wwPDB X-ray Structure Validation Summary Report (i)

Oct 9, 2023 – 12:30 PM EDT

PDB ID	:	6WPJ
Title	:	Structure of HIV-1 Reverse Transcriptase (RT) in complex with dsDNA and d4T
Authors	:	Bertoletti, N.; Anderson, K.S.
Deposited on	:	2020-04-27
Resolution	:	2.73 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:

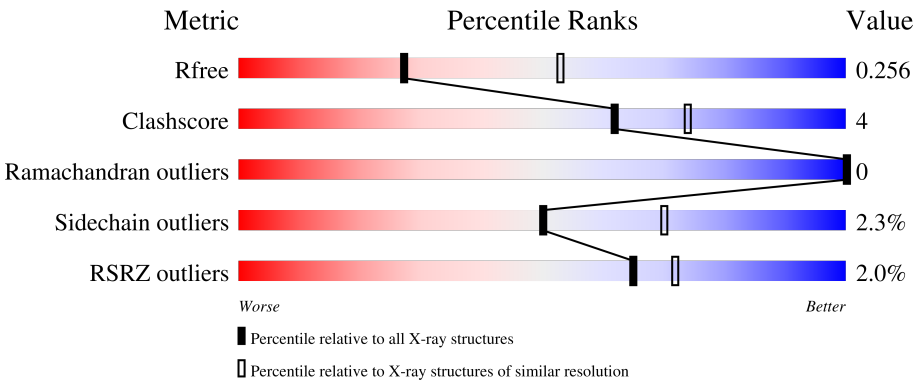
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 2.73 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	1271 (2.76-2.72)
Clashscore	141614	1322 (2.76-2.72)
Ramachandran outliers	138981	1297 (2.76-2.72)
Sidechain outliers	138945	1298 (2.76-2.72)
RSRZ outliers	127900	1243 (2.76-2.72)

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	A	561	2%, 88%, 11%, ••
2	B	452	2%, 76%, 10%, 13%
3	P	21	43%, 38%, 5%, 14%
4	T	27	48%, 33%, •, 15%

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 8423 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 Reverse transcriptase/ribonuclease H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	557	4428	2864	744	812	8	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	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 RT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	394	3110	2024	514	567	5	0	0	0

There are 13 discrepancies between the modelled and reference sequences:

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

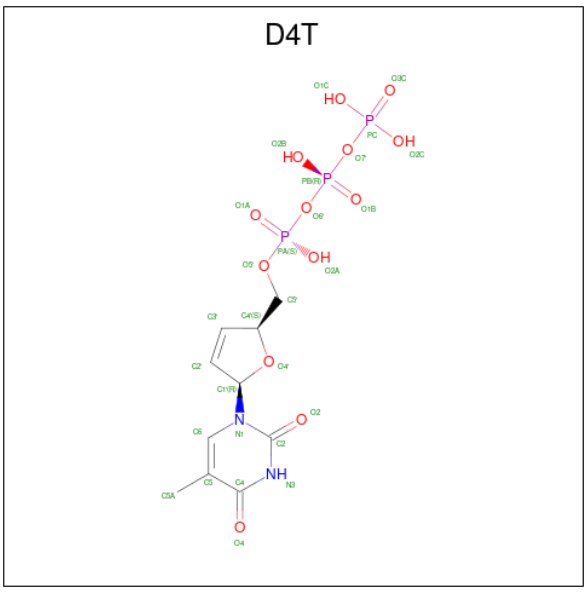
- Molecule 3 is a DNA chain called DNA Primer 21-mer.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	P	S			
3	P	18	366	175	64	109	17	1	0	0	0

- Molecule 4 is a DNA chain called DNA template 27-mer.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	T	23	478	223	98	134	23	0	0	0

- Molecule 5 is 2',3'-DEHYDRO-2',3'-DEOXY-THYMIDINE 5'-TRIPHOSPHATE (three-letter code: D4T) (formula: $C_{10}H_{15}N_2O_{13}P_3$) (labeled as "Ligand of Interest" by depositor).

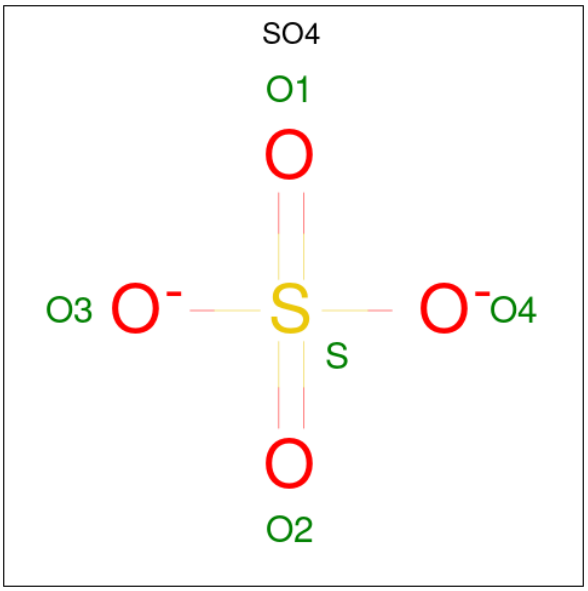


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
5	A	1	28	10	2	13	3	0	0

- Molecule 6 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

- Molecule 7 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



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

- Molecule 8 is water.

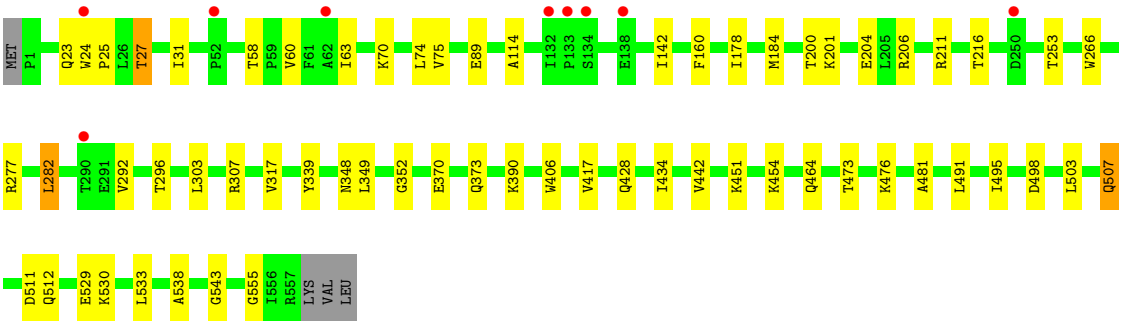
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	A	5	Total O 5 5	0	0
8	B	1	Total O 1 1	0	0

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.

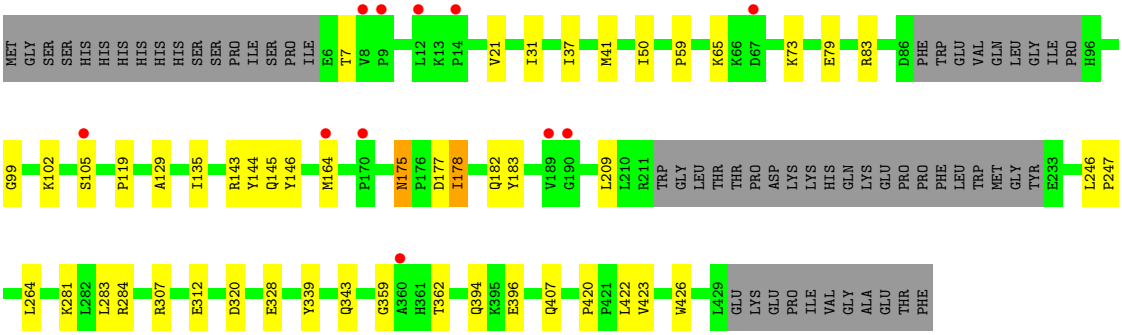
• Molecule 1: Reverse transcriptase/ribonuclease H

Chain A: 2% 88% 11%



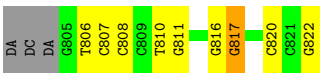
• Molecule 2: p51 RT

Chain B: 2% 76% 10% 13%



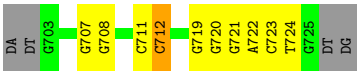
• Molecule 3: DNA Primer 21-mer

Chain P: 43% 38% 5% 14%



• Molecule 4: DNA template 27-mer

Chain T: 48% 33% 15%


DA
DT
G703
G707
G708
C711
C712
G719
G720
G721
A722
C723
T724
G725
DT
DG

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	167.59Å 170.06Å 102.46Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.56 – 2.73 47.56 – 2.73	Depositor EDS
% Data completeness (in resolution range)	99.1 (47.56-2.73) 99.1 (47.56-2.73)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.15	Depositor
$< I/\sigma(I) >$ [1]	1.04 (at 2.73Å)	Xtriage
Refinement program	PHENIX 1.14_3260	Depositor
R, R_{free}	0.203 , 0.257 0.203 , 0.256	Depositor DCC
R_{free} test set	1945 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	62.4	Xtriage
Anisotropy	0.337	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 47.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.008 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	8423	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.94% 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: MG, G47, DDG, D4T, SO4

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	0.40	0/4545	0.57	0/6193
2	B	0.35	0/3196	0.53	0/4366
3	P	0.85	0/355	1.02	0/543
4	T	0.82	0/538	0.88	1/829 (0.1%)
All	All	0.45	0/8634	0.61	1/11931 (0.0%)

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
2	B	0	1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	T	712	DC	O5'-P-OP2	-5.67	100.59	105.70

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	183	TYR	Peptide

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	4428	0	4372	35	0
2	B	3110	0	3012	25	0
3	P	366	0	207	8	0
4	T	478	0	255	7	0
5	A	28	0	11	0	0
6	A	2	0	0	0	0
7	A	5	0	0	0	0
8	A	5	0	0	0	0
8	B	1	0	0	0	0
All	All	8423	0	7857	71	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 4.

The worst 5 of 71 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:320:ASP:H	2:B:343:GLN:HE22	1.27	0.81
2:B:281:LYS:HE2	2:B:284:ARG:HH22	1.51	0.73
3:P:816:DG:H2'	3:P:817:G47:H8	1.73	0.70
2:B:359:GLY:HA3	2:B:362:THR:HB	1.74	0.69
3:P:806:DT:H2"	3:P:807:DC:O5'	1.92	0.68

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	555/561 (99%)	534 (96%)	21 (4%)	0	100	100
2	B	388/452 (86%)	377 (97%)	11 (3%)	0	100	100
All	All	943/1013 (93%)	911 (97%)	32 (3%)	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	466/501 (93%)	453 (97%)	13 (3%)	43	63
2	B	321/411 (78%)	316 (98%)	5 (2%)	62	78
All	All	787/912 (86%)	769 (98%)	18 (2%)	50	70

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	175	ASN
2	B	209	LEU
2	B	178	ILE
1	A	349	LEU
2	B	105	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	182	GLN
2	B	343	GLN
2	B	407	GLN
1	A	428	GLN
1	A	507	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)

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	G47	P	817	1,4,3	19,27,28	0.85	0	17,38,41	1.19	2 (11%)
3	DDG	P	822	4,3	17,23,24	1.07	2 (11%)	15,33,36	1.38	3 (20%)

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	G47	P	817	1,4,3	-	0/7/25/26	0/3/3/3
3	DDG	P	822	4,3	-	2/3/18/19	0/3/3/3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	P	822	DDG	C6-N1	-2.46	1.34	1.37
3	P	822	DDG	C5-C4	2.05	1.48	1.43

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	P	822	DDG	C3'-C2'-C1'	2.71	105.92	102.78
3	P	822	DDG	C5-C6-N1	2.27	117.96	113.95
3	P	817	G47	O6-C6-C5	-2.25	119.97	124.37
3	P	822	DDG	C2'-C1'-N9	-2.09	108.54	112.48

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	P	817	G47	C5-C6-N1	2.03	117.54	113.95

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	P	822	DDG	C3'-C4'-C5'-O5'
3	P	822	DDG	O4'-C4'-C5'-O5'

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	P	817	G47	2	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 4 ligands modelled in this entry, 2 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	$\#\|Z\| > 2$
5	D4T	A	601	6	22,29,29	4.62	6 (27%)	27,45,45	2.18	3 (11%)
7	SO4	A	604	-	4,4,4	0.18	0	6,6,6	0.48	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	D4T	A	601	6	-	5/19/31/31	0/2/2/2

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	601	D4T	C1'-C2'	-13.82	1.31	1.49
5	A	601	D4T	C4'-C3'	-12.39	1.27	1.50
5	A	601	D4T	O4'-C1'	6.43	1.53	1.42
5	A	601	D4T	O4'-C4'	6.41	1.58	1.44
5	A	601	D4T	C3'-C2'	4.93	1.46	1.32

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	A	601	D4T	C4'-O4'-C1'	-7.22	103.08	109.99
5	A	601	D4T	C2-N3-C4	6.94	121.00	115.14
5	A	601	D4T	C5'-C4'-C3'	-2.45	111.13	115.61

There are no chirality outliers.

All (5) torsion outliers are listed below:

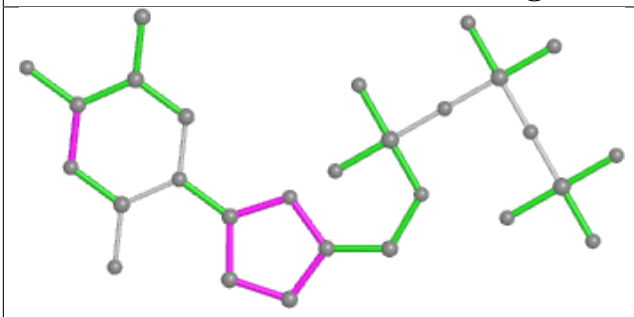
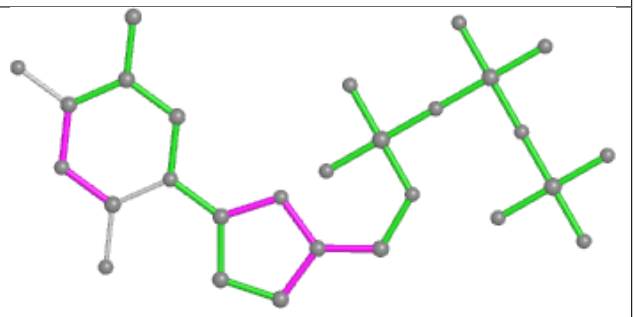
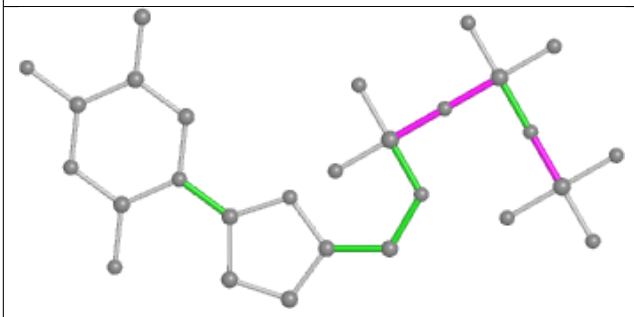
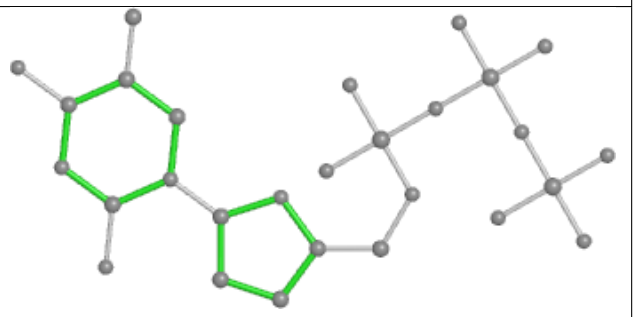
Mol	Chain	Res	Type	Atoms
5	A	601	D4T	PB-O7'-PC-O2C
5	A	601	D4T	PB-O6'-PA-O1A
5	A	601	D4T	PA-O6'-PB-O2B
5	A	601	D4T	PB-O7'-PC-O3C
5	A	601	D4T	PB-O6'-PA-O2A

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 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.

Ligand D4T A 601	
 Bond lengths	 Bond angles
 Torsions	 Rings

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	557/561 (99%)	0.04	9 (1%) 72 78	30, 46, 75, 93	0
2	B	394/452 (87%)	0.20	11 (2%) 53 60	32, 58, 85, 104	0
3	P	16/21 (76%)	-0.36	0 100 100	42, 63, 68, 69	0
4	T	23/27 (85%)	-0.02	0 100 100	39, 67, 93, 104	0
All	All	990/1061 (93%)	0.10	20 (2%) 65 72	30, 52, 82, 104	0

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	360	ALA	3.4
1	A	132	ILE	2.9
1	A	134	SER	2.9
1	A	133	PRO	2.8
2	B	164	MET	2.7

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	G47	P	817	25/26	0.94	0.13	54,60,76,83	0
3	DDG	P	822	21/22	0.97	0.17	37,40,43,45	0

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

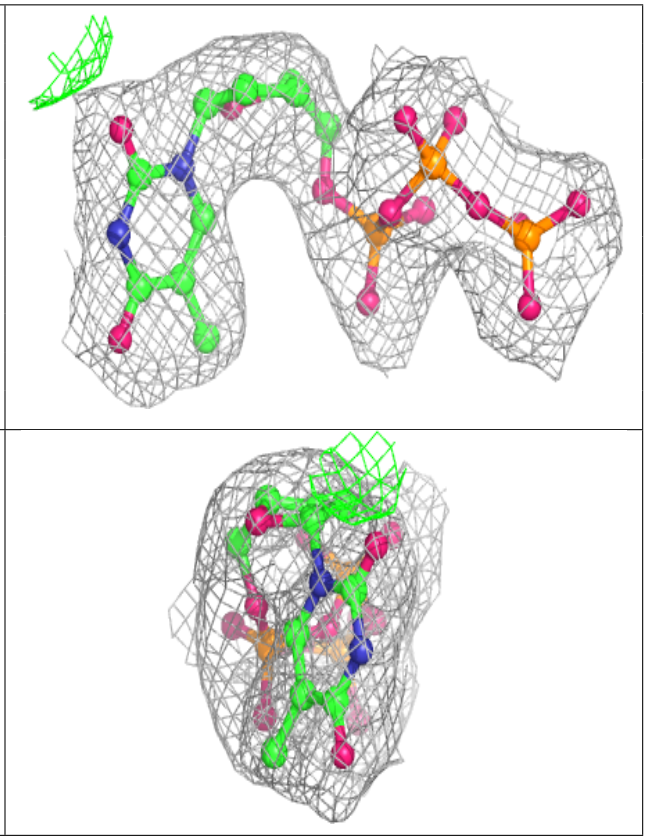
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists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

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
6	MG	A	602	1/1	0.94	0.12	35,35,35,35	0
6	MG	A	603	1/1	0.96	0.24	17,17,17,17	0
5	D4T	A	601	28/28	0.97	0.15	30,42,45,46	0
7	SO4	A	604	5/5	0.99	0.11	43,46,54,59	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 D4T A 601:

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.