



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 8, 2026 – 02:56 PM UTC

PDB ID : 4FYD / pdb_00004fyd
Title : Crystal structure of yeast DNA polymerase alpha bound to DNA/RNA and dGTP
Authors : Perera, R.L.; Pellegrini, L.
Deposited on : 2012-07-04
Resolution : 3.10 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

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	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Buster-report	:	wwPDB partial adaption of 1.1.7 (2018)
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

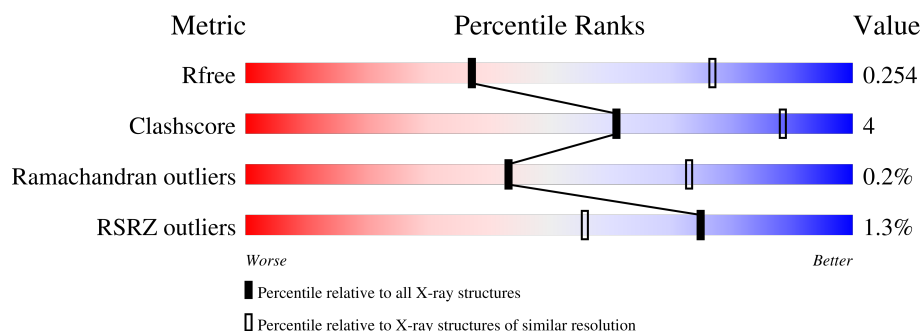
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.10 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}	180053	1456 (3.10-3.10)
Clashscore	190562	1539 (3.10-3.10)
Ramachandran outliers	187476	1467 (3.10-3.10)
RSRZ outliers	180081	1456 (3.10-3.10)

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 $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	910	84% 9% 7%
1	B	910	83% 10% 7%
2	C	25	56% 8% 36%
2	D	25	48% 16% 36%
3	E	12	67% 33%
3	F	12	67% 33%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 29108 atoms, of which 14364 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 alpha catalytic subunit A.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	848	Total	C	H	N	O	S	0	0	0
			13634	4285	6869	1168	1265	47			
1	B	848	Total	C	H	N	O	S	0	0	0
			13634	4285	6869	1168	1265	47			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	508	ALA	ARG	engineered mutation	UNP P13382
A	509	ALA	ASN	engineered mutation	UNP P13382
A	998	ASN	ASP	engineered mutation	UNP P13382
B	508	ALA	ARG	engineered mutation	UNP P13382
B	509	ALA	ASN	engineered mutation	UNP P13382
B	998	ASN	ASP	engineered mutation	UNP P13382

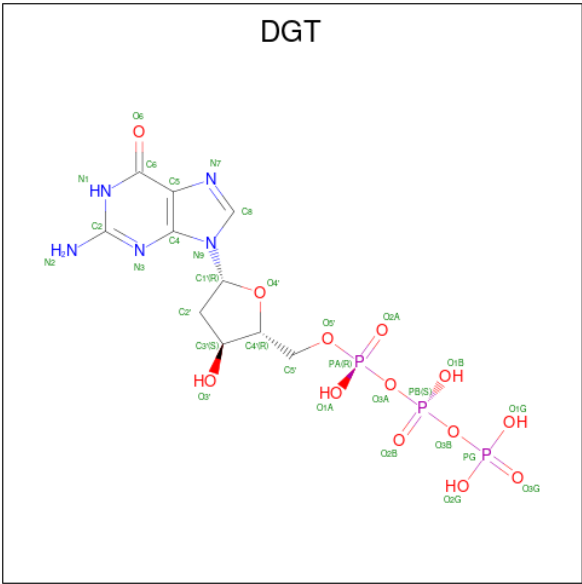
- Molecule 2 is a DNA chain called DNA (5'-D(*TP*GP*AP*GP*CP*GP*TP*G*TP*GP*T
P*AP*CP*CP*CP*CP*TP*GP*CP*CP*CP*GP*CP*CP*G)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	C	16	Total	C	H	N	O	P	0	0	0
			494	151	179	53	96	15			
2	D	16	Total	C	H	N	O	P	0	0	0
			494	151	179	53	96	15			

- Molecule 3 is DNA/RNA hybrid called DNA/RNA (5'-R(*CP*GP*GP*CP*GP*GP*GP*C
P*AP*G)-D(P*GP*G)-3').

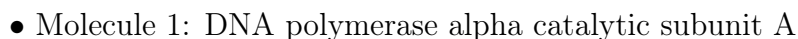
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	E	12	Total	C	H	N	O	P	0	0	0
			395	117	134	54	79	11			
3	F	12	Total	C	H	N	O	P	0	0	0
			395	117	134	54	79	11			

- Molecule 4 is 2'-DEOXYGUANOSINE-5'-TRIPHOSPHATE (CCD ID: DGT) (formula: C₁₀H₁₆N₅O₁₃P₃).



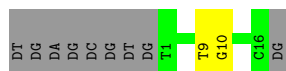
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	A	1	Total	C	N	O	P	0	0
			31	10	5	13	3		
4	B	1	Total	C	N	O	P	0	0
			31	10	5	13	3		

- Molecule 1: DNA polymerase alpha catalytic subunit A



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

Chain C:  56% 8% 36%



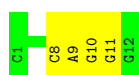
- Molecule 2: DNA (5'-D(*TP*GP*AP*GP*CP*GP*TP*G*TP*GP*TP*AP*CP*CP*CP*CP*TP*GP*CP*CP*CP*GP*CP*CP*G)-3')

Chain D:  48% 16% 36%



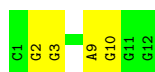
- Molecule 3: DNA/RNA (5'-R(*CP*GP*GP*CP*GP*GP*GP*CP*AP*G)-D(P*GP*G)-3')

Chain E:  67% 33%



- Molecule 3: DNA/RNA (5'-R(*CP*GP*GP*CP*GP*GP*GP*CP*AP*G)-D(P*GP*G)-3')

Chain F:  67% 33%



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	111.69Å 145.69Å 197.20Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.80 – 3.10 48.80 – 3.10	Depositor EDS
% Data completeness (in resolution range)	99.2 (48.80-3.10) 99.8 (48.80-3.10)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	0.10	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.12 (at 3.12Å)	Xtriage
Refinement program	PHENIX (phenix.refine: dev_1078)	Depositor
R, R_{free}	0.210 , 0.247 0.221 , 0.254	Depositor DCC
R_{free} test set	2987 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å ²)	98.4	Xtriage
Anisotropy	0.488	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 63.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	29108	wwPDB-VP
Average B, all atoms (Å ²)	121.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.54% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: 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	0.29	0/6888	0.66	0/9312
1	B	0.29	0/6888	0.66	0/9312
2	C	0.16	0/350	0.56	0/536
2	D	0.13	0/350	0.53	0/536
3	E	0.14	0/293	0.38	0/457
3	F	0.13	0/293	0.37	0/457
All	All	0.28	0/15062	0.64	0/20610

There are no bond length outliers.

There are no bond angle outliers.

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	6765	6869	6855	54	0
1	B	6765	6869	6855	59	0
2	C	315	179	181	1	0
2	D	315	179	181	3	0
3	E	261	134	134	5	0
3	F	261	134	134	4	0
4	A	31	0	12	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	31	0	12	2	0
All	All	14744	14364	14364	122	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 4.

The worst 5 of 122 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)
4:A:1301:DGT:O2B	4:A:1301:DGT:H5'A	1.89	0.73
4:B:1301:DGT:H5'A	4:B:1301:DGT:O2B	1.97	0.64
4:A:1301:DGT:H5'A	4:A:1301:DGT:PB	2.40	0.60
1:B:587:ARG:HD3	1:B:616:GLU:HA	1.85	0.58
2:D:9:DT:H2''	2:D:10:DG:H5''	1.86	0.58

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	838/910 (92%)	789 (94%)	47 (6%)	2 (0%)	43	73
1	B	838/910 (92%)	787 (94%)	49 (6%)	2 (0%)	43	73
All	All	1676/1820 (92%)	1576 (94%)	96 (6%)	4 (0%)	43	73

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	589	PRO
1	A	1120	ASN
1	B	589	PRO

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	1120	ASN

5.3.2 Protein sidechains [i](#)

There are no protein residues with a non-rotameric sidechain to report in this entry.

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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

5.5 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

2 ligands 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	DGT	A	1301	-	32,33,33	1.79	5 (15%)	48,52,52	1.70	10 (20%)
4	DGT	B	1301	-	32,33,33	1.78	5 (15%)	48,52,52	1.68	9 (18%)

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	1301	-	-	3/22/34/34	0/3/3/3
4	DGT	B	1301	-	-	5/22/34/34	0/3/3/3

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	1301	DGT	C4-N3	6.38	1.48	1.34
4	A	1301	DGT	C4-N3	6.36	1.48	1.34
4	B	1301	DGT	C2-N2	4.08	1.43	1.34
4	A	1301	DGT	C2-N2	4.05	1.43	1.34
4	B	1301	DGT	C2-N1	2.67	1.44	1.37

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	1301	DGT	C5-C4-N3	-5.51	119.62	128.39
4	B	1301	DGT	C5-C4-N3	-5.43	119.75	128.39
4	A	1301	DGT	C2-N3-C4	4.63	120.28	112.30
4	B	1301	DGT	C2-N3-C4	4.61	120.23	112.30
4	A	1301	DGT	N9-C4-N3	4.08	134.12	125.95

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	1301	DGT	PB-O3A-PA-O5'
4	B	1301	DGT	PB-O3B-PG-O1G
4	A	1301	DGT	O4'-C4'-C5'-O5'
4	B	1301	DGT	O4'-C4'-C5'-O5'
4	B	1301	DGT	C3'-C4'-C5'-O5'

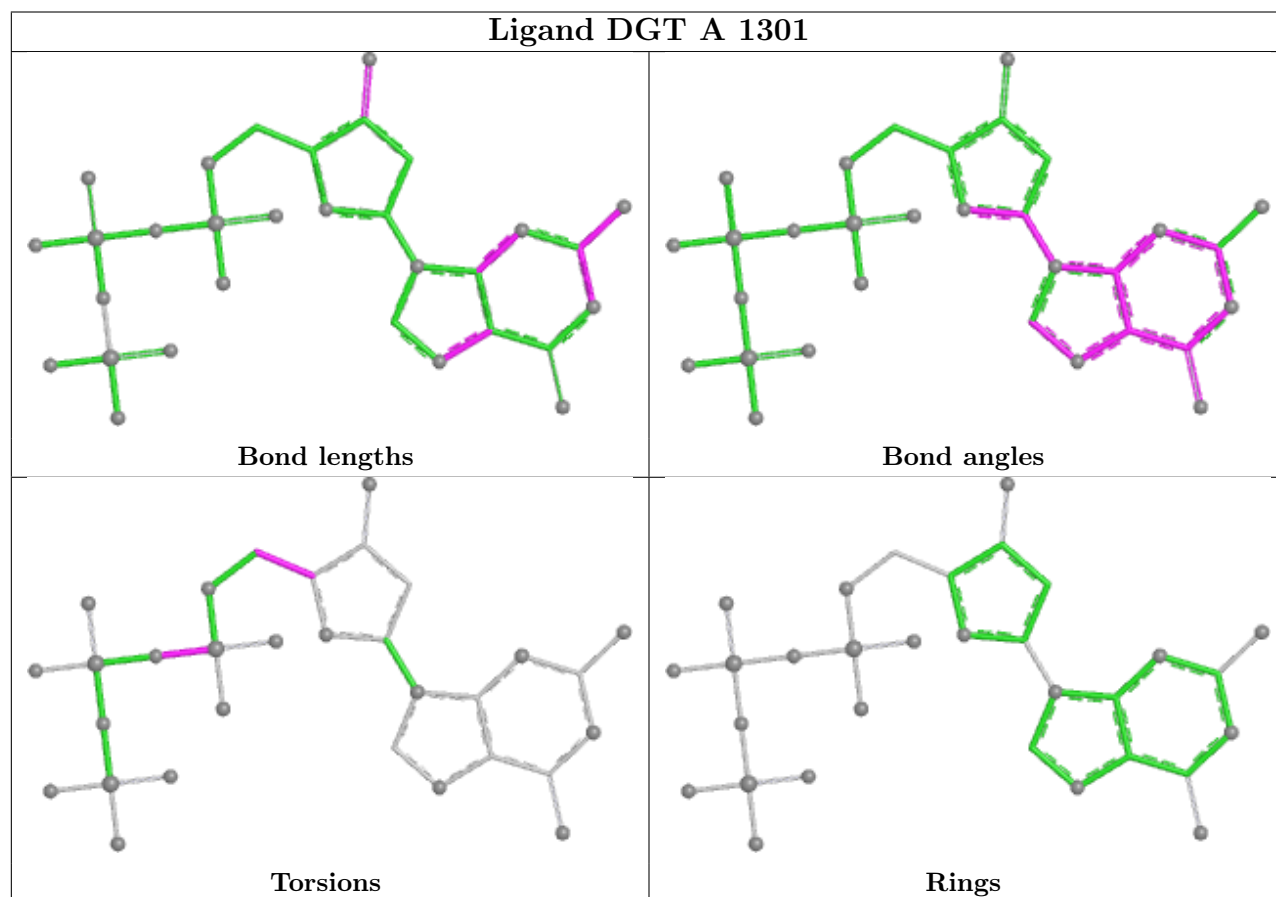
There are no ring outliers.

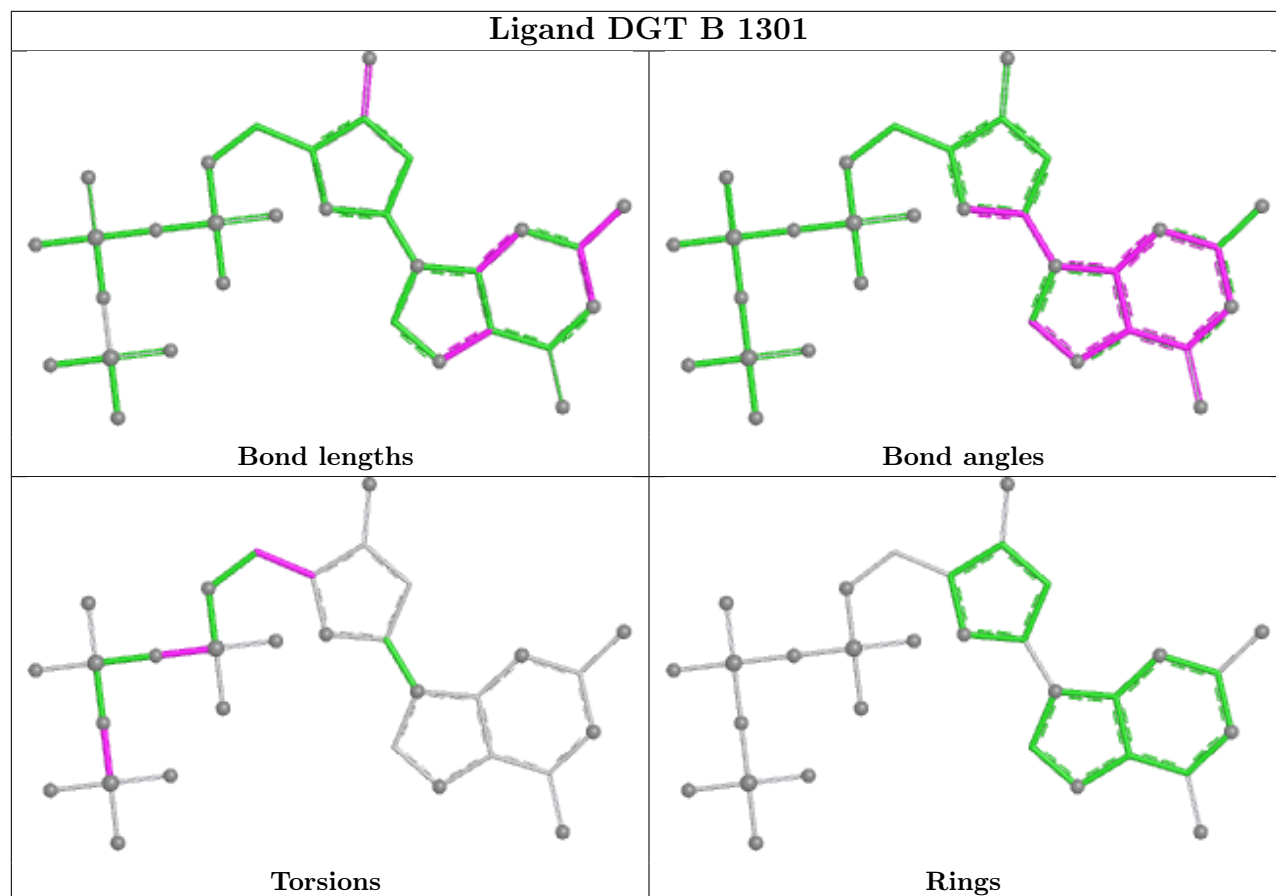
2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1301	DGT	2	0
4	B	1301	DGT	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 than 5% of the Mogul distribution of torsion angles is

within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.





5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 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	848/910 (93%)	0.02	10 (1%) 76 58	66, 113, 172, 209	0
1	B	848/910 (93%)	-0.00	13 (1%) 72 52	67, 118, 189, 226	0
2	C	16/25 (64%)	-0.04	0 100 100	76, 95, 184, 194	0
2	D	16/25 (64%)	0.07	0 100 100	87, 116, 175, 186	0
3	E	12/12 (100%)	-0.59	0 100 100	78, 91, 100, 101	0
3	F	12/12 (100%)	-0.06	0 100 100	88, 107, 124, 124	0
All	All	1752/1894 (92%)	0.01	23 (1%) 75 56	66, 114, 184, 226	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	610	VAL	4.9
1	A	676	GLY	4.1
1	A	1242	GLY	4.1
1	B	1193	ALA	4.0
1	B	676	GLY	3.7

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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

6.3 Carbohydrates [i](#)

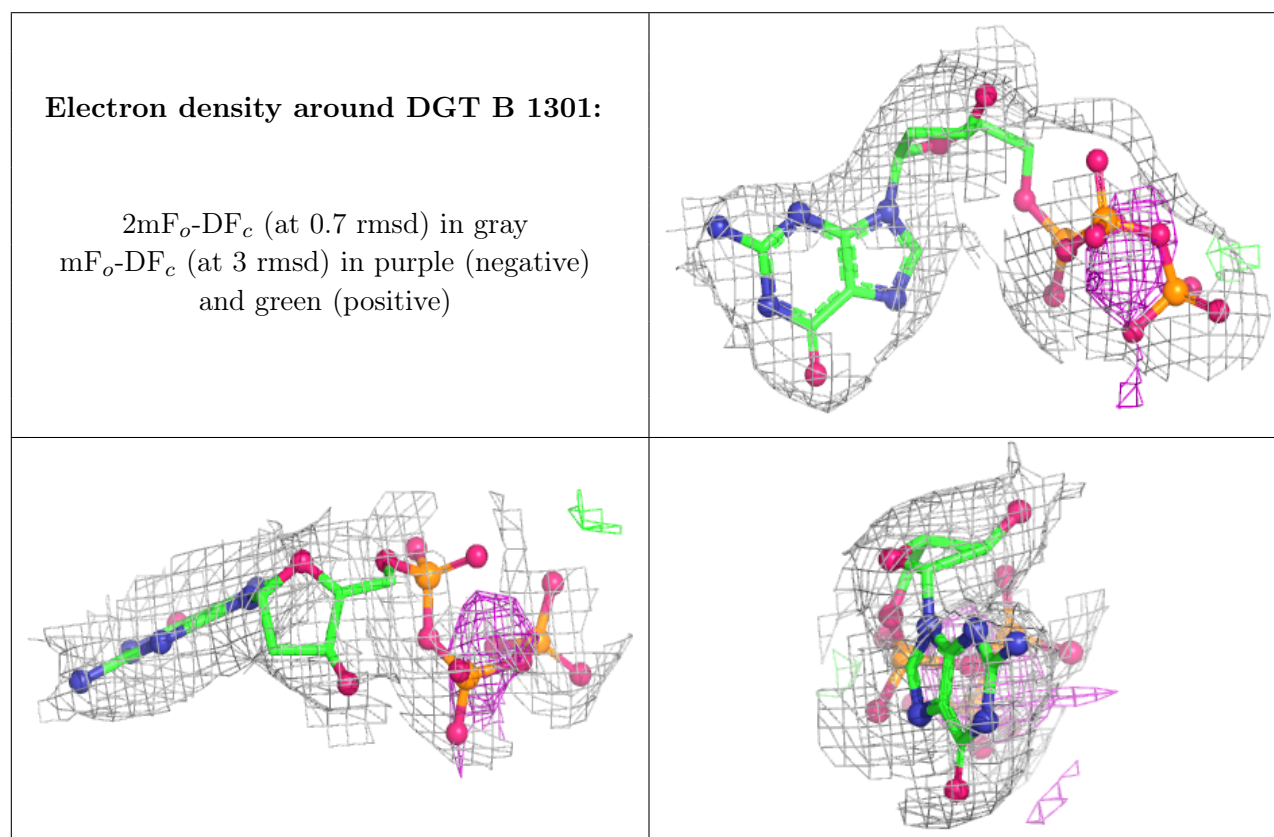
There are no oligosaccharides in this entry.

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.

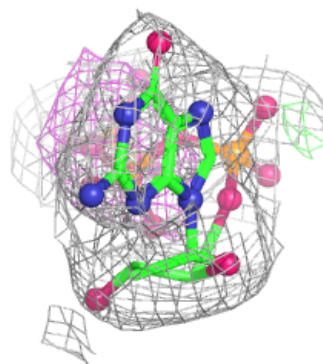
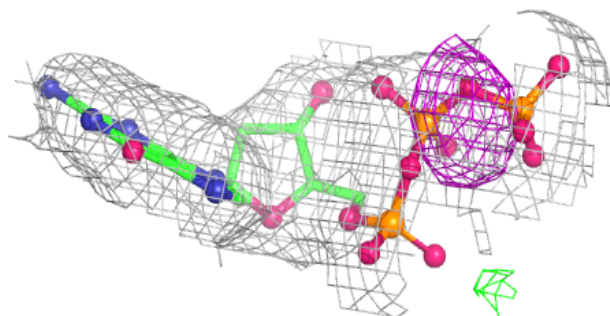
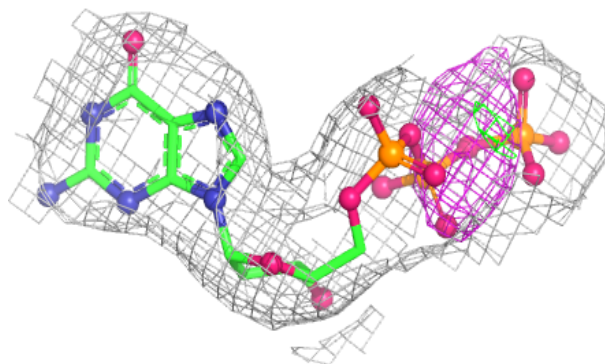
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($\text{\AA}^2$)	Q<0.9
4	DGT	B	1301	31/31	0.89	0.09	81,101,171,179	0
4	DGT	A	1301	31/31	0.91	0.10	68,85,155,163	0

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 DGT A 1301:

$2mF_o-DF_c$ (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.