



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 9, 2020 – 05:43 AM BST

PDB ID : 6NKW
Title : Ternary complex crystal structure of DNA polymerase Beta with "hot-spot sequence" with beta-gamma-methylene dGTP
Authors : Batra, V.K.; Wilson, S.H.
Deposited on : 2019-01-07
Resolution : 1.98 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 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 : 2.13.1
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) : 2.13.1

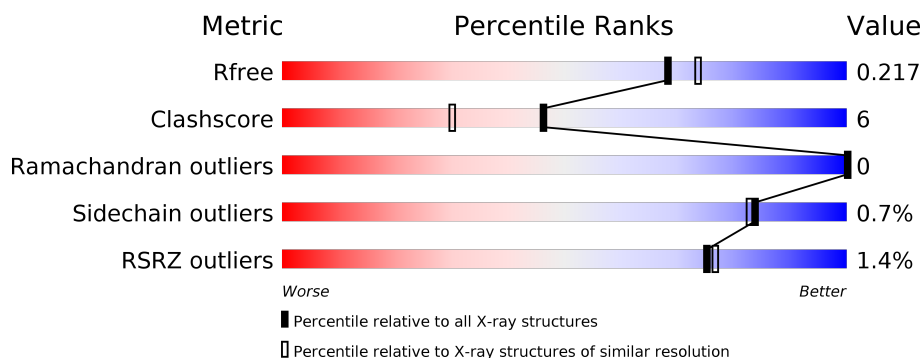
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.98 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	11647 (2.00-1.96)
Clashscore	141614	1014 (1.98-1.98)
Ramachandran outliers	138981	1006 (1.98-1.98)
Sidechain outliers	138945	1006 (1.98-1.98)
RSRZ outliers	127900	11410 (2.00-1.96)

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 on 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	T	16	<div> <div></div> <div>75%25%</div> </div>
2	P	10	<div> <div></div> <div>50%40%10%</div> </div>
3	D	5	<div> <div></div> <div>80%20%</div> </div>
4	A	335	<div> <div></div> <div>%89%8%.</div> </div>

2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 3727 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 DNA chain called DNA (5'-D(*CP*CP*GP*AP*AP*CP*AP*AP*GP*CP*AP*TP*CP*AP*GP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	T	16	Total	C	N	O	P	0	0	0
			323	154	65	89	15			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	P	10	Total	C	N	O	P	0	0	0
			201	98	34	60	9			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	5	Total	C	N	O	P	0	0	0
			104	49	17	33	5			

- Molecule 4 is a protein called DNA polymerase beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	A	326	Total	C	N	O	S	0	3	0
			2632	1664	460	498	10			

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total	Mg	0	0
			1	1		
5	T	1	Total	Mg	0	0
			1	1		

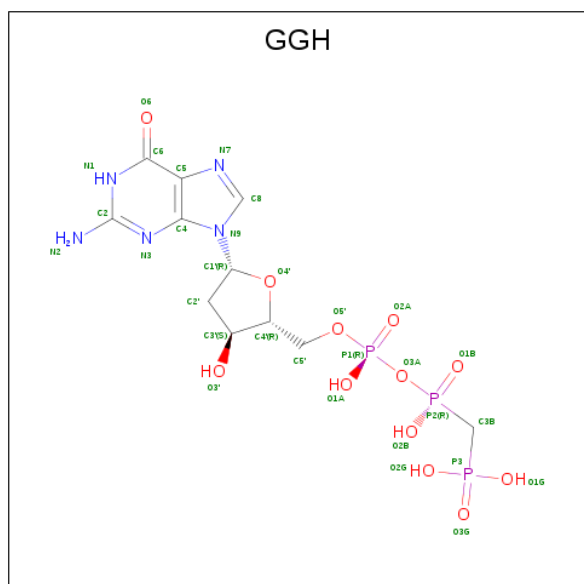
- Molecule 6 is SODIUM ION (three-letter code: NA) (formula: Na).

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

- Molecule 7 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	4	Total Cl 4 4	0	0

- Molecule 8 is 2'-DEOXY-5'-O-(HYDROXY{[HYDROXY(PHOSPHONOMETHYL)PHOSPHORYL]OXY}PHOSPHORYL)GUANOSINE (three-letter code: GGH) (formula: $C_{11}H_{18}N_5O_{12}P_3$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
8	A	1	Total	C	N	O	P	0	0
			31	11	5	12	3		

- Molecule 9 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
9	T	46	Total O 46 46	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	P	26	Total 26	O 26	0	0
9	D	19	Total 19	O 19	0	0
9	A	336	Total 336	O 336	0	0

- Molecule 1: DNA (5'-D(*CP*CP*GP*AP*AP*CP*AP*AP*GP*CP*AP*TP*CP*AP*GP*C)-3')

Diagram illustrating a sequence of nodes: C1, C6, C13, A14, G15, C16. The nodes are arranged vertically, with C1 and C6 highlighted in yellow, C13 and A14 in yellow, and G15 and C16 in green. Green bars connect C1 to C6 and C6 to C13.

A diagram showing a yellow box labeled T1 connected to a green box labeled G5.

Amino Acid	Number of Mutations (approx.)
MET	100
SER	10
LYS	10
ARG	10
LYS	10
ALA	10
PRO	10
GLN	10
GLU	10
T10	10
E86	10
Q90	10
A111	10
F114	10
V115	10
I119	10
K120	10
E123	10
K127	10
H128	10
E129	10
D130	10
K131	10
Q157	10
E172	10
C178	10
R182	10
M191	10
E216	10
L228	10
M236	10
K244	10
E247	10
I255	10
F272	10
K280	10
G302	10
V303	10
T304	10

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	50.44Å 79.19Å 55.35Å 90.00° 107.60° 90.00°	Depositor
Resolution (Å)	31.16 – 1.98 31.67 – 1.98	Depositor EDS
% Data completeness (in resolution range)	96.8 (31.16-1.98) 88.8 (31.67-1.98)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.62 (at 1.98Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.174 , 0.217 0.174 , 0.217	Depositor DCC
R_{free} test set	2000 reflections (7.09%)	wwPDB-VP
Wilson B-factor (Å ²)	22.8	Xtriage
Anisotropy	0.234	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 38.9	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	0.95	EDS
Total number of atoms	3727	wwPDB-VP
Average B, all atoms (Å ²)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.21% 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: GGH, 2DT, MG, NA, CL

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	T	0.73	0/363	0.90	0/557
2	P	0.77	0/203	0.97	0/312
3	D	1.18	1/115 (0.9%)	0.99	0/174
4	A	0.41	1/2690 (0.0%)	0.56	2/3614 (0.1%)
All	All	0.52	2/3371 (0.1%)	0.66	2/4657 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	1	DT	OP3-P	-9.54	1.49	1.61
4	A	120	LYS	CD-CE	-6.57	1.34	1.51

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	120	LYS	CB-CG-CD	-7.05	93.27	111.60
4	A	120	LYS	CG-CD-CE	5.32	127.87	111.90

There are no chirality outliers.

There are no planarity outliers.

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	T	323	0	179	3	0
2	P	201	0	116	5	0
3	D	104	0	58	0	0
4	A	2632	0	2661	32	0
5	A	1	0	0	0	0
5	T	1	0	0	0	0
6	A	1	0	0	0	0
6	D	1	0	0	0	0
6	P	1	0	0	0	0
7	A	4	0	0	0	0
8	A	31	0	14	0	0
9	A	336	0	0	9	1
9	D	19	0	0	0	0
9	P	26	0	0	3	0
9	T	46	0	0	2	0
All	All	3727	0	3028	40	1

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 6.

The worst 5 of 40 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:115:VAL:HA	4:A:120:LYS:CE	1.83	1.07
4:A:115:VAL:HG22	4:A:120:LYS:HE3	1.54	0.89
1:T:6:DC:OP1	9:T:201:HOH:O	1.90	0.88
4:A:119:ILE:O	4:A:120:LYS:NZ	2.08	0.86
2:P:2:DC:OP1	9:P:201:HOH:O	1.93	0.85

All (1) 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)
9:A:766:HOH:O	9:A:792:HOH:O[1_656]	2.06	0.14

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	
4	A	327/335 (98%)	321 (98%)	6 (2%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains [i](#)

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

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

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
4	A	290/295 (98%)	288 (99%)	2 (1%)	84	83

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	A	272	PHE
4	A	314	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

1 non-standard protein/DNA/RNA residue is 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	2DT	P	10	1,2	14,20,21	1.12	1 (7%)	12,28,31	4.32	1 (8%)

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	2DT	P	10	1,2	-	0/4/18/19	0/2/2/2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	P	10	2DT	C4-N3	3.04	1.38	1.33

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	P	10	2DT	C4-N3-C2	14.69	127.55	115.14

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	P	10	2DT	1	0

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 10 ligands modelled in this entry, 9 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
8	GGH	A	407	5,6	27,33,33	3.58	11 (40%)	33,52,52	1.68	7 (21%)

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
8	GGH	A	407	5,6	-	3/15/34/34	0/3/3/3

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
8	A	407	GGH	P2-O3A	8.43	1.67	1.58
8	A	407	GGH	C2'-C3'	-7.71	1.32	1.52
8	A	407	GGH	O4'-C4'	-7.42	1.28	1.45
8	A	407	GGH	O6-C6	6.18	1.40	1.24
8	A	407	GGH	O4'-C1'	5.73	1.55	1.42

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
8	A	407	GGH	C2-N3-C4	4.61	120.62	115.36
8	A	407	GGH	N3-C2-N1	-3.85	122.08	127.22
8	A	407	GGH	C6-C5-C4	-2.95	117.98	120.80
8	A	407	GGH	C5-C6-N1	-2.89	119.48	123.43

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
8	A	407	GGH	C6-N1-C2	2.79	120.36	115.93

There are no chirality outliers.

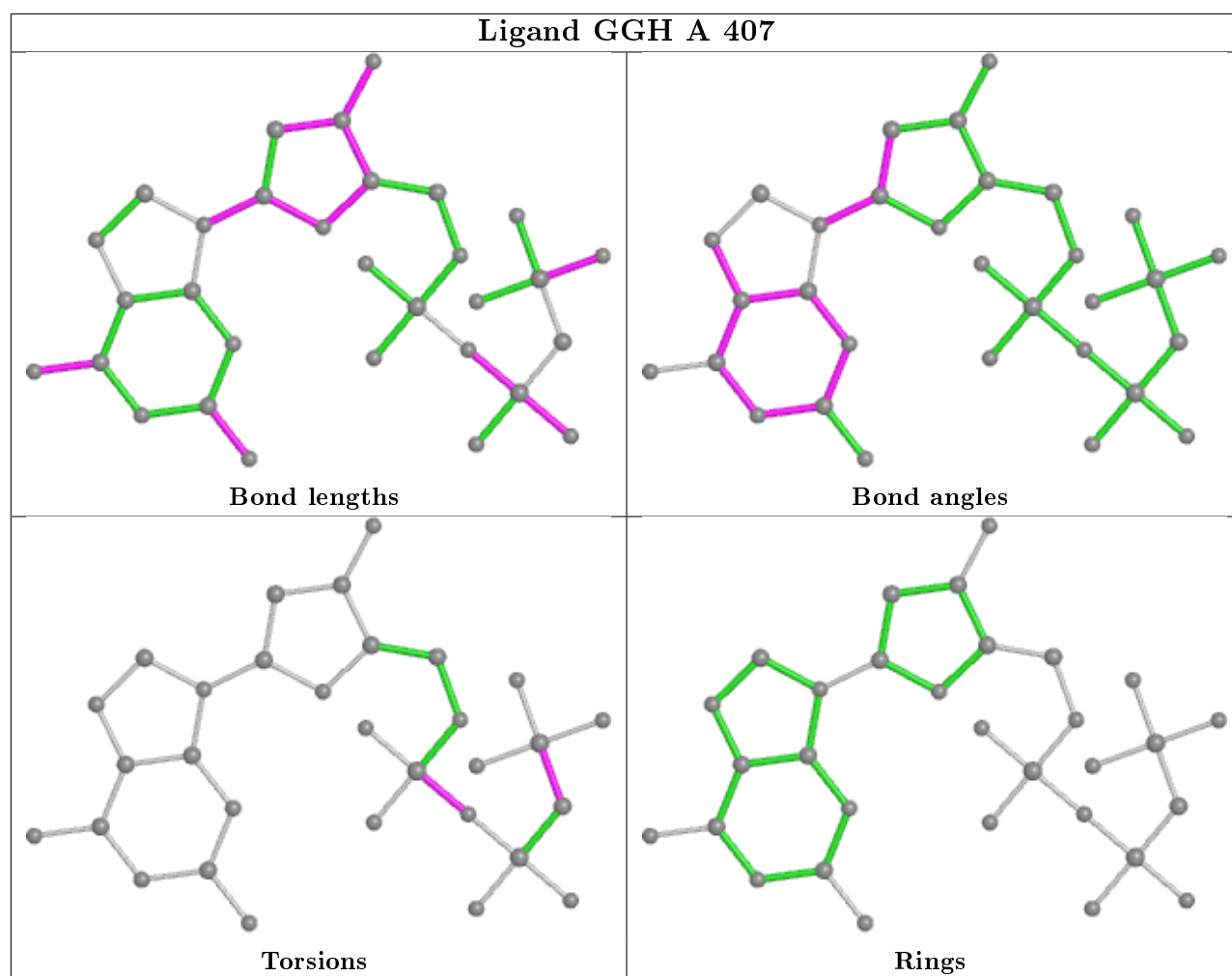
All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
8	A	407	GGH	P2-C3B-P3-O3G
8	A	407	GGH	P2-C3B-P3-O2G
8	A	407	GGH	P2-O3A-P1-O2A

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 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(Å²)	Q<0.9	
1	T	16/16 (100%)	-0.38	0	100	100	21, 32, 45, 45	0
2	P	9/10 (90%)	-0.21	0	100	100	21, 33, 40, 41	0
3	D	5/5 (100%)	-0.42	0	100	100	25, 26, 40, 42	0
4	A	326/335 (97%)	-0.29	5 (1%)	73	75	16, 24, 39, 50	0
All	All	356/366 (97%)	-0.29	5 (1%)	75	77	16, 25, 41, 50	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	A	303	VAL	4.0
4	A	120	LYS	3.4
4	A	302	GLY	2.8
4	A	305	GLY	2.5
4	A	304	THR	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
2	2DT	P	10	19/20	0.98	0.11	20,23,26,26	0

6.3 Carbohydrates [i](#)

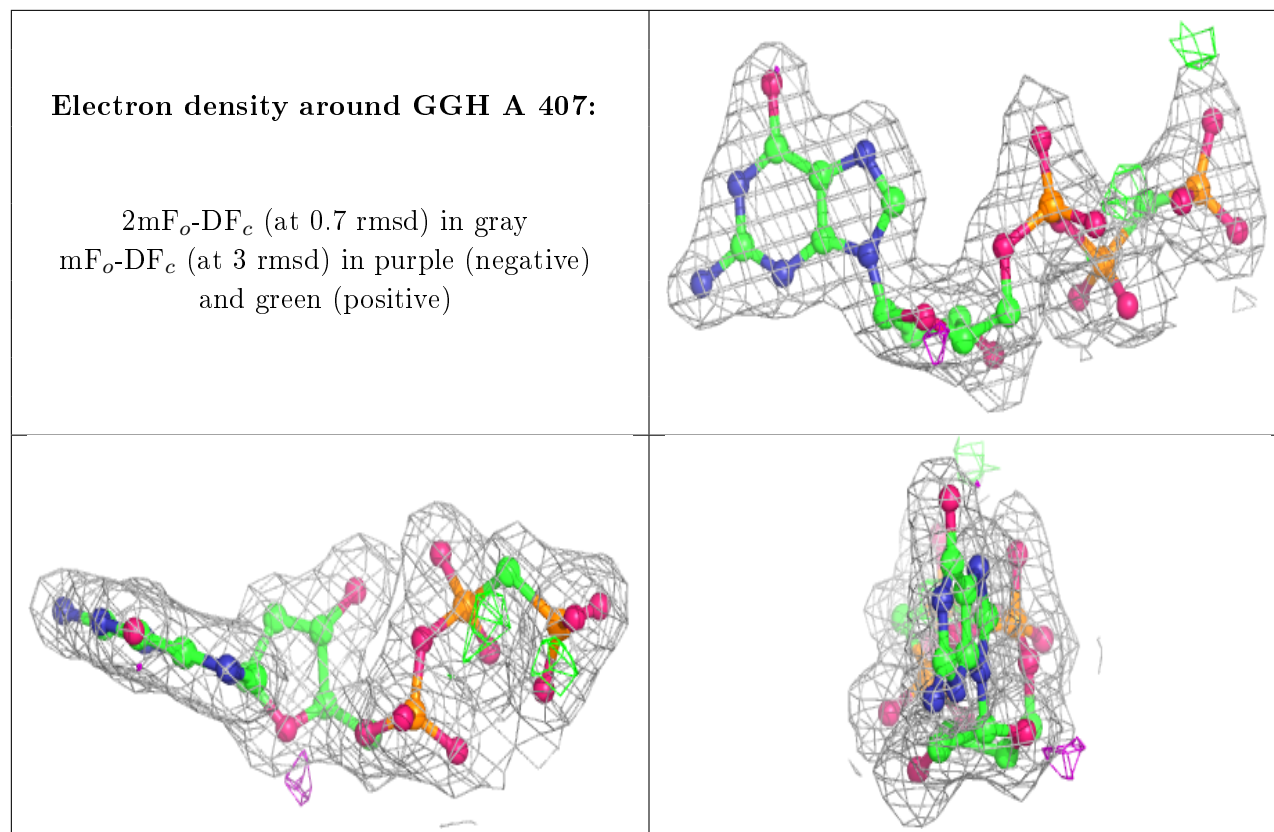
There are no monosaccharides in this entry.

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(\AA^2)	Q<0.9
7	CL	A	405	1/1	0.89	0.06	42,42,42,42	0
7	CL	A	403	1/1	0.93	0.10	37,37,37,37	0
5	MG	T	101	1/1	0.95	0.08	35,35,35,35	0
6	NA	A	402	1/1	0.97	0.08	24,24,24,24	0
7	CL	A	404	1/1	0.98	0.05	31,31,31,31	0
6	NA	D	101	1/1	0.98	0.05	30,30,30,30	0
5	MG	A	401	1/1	0.98	0.05	16,16,16,16	0
8	GGH	A	407	31/31	0.98	0.10	15,18,22,24	0
6	NA	P	101	1/1	0.99	0.14	22,22,22,22	0
7	CL	A	406	1/1	1.00	0.02	30,30,30,30	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.