1:A:574:ILE:HD12	2.02	0.41
1:D:493[B]:VAL:HG12	1:D:825:LEU:HD13	2.02	0.41
1:A:755:GLN:HE21	1:A:759:GLN:NE2	2.19	0.41
1:A:347:ARG:HA	1:A:348:PRO:HD3	1.96	0.40
1:A:348:PRO:HD2	1:A:349:GLU:OE2	2.21	0.40
1:A:766:LEU:HD12	1:A:799:SER:HB3	2.03	0.40
1:D:519:TYR:O	1:D:523:GLY:N	2.55	0.40
2:E:21:DC:H2'	2:E:22:DG:C8	2.56	0.40

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	580/580 (100%)	567 (98%)	13 (2%)	0	100	100
1	D	585/580 (101%)	572 (98%)	12 (2%)	1 (0%)	47	33
All	All	1165/1160 (100%)	1139 (98%)	25 (2%)	1 (0%)	51	37

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	628	ILE

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	497/494 (101%)	484 (97%)	13 (3%)	46 32
1	D	502/494 (102%)	501 (100%)	1 (0%)	93 92
All	All	999/988 (101%)	985 (99%)	14 (1%)	67 58

All (14) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	298	LYS
1	A	303	LEU
1	A	507	LEU
1	A	511	LEU
1	A	544	LEU
1	A	728	SER
1	A	747	LYS
1	A	779	ARG
1	A	819	ARG
1	A	820	LEU
1	A	841	MET
1	A	852	MET
1	A	876	LYS
1	D	298	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	405	GLN
1	A	759	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)

4 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	DDG	E	29	3,2	17,23,24	0.90	0	15,33,36	1.52	3 (20%)
1	CME	D	388	1	8,9,10	0.69	0	5,9,11	2.06	2 (40%)
2	DDG	B	29	3,2	17,23,24	0.94	0	15,33,36	1.73	3 (20%)
1	CME	A	388	1	8,9,10	0.68	0	5,9,11	2.69	2 (40%)

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	DDG	E	29	3,2	-	0/3/18/19	0/3/3/3
1	CME	D	388	1	-	1/5/8/10	-
2	DDG	B	29	3,2	-	0/3/18/19	0/3/3/3
1	CME	A	388	1	-	2/5/8/10	-

There are no bond length outliers.

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	388	CME	CE-SD-SG	4.25	123.05	103.45
2	B	29	DDG	C2'-C1'-N9	-4.00	104.95	112.48
1	A	388	CME	CB-SG-SD	-3.88	93.77	103.82
1	D	388	CME	CE-SD-SG	3.84	121.14	103.45
2	E	29	DDG	O6-C6-C5	-2.74	119.01	124.37

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	29	DDG	O6-C6-C5	-2.62	119.25	124.37
2	E	29	DDG	C5-C6-N1	2.47	118.32	113.95
2	B	29	DDG	C5-C6-N1	2.38	118.16	113.95
2	E	29	DDG	C2'-C1'-N9	-2.14	108.45	112.48
1	D	388	CME	CB-SG-SD	-2.05	98.52	103.82

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	D	388	CME	CZ-CE-SD-SG
1	A	388	CME	SD-CE-CZ-OH
1	A	388	CME	CZ-CE-SD-SG

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	0.45	0	15,15,17	1.24	1 (6%)
4	FRU	G	2	4	11,12,12	0.57	0	10,18,18	1.04	1 (10%)
4	GLC	H	1	4	11,11,12	0.36	0	15,15,17	0.90	1 (6%)
4	FRU	H	2	4	11,12,12	0.71	0	10,18,18	0.75	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

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	G	1	GLC	C1-O5-C5	4.25	117.94	112.19
4	H	1	GLC	C1-O5-C5	2.34	115.37	112.19
4	G	2	FRU	O6-C6-C5	-2.04	104.29	111.29

There are no chirality outliers.

There are no torsion outliers.

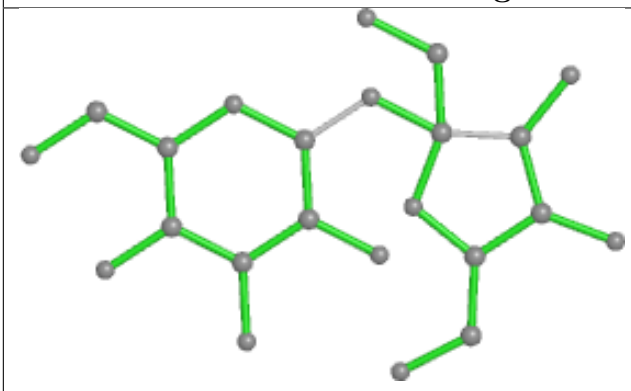
There are no ring outliers.

2 monomers are involved in 2 short contacts:

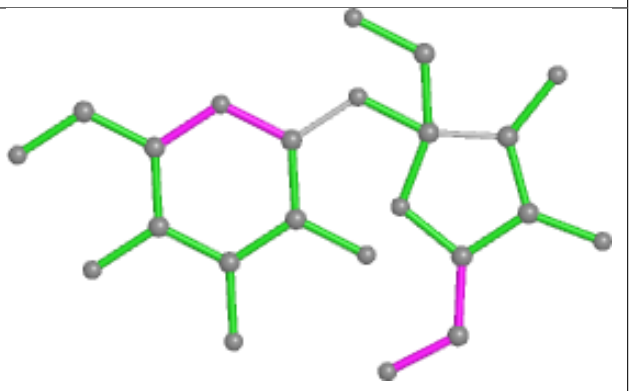
Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	H	2	FRU	1	0
4	G	2	FRU	1	0

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

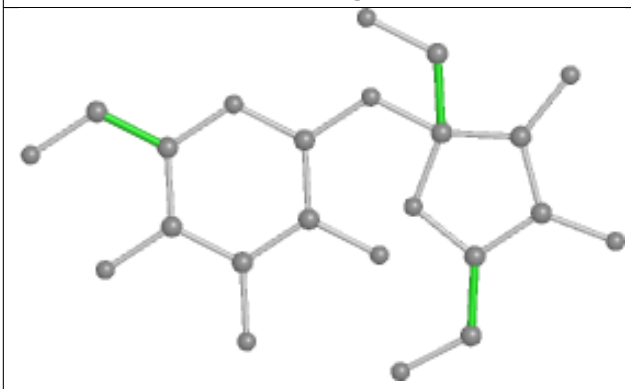
Oligosaccharide Chain G



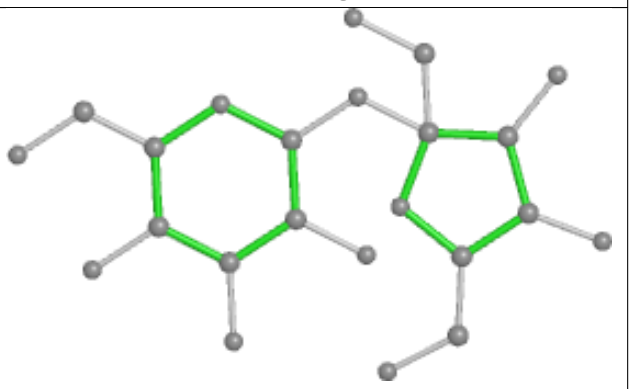
Bond lengths



Bond angles

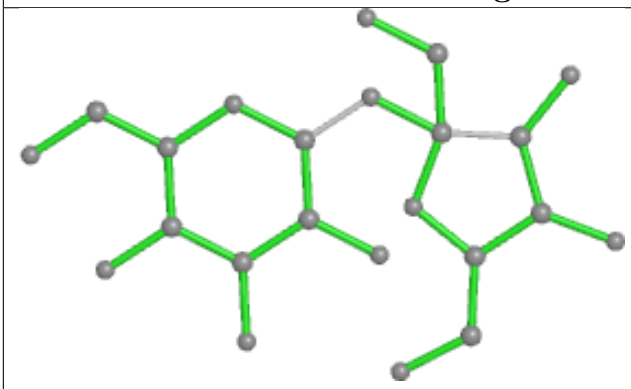


Torsions

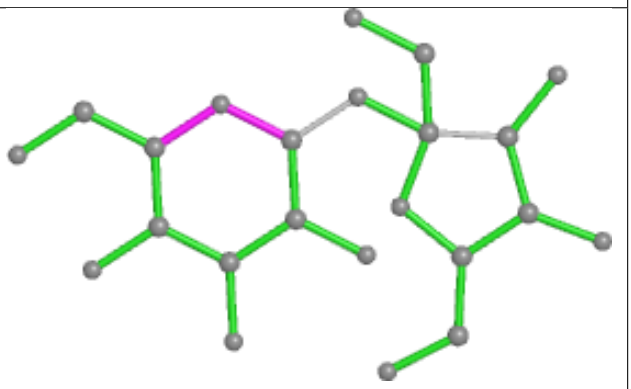


Rings

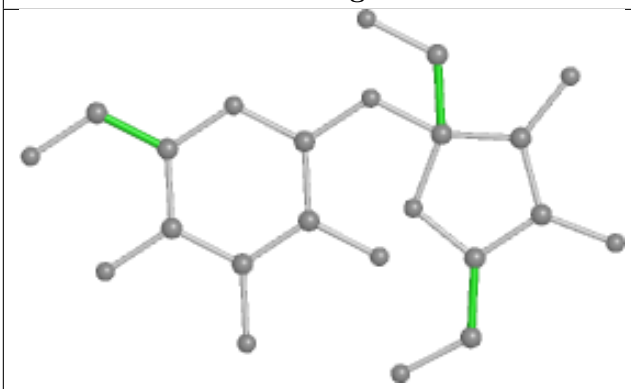
Oligosaccharide Chain H



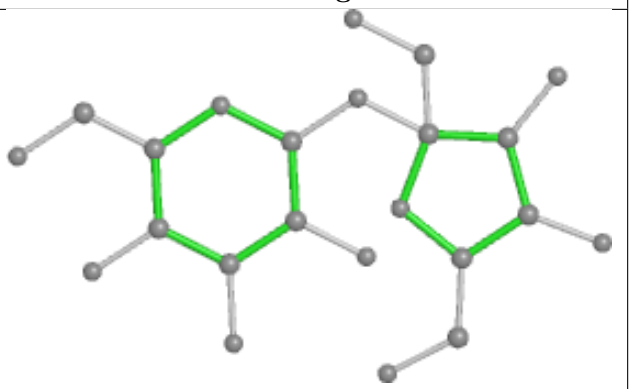
Bond lengths



Bond angles



Torsions



Rings

5.6 Ligand geometry (i)

Of 7 ligands modelled in this entry, 2 are 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	SO4	D	1	-	4,4,4	0.15	0	6,6,6	0.11	0
7	D3T	D	202	6	25,29,29	2.25	4 (16%)	35,45,45	2.07	10 (28%)
7	D3T	A	201	6	25,29,29	2.27	4 (16%)	35,45,45	2.07	9 (25%)
5	SO4	A	2	-	4,4,4	0.14	0	6,6,6	0.11	0
5	SO4	D	878	-	4,4,4	0.16	0	6,6,6	0.56	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
7	D3T	A	201	6	-	2/22/31/31	0/2/2/2
7	D3T	D	202	6	-	2/22/31/31	0/2/2/2

All (8) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	D	202	D3T	O2-C2	7.30	1.36	1.23
7	A	201	D3T	O2-C2	7.17	1.36	1.23
7	A	201	D3T	O4-C4	6.57	1.36	1.23
7	D	202	D3T	O4-C4	6.39	1.35	1.23
7	D	202	D3T	C6-C5	3.16	1.39	1.34
7	A	201	D3T	C2-N1	-3.09	1.33	1.38
7	A	201	D3T	C6-C5	2.75	1.39	1.34
7	D	202	D3T	C4-C5	2.49	1.48	1.44

All (19) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
7	A	201	D3T	C4-N3-C2	-5.73	119.94	127.35
7	D	202	D3T	N3-C2-N1	5.52	122.22	114.89
7	D	202	D3T	C4-N3-C2	-5.50	120.22	127.35
7	A	201	D3T	N3-C2-N1	5.20	121.79	114.89
7	D	202	D3T	C5-C4-N3	4.55	119.20	115.31
7	A	201	D3T	C5-C4-N3	4.14	118.84	115.31
7	A	201	D3T	C5-C6-N1	-3.41	119.83	123.34
7	D	202	D3T	C5-C6-N1	-3.13	120.12	123.34
7	A	201	D3T	O2-C2-N1	-2.86	118.98	122.79
7	A	201	D3T	C5M-C5-C4	2.86	121.92	118.77
7	D	202	D3T	O2-C2-N3	-2.77	116.35	121.50
7	A	201	D3T	O4-C4-C5	-2.68	121.79	124.90
7	D	202	D3T	O2A-PA-O1A	2.62	125.18	112.24
7	D	202	D3T	O3G-PG-O2G	2.37	116.69	107.64
7	D	202	D3T	C5M-C5-C4	2.33	121.33	118.77
7	A	201	D3T	O3B-PG-O1G	-2.27	98.59	111.19
7	A	201	D3T	PB-O3B-PG	-2.22	125.20	132.83
7	D	202	D3T	O4-C4-C5	-2.14	122.42	124.90
7	D	202	D3T	PB-O3B-PG	-2.03	125.87	132.83

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
7	A	201	D3T	PG-O3B-PB-O1B
7	D	202	D3T	PG-O3B-PB-O1B
7	A	201	D3T	PG-O3B-PB-O2B
7	D	202	D3T	PG-O3B-PB-O2B

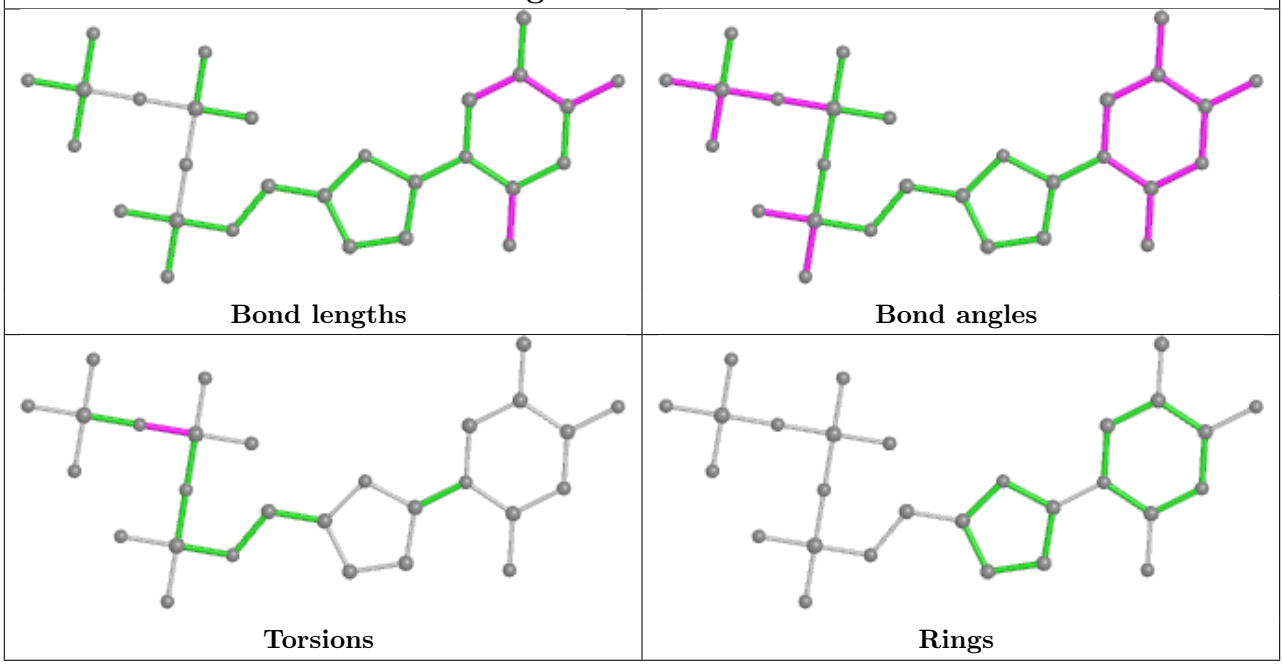
There are no ring outliers.

No monomer is involved in short contacts.

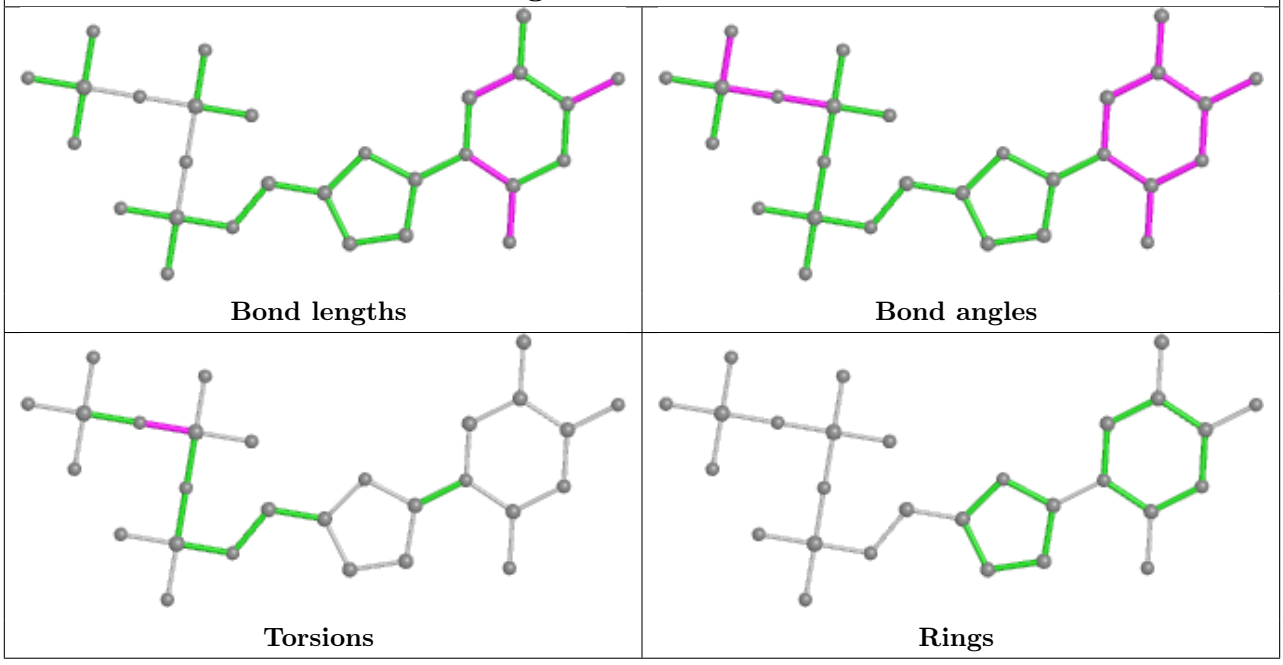
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less 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 D3T D 202



Ligand D3T A 201



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^{th} 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	579/580 (99%)	0.51	46 (7%) 12 9	16, 33, 49, 57	0
1	D	579/580 (99%)	0.26	15 (2%) 56 51	10, 22, 43, 50	0
2	B	8/9 (88%)	0.03	0 100 100	18, 29, 46, 51	0
2	E	8/9 (88%)	0.06	0 100 100	13, 30, 55, 62	0
3	C	12/12 (100%)	0.18	1 (8%) 11 8	16, 31, 57, 75	0
3	F	12/12 (100%)	0.21	1 (8%) 11 8	13, 24, 49, 67	0
All	All	1198/1202 (99%)	0.38	63 (5%) 26 21	10, 28, 47, 75	0

All (63) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	433	ALA	6.7
1	A	820	LEU	5.5
1	A	551	LYS	5.4
1	A	647	TRP	5.2
1	A	645	SER	5.2
1	A	297	ALA	5.0
1	D	695	ASP	4.7
1	A	843	ARG	4.5
1	D	298	LYS	4.5
3	C	1	DA	4.3
1	A	819	ARG	4.0
1	D	692	VAL	4.0
1	A	781	PHE	3.6
1	A	298	LYS	3.5
1	A	549	LYS	3.5
1	A	550	THR	3.5
1	A	433	ALA	3.3
1	A	777	THR	3.3
1	A	554	TYR	3.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	442	VAL	3.3
1	A	816	LYS	3.2
1	A	844	LEU	3.2
1	A	509	GLU	3.2
1	A	646	ASP	3.2
1	D	305	ASP	3.1
1	A	842	GLU	3.0
1	A	821	GLN	3.0
1	A	815	LEU	2.9
1	D	431	LYS	2.9
1	D	554	TYR	2.9
1	D	297	ALA	2.8
1	A	846	ARG	2.8
1	A	876	LYS	2.8
1	D	519	TYR	2.8
1	A	519	TYR	2.7
1	A	505	LYS	2.6
1	A	870	SER	2.6
1	A	839	GLU	2.6
1	A	300	ALA	2.6
1	A	779	ARG	2.6
1	A	648	LEU	2.5
3	F	1	DA	2.5
1	D	434	LYS	2.5
1	A	837	PRO	2.5
1	D	432	GLY	2.4
1	A	644	GLU	2.4
1	A	840	GLU	2.3
1	D	551	LYS	2.2
1	A	847	LEU	2.2
1	A	498	LYS	2.2
1	D	520	GLU	2.2
1	A	838	LYS	2.2
1	A	591	LEU	2.2
1	A	822	ALA	2.2
1	A	782	ASN	2.2
1	A	525	GLU	2.1
1	A	734	GLU	2.1
1	A	836	ALA	2.1
1	D	306	ARG	2.1
1	A	868	TYR	2.1
1	A	299	MET	2.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	523	GLY	2.0
1	D	696	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^{th} 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
1	CME	A	388	10/11	0.92	0.15	26,30,47,49	0
1	CME	D	388	10/11	0.95	0.12	17,20,41,44	0
2	DDG	B	29	21/22	0.98	0.10	15,18,22,26	0
2	DDG	E	29	21/22	0.98	0.10	11,14,16,17	0

6.3 Carbohydrates (i)

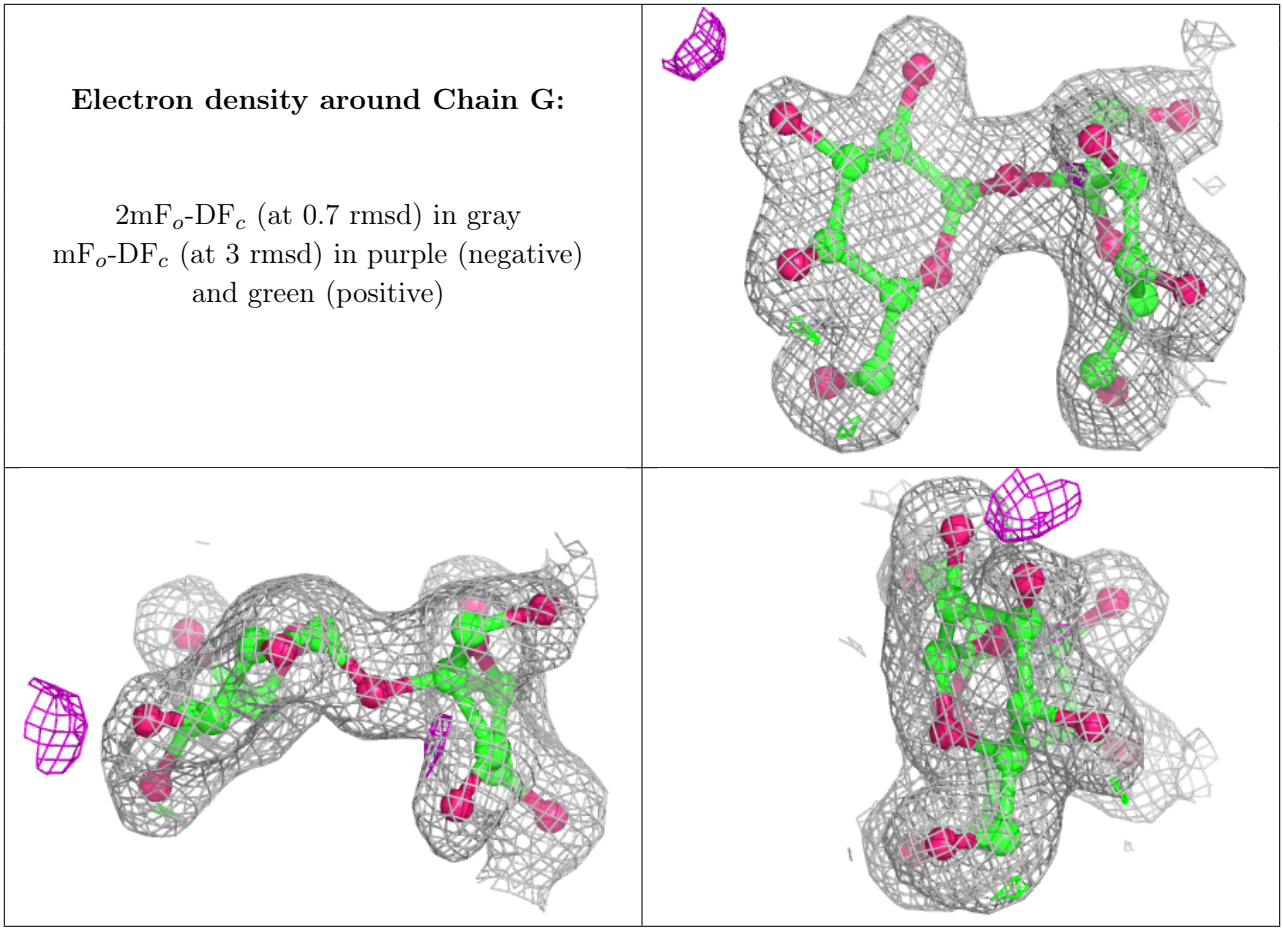
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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	0.89	0.13	31,32,33,35	0
4	FRU	G	2	12/12	0.91	0.12	26,30,32,36	0
4	GLC	H	1	11/12	0.94	0.08	19,20,23,23	0
4	FRU	H	2	12/12	0.94	0.08	15,18,19,19	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.

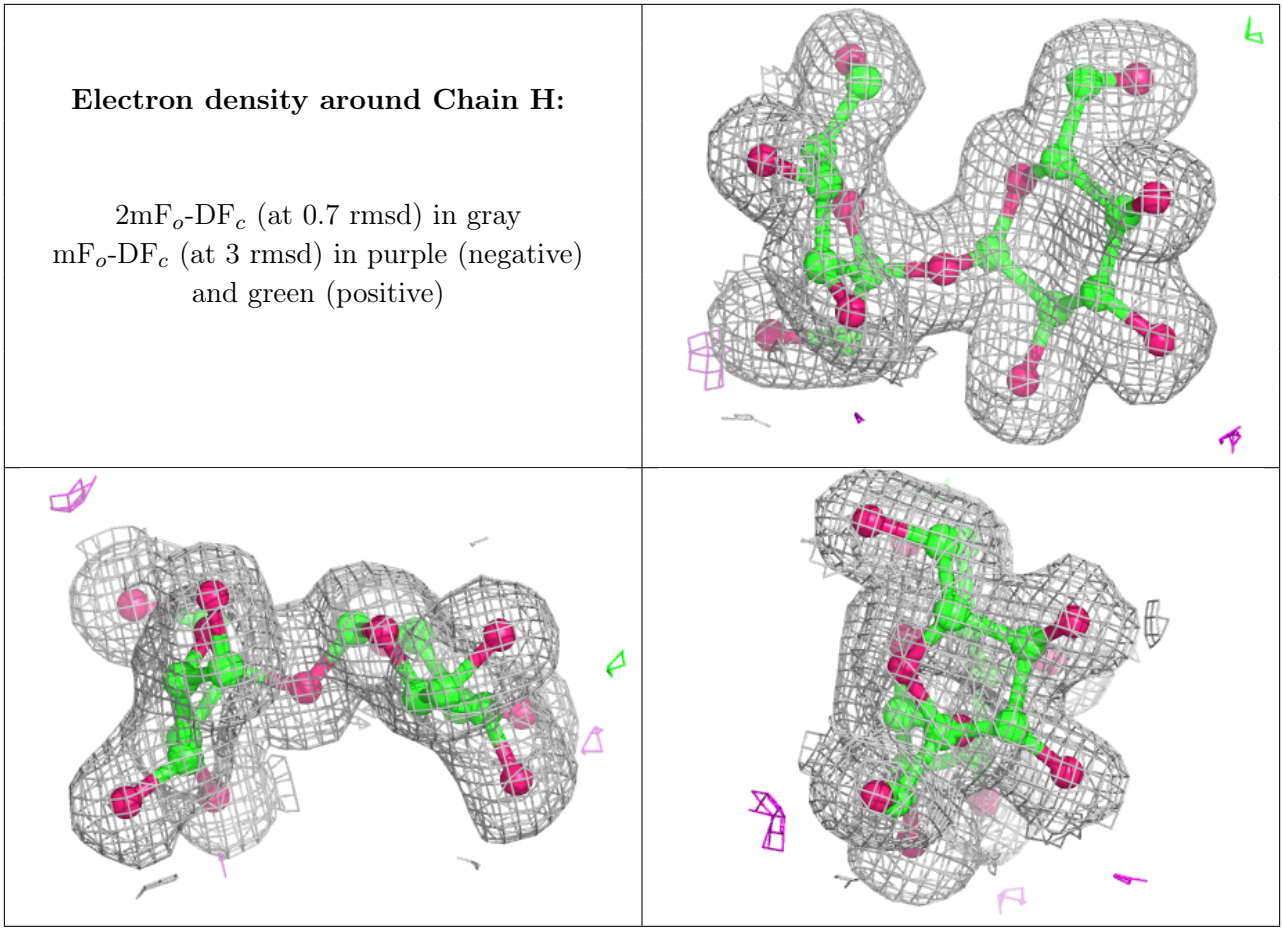
Electron density around Chain G:

$2mF_o$-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

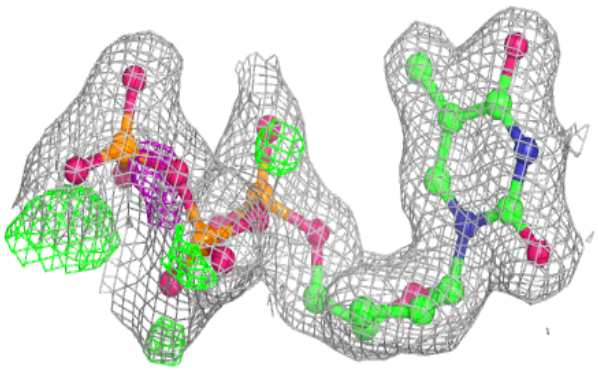
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9'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

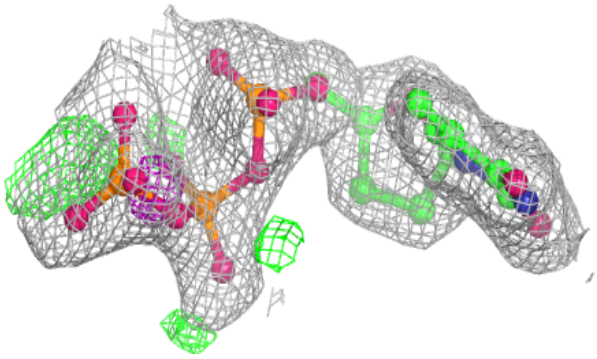
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
5	SO4	D	1	5/5	0.88	0.23	72,72,73,73	0
5	SO4	A	2	5/5	0.92	0.24	87,88,88,88	0
5	SO4	D	878	5/5	0.93	0.15	38,42,42,44	0
6	MG	D	200	1/1	0.94	0.06	30,30,30,30	0
7	D3T	D	202	28/28	0.94	0.10	19,24,34,36	0
7	D3T	A	201	28/28	0.96	0.08	22,26,36,36	0
6	MG	A	200	1/1	0.96	0.07	35,35,35,35	0

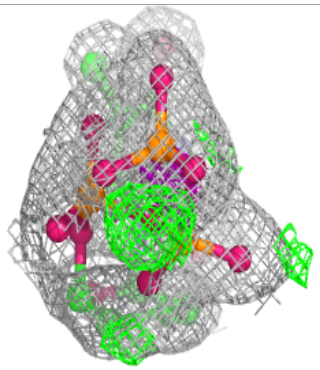
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around D3T D 202:

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

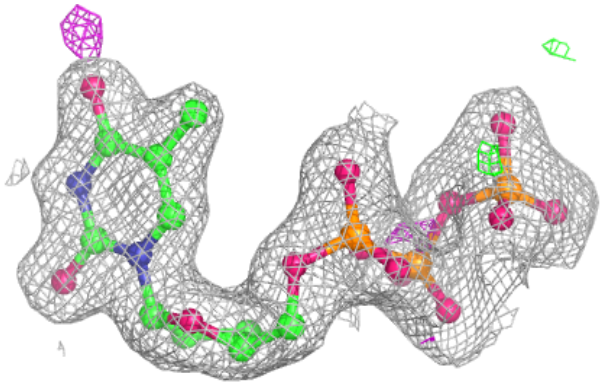


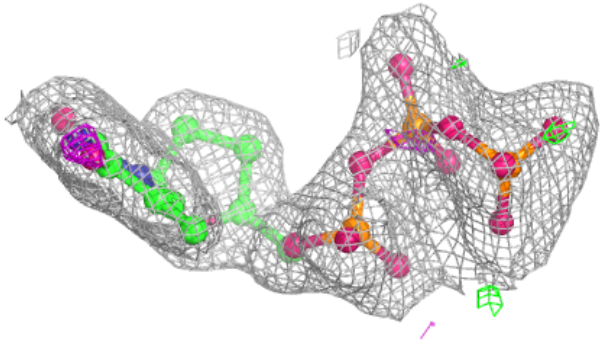


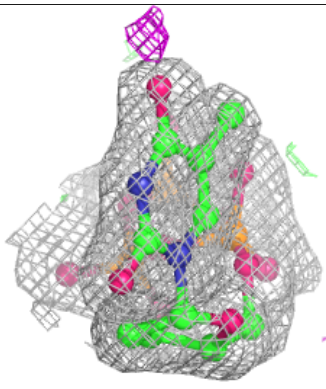


Electron density around D3T A 201:

$2mF_o$-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.