



wwPDB X-ray Structure Validation Summary Report ⓘ

Apr 25, 2026 – 02:09 PM EDT

PDB ID : 5YTG / pdb_00005ytg
Title : Structure of large fragment of DNA Polymerase I from *Thermus aquaticus*
Host-Guest complex with the unnatural base I-fC pair with dA
Authors : Zeng, H.; Mondal, M.; Song, R.Y.; Zhang, J.; Xia, B.; Gao, Y.Q.; Yi, C.Q.
Deposited on : 2017-11-17
Resolution : 2.07 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

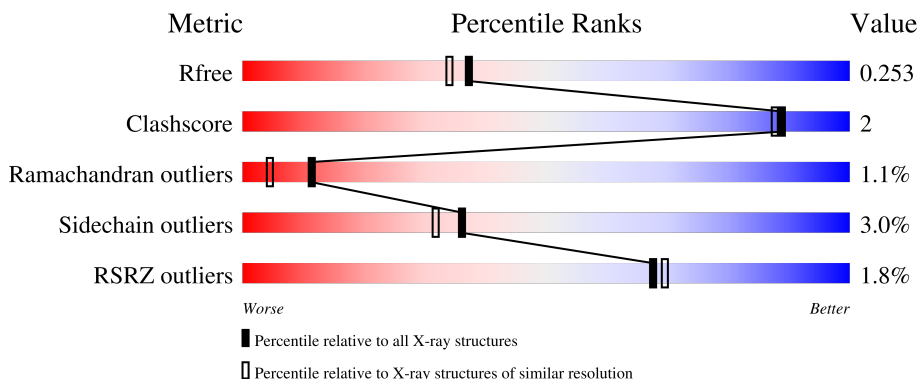
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.07 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	3774 (2.08-2.04)
Clashscore	190562	3883 (2.08-2.04)
Ramachandran outliers	187476	3860 (2.08-2.04)
Sidechain outliers	187428	3860 (2.08-2.04)
RSRZ outliers	180081	3775 (2.08-2.04)

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	539	19% 92% 7%
2	B	12	67% 25% 8%
3	C	16	94% 6%

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 5165 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, thermostable.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	539	Total	C	N	O	S	0	1	0
			4277	2718	770	776	13			

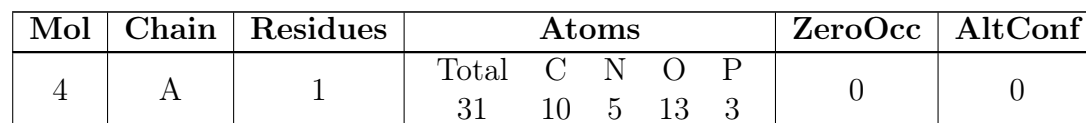
- Molecule 2 is a DNA chain called DNA (5'-D(*GP*AP*CP*CP*AP*CP*GP*GP*CP*GP*CP*(DOC))-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	12	Total	C	N	O	P	0	0	0
			240	114	48	67	11			

- Molecule 3 is a DNA chain called DNA (5'-D(*AP*AP*AP*CP*GP*GP*CP*GP*CP*CP*GP*(94O)P*GP*GP*TP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	16	Total	C	N	O	P	0	0	0
			337	164	65	93	15			

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



- | Mol | Chain | Residues | Atoms | ZeroOcc | AltConf |
|-----|-------|----------|-----------------|---------|---------|
| 5 | A | 2 | Total Mg
2 2 | 0 | 0 |
| 5 | B | 1 | Total Mg
1 1 | 0 | 0 |

- GOL
-
- The diagram shows the chemical structure of glycerol (1,2,3-propanetriol). It consists of a three-carbon chain. The first carbon (C1) is bonded to a hydroxyl group (HO) labeled O1. The second carbon (C2) is bonded to two hydroxyl groups, one labeled O2 and the other labeled O3. The third carbon (C3) is bonded to a hydroxyl group labeled O3. The atoms are labeled in green, and the bonds are in red.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	C	O	0	0
			6	3	3		

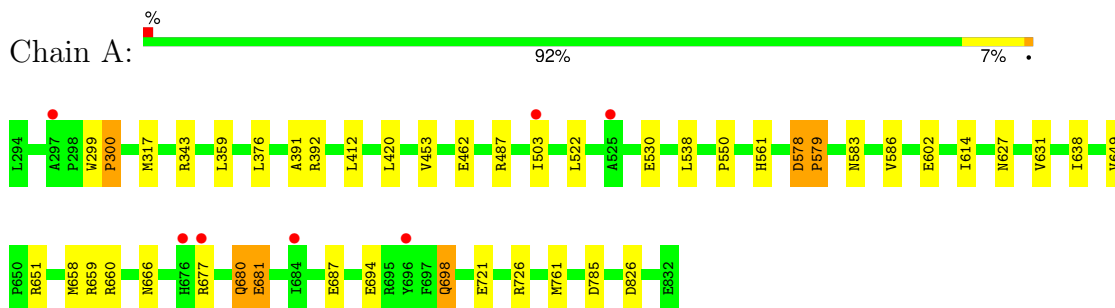
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	212	Total	O	0	0
			212	212		
7	B	24	Total	O	0	0
			24	24		
7	C	35	Total	O	0	0
			35	35		

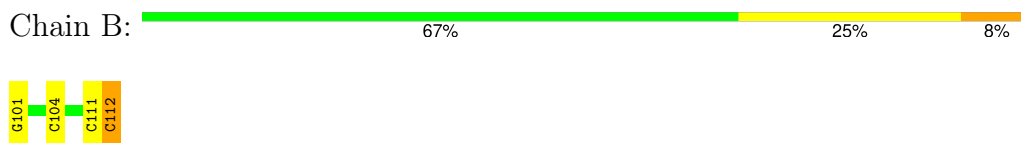
3 Residue-property plots [i](#)

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.

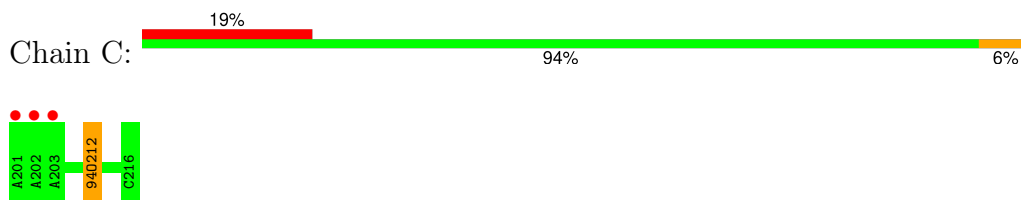
- Molecule 1: DNA polymerase I, thermostable



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



- Molecule 3: DNA (5'-D(*AP*AP*AP*CP*GP*GP*CP*GP*CP*CP*GP*(94O)P*GP*GP*TP*C)-3')



4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α , β , γ	109.06Å 109.06Å 90.94Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	94.45 – 2.07 94.45 – 2.07	Depositor EDS
% Data completeness (in resolution range)	99.7 (94.45-2.07) 99.7 (94.45-2.07)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.73 (at 2.06Å)	Xtriage
Refinement program	REFMAC 5.8.0103	Depositor
R, R_{free}	0.198 , 0.252 0.203 , 0.253	Depositor DCC
R_{free} test set	1860 reflections (4.84%)	wwPDB-VP
Wilson B-factor (Å ²)	34.9	Xtriage
Anisotropy	0.407	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 33.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.026 for -h,-k,l	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	5165	wwPDB-VP
Average B, all atoms (Å ²)	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.26% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: DGT, GOL, 94O, MG, DOC

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	0.67	0/4370	0.89	4/5921 (0.1%)
2	B	0.50	0/249	0.97	2/382 (0.5%)
3	C	0.47	0/344	0.80	0/527
All	All	0.65	0/4963	0.89	6/6830 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	4

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	101	DG	C5'-C4'-O4'	7.08	120.02	109.40
1	A	300	PRO	N-CA-C	-6.67	102.57	110.70
1	A	614	ILE	N-CA-C	6.05	116.22	110.42
1	A	579	PRO	N-CA-C	-5.81	100.49	112.47
2	B	101	DG	C5'-C4'-C3'	5.38	122.97	114.90

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	299	TRP	Peptide
1	A	578	ASP	Mainchain,Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	A	698	GLN	Peptide

5.2 Too-close contacts ⓘ

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

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	4277	0	4322	16	0
2	B	240	0	134	2	0
3	C	337	0	168	1	0
4	A	31	0	12	1	0
5	A	2	0	0	0	0
5	B	1	0	0	0	0
6	A	6	0	8	0	0
7	A	212	0	0	1	2
7	B	24	0	0	0	0
7	C	35	0	0	0	0
All	All	5165	0	4644	18	2

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 2.

The worst 5 of 18 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:677:ARG:O	1:A:680:GLN:O	2.05	0.75
1:A:680:GLN:O	1:A:681:GLU:HB2	1.89	0.70
1:A:659:ARG:NH1	4:A:901:DGT:O1G	2.35	0.59
2:B:111:DC:H2'	2:B:112:DOC:H6	1.85	0.58
1:A:631:VAL:HG11	1:A:638:ILE:HD11	1.90	0.53

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

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
7:A:1198:HOH:O	7:A:1198:HOH:O[5_554]	1.08	1.12

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
7:A:1211:HOH:O	7:A:1211:HOH:O[6_554]	1.71	0.49

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	538/539 (100%)	511 (95%)	21 (4%)	6 (1%)	11 4

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	300	PRO
1	A	578	ASP
1	A	681	GLU
1	A	698	GLN
1	A	579	PRO

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	439/441 (100%)	426 (97%)	13 (3%)	36 32

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	602	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	651	ARG
1	A	826	ASP
1	A	687	GLU
1	A	694	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	384	ASN
1	A	583	ASN
1	A	633	GLN
1	A	666	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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
3	94O	C	212	2,3	29,34,35	2.09	5 (17%)	40,51,54	2.57	9 (22%)
2	DOC	B	112	2,3	16,19,20	1.34	4 (25%)	20,26,29	1.20	2 (10%)

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
3	94O	C	212	2,3	-	2/7/33/34	0/5/5/5
2	DOC	B	112	2,3	-	0/7/18/19	0/2/2/2

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	212	94O	C21-C22	6.23	1.49	1.40
3	C	212	94O	C4-C5	5.39	1.50	1.41
3	C	212	94O	C25-C26	5.08	1.48	1.41
3	C	212	94O	O31-C24	3.17	1.28	1.22
2	B	112	DOC	C6-N1	-2.50	1.32	1.38

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	212	94O	C21-C24-C25	8.24	113.37	105.31
3	C	212	94O	C26-C25-C24	-6.99	103.34	108.94
3	C	212	94O	C21-C22-N23	-6.44	121.10	125.48
3	C	212	94O	C22-C21-C24	-4.28	103.69	108.57
3	C	212	94O	O2-C2-N3	-4.03	115.97	122.33

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	C	212	94O	O4'-C4'-C5'-O5'
3	C	212	94O	C3'-C4'-C5'-O5'

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	212	94O	1	0
2	B	112	DOC	1	0

5.5 Carbohydrates

There are no oligosaccharides in this entry.

5.6 Ligand geometry

Of 5 ligands modelled in this entry, 3 are monoatomic - leaving 2 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
6	GOL	A	904	-	5,5,5	0.22	0	5,5,5	0.46	0
4	DGT	A	901	5	32,33,33	1.16	3 (9%)	48,52,52	1.83	12 (25%)

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
6	GOL	A	904	-	-	1/4/4/4	-
4	DGT	A	901	5	-	3/22/34/34	0/3/3/3

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	901	DGT	PB-O3A	2.65	1.62	1.59
4	A	901	DGT	C5-C4	2.44	1.45	1.38
4	A	901	DGT	C5-N7	-2.28	1.34	1.39

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	901	DGT	C5-C4-N3	-4.94	120.52	128.39
4	A	901	DGT	C2-N3-C4	4.59	120.20	112.30
4	A	901	DGT	C2'-C1'-N9	-4.09	103.59	113.81
4	A	901	DGT	N9-C4-N3	3.70	133.36	125.95
4	A	901	DGT	O6-C6-C5	-3.37	117.64	126.53

There are no chirality outliers.

All (4) torsion outliers are listed below:

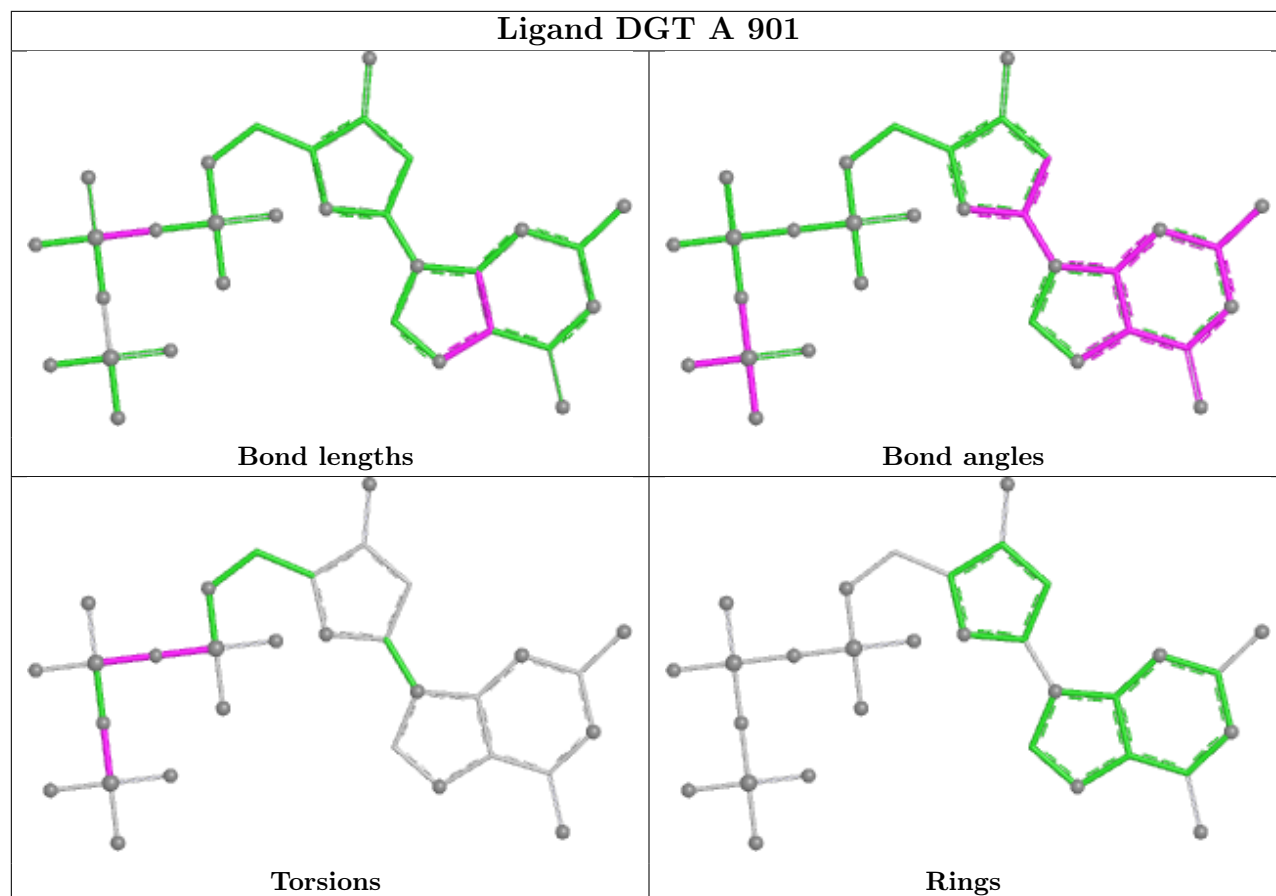
Mol	Chain	Res	Type	Atoms
4	A	901	DGT	PB-O3B-PG-O1G
4	A	901	DGT	PA-O3A-PB-O1B
6	A	904	GOL	C1-C2-C3-O3
4	A	901	DGT	PB-O3A-PA-O1A

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	901	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 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	539/539 (100%)	0.13	7 (1%) 75 77	21, 46, 75, 89	1 (0%)
2	B	11/12 (91%)	-0.23	0 100 100	32, 36, 62, 64	0
3	C	15/16 (93%)	0.18	3 (20%) 3 2	29, 41, 106, 122	1 (6%)
All	All	565/567 (99%)	0.12	10 (1%) 67 69	21, 46, 77, 122	2 (0%)

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	202	DA	3.5
3	C	201	DA	2.8
1	A	677	ARG	2.7
3	C	203	DA	2.5
1	A	503	ILE	2.3

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
3	94O	C	212	30/31	0.87	0.12	54,59,64,68	0
2	DOC	B	112	18/19	0.95	0.07	31,33,37,37	0

6.3 Carbohydrates [i](#)

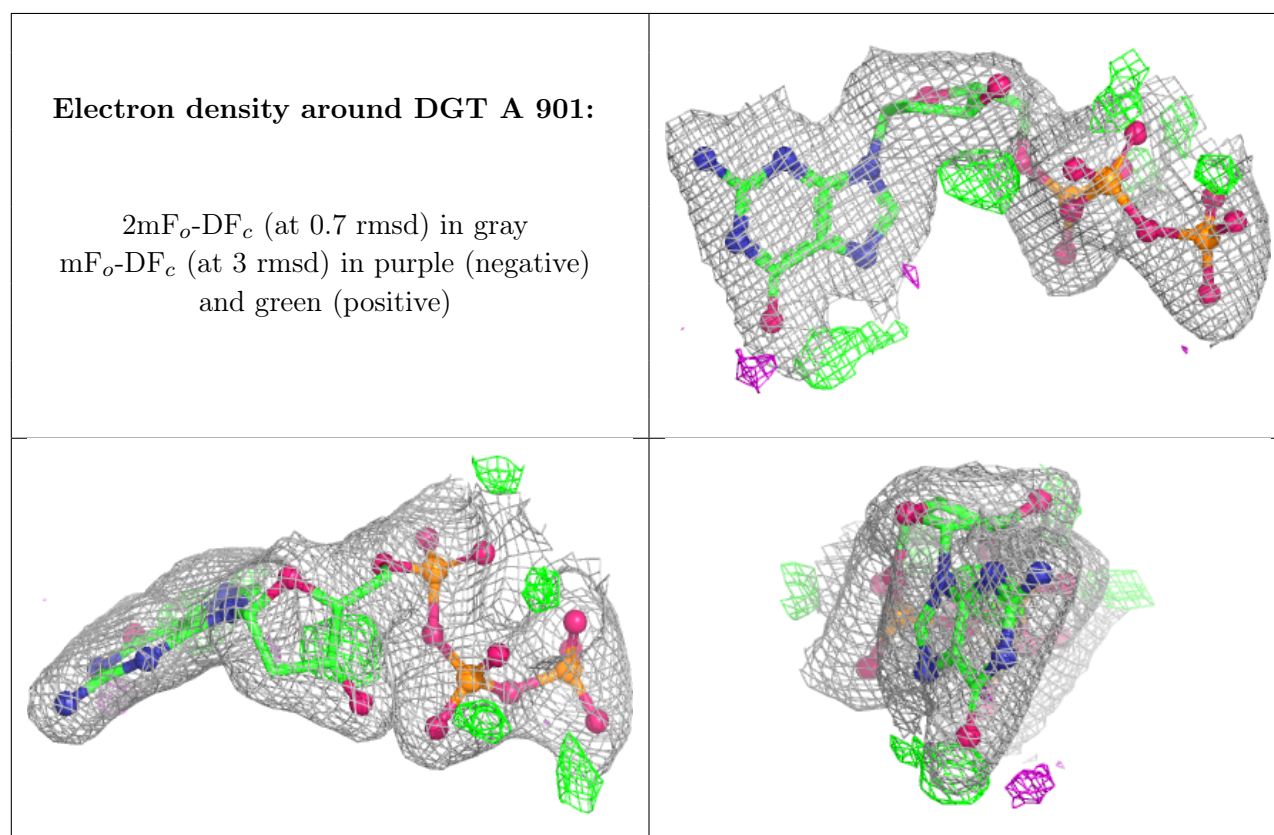
There are no oligosaccharides in this entry.

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(Å ²)	Q<0.9
5	MG	B	201	1/1	0.48	0.20	81,81,81,81	0
5	MG	A	903	1/1	0.74	0.17	69,69,69,69	0
6	GOL	A	904	6/6	0.81	0.12	63,67,72,76	0
5	MG	A	902	1/1	0.94	0.10	71,71,71,71	0
4	DGT	A	901	31/31	0.95	0.07	32,37,43,43	0

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



6.5 Other polymers [i](#)

There are no such residues in this entry.