wwPDB X-ray Structure Validation Summary Report (i)

Jul 25, 2023 – 12:53 AM EDT

PDB ID	:	1BPY
Title	:	HUMAN DNA POLYMERASE BETA COMPLEXED WITH GAPPED DNA AND DDCTP
Authors	:	Sawaya, M.R.; Pelletier, H.; Prasad, R.; Wilson, S.H.; Kraut, J.
Deposited on	:	1997-04-15
Resolution	:	2.20 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 (i) 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 (i)) were used in the production of this report:

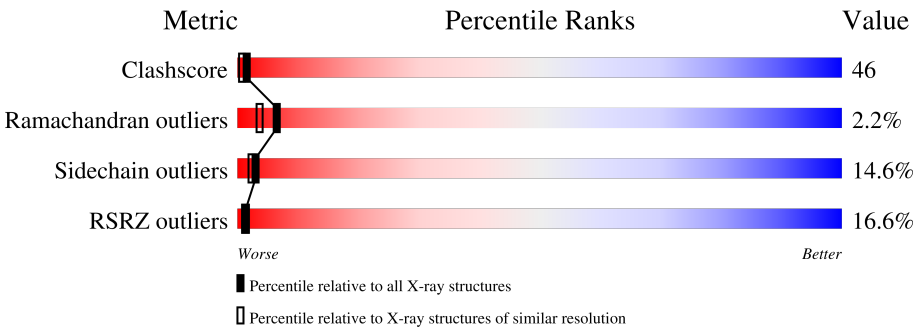
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.34
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.34

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.20 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))
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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 <=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	19% 56% 44%
2	P	10	10% 30% 70%
3	D	5	20% 40% 40%
4	A	335	16% 31% 47% 15% · ·

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 3597 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*CP*GP*GP*CP*GP*CP*AP*TP*CP*AP*GP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	T	16	323	153	63	92	15	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	P	10	202	97	38	58	9	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	D	5	106	49	20	32	5	0	0	0

- Molecule 4 is a protein called PROTEIN (DNA POLYMERASE BETA).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	A	326	2614	1652	457	496	9	0	0	0

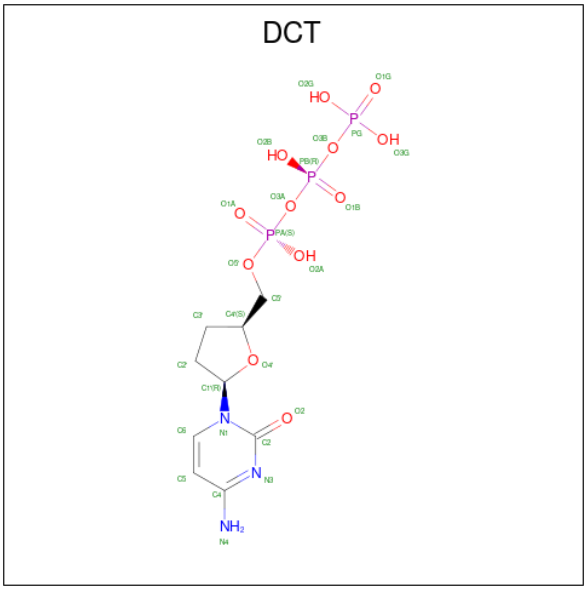
- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mg		
5	A	2	2	2	0	0

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Na		
6	A	2	2	2	0	0

- Molecule 7 is 2',3'-DIDEOXYCYTIDINE 5'-TRIPHOSPHATE (three-letter code: DCT) (formula: $C_9H_{16}N_3O_{12}P_3$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
7	A	1	27	9	3	12	3	0	0

- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
8	T	50	50	50	0	0
			Total	O		
8	P	21	21	21	0	0
			Total	O		
8	D	27	27	27	0	0
			Total	O		
8	A	223	223	223	0	0

3 Residue-property plots (i)

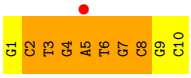
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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 (5'-D(*CP*CP*GP*AP*CP*GP*GP*CP*GP*CP*AP*TP*CP*AP*GP*C)-3')

Chain T:

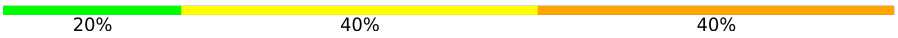
• Molecule 2: DNA (5'-D(*GP*CP*TP*GP*AP*TP*GP*CP*GP*DOC)-3')

Chain P:



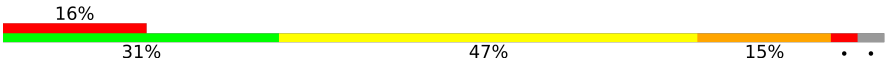
• Molecule 3: DNA (5'-D(*GP*TP*CP*GP*G)-3')

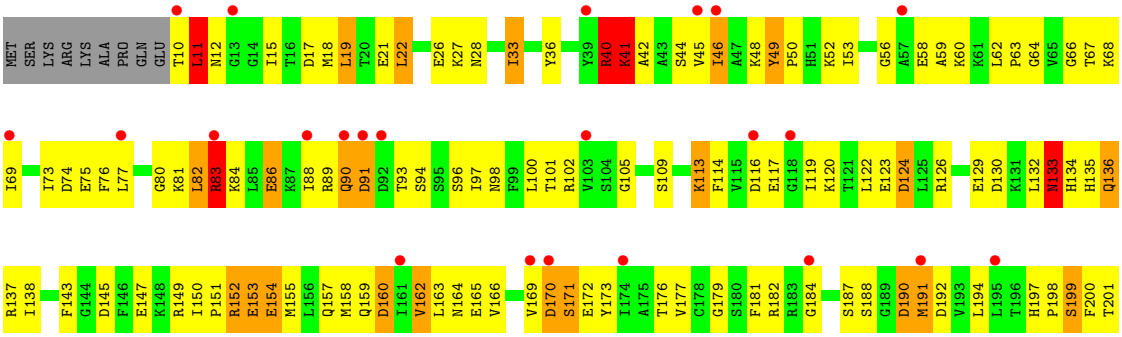
Chain D:

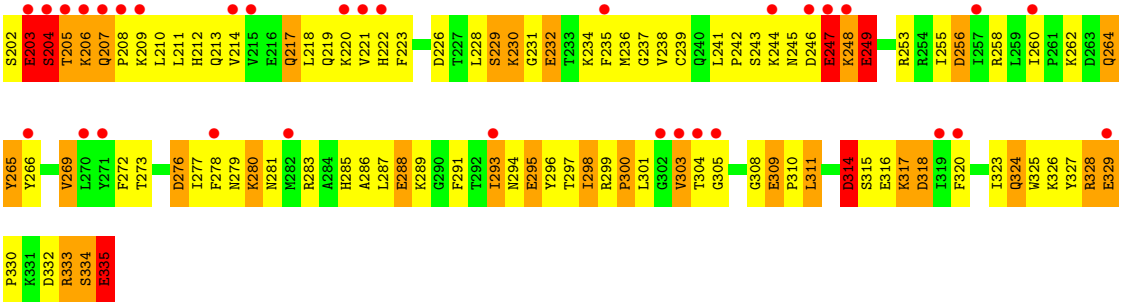


• Molecule 4: PROTEIN (DNA POLYMERASE BETA)

Chain A:






S202 E203 S204 T205 K206 Q207 P208 K209 L210 L211 H212 Q213 V214 V215 E216 Q217 L218 Q219 K220 V221 H222 F223
D226 T227 L228 S229 K230 G231 E232 T233 K234 F235 M236 G237 V238 C239 Q240 L241 P242 S243 K244 N245 D246 E247 K248 E249
R253 R254 I255 D256 I257 R258 L259 I260 P261 K262 D263 Q264
Y265 Y266 V269 L270 Y271 F272 T273 D276 I277 F278 N279 K280 N281 M282 R283 A284 H285 A286 L287 E288 K289 G290 F291 T292 I293 N294 E295 Y296 T297 I298 R299 P300 L301 G302 V303 T304 G305
G308 E309 P310 L311 D314 S315 E316 K317 D318 I319 F320 I323 Q324 W325 K326 Y327 R328 E329
P330 K331 D332 R333 S334 E335

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	50.54Å 79.82Å 54.94Å 90.00° 107.11° 90.00°	Depositor
Resolution (Å)	20.00 – 2.20 19.84 – 2.20	Depositor EDS
% Data completeness (in resolution range)	95.4 (20.00-2.20) 91.8 (19.84-2.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$< I/\sigma(I) >$ [1]	2.12 (at 2.19Å)	Xtriage
Refinement program	TNT 5D	Depositor
R, R_{free}	0.232 , 0.330 0.210 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	20.9	Xtriage
Anisotropy	0.849	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.22 , 85.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	3597	wwPDB-VP
Average B, all atoms (Å^2)	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.05% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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, NA, DCT

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	2.01	11/362 (3.0%)	2.28	21/556 (3.8%)
2	P	2.09	6/206 (2.9%)	2.47	14/317 (4.4%)
3	D	2.14	3/118 (2.5%)	2.36	6/179 (3.4%)
4	A	1.22	21/2663 (0.8%)	1.75	63/3578 (1.8%)
All	All	1.43	41/3349 (1.2%)	1.90	104/4630 (2.2%)

The worst 5 of 41 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	P	5	DA	N9-C4	-10.09	1.31	1.37
3	D	1	DG	OP3-P	-8.54	1.50	1.61
4	A	232	GLU	CD-OE2	7.89	1.34	1.25
4	A	26	GLU	CD-OE1	7.70	1.34	1.25
4	A	117	GLU	CD-OE2	7.58	1.33	1.25

The worst 5 of 104 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	333	ARG	NE-CZ-NH1	11.27	125.94	120.30
2	P	5	DA	C8-N9-C4	11.02	110.21	105.80
4	A	137	ARG	NE-CZ-NH1	10.54	125.57	120.30
1	T	3	DG	O4'-C1'-N9	-10.51	100.64	108.00
4	A	204	SER	N-CA-C	9.91	137.75	111.00

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	T	323	0	179	13	0
2	P	202	0	114	5	0
3	D	106	0	57	9	0
4	A	2614	0	2635	269	0
5	A	2	0	0	0	0
6	A	2	0	0	0	0
7	A	27	0	12	0	0
8	A	223	0	0	20	0
8	D	27	0	0	2	0
8	P	21	0	0	2	0
8	T	50	0	0	3	0
All	All	3597	0	2997	286	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 46.

The worst 5 of 286 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:150:ILE:HD12	4:A:155:MET:HE1	1.28	1.12
4:A:285:HIS:HD2	4:A:323:ILE:HD12	1.17	1.07
4:A:230:LYS:HG3	4:A:235:PHE:HD1	1.13	1.06
4:A:316:GLU:H	4:A:317:LYS:NZ	1.53	1.06
4:A:207:GLN:HB3	4:A:209:LYS:HG2	1.41	1.03

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
4	A	324/335 (97%)	285 (88%)	32 (10%)	7 (2%)	6 4

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	A	11	LEU
4	A	179	GLY
4	A	202	SER
4	A	220	LYS
4	A	143	PHE

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	287/295 (97%)	245 (85%)	42 (15%)	3 2

5 of 42 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	A	217	GLN
4	A	280	LYS
4	A	218	LEU
4	A	264	GLN
4	A	314	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
4	A	245	ASN
4	A	264	GLN
4	A	285	HIS
4	A	279	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
4	A	133	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)

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	DOC	P	10	2,1	16,19,20	0.79	0	20,26,29	1.58	3 (15%)

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	DOC	P	10	2,1	-	0/7/18/19	0/2/2/2

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	P	10	DOC	C2'-C1'-N1	4.37	120.69	112.40
2	P	10	DOC	O4'-C1'-C2'	4.06	111.07	106.67
2	P	10	DOC	C2'-C3'-C4'	2.39	107.21	102.72

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 5 ligands modelled in this entry, 4 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	$\#\|Z\| > 2$
7	DCT	A	338	5	24,28,28	1.59	4 (16%)	33,43,43	1.28	4 (12%)

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
7	DCT	A	338	5	-	5/22/31/31	0/2/2/2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	A	338	DCT	C1'-N1	-3.86	1.38	1.48
7	A	338	DCT	O4'-C1'	-3.05	1.35	1.42
7	A	338	DCT	C2-N1	-2.83	1.33	1.40
7	A	338	DCT	PA-O2A	-2.67	1.42	1.55

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
7	A	338	DCT	O3G-PG-O3B	-2.85	95.07	104.64
7	A	338	DCT	C4'-O4'-C1'	2.74	112.40	109.81
7	A	338	DCT	O4'-C1'-N1	2.60	112.50	107.86
7	A	338	DCT	O2-C2-N3	-2.17	118.81	122.33

There are no chirality outliers.

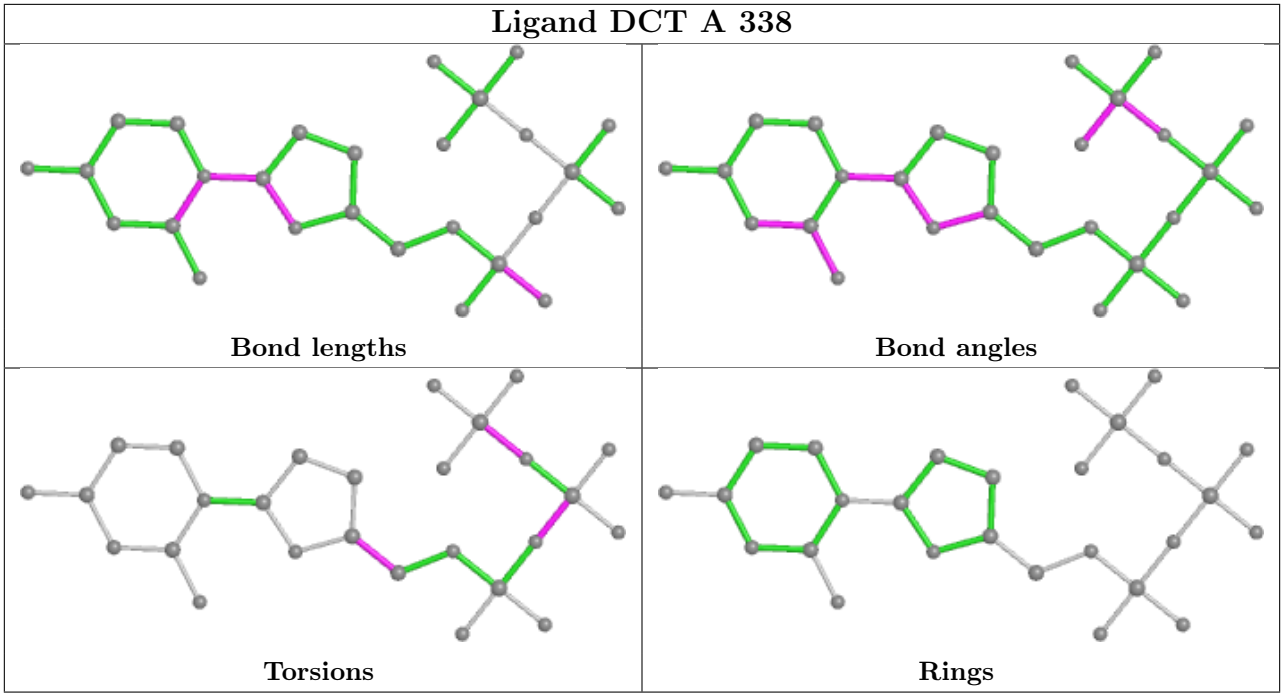
All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
7	A	338	DCT	O4'-C4'-C5'-O5'
7	A	338	DCT	PB-O3B-PG-O3G
7	A	338	DCT	C3'-C4'-C5'-O5'
7	A	338	DCT	PA-O3A-PB-O1B
7	A	338	DCT	PA-O3A-PB-O2B

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 then 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	T	16/16 (100%)	0.76	3 (18%)	1	1	41, 49, 68, 69	0
2	P	9/10 (90%)	1.17	1 (11%)	5	4	49, 53, 64, 66	0
3	D	5/5 (100%)	0.45	0	100	100	38, 41, 44, 47	0
4	A	326/335 (97%)	1.17	55 (16%)	1	1	32, 48, 70, 97	0
All	All	356/366 (97%)	1.15	59 (16%)	1	1	32, 49, 70, 97	0

The worst 5 of 59 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	A	205	THR	9.3
4	A	208	PRO	9.0
4	A	204	SER	6.4
4	A	303	VAL	4.7
4	A	329	GLU	4.4

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
2	DOC	P	10	18/19	0.92	0.15	32,47,56,56	0

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

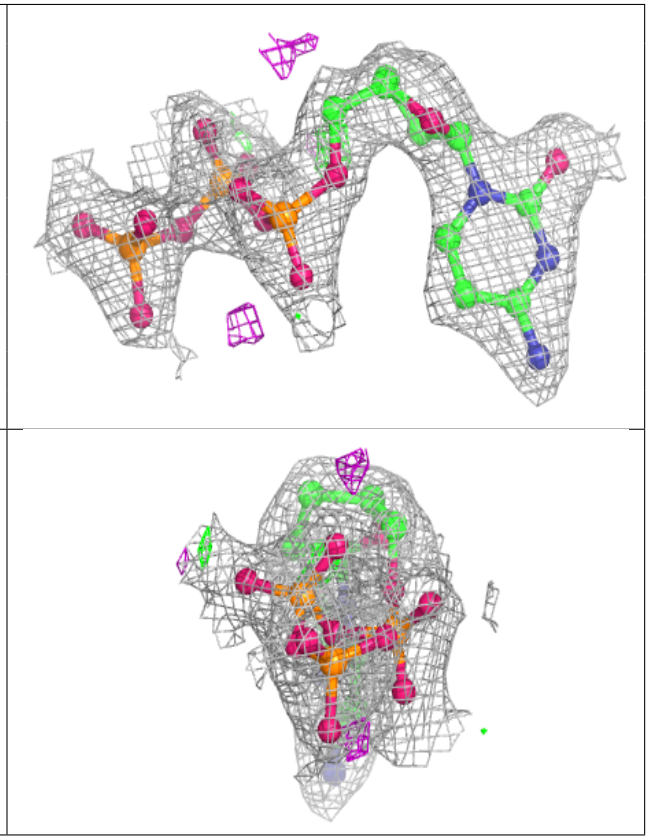
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
5	MG	A	340	1/1	0.87	0.18	49,49,49,49	0
5	MG	A	339	1/1	0.90	0.15	45,45,45,45	0
7	DCT	A	338	27/27	0.92	0.16	27,41,48,57	0
6	NA	A	342	1/1	0.96	0.10	38,38,38,38	0
6	NA	A	341	1/1	0.99	0.12	29,29,29,29	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 DCT A 338:

$2mF_o$-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.