



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 13, 2024 – 04:45 PM EDT

PDB ID : 8GD7
Title : Loop Deleted DNA Polymerase Theta Polymerase Domain in Complex with
Double Strand DNA Overhang and Inhibitor
Authors : Fried, W.A.; Chen, X.S.; Li, S.X.
Deposited on : 2023-03-03
Resolution : 3.24 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 : 4.02b-467
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.36.2
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) : 2.36.2

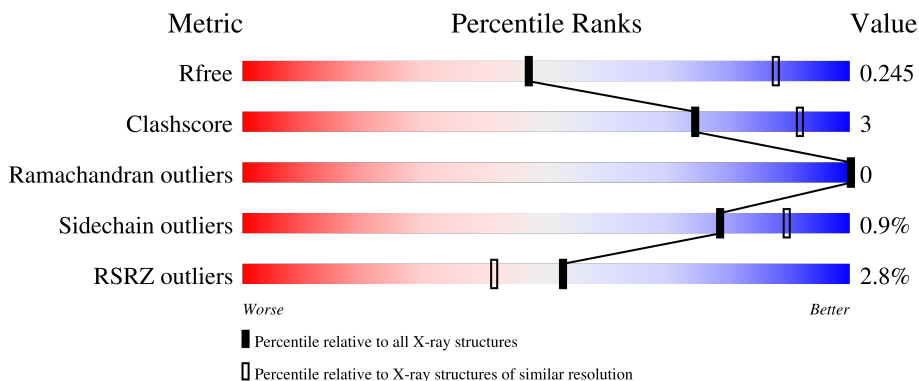
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.24 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	1619 (3.28-3.20)
Clashscore	141614	1755 (3.28-3.20)
Ramachandran outliers	138981	1728 (3.28-3.20)
Sidechain outliers	138945	1727 (3.28-3.20)
RSRZ outliers	127900	1567 (3.28-3.20)

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	652	 2% 90% 7%
2	E	24	 8% 38% 17% 46%
3	F	14	 36% 36% 29%

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 5514 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 theta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	631	4983	3167	854	933	29	0	0	0

There are 141 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	SER	deletion	UNP O75417
A	?	-	LEU	deletion	UNP O75417
A	?	-	THR	deletion	UNP O75417
A	?	-	SER	deletion	UNP O75417
A	?	-	SER	deletion	UNP O75417
A	?	-	LYS	deletion	UNP O75417
A	?	-	THR	deletion	UNP O75417
A	?	-	ALA	deletion	UNP O75417
A	?	-	THR	deletion	UNP O75417
A	?	-	ILE	deletion	UNP O75417
A	?	-	GLY	deletion	UNP O75417
A	?	-	SER	deletion	UNP O75417
A	?	-	ARG	deletion	UNP O75417
A	?	-	PHE	deletion	UNP O75417
A	?	-	LYS	deletion	UNP O75417
A	?	-	GLN	deletion	UNP O75417
A	?	-	ALA	deletion	UNP O75417
A	?	-	SER	deletion	UNP O75417
A	?	-	SER	deletion	UNP O75417
A	?	-	PRO	deletion	UNP O75417
A	?	-	GLN	deletion	UNP O75417
A	?	-	GLU	deletion	UNP O75417
A	?	-	ILE	deletion	UNP O75417
A	?	-	PRO	deletion	UNP O75417
A	?	-	ILE	deletion	UNP O75417
A	?	-	ARG	deletion	UNP O75417
A	?	-	ASP	deletion	UNP O75417

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	ASP	deletion	UNP O75417
A	?	-	GLY	deletion	UNP O75417
A	?	-	PHE	deletion	UNP O75417
A	?	-	PRO	deletion	UNP O75417
A	?	-	ILE	deletion	UNP O75417
A	1893	GLY	LYS	conflict	UNP O75417
A	1894	SER	GLY	conflict	UNP O75417
A	1895	GLY	CYS	conflict	UNP O75417
A	?	-	GLN	deletion	UNP O75417
A	?	-	LYS	deletion	UNP O75417
A	?	-	GLU	deletion	UNP O75417
A	?	-	GLN	deletion	UNP O75417
A	?	-	LYS	deletion	UNP O75417
A	?	-	HIS	deletion	UNP O75417
A	?	-	SER	deletion	UNP O75417
A	?	-	GLU	deletion	UNP O75417
A	?	-	ILE	deletion	UNP O75417
A	?	-	SER	deletion	UNP O75417
A	?	-	ALA	deletion	UNP O75417
A	?	-	SER	deletion	UNP O75417
A	1930	GLY	LEU	conflict	UNP O75417
A	1931	GLY	VAL	conflict	UNP O75417
A	1932	SER	PRO	conflict	UNP O75417
A	1933	GLY	PRO	conflict	UNP O75417
A	1934	GLY	SER	conflict	UNP O75417
A	?	-	ASN	deletion	UNP O75417
A	?	-	ARG	deletion	UNP O75417
A	?	-	GLU	deletion	UNP O75417
A	?	-	MET	deletion	UNP O75417
A	?	-	LYS	deletion	UNP O75417
A	?	-	ASN	deletion	UNP O75417
A	?	-	GLN	deletion	UNP O75417
A	?	-	GLY	deletion	UNP O75417
A	?	-	SER	deletion	UNP O75417
A	?	-	LYS	deletion	UNP O75417
A	?	-	LYS	deletion	UNP O75417
A	?	-	THR	deletion	UNP O75417
A	?	-	LEU	deletion	UNP O75417
A	?	-	GLY	deletion	UNP O75417
A	?	-	SER	deletion	UNP O75417
A	?	-	THR	deletion	UNP O75417
A	?	-	ARG	deletion	UNP O75417

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	ARG	deletion	UNP O75417
A	?	-	GLY	deletion	UNP O75417
A	?	-	ILE	deletion	UNP O75417
A	?	-	ASP	deletion	UNP O75417
A	?	-	ASN	deletion	UNP O75417
A	?	-	GLY	deletion	UNP O75417
A	?	-	ARG	deletion	UNP O75417
A	?	-	LYS	deletion	UNP O75417
A	2171	GLY	LEU	conflict	UNP O75417
A	2172	GLY	ARG	conflict	UNP O75417
A	2173	SER	LEU	conflict	UNP O75417
A	2175	GLY	ARG	conflict	UNP O75417
A	?	-	PRO	deletion	UNP O75417
A	?	-	THR	deletion	UNP O75417
A	?	-	LEU	deletion	UNP O75417
A	?	-	VAL	deletion	UNP O75417
A	?	-	GLY	deletion	UNP O75417
A	?	-	GLU	deletion	UNP O75417
A	?	-	SER	deletion	UNP O75417
A	?	-	PRO	deletion	UNP O75417
A	?	-	PRO	deletion	UNP O75417
A	?	-	SER	deletion	UNP O75417
A	?	-	GLN	deletion	UNP O75417
A	?	-	ALA	deletion	UNP O75417
A	?	-	VAL	deletion	UNP O75417
A	?	-	GLY	deletion	UNP O75417
A	?	-	LYS	deletion	UNP O75417
A	?	-	GLY	deletion	UNP O75417
A	?	-	LEU	deletion	UNP O75417
A	?	-	LEU	deletion	UNP O75417
A	?	-	PRO	deletion	UNP O75417
A	?	-	MET	deletion	UNP O75417
A	?	-	GLY	deletion	UNP O75417
A	?	-	ARG	deletion	UNP O75417
A	?	-	GLY	deletion	UNP O75417
A	?	-	LYS	deletion	UNP O75417
A	?	-	TYR	deletion	UNP O75417
A	?	-	LYS	deletion	UNP O75417
A	?	-	LYS	deletion	UNP O75417
A	?	-	GLY	deletion	UNP O75417
A	?	-	PHE	deletion	UNP O75417
A	?	-	SER	deletion	UNP O75417

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	VAL	deletion	UNP O75417
A	?	-	ASN	deletion	UNP O75417
A	?	-	PRO	deletion	UNP O75417
A	?	-	ARG	deletion	UNP O75417
A	?	-	CYS	deletion	UNP O75417
A	?	-	GLN	deletion	UNP O75417
A	?	-	ALA	deletion	UNP O75417
A	?	-	GLN	deletion	UNP O75417
A	?	-	MET	deletion	UNP O75417
A	?	-	GLU	deletion	UNP O75417
A	?	-	GLU	deletion	UNP O75417
A	?	-	ARG	deletion	UNP O75417
A	2303	GLY	ALA	conflict	UNP O75417
A	2304	GLY	ALA	conflict	UNP O75417
A	2305	SER	ASP	conflict	UNP O75417
A	2306	GLY	ARG	conflict	UNP O75417
A	?	-	GLN	deletion	UNP O75417
A	?	-	THR	deletion	UNP O75417
A	?	-	GLY	deletion	UNP O75417
A	?	-	LEU	deletion	UNP O75417
A	?	-	SER	deletion	UNP O75417
A	?	-	ARG	deletion	UNP O75417
A	?	-	LYS	deletion	UNP O75417
A	?	-	ARG	deletion	UNP O75417
A	?	-	LYS	deletion	UNP O75417
A	2522	GLY	LEU	conflict	UNP O75417
A	2523	GLY	GLN	conflict	UNP O75417
A	2524	SER	GLY	conflict	UNP O75417
A	2525	GLY	MET	conflict	UNP O75417
A	2526	GLY	PHE	conflict	UNP O75417

- Molecule 2 is a DNA chain called DNA Template.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	13	Total	C	N	O	P	0	0	0
			263	124	47	79	13			

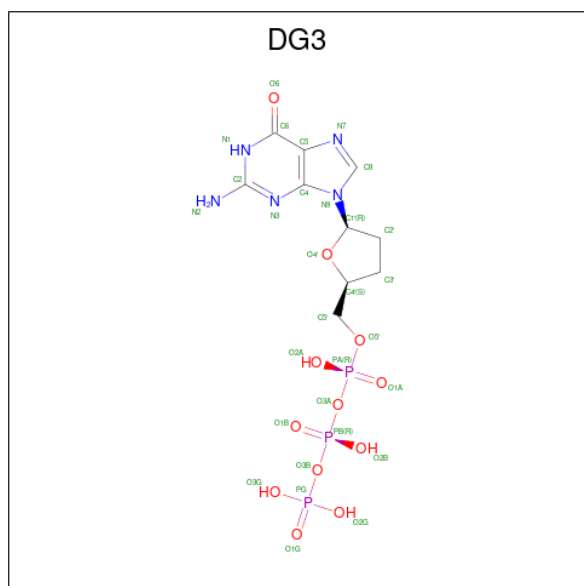
- Molecule 3 is a DNA chain called DNA Primer.

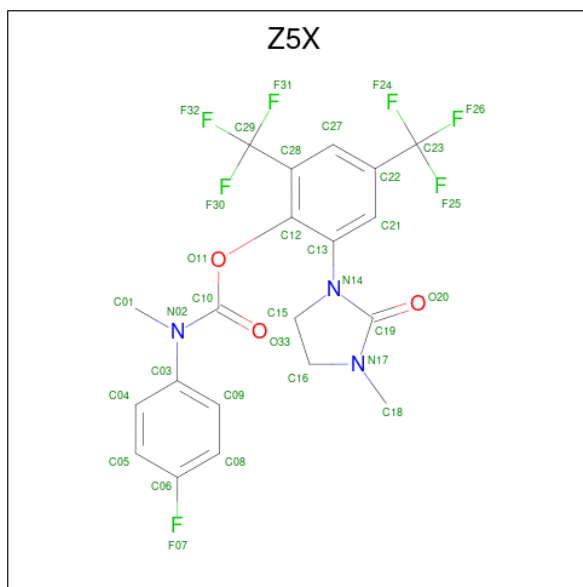
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	10	Total	C	N	O	P	0	0	0
			204	96	39	59	10			

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mg		
4	A	1	1	1	0	0

- Molecule 5 is 2'-3'-DIDEOXYGUANOSINE-5'-TRIPHOSPHATE (three-letter code: DG3) (formula: C₁₀H₁₆N₅O₁₂P₃).





Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	F	N	O		
6	A	1	33	20	7	3	3	0	0

4 Data and refinement statistics i

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α , β , γ	171.43Å 171.43Å 63.17Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	42.86 – 3.24 42.86 – 3.01	Depositor EDS
% Data completeness (in resolution range)	98.4 (42.86-3.24) 93.3 (42.86-3.01)	Depositor EDS
R_{merge}	0.26	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.17 (at 3.01Å)	Xtrriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.228 , 0.244 0.229 , 0.245	Depositor DCC
R_{free} test set	1986 reflections (10.00%)	wwPDB-VP
Wilson B-factor (Å ²)	89.1	Xtrriage
Anisotropy	0.559	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 66.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	0.048 for -h,-k,l	Xtrriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	5514	wwPDB-VP
Average B, all atoms (Å ²)	121.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.40% 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: Z5X, DG3, 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	0.25	0/5077	0.47	0/6845
2	E	0.52	0/293	0.90	0/449
3	F	0.48	0/228	0.80	0/349
All	All	0.28	0/5598	0.52	0/7643

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

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	4983	0	4998	27	0
2	E	263	0	146	3	0
3	F	204	0	110	5	0
4	A	1	0	0	0	0
5	A	30	0	12	0	0
6	A	33	0	0	0	0
All	All	5514	0	5266	34	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.

The worst 5 of 34 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:1827:ILE:HG21	1:A:2043:ARG:HH11	1.68	0.58
3:F:12:DC:H2'	3:F:13:DG:H8	1.69	0.57
1:A:1852:ILE:HD12	1:A:1903:LEU:HD21	1.86	0.57
1:A:2210:VAL:HG21	1:A:2253:PRO:HD3	1.88	0.56
1:A:2420:TYR:HB3	1:A:2423:ILE:HD12	1.87	0.55

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	621/652 (95%)	588 (95%)	33 (5%)	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
1	A	550/561 (98%)	545 (99%)	5 (1%)	78 89

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	2031	SER
1	A	2040	HIS
1	A	2248	ASN
1	A	2380	GLN
1	A	2501	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	2248	ASN

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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 3 ligands modelled in this entry, 1 is 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	Z5X	A	2603	-	34,35,35	1.86	4 (11%)	49,54,54	1.64	8 (16%)
5	DG3	A	2602	4	25,32,32	1.89	4 (16%)	28,50,50	1.97	6 (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
6	Z5X	A	2603	-	-	0/28/41/41	0/3/3/3
5	DG3	A	2602	4	-	5/18/31/31	0/3/3/3

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	A	2603	Z5X	C19-N14	7.88	1.47	1.38
5	A	2602	DG3	PA-O5'	5.66	1.82	1.59
6	A	2603	Z5X	C19-N17	4.59	1.47	1.35
5	A	2602	DG3	C2-N2	4.18	1.44	1.34
5	A	2602	DG3	C6-N1	3.98	1.43	1.37

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	A	2603	Z5X	C15-N14-C19	-5.25	107.78	111.81
5	A	2602	DG3	PA-O3A-PB	4.45	148.10	132.83
5	A	2602	DG3	PB-O3B-PG	4.33	147.68	132.83
5	A	2602	DG3	O6-C6-C5	4.08	132.35	124.37
6	A	2603	Z5X	C16-C15-N14	3.97	108.21	102.95

There are no chirality outliers.

All (5) torsion outliers are listed below:

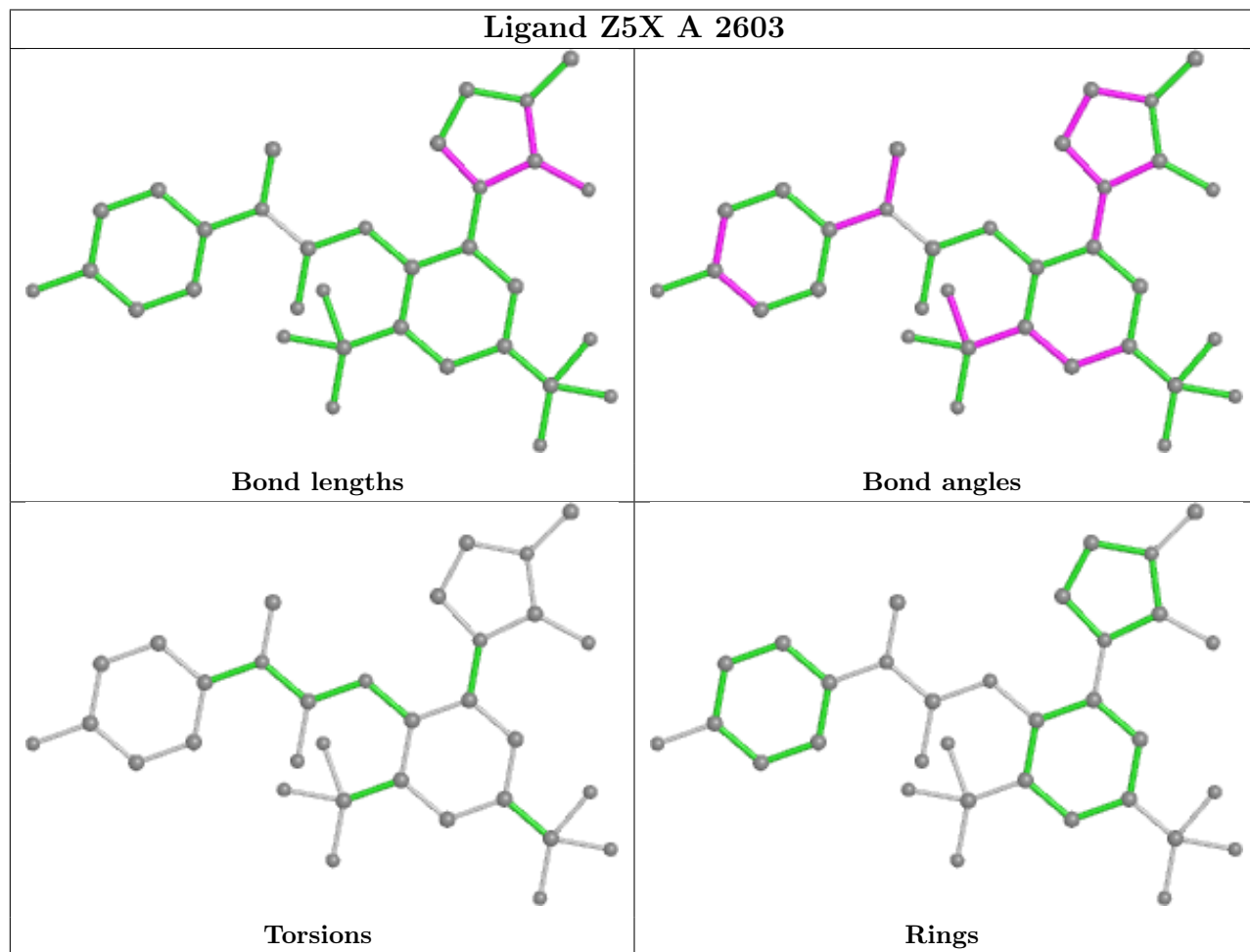
Mol	Chain	Res	Type	Atoms
5	A	2602	DG3	PB-O3B-PG-O3G
5	A	2602	DG3	O4'-C4'-C5'-O5'
5	A	2602	DG3	PB-O3B-PG-O1G
5	A	2602	DG3	C3'-C4'-C5'-O5'
5	A	2602	DG3	PB-O3B-PG-O2G

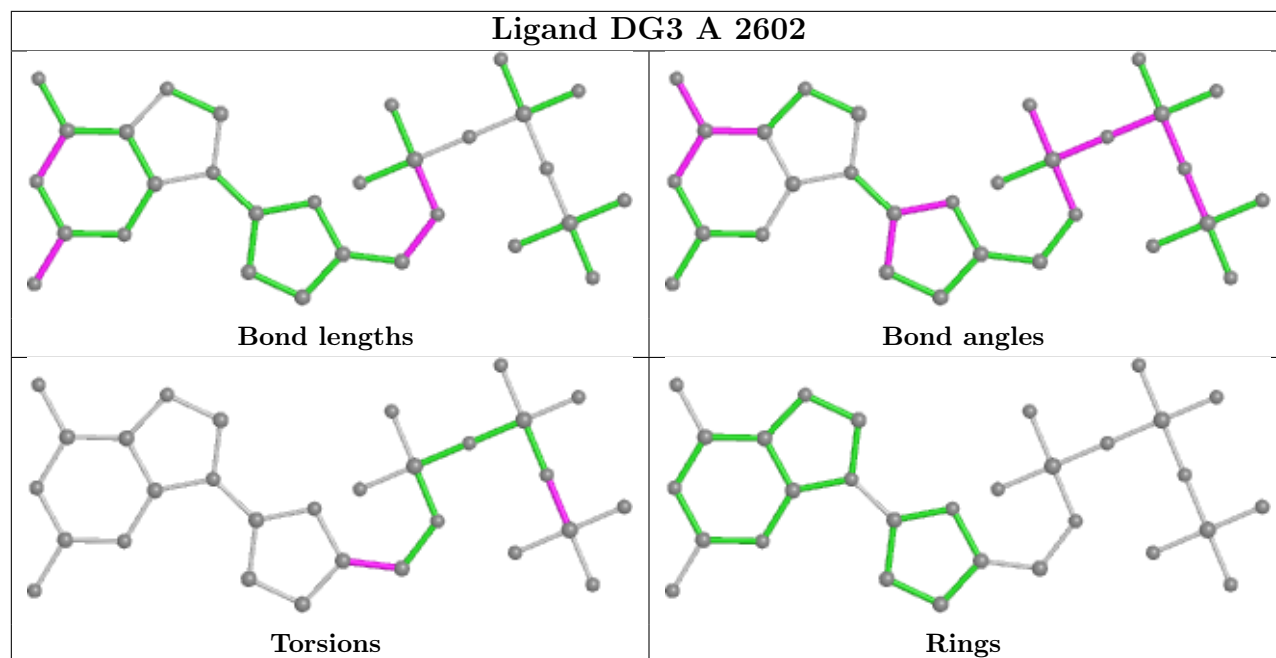
There are no ring outliers.

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(Å ²)	Q<0.9
1	A	631/652 (96%)	0.21	16 (2%) 57 46	66, 114, 170, 224	0
2	E	13/24 (54%)	0.72	2 (15%) 2 1	99, 113, 222, 224	0
3	F	10/14 (71%)	0.56	0 100 100	90, 128, 211, 245	0
All	All	654/690 (94%)	0.23	18 (2%) 53 41	66, 115, 175, 245	0

The worst 5 of 18 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	2590	VAL	4.8
1	A	2458	PRO	3.5
1	A	2039	GLU	3.2
1	A	2178	SER	3.0
2	E	8	DT	2.8

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](#)

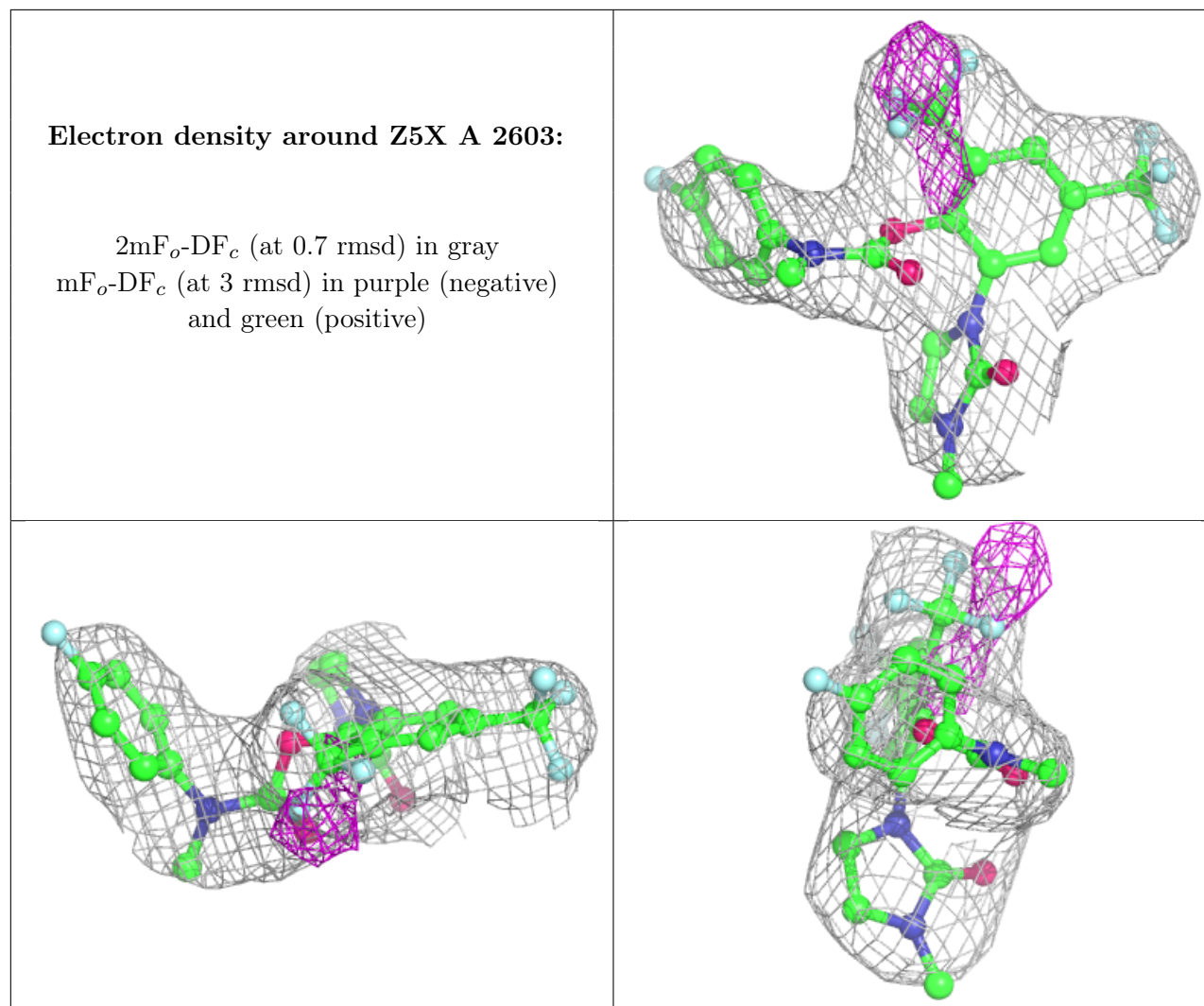
There are no monosaccharides in this entry.

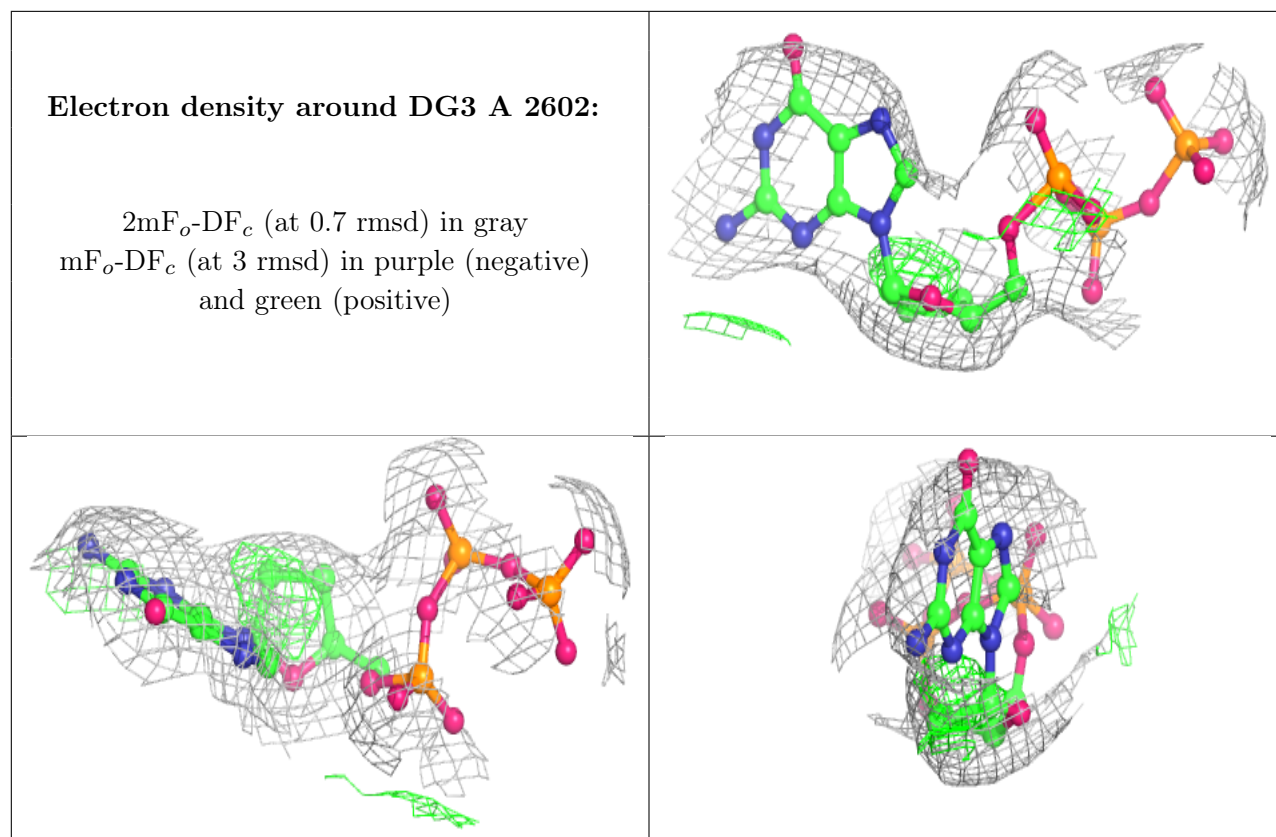
6.4 Ligands [i](#)

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
6	Z5X	A	2603	33/33	0.91	0.28	115,120,122,123	0
5	DG3	A	2602	30/30	0.95	0.33	101,107,110,110	0
4	MG	A	2601	1/1	0.97	0.41	82,82,82,82	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.