



wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 16, 2019 – 01:06 PM EST

PDB ID : 6UIT
Title : HIV-1 wild-type reverse transcriptase-DNA complex with dCTP
Authors : Lansdon, E.B.
Deposited on : 2019-10-01
Resolution : 2.81 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 : 1.8.0 (224370), CSD as540be (2019)
Xtriage (Phenix) : 1.13
EDS : 2.4
buster-report : 1.1.7 (2018)
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac : 5.8.0158
CCP4 : 7.0 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.4

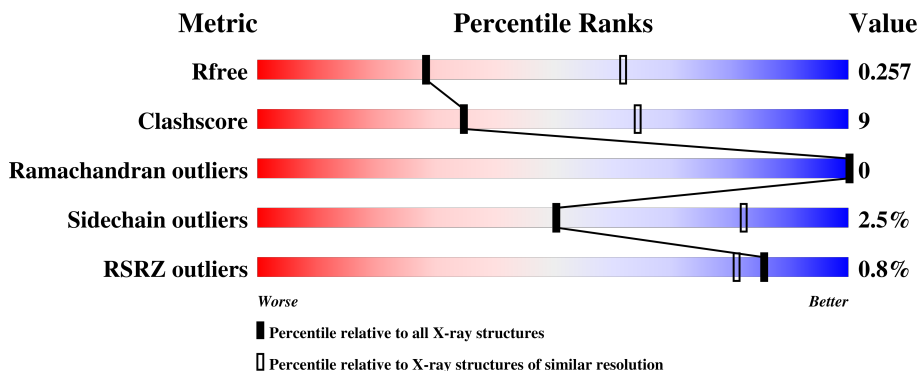
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.81 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}	111664	2792 (2.80-2.80)
Clashscore	122126	3209 (2.80-2.80)
Ramachandran outliers	120053	3158 (2.80-2.80)
Sidechain outliers	120020	3160 (2.80-2.80)
RSRZ outliers	108989	2726 (2.80-2.80)

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. 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	572	<div> <div></div> <div>75%20%5%</div> </div>
2	B	440	<div> <div>%</div> <div>68%21%10%</div> </div>
3	P	21	<div> <div></div> <div>62%24%14%</div> </div>
4	T	27	<div> <div>4%</div> <div>37%44%15%</div> </div>

2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 8624 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 p66 Reverse transcriptase/RNaseH.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	545	Total	C	N	O	S	0	0	0
			4441	2874	740	819	8			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-11	MET	-	initiating methionine	UNP P04585
A	-10	GLY	-	expression tag	UNP P04585
A	-9	SER	-	expression tag	UNP P04585
A	-8	SER	-	expression tag	UNP P04585
A	-7	HIS	-	expression tag	UNP P04585
A	-6	HIS	-	expression tag	UNP P04585
A	-5	HIS	-	expression tag	UNP P04585
A	-4	HIS	-	expression tag	UNP P04585
A	-3	HIS	-	expression tag	UNP P04585
A	-2	HIS	-	expression tag	UNP P04585
A	-1	SER	-	expression tag	UNP P04585
A	0	SER	-	expression tag	UNP P04585
A	258	CYS	GLN	engineered mutation	UNP P04585
A	280	SER	CYS	engineered mutation	UNP P04585

- Molecule 2 is a protein called p51 Reverse transcriptase/RNaseH.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	394	Total	C	N	O	S	0	0	0
			3263	2122	540	595	6			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	280	SER	CYS	engineered mutation	UNP P04585

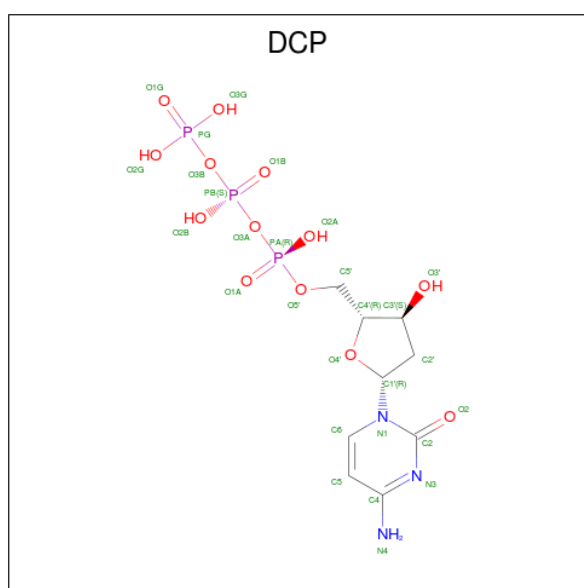
- Molecule 3 is a DNA chain called DNA primer.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	P	18	Total	C	N	O	P	0	0	0
			360	172	62	109	17			

- Molecule 4 is a DNA chain called DNA template.

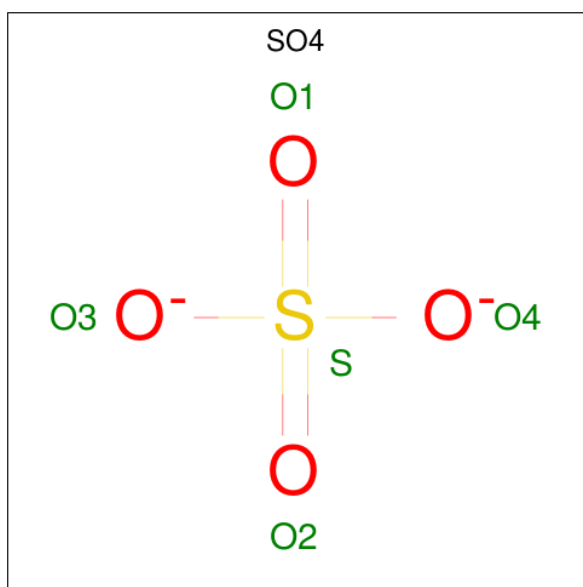
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	T	23	Total	C	N	O	P	0	0	0
			482	224	100	135	23			

- Molecule 5 is 2'-DEOXYCYTIDINE-5'-TRIPHOSPHATE (three-letter code: DCP) (formula: $C_9H_{16}N_3O_{13}P_3$) (labeled as "Ligand of Interest" by author).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	A	1	Total	C	N	O	P	0	0
			28	9	3	13	3		

- Molecule 6 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	O	S	0	0
			5	4	1		
6	A	1	Total	O	S	0	0
			5	4	1		
6	P	1	Total	O	S	0	0
			5	4	1		

- Molecule 7 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

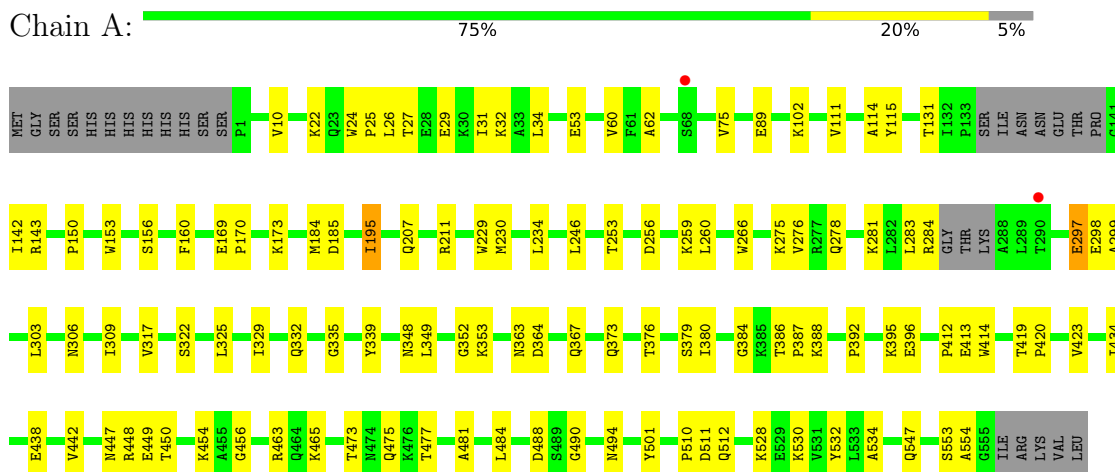
- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	19	Total	O	0	0
			19	19		
8	B	14	Total	O	0	0
			14	14		
8	T	1	Total	O	0	0
			1	1		

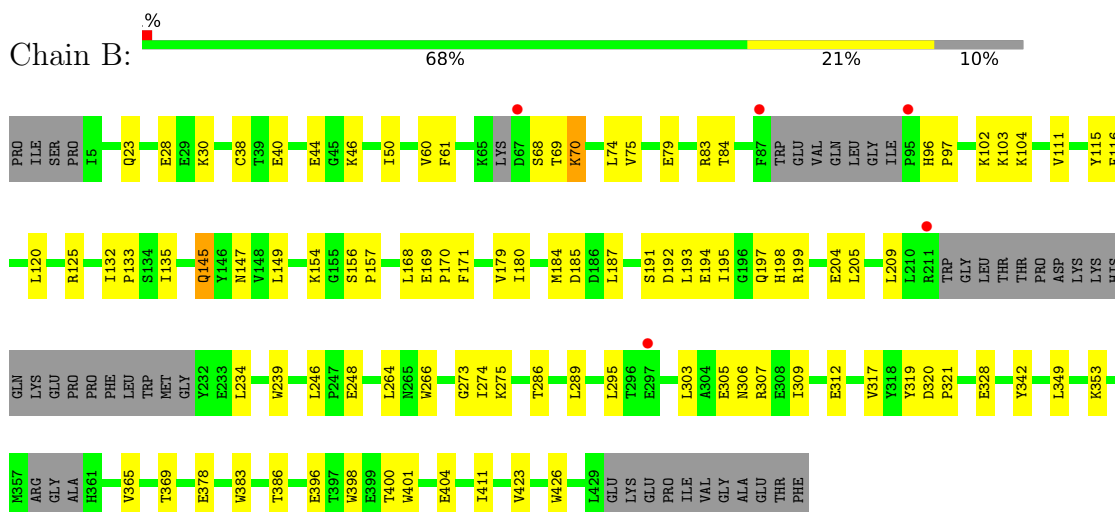
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA 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: p66 Reverse transcriptase/RNaseH

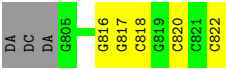


- Molecule 2: p51 Reverse transcriptase/RNaseH

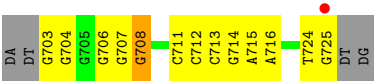


- Molecule 3: DNA primer





● Molecule 4: DNA template



4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	167.53Å 167.79Å 102.29Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.09 – 2.81 47.09 – 2.81	Depositor EDS
% Data completeness (in resolution range)	85.4 (47.09-2.81) 90.4 (47.09-2.81)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.52 (at 2.81Å)	Xtriage
Refinement program	PHENIX 1.9_1692, PHENIX 1.9_1692	Depositor
R, R_{free}	0.185 , 0.256 0.186 , 0.257	Depositor DCC
R_{free} test set	2000 reflections (5.69%)	wwPDB-VP
Wilson B-factor (Å ²)	44.6	Xtriage
Anisotropy	0.122	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 33.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.014 for -k,-h,-l	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	8624	wwPDB-VP
Average B, all atoms (Å ²)	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.09% 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: DOC, MG, SO4, DCP

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	0.50	0/4555	0.60	0/6185
2	B	0.45	0/3351	0.57	0/4547
3	P	0.76	0/381	1.00	0/586
4	T	0.95	1/543 (0.2%)	0.85	0/838
All	All	0.53	1/8830 (0.0%)	0.63	0/12156

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	T	708	DG	C3'-O3'	-5.80	1.36	1.44

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	4441	0	4490	73	0
2	B	3263	0	3294	66	0
3	P	360	0	203	7	0
4	T	482	0	255	15	0
5	A	28	0	12	0	0
6	A	10	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	P	5	0	0	0	0
7	A	1	0	0	0	0
8	A	19	0	0	0	0
8	B	14	0	0	0	0
8	T	1	0	0	0	0
All	All	8624	0	8254	149	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 9.

The worst 5 of 149 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)
2:B:365:VAL:O	2:B:369:THR:HG23	1.88	0.73
2:B:423:VAL:HA	2:B:426:TRP:CD1	2.25	0.72
1:A:465:LYS:HE2	1:A:484:LEU:HD11	1.70	0.72
2:B:274:ILE:C	2:B:275:LYS:HD3	2.10	0.72
1:A:131:THR:HG23	1:A:143:ARG:HG2	1.75	0.69

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	539/572 (94%)	505 (94%)	34 (6%)	0	100	100
2	B	384/440 (87%)	365 (95%)	19 (5%)	0	100	100
All	All	923/1012 (91%)	870 (94%)	53 (6%)	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	486/511 (95%)	473 (97%)	13 (3%)	48	81
2	B	361/400 (90%)	353 (98%)	8 (2%)	55	85
All	All	847/911 (93%)	826 (98%)	21 (2%)	50	82

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	322	SER
1	A	396	GLU
2	B	209	LEU
1	A	297	GLU
2	B	234	LEU

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

Mol	Chain	Res	Type
1	A	464	GLN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

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
3	DOC	P	822	3,4	13,19,20	2.66	5 (38%)	13,26,29	4.39	6 (46%)

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	DOC	P	822	3,4	-	1/3/18/19	0/2/2/2

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	P	822	DOC	O5'-C5'	-5.43	1.37	1.44
3	P	822	DOC	C4-N3	4.80	1.43	1.35
3	P	822	DOC	C5-C4	-3.28	1.33	1.41
3	P	822	DOC	O4'-C1'	3.23	1.49	1.42
3	P	822	DOC	C4-N4	2.21	1.41	1.35

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	P	822	DOC	C2'-C1'-N1	12.29	135.64	112.48
3	P	822	DOC	O4'-C4'-C5'	6.92	120.82	109.59
3	P	822	DOC	C2-N3-C4	4.41	120.47	116.26
3	P	822	DOC	O4'-C1'-C2'	-3.87	102.48	106.67
3	P	822	DOC	N4-C4-N3	2.31	120.12	116.48

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	P	822	DOC	C3'-C4'-C5'-O5'

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	P	822	DOC	3	0

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 5 ligands modelled in this entry, 1 is monoatomic - leaving 4 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$
5	DCP	A	601	7	23,29,29	1.19	3 (13%)	29,45,45	2.66	6 (20%)
6	SO4	A	602	-	4,4,4	0.27	0	6,6,6	0.11	0
6	SO4	A	603	-	4,4,4	0.24	0	6,6,6	0.13	0
6	SO4	P	901	-	4,4,4	0.28	0	6,6,6	0.35	0

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
5	DCP	A	601	7	-	4/18/34/34	0/2/2/2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	601	DCP	PG-O2G	-3.02	1.43	1.54
5	A	601	DCP	PB-O2B	-2.16	1.44	1.55
5	A	601	DCP	PA-O2A	-2.04	1.45	1.55

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	601	DCP	C2-N3-C4	11.59	127.34	116.26
5	A	601	DCP	C5-C4-N3	-3.88	117.22	121.72
5	A	601	DCP	PB-O3B-PG	3.26	142.95	132.57

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	601	DCP	O4'-C1'-N1	2.43	111.88	107.78
5	A	601	DCP	N4-C4-N3	2.29	120.09	116.48

There are no chirality outliers.

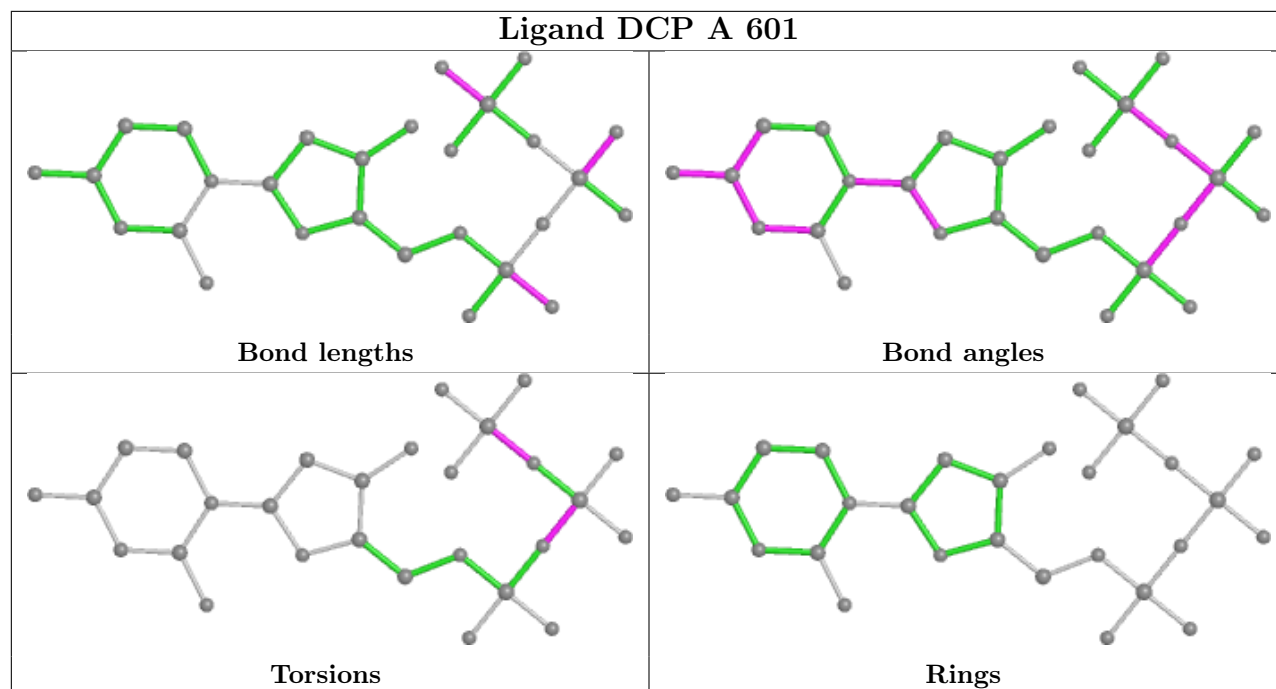
All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	601	DCP	PB-O3B-PG-O2G
5	A	601	DCP	PA-O3A-PB-O1B
5	A	601	DCP	PA-O3A-PB-O2B
5	A	601	DCP	PB-O3B-PG-O1G

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	545/572 (95%)	-0.31	2 (0%) 92 90	20, 38, 66, 97	0
2	B	394/440 (89%)	-0.10	5 (1%) 77 71	22, 51, 83, 99	0
3	P	17/21 (80%)	-0.71	0 100 100	29, 50, 67, 68	0
4	T	23/27 (85%)	-0.27	1 (4%) 35 25	29, 62, 99, 112	0
All	All	979/1060 (92%)	-0.23	8 (0%) 86 81	20, 43, 78, 112	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	68	SER	4.9
2	B	87	PHE	4.0
4	T	725	DG	3.3
2	B	67	ASP	2.7
1	A	290	THR	2.4

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	DOC	P	822	18/19	0.98	0.13	20,26,31,34	0

6.3 Carbohydrates [i](#)

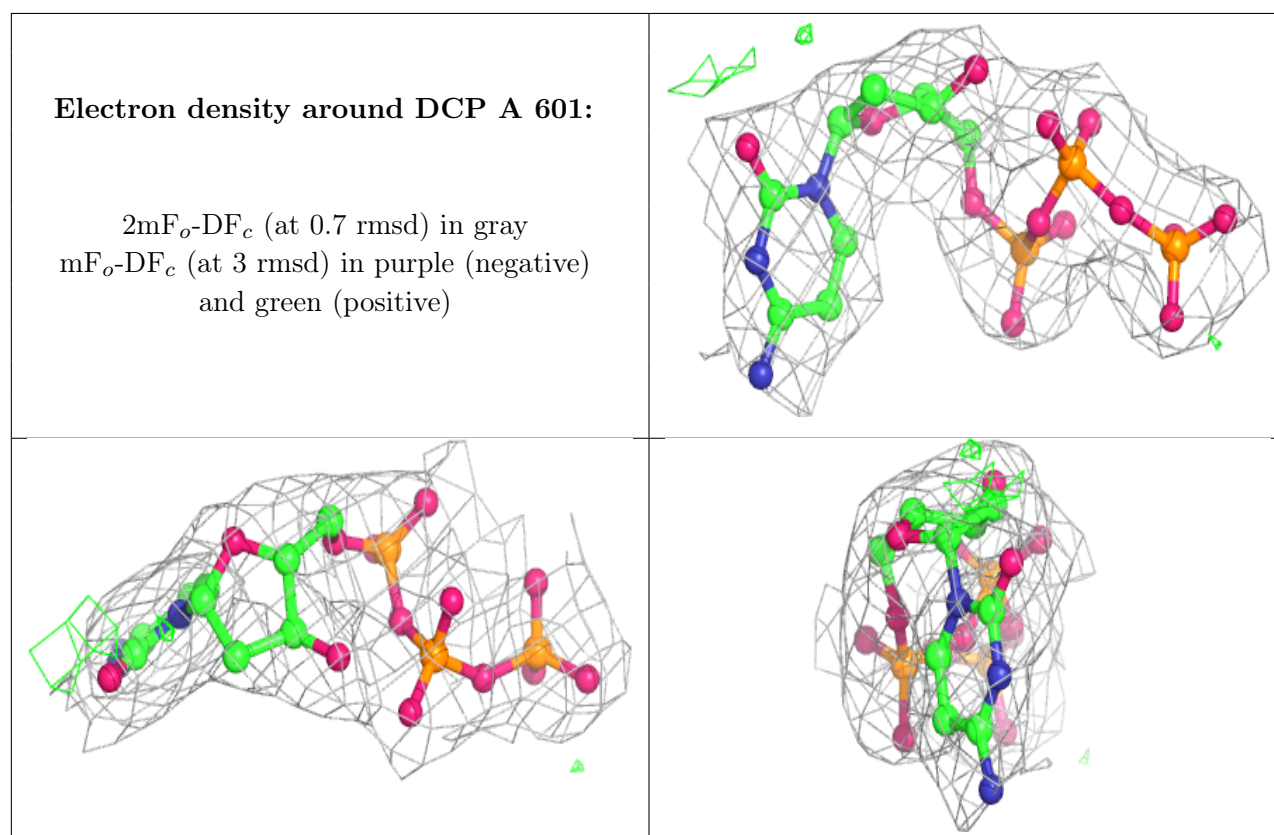
There are no carbohydrates in this entry.

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
6	SO4	P	901	5/5	0.86	0.23	66,73,76,100	0
7	MG	A	604	1/1	0.94	0.11	24,24,24,24	0
6	SO4	A	603	5/5	0.95	0.11	74,77,82,99	0
5	DCP	A	601	28/28	0.98	0.14	27,32,39,41	0
6	SO4	A	602	5/5	0.98	0.10	42,43,54,55	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.