WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Sep 5, 2023 – 03:02 PM EDT

PDB ID	:	4DSJ
Title	:	Crystal structure of fragment DNA polymerase I from Bacillus stearother-mophilus with duplex DNA, dGTP and Calcium
Authors	:	Gan, J.H.; Abdur, R.; Liu, H.H.; Sheng, J.; Caton-Willians, J.; Soares, A.S.; Huang, Z.
Deposited on	:	2012-02-19
Resolution	:	2.86 Å(reported)

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

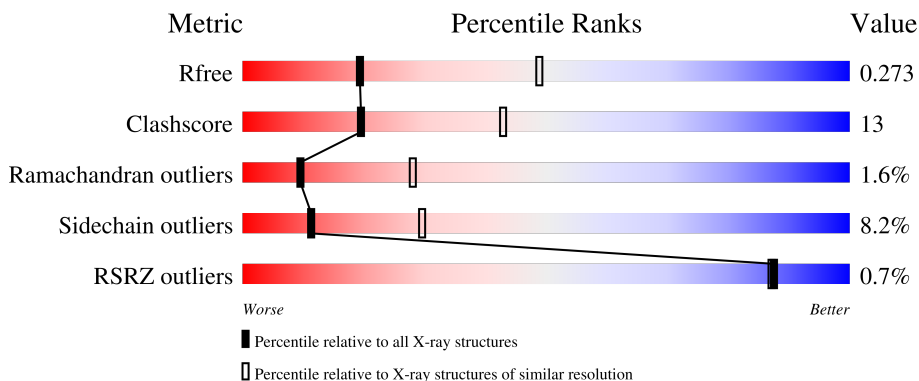|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh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

1 Overall quality at a glance ⓘ

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

The reported resolution of this entry is 2.86 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.
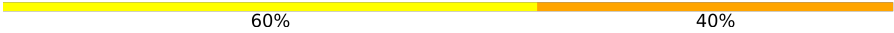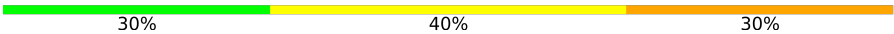
Metric	Percentile Ranks	Value
Rfree		0.273
Clashscore		13
Ramachandran outliers		1.6%
Sidechain outliers		8.2%
RSRZ outliers		0.7%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3168 (2.90-2.82)
Clashscore	141614	3438 (2.90-2.82)
Ramachandran outliers	138981	3348 (2.90-2.82)
Sidechain outliers	138945	3351 (2.90-2.82)
RSRZ outliers	127900	3103 (2.90-2.82)

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	579	66% 30% ·
1	B	579	.% 68% 28% ··
2	C	12	8% 50% 42%
2	E	12	8% 50% 42%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	D	10	60% 40%
3	F	10	30% 40% 30%

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 10210 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	579	Total 4655	C 2960	N 807	O 871	S 17	0	1	0
1	B	575	Total 4623	C 2940	N 801	O 865	S 17	0	1	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	598	ASP	ALA	engineered mutation	UNP D9N168
A	713	VAL	PRO	engineered mutation	UNP D9N168
B	598	ASP	ALA	engineered mutation	UNP D9N168
B	713	VAL	PRO	engineered mutation	UNP D9N168

- Molecule 2 is a DNA chain called DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	12	Total 245	C 117	N 48	O 69	P 11	0	0	0
2	E	12	Total 245	C 117	N 48	O 69	P 11	0	0	0

- Molecule 3 is a DNA chain called dGTP.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	10	Total 200	C 96	N 36	O 59	P 9	0	0	0
3	F	10	Total 200	C 96	N 36	O 59	P 9	0	0	0

- Molecule 4 is 2'-DEOXYGUANOSINE-5'-TRIPHOSPHATE (three-letter code: DGT) (formula: $C_{10}H_{16}N_5O_{13}P_3$).

DGT

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
4	A	1	31	10	5	13	3	0	0

- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Ca		
5	A	2	2	2	0	0

- Molecule 6 is water.

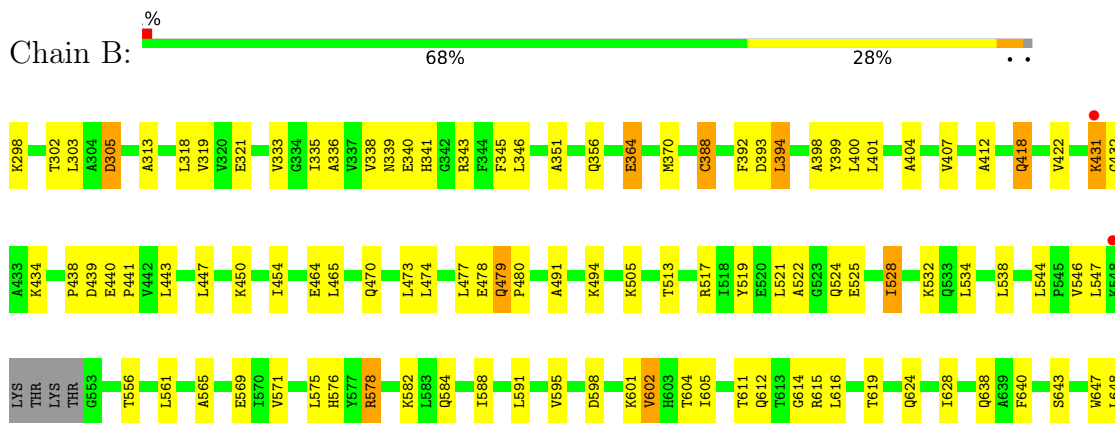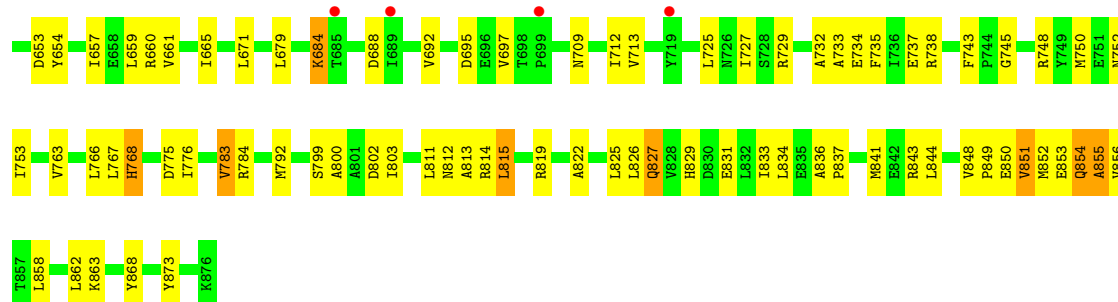| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	A	8	8	8	0	0
6	B	1	1	1	0	0

3 Residue-property plots ⓘ

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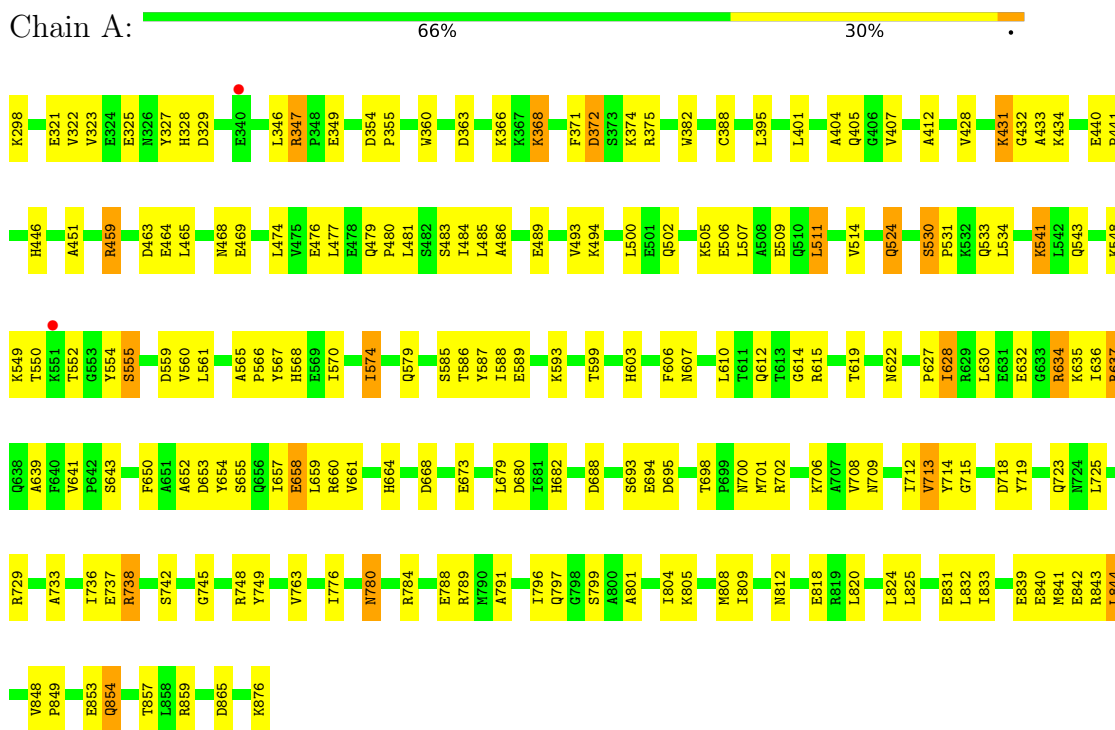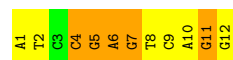
- Molecule 1: DNA polymerase

- Molecule 1: DNA polymerase

D653 Y654 I657 E658 L659 R660 V661 I665 L671 L679 K684 T685 D688 I689 V692 D695 E696 V697 T698 P699 N709 I712 V713 Y719 L725 N726 I727 S728 R729 A732 A733 E734 F735 I736 E737 R738 F743 P744 G745 R748 Y749 M750 E751 N752

I753 V763 L766 L767 H768 D775 I776 V783 R784 M792 S799 A800 A801 D802 I803 L811 N812 A813 R814 L815 R819 A822 L825 L826 Q827 V828 H829 D830 E831 L832 I833 L834 E835 A836 P837 M841 E842 R843 L844 V848 P849 E850 V851 M852 E853 Q854 A855 V856

T857 L858 L862 K863 Y868 Y873 K876

• Molecule 2: DNA

Chain C:

8%	50%	42%

A1 T2 C3 C4 G5 A6 G7 T8 C9 A10 G11 G12

• Molecule 2: DNA

Chain E:

8%	50%	42%

A1 T2 C3 C4 G5 A6 G7 T8 C9 A10 G11 G12

• Molecule 3: dGTP

Chain D:

60%	40%

C21 C22 T23 G24 A25 C26 T27 C28 G29 G30

• Molecule 3: dGTP

Chain F:

30%	40%	30%

C21 C22 T23 G24 A25 C26 T27 G30

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	60.02Å 112.07Å 229.45Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	114.73 – 2.86 29.75 – 2.86	Depositor EDS
% Data completeness (in resolution range)	89.8 (114.73-2.86) 89.9 (29.75-2.86)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$< I/\sigma(I) >$ [1]	1.74 (at 2.85Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.213 , 0.277 0.206 , 0.273	Depositor DCC
R_{free} test set	3269 reflections (9.93%)	wwPDB-VP
Wilson B-factor (Å^2)	68.3	Xtriage
Anisotropy	0.268	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 44.6	EDS
L-test for twinning[2]	$< |L| > = 0.43, < L^2 > = 0.26$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	10210	wwPDB-VP
Average B, all atoms (Å^2)	86.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.82% 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 ⓘ

5.1   Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: CA, DGT

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	0.68	0/4742	0.77	0/6407
1	B	0.51	0/4709	0.64	1/6362 (0.0%)
2	C	1.32	1/275 (0.4%)	2.47	20/423 (4.7%)
2	E	1.04	0/275	2.05	18/423 (4.3%)
3	D	1.28	1/223 (0.4%)	2.20	12/342 (3.5%)
3	F	1.30	0/223	2.01	7/342 (2.0%)
All	All	0.69	2/10447 (0.0%)	0.98	58/14299 (0.4%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	21	DC	C3'-O3'	-5.24	1.37	1.44
2	C	7	DG	C5-C6	-5.04	1.37	1.42

The worst 5 of 58 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	11	DG	O4'-C4'-C3'	-12.44	98.54	106.00
2	C	7	DG	P-O3'-C3'	12.21	134.36	119.70
3	F	27	DT	O4'-C4'-C3'	-12.13	98.72	106.00
3	D	29	DG	O4'-C1'-N9	12.04	116.43	108.00
2	C	11	DG	O4'-C4'-C3'	-11.54	99.07	106.00

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	4655	0	4710	129	0
1	B	4623	0	4669	110	0
2	C	245	0	136	11	0
2	E	245	0	136	5	0
3	D	200	0	114	7	0
3	F	200	0	114	5	0
4	A	31	0	12	1	0
5	A	2	0	0	0	0
6	A	8	0	0	0	0
6	B	1	0	0	0	0
All	All	10210	0	9891	254	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 13.

The worst 5 of 254 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:A:653:ASP:HB3	1:A:831:GLU:HG2	1.35	1.08
1:B:647:TRP:HA	1:B:837:PRO:HA	1.46	0.95
1:B:737[B]:GLU:OE1	1:B:737[B]:GLU:HA	1.70	0.91
1:B:473:LEU:HD12	1:B:477:LEU:HD23	1.55	0.88
2:E:5:DG:H2'	2:E:6:DA:O4'	1.74	0.86

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	578/579 (100%)	516 (89%)	54 (9%)	8 (1%)	11	31
1	B	572/579 (99%)	492 (86%)	70 (12%)	10 (2%)	9	27
All	All	1150/1158 (99%)	1008 (88%)	124 (11%)	18 (2%)	9	28

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	418	GLN
1	A	433	ALA
1	A	628	ILE
1	A	694	GLU
1	B	431	LYS

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	497/496 (100%)	454 (91%)	43 (9%)	10	27
1	B	493/496 (99%)	455 (92%)	38 (8%)	13	32
All	All	990/992 (100%)	909 (92%)	81 (8%)	11	29

5 of 81 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	546	VAL
1	B	763	VAL
1	B	569	GLU
1	B	659	LEU
1	B	815	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	612	GLN
1	B	827	GLN
1	B	823	HIS
1	A	782	ASN
1	B	579	GLN

5.3.3   RNA ⓘ

There are no RNA molecules in this entry.

5.4   Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5   Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6   Ligand geometry ⓘ

Of 3 ligands modelled in this entry, 2 are monoatomic - leaving 1 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	#$|Z| > 2$
4	DGT	A	1001	5	26,33,33	0.87	1 (3%)	32,52,52	1.27	3 (9%)

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	DGT	A	1001	5	-	3/18/34/34	0/3/3/3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	1001	DGT	C6-N1	-2.03	1.34	1.37

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	1001	DGT	PB-O3B-PG	-3.12	122.13	132.83
4	A	1001	DGT	C5-C6-N1	2.91	119.08	113.95
4	A	1001	DGT	PA-O3A-PB	-2.38	124.65	132.83

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	1001	DGT	PA-O3A-PB-O2B
4	A	1001	DGT	C4'-C5'-O5'-PA
4	A	1001	DGT	C5'-O5'-PA-O2A

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1001	DGT	1	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less 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 DGT A 1001

Bond lengths

Bond angles

Torsions

Rings
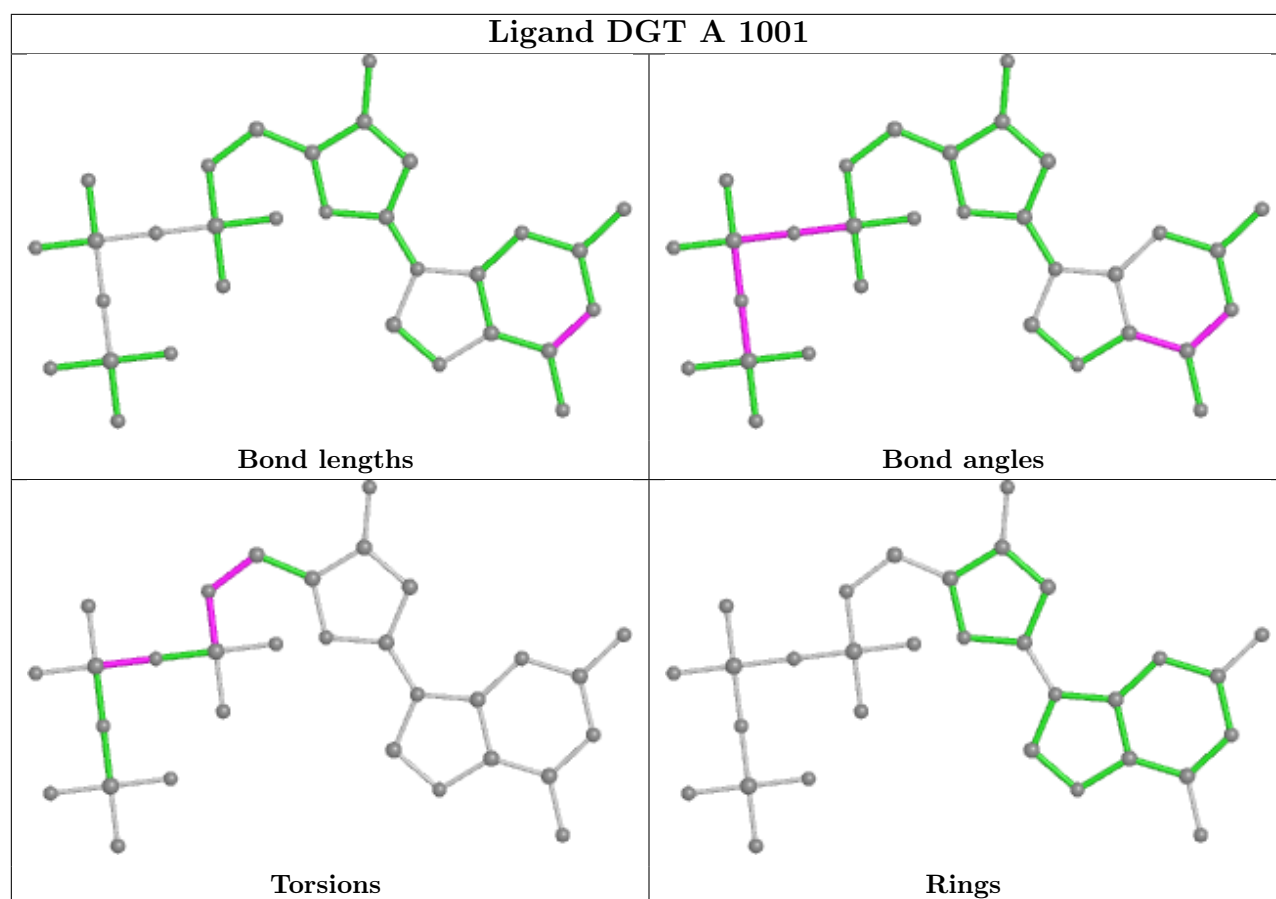
5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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, 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(Å²)	Q<0.9
1	A	579/579 (100%)	-0.43	2 (0%)	94	94	40, 68, 94, 106	0
1	B	575/579 (99%)	-0.06	6 (1%)	82	81	71, 105, 136, 155	0
2	C	12/12 (100%)	-0.69	0	100	100	46, 57, 102, 109	0
2	E	12/12 (100%)	-0.63	0	100	100	74, 87, 98, 99	0
3	D	10/10 (100%)	-1.00	0	100	100	45, 52, 68, 87	0
3	F	10/10 (100%)	-0.90	0	100	100	49, 60, 83, 84	0
All	All	1198/1202 (99%)	-0.27	8 (0%)	87	87	40, 85, 130, 155	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	719	TYR	3.8
1	B	548	LYS	3.0
1	B	431	LYS	2.7
1	B	685	THR	2.7
1	B	689	ILE	2.6

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.
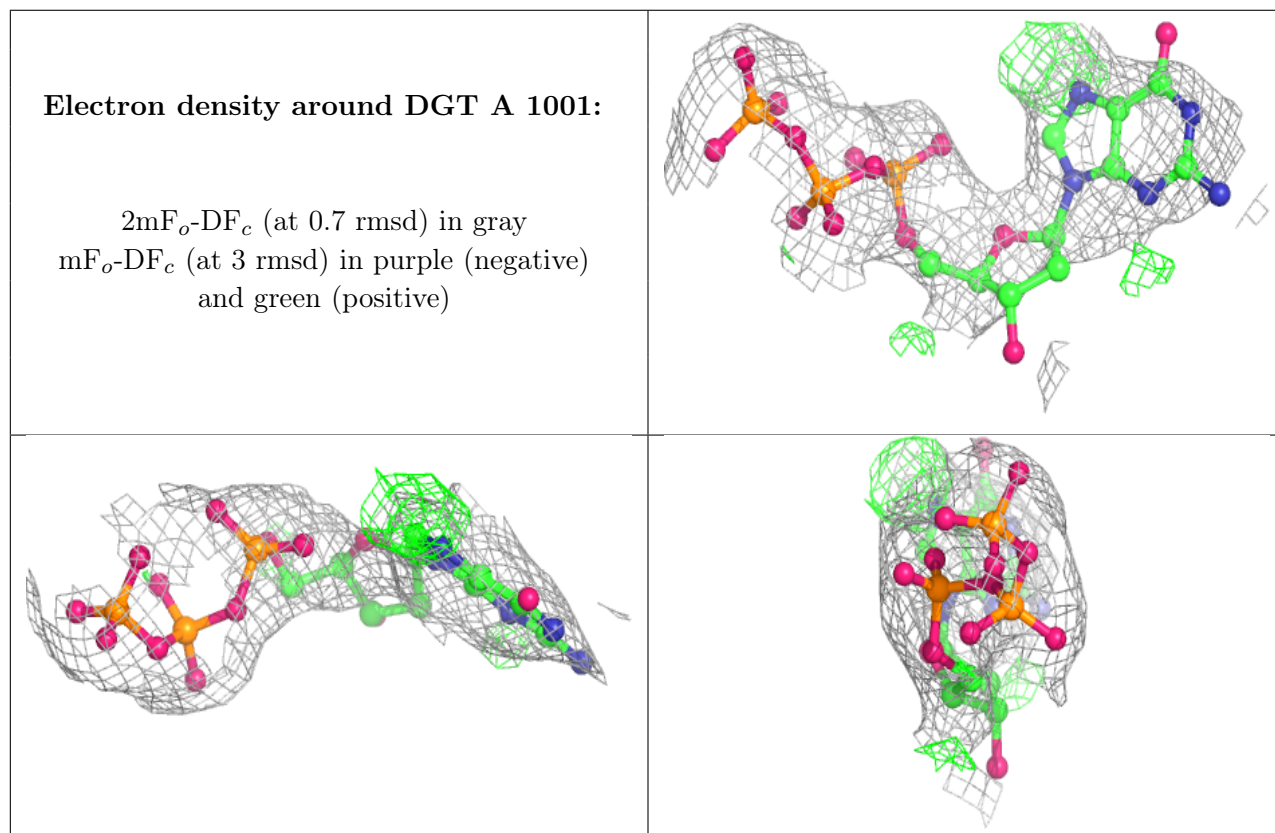
6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

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
4	DGT	A	1001	31/31	0.92	0.21	110,113,115,115	31
5	CA	A	1003	1/1	0.93	0.25	104,104,104,104	1
5	CA	A	1002	1/1	0.94	0.41	107,107,107,107	1

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 DGT A 1001:

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.