



# Full wwPDB X-ray Structure Validation Report ⓘ

Oct 27, 2020 – 10:04 AM EDT

PDB ID : 6WPH  
Title : Structure of HIV-1 Reverse Transcriptase (RT) in complex with dsDNA and (-)-FTC  
Authors : Bertoletti, N.; Anderson, K.S.  
Deposited on : 2020-04-27  
Resolution : 2.72 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto: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 : 2.14.6  
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) : 2.14.6

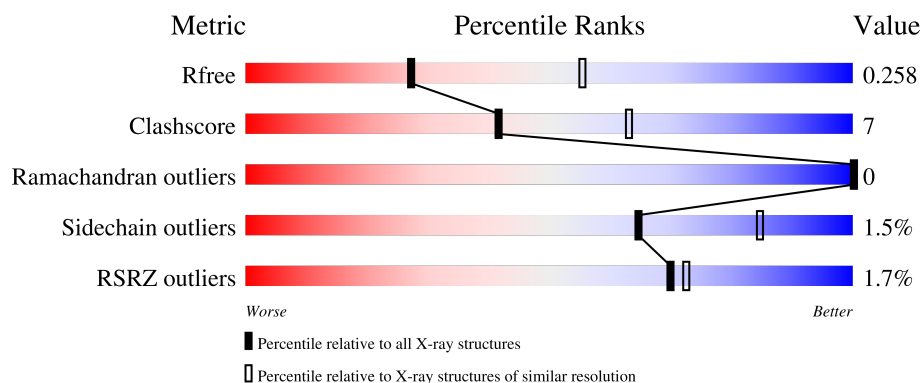
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.72 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	3359 (2.74-2.70)
Clashscore	141614	3686 (2.74-2.70)
Ramachandran outliers	138981	3622 (2.74-2.70)
Sidechain outliers	138945	3623 (2.74-2.70)
RSRZ outliers	127900	3276 (2.74-2.70)

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  $\geq 3$ , 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	560	<div> <div>81%</div> <div>16%</div> <div>..</div> </div>
2	B	452	<div> <div>2%</div> <div>71%</div> <div>15%</div> <div>13%</div> </div>
3	P	21	<div> <div>52%</div> <div>24%</div> <div>10%</div> <div>14%</div> </div>
4	T	27	<div> <div>4%</div> <div>33%</div> <div>48%</div> <div>19%</div> </div>

## 2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 8397 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
1	A	548	Total	C	N	O	S	0	0	0
			4359	2823	725	803	8			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
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
2	B	391	Total	C	N	O	S	0	0	0
			3121	2027	515	574	5			

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
3	P	18	Total	C	N	O	P	S	0	0	0
			367	175	64	110	17	1			

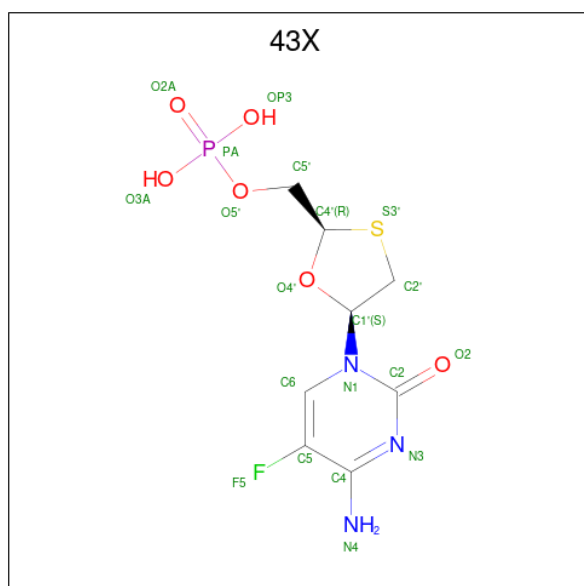
- Molecule 4 is a DNA chain called DNA template 27-mer.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
4	T	22	Total	C	N	O	P		0	0	0
			457	213	93	129	22				

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

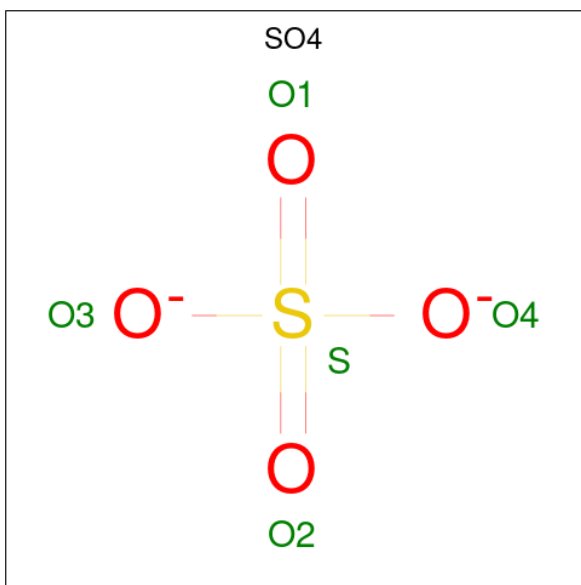
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	1	Total	Mg	0	0
			1	1		
5	A	1	Total	Mg	0	0
			1	1		

- Molecule 6 is [(2R,5S)-5-(4-amino-5-fluoro-2-oxypyrimidin-1(2H)-yl)-1,3-oxathiolan-2-yl]methyl dihydrogen phosphate (three-letter code: 43X) (formula: C<sub>8</sub>H<sub>11</sub>FN<sub>3</sub>O<sub>6</sub>PS) (labeled as "Ligand of Interest" by author).



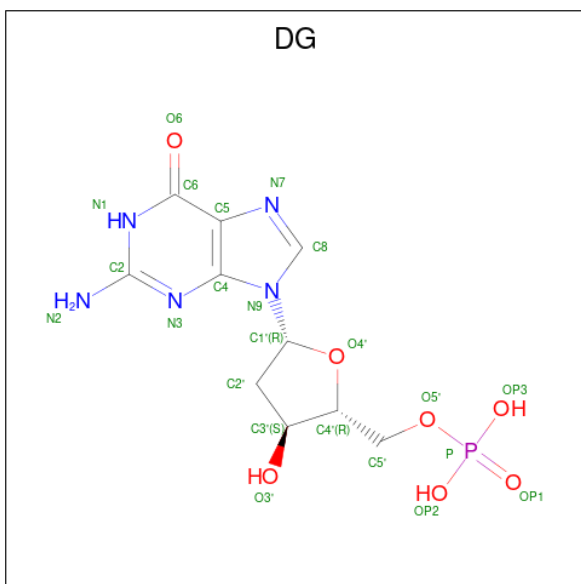
Mol	Chain	Residues	Atoms							ZeroOcc	AltConf
6	P	1	Total	C	F	N	O	P	S	0	1
			37	16	2	6	10	1	2		

- Molecule 7 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



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

- Molecule 8 is 2'-DEOXYGUANOSINE-5'-MONOPHOSPHATE (three-letter code: DG) (formula:  $C_{10}H_{14}N_5O_7P$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
8	T	1	Total	C	N	O	P	0
			22	10	5	6	1	0

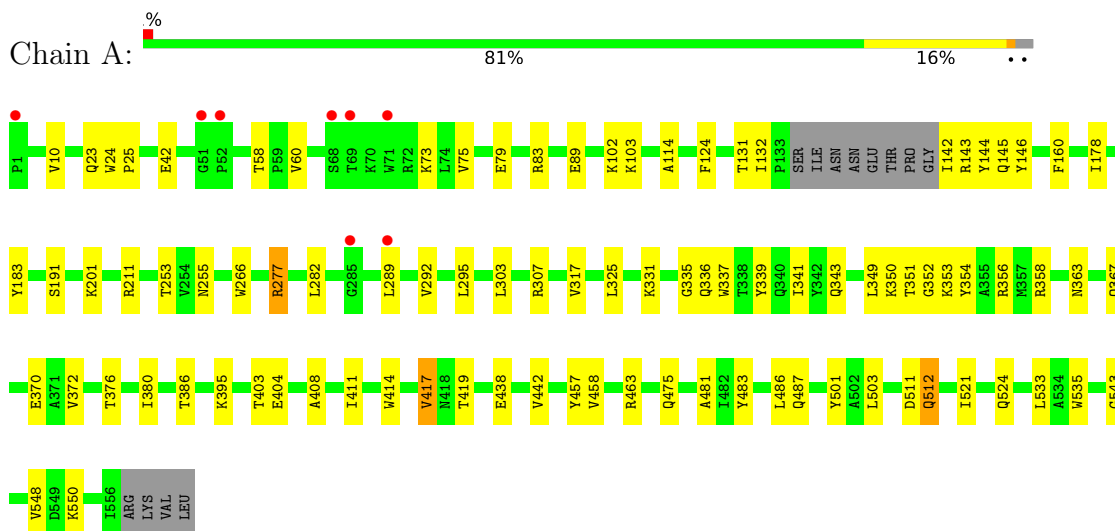
- Molecule 9 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	A	14	Total 14	O 14	0	0
9	B	4	Total 4	O 4	0	0
9	P	3	Total 3	O 3	0	0
9	T	1	Total 1	O 1	0	0

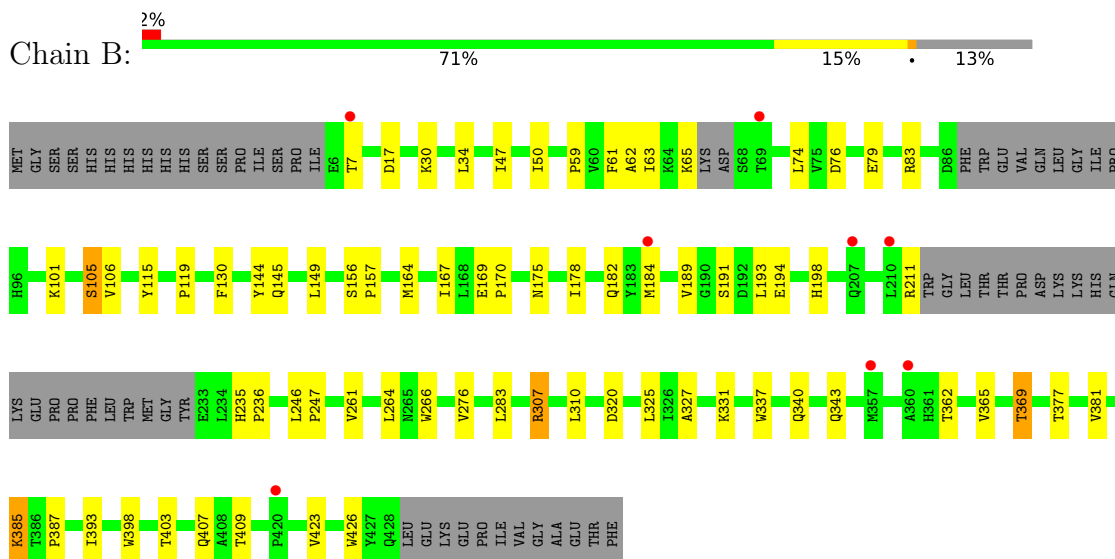
### 3 Residue-property plots [i](#)

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



- Molecule 2: p51 RT

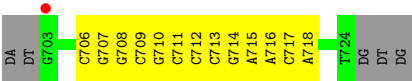


- Molecule 3: DNA Primer 21-mer





● Molecule 4: DNA template 27-mer



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	165.98Å 168.05Å 101.92Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.52 – 2.72 29.52 – 2.72	Depositor EDS
% Data completeness (in resolution range)	98.1 (29.52-2.72) 98.1 (29.52-2.72)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.16	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.95 (at 2.72Å)	Xtriage
Refinement program	PHENIX 1.14 _3260	Depositor
R, $R_{free}$	0.197 , 0.257 0.198 , 0.258	Depositor DCC
$R_{free}$ test set	1887 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	64.6	Xtriage
Anisotropy	0.510	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 36.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.001 for -k,-h,-l	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	8397	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	65.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.23% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: MG, SO4, 43X, G47

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	0/4474	0.56	0/6101
2	B	0.38	0/3206	0.56	1/4374 (0.0%)
3	P	0.91	0/380	1.10	2/582 (0.3%)
4	T	0.83	0/514	0.85	0/792
All	All	0.46	0/8574	0.62	3/11849 (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
1	A	0	1

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	385	LYS	C-N-CA	-6.78	104.75	121.70
3	P	822	DG	O4'-C1'-N9	5.37	111.76	108.00
3	P	820	DC	O4'-C4'-C3'	-5.29	102.38	104.50

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	183	TYR	Peptide

## 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	4359	0	4286	56	0
2	B	3121	0	3033	43	0
3	P	367	0	206	8	0
4	T	457	0	244	12	0
5	A	1	0	0	0	0
5	B	1	0	0	0	0
6	P	37	0	0	0	0
7	T	10	0	0	0	0
8	T	22	0	12	0	0
9	A	14	0	0	0	0
9	B	4	0	0	0	0
9	P	3	0	0	0	0
9	T	1	0	0	0	0
All	All	8397	0	7781	113	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 7.

All (113) 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	1.20	0.89
2:B:79:GLU:HG3	2:B:83:ARG:HE	1.53	0.73
2:B:7:THR:HG22	2:B:119:PRO:HB2	1.72	0.71
2:B:167:ILE:HG12	2:B:211:ARG:HD2	1.73	0.70
1:A:343:GLN:HG3	1:A:349:LEU:HD21	1.74	0.68
2:B:365:VAL:O	2:B:369:THR:HG23	1.94	0.68
1:A:23:GLN:HE22	1:A:60:VAL:H	1.43	0.66
4:T:716:DA:H2'	4:T:717:DC:O4'	1.96	0.66
1:A:60:VAL:HG22	1:A:75:VAL:HG22	1.76	0.65
2:B:261:VAL:HG13	2:B:276:VAL:HG11	1.79	0.63
3:P:816:DG:H2'	3:P:817:G47:C8	2.29	0.62
4:T:711:DC:H2'	4:T:712:DC:C6	2.34	0.62
2:B:59:PRO:HG2	2:B:76:ASP:HB3	1.82	0.60
3:P:817:G47:H2'2	3:P:818:DC:C6	2.37	0.60

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:511:ASP:OD1	1:A:512:GLN:HG3	2.03	0.59
1:A:372:VAL:HG11	1:A:411:ILE:HG23	1.83	0.58
1:A:79:GLU:O	1:A:83:ARG:HG2	2.03	0.58
1:A:331:LYS:HB2	1:A:337:TRP:CZ3	2.38	0.58
1:A:73:LYS:HE3	1:A:146:TYR:OH	2.05	0.57
1:A:363:ASN:HA	1:A:511:ASP:OD1	2.06	0.56
3:P:816:DG:H2'	3:P:817:G47:H8	1.88	0.55
1:A:131:THR:HG22	1:A:143:ARG:HD2	1.89	0.55
4:T:713:DC:H2''	4:T:714:DG:C8	2.41	0.55
1:A:132:ILE:HB	1:A:142:ILE:HG13	1.88	0.54
2:B:157:PRO:HG3	2:B:184:MET:HA	1.88	0.54
1:A:114:ALA:HB1	1:A:160:PHE:CZ	2.42	0.54
4:T:706:DC:H2'	4:T:707:DG:C8	2.43	0.54
1:A:42:GLU:HG3	1:A:144:TYR:HE1	1.74	0.53
2:B:175:ASN:HB3	2:B:178:ILE:HG12	1.89	0.53
1:A:442:VAL:HB	1:A:481:ALA:HB1	1.90	0.53
1:A:303:LEU:O	1:A:307:ARG:HG3	2.08	0.53
2:B:191:SER:OG	2:B:198:HIS:ND1	2.35	0.53
2:B:65:LYS:HA	2:B:407:GLN:HE22	1.73	0.52
1:A:178:ILE:HD11	1:A:201:LYS:HG2	1.91	0.52
2:B:247:PRO:O	2:B:307:ARG:NH2	2.43	0.52
1:A:438:GLU:OE2	1:A:463:ARG:HD3	2.09	0.52
2:B:149:LEU:HD13	2:B:156:SER:HA	1.93	0.51
1:A:543:GLY:HA3	2:B:283:LEU:O	2.11	0.51
2:B:47:ILE:HD12	2:B:144:TYR:CD2	2.47	0.50
2:B:130:PHE:CZ	2:B:144:TYR:HB2	2.45	0.50
2:B:74:LEU:HD11	2:B:409:THR:HA	1.94	0.49
2:B:246:LEU:HD11	2:B:264:LEU:HD21	1.93	0.49
2:B:61:PHE:CZ	2:B:74:LEU:HD23	2.47	0.49
1:A:266:TRP:CE2	3:P:820:DC:H4'	2.47	0.49
1:A:503:LEU:HD22	1:A:535:TRP:HB2	1.94	0.49
2:B:193:LEU:HD12	2:B:194:GLU:H	1.78	0.48
2:B:17:ASP:O	2:B:83:ARG:HD3	2.13	0.48
1:A:339:TYR:CZ	1:A:352:GLY:HA3	2.48	0.48
2:B:423:VAL:HA	2:B:426:TRP:CD1	2.48	0.48
1:A:266:TRP:CD2	3:P:820:DC:H4'	2.49	0.48
1:A:178:ILE:CD1	1:A:201:LYS:HG2	2.44	0.48
2:B:30:LYS:NZ	2:B:62:ALA:O	2.39	0.47
1:A:475:GLN:HB3	1:A:501:TYR:CE2	2.50	0.47
4:T:717:DC:H2'	4:T:718:DA:C8	2.48	0.47
2:B:164:MET:HG2	2:B:182:GLN:NE2	2.30	0.47

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:T:714:DG:H2''	4:T:715:DA:C8	2.50	0.47
2:B:50:ILE:HD13	2:B:145:GLN:HB3	1.97	0.46
1:A:42:GLU:HG3	1:A:144:TYR:CE1	2.50	0.46
2:B:377:THR:O	2:B:381:VAL:HG23	2.15	0.46
1:A:24:TRP:HB2	1:A:25:PRO:HD2	1.97	0.46
2:B:169:GLU:HB3	2:B:170:PRO:HD3	1.97	0.46
2:B:369:THR:HG22	2:B:398:TRP:CZ3	2.50	0.46
1:A:102:LYS:O	1:A:103:LYS:HD2	2.16	0.46
1:A:289:LEU:HD11	3:P:817:G47:H4'	1.97	0.46
1:A:343:GLN:HG3	1:A:349:LEU:CD2	2.43	0.45
1:A:376:THR:HG23	1:A:386:THR:HG23	1.99	0.45
2:B:266:TRP:CZ3	2:B:426:TRP:HB3	2.52	0.45
4:T:717:DC:H2''	4:T:718:DA:H5'	1.97	0.45
2:B:115:TYR:HB3	2:B:149:LEU:HB2	1.97	0.45
1:A:483:TYR:HB2	1:A:521:ILE:HG12	1.99	0.45
1:A:550:LYS:HB2	1:A:550:LYS:HE3	1.87	0.44
3:P:817:G47:H2'2	3:P:818:DC:H6	1.81	0.44
1:A:255:ASN:HD22	1:A:289:LEU:HD23	1.83	0.44
1:A:253:THR:HG22	1:A:292:VAL:HG22	1.99	0.44
2:B:369:THR:HG22	2:B:398:TRP:CH2	2.53	0.43
2:B:310:LEU:HD23	2:B:310:LEU:HA	1.78	0.43
2:B:331:LYS:HB2	2:B:337:TRP:CZ3	2.54	0.43
1:A:417:VAL:HG13	1:A:419:THR:HG23	2.00	0.43
4:T:709:DC:H2'	4:T:710:DG:H8	1.84	0.43
1:A:277:ARG:HB3	1:A:336:GLN:NE2	2.34	0.43
1:A:442:VAL:HG12	1:A:457:TYR:HB3	2.00	0.43
3:P:813:DT:H2''	3:P:814:DC:C6	2.54	0.43
4:T:716:DA:H2'	4:T:717:DC:C6	2.53	0.42
1:A:354:TYR:CE2	1:A:370:GLU:HG3	2.54	0.42
1:A:458:VAL:HG22	1:A:548:VAL:HG22	2.01	0.42
2:B:63:ILE:HD13	2:B:74:LEU:HD22	2.00	0.42
2:B:61:PHE:CE2	2:B:403:THR:HG22	2.54	0.42
2:B:101:LYS:O	2:B:236:PRO:HB2	2.20	0.42
1:A:10:VAL:HG23	1:A:124:PHE:CD1	2.54	0.42
2:B:105:SER:HB3	2:B:235:HIS:CE1	2.55	0.42
1:A:483:TYR:O	1:A:487:GLN:HG3	2.20	0.42
1:A:486:LEU:HB3	1:A:524:GLN:HB3	2.02	0.41
1:A:356:ARG:CZ	1:A:358:ARG:HD3	2.49	0.41
4:T:716:DA:H2'	4:T:717:DC:H6	1.85	0.41
1:A:335:GLY:HA2	1:A:367:GLN:OE1	2.21	0.41
1:A:395:LYS:HD3	1:A:414:TRP:CZ2	2.55	0.41

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:367:GLN:NE2	1:A:512:GLN:HE22	2.19	0.41
1:A:89:GLU:HG3	4:T:708:DG:OP1	2.20	0.41
2:B:327:ALA:HA	2:B:340:GLN:O	2.21	0.41
1:A:408:ALA:O	2:B:393:ILE:HG13	2.20	0.41
1:A:376:THR:O	1:A:380:ILE:HG13	2.21	0.41
2:B:34:LEU:HD23	2:B:34:LEU:HA	1.87	0.41
1:A:325:LEU:HD23	1:A:325:LEU:HA	1.93	0.41
1:A:533:LEU:HA	1:A:533:LEU:HD12	1.91	0.41
1:A:282:LEU:HD21	1:A:295:LEU:HD23	2.03	0.40
1:A:317:VAL:HG23	1:A:349:LEU:HD13	2.03	0.40
2:B:106:VAL:HA	2:B:189:VAL:O	2.21	0.40
1:A:341:ILE:O	1:A:349:LEU:N	2.54	0.40
1:A:403:THR:HG22	1:A:404:GLU:HG3	2.03	0.40
1:A:350:LYS:NZ	1:A:351:THR:O	2.54	0.40
2:B:325:LEU:HB2	2:B:387:PRO:HA	2.04	0.40
4:T:707:DG:H2'	4:T:708:DG:C8	2.56	0.40
2:B:325:LEU:HD12	2:B:385:LYS:HB3	2.03	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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	544/560 (97%)	531 (98%)	13 (2%)	0	100	100
2	B	383/452 (85%)	368 (96%)	15 (4%)	0	100	100
All	All	927/1012 (92%)	899 (97%)	28 (3%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains ⓘ

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	458/500 (92%)	450 (98%)	8 (2%)	60	83
2	B	327/411 (80%)	323 (99%)	4 (1%)	71	88
All	All	785/911 (86%)	773 (98%)	12 (2%)	65	85

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	58	THR
1	A	145	GLN
1	A	191	SER
1	A	211	ARG
1	A	277	ARG
1	A	353	LYS
1	A	417	VAL
1	A	512	GLN
2	B	105	SER
2	B	307	ARG
2	B	362	THR
2	B	369	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (8) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	23	GLN
1	A	197	GLN
1	A	340	GLN
1	A	512	GLN
1	A	520	GLN
2	B	182	GLN
2	B	343	GLN
2	B	407	GLN

### 5.3.3 RNA ⓘ

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$
3	G47	P	817	1,3,4	20,27,28	1.17	2 (10%)	21,38,41	2.26	8 (38%)

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	1,3,4	-	2/7/25/26	0/3/3/3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	P	817	G47	C6-C5	4.10	1.48	1.41
3	P	817	G47	C5-C4	2.48	1.47	1.40

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	P	817	G47	C2-N3-C4	5.22	121.20	115.28
3	P	817	G47	C5-C6-N1	-4.10	117.82	123.43
3	P	817	G47	C6-N1-C2	3.99	122.33	115.18
3	P	817	G47	C6-C5-C4	-3.71	117.25	120.80
3	P	817	G47	C4-C5-N7	-3.20	106.07	109.40
3	P	817	G47	C2'-C1'-N9	-2.86	107.66	114.27
3	P	817	G47	C6A-N2-C2	-2.59	119.26	123.75

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	P	817	G47	N3-C2-N1	-2.04	123.01	126.23

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	P	817	G47	N1-C2-N2-C6A
3	P	817	G47	N3-C2-N2-C6A

There are no ring outliers.

1 monomer is involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	P	817	G47	5	0

## 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 7 ligands modelled in this entry, 2 are monoatomic - leaving 5 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	DG	T	902	-	18,24,25	1.21	1 (5%)	19,35,38	2.60	4 (21%)
7	SO4	T	903	-	4,4,4	0.15	0	6,6,6	0.08	0
6	43X	P	901[B]	-	14,20,21	1.17	2 (14%)	16,28,31	1.66	3 (18%)
7	SO4	T	901	-	4,4,4	0.13	0	6,6,6	0.09	0
6	43X	P	901[A]	-	14,20,21	1.12	2 (14%)	16,28,31	1.38	2 (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
8	DG	T	902	-	-	0/3/21/22	0/3/3/3
6	43X	P	901[A]	-	-	2/3/18/19	0/2/2/2
6	43X	P	901[B]	-	-	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)
8	T	902	DG	C6-N1	4.04	1.40	1.33
6	P	901[B]	43X	C4-C5	3.03	1.42	1.40
6	P	901[A]	43X	C4-C5	2.99	1.42	1.40
6	P	901[B]	43X	C6-C5	-2.98	1.33	1.38
6	P	901[A]	43X	C6-C5	-2.71	1.33	1.38

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
8	T	902	DG	C5-C6-N1	-8.61	111.65	123.43
8	T	902	DG	C6-N1-C2	5.66	124.92	115.93
6	P	901[B]	43X	C2-N3-C4	4.08	120.94	116.02
6	P	901[A]	43X	C2-N3-C4	4.04	120.89	116.02
6	P	901[B]	43X	O5'-C5'-C4'	4.00	114.69	107.75
8	T	902	DG	C2-N3-C4	-2.92	112.02	115.36
8	T	902	DG	N3-C2-N1	-2.38	124.04	127.22
6	P	901[A]	43X	F5-C5-C6	2.07	121.58	117.51
6	P	901[B]	43X	F5-C5-C6	2.05	121.54	117.51

There are no chirality outliers.

All (3) torsion outliers are listed below:

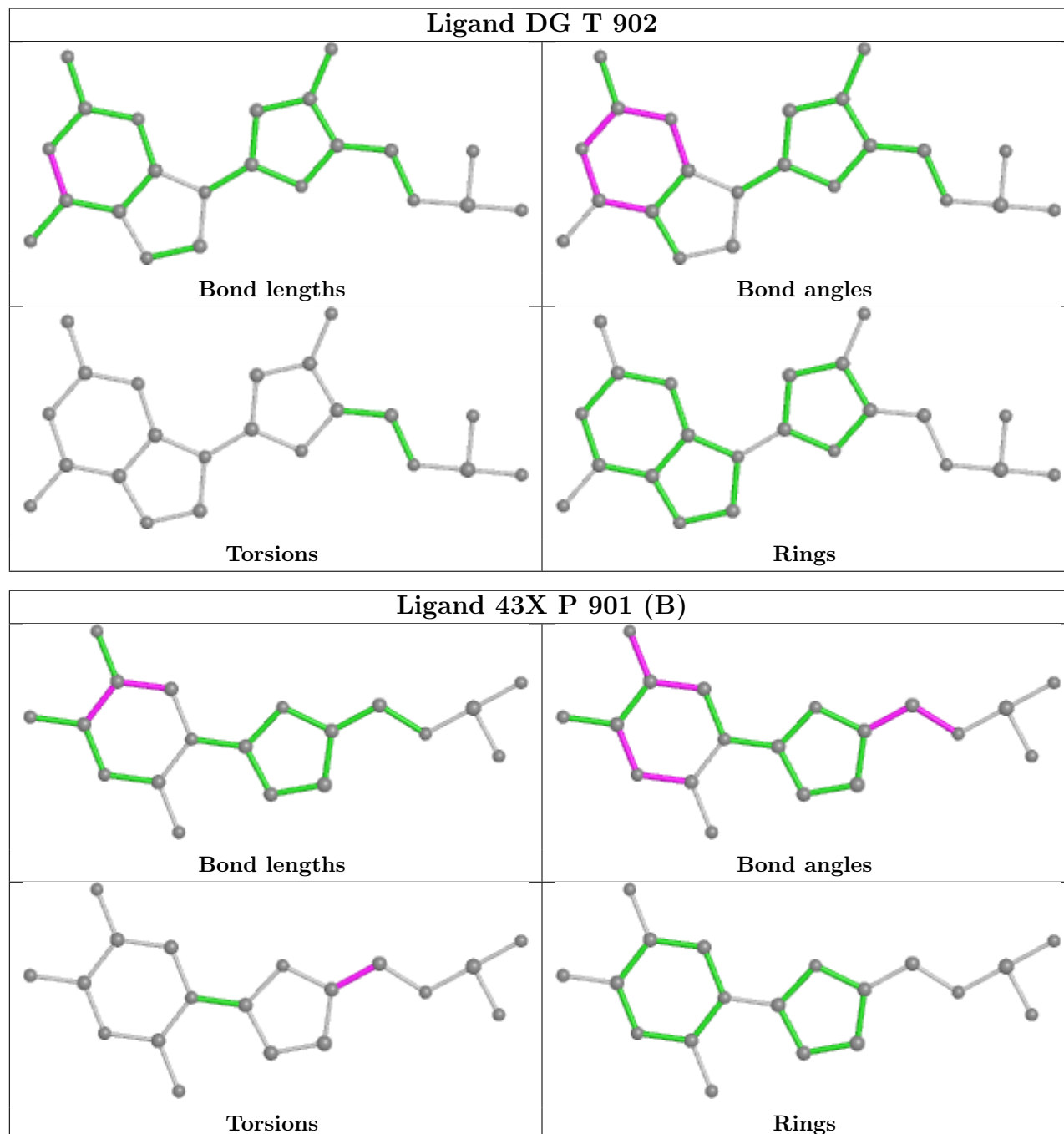
Mol	Chain	Res	Type	Atoms
6	P	901[A]	43X	O4'-C4'-C5'-O5'
6	P	901[A]	43X	S3'-C4'-C5'-O5'
6	P	901[B]	43X	O4'-C4'-C5'-O5'

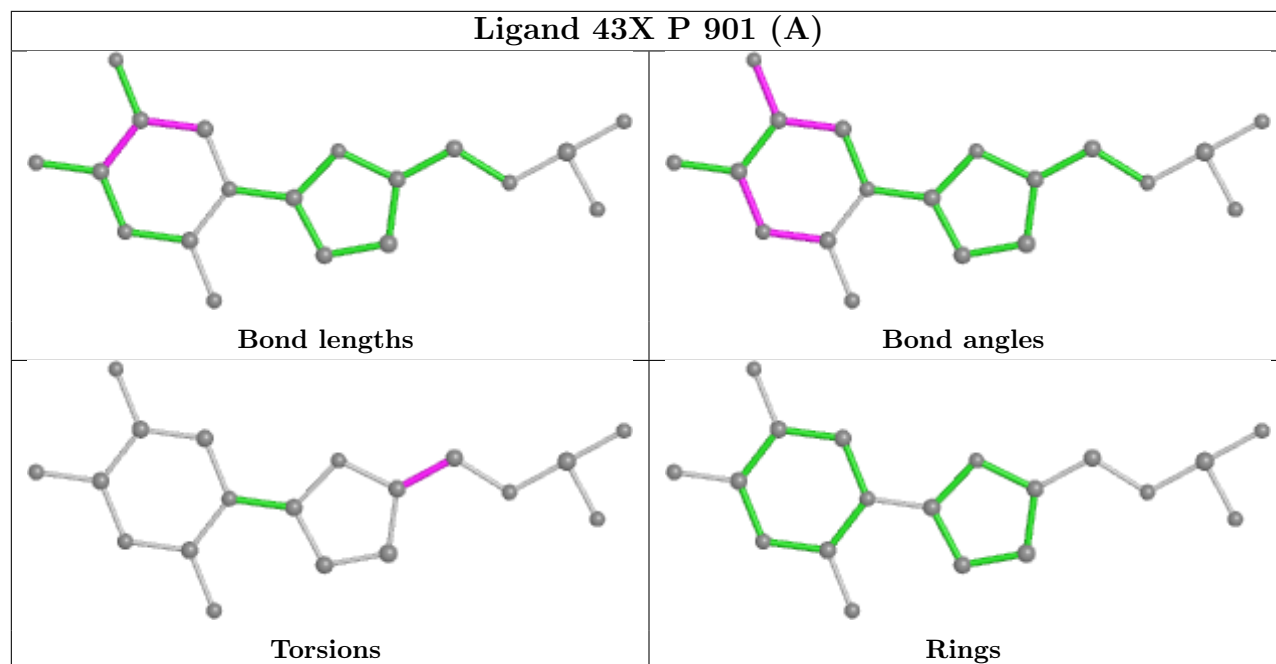
There are no ring outliers.

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

### 6.1 Protein, DNA and RNA chains

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	548/560 (97%)	-0.18	8 (1%) 73 76	36, 56, 104, 131	0
2	B	391/452 (86%)	-0.16	8 (2%) 65 67	36, 64, 110, 128	0
3	P	17/21 (80%)	-0.47	0 100 100	50, 63, 85, 90	0
4	T	22/27 (81%)	-0.13	1 (4%) 33 32	50, 73, 97, 154	0
All	All	978/1060 (92%)	-0.18	17 (1%) 70 72	36, 59, 107, 154	0

All (17) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	68	SER	4.7
1	A	52	PRO	4.1
1	A	69	THR	3.8
2	B	357	MET	3.2
2	B	184	MET	3.1
4	T	703	DG	2.8
2	B	210	LEU	2.7
2	B	360	ALA	2.6
1	A	51	GLY	2.6
1	A	289	LEU	2.5
2	B	69	THR	2.5
2	B	207	GLN	2.4
1	A	285	GLY	2.4
1	A	1	PRO	2.2
1	A	71	TRP	2.1
2	B	420	PRO	2.0
2	B	7	THR	2.0

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
3	G47	P	817	25/26	0.94	0.17	54,67,73,83	0

## 6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 6.4 Ligands [i](#)

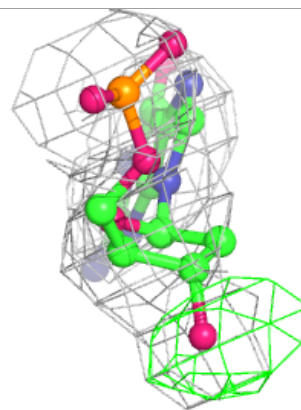
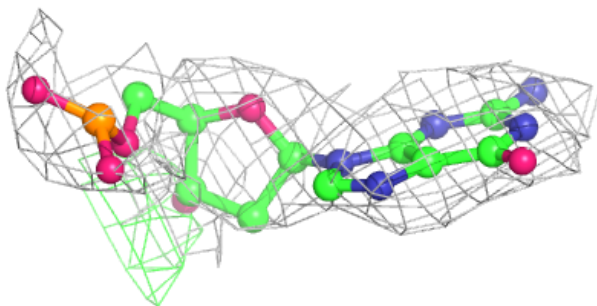
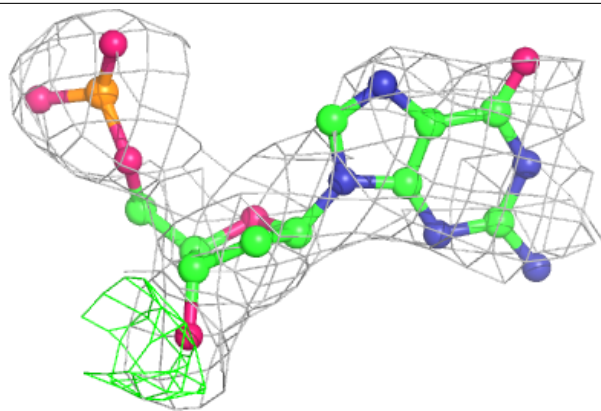
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q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

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
8	DG	T	902	22/23	0.84	0.26	101,112,115,116	0
6	43X	P	901[B]	19/20	0.86	0.29	79,83,86,86	18
6	43X	P	901[A]	19/20	0.86	0.29	80,84,86,86	18
7	SO4	T	903	5/5	0.89	0.20	114,115,116,116	0
7	SO4	T	901	5/5	0.89	0.18	137,138,139,139	0
5	MG	A	601	1/1	0.94	0.28	56,56,56,56	0
5	MG	B	501	1/1	0.94	0.38	62,62,62,62	1

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.

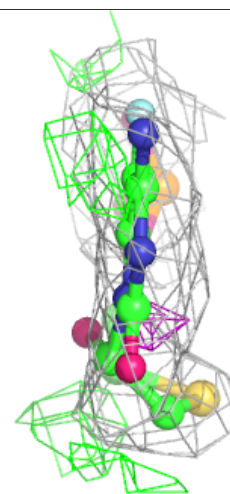
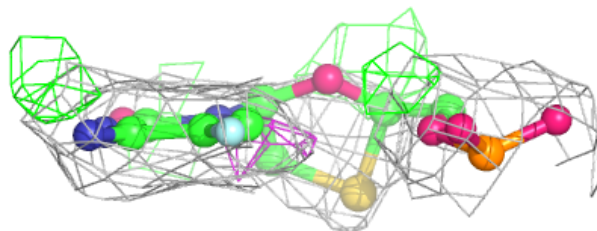
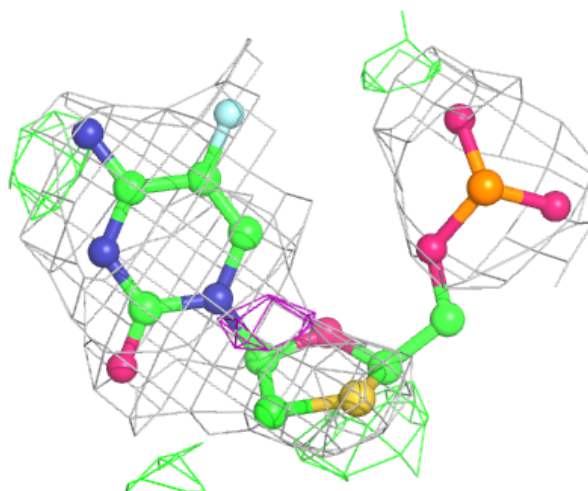
**Electron density around DG T 902:**

$2mF_o - DF_c$  (at 0.7 rmsd) in gray  
 $mF_o - DF_c$  (at 3 rmsd) in purple (negative)  
and green (positive)



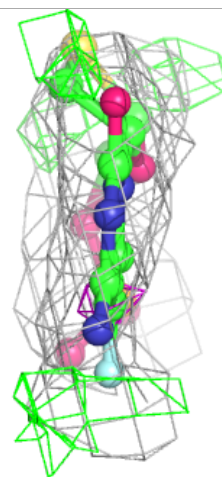
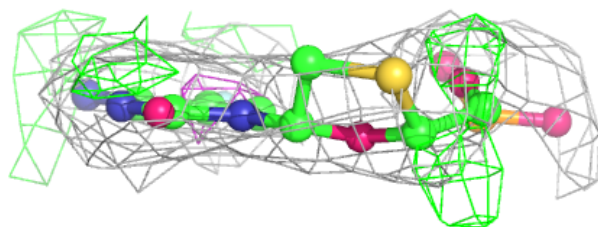
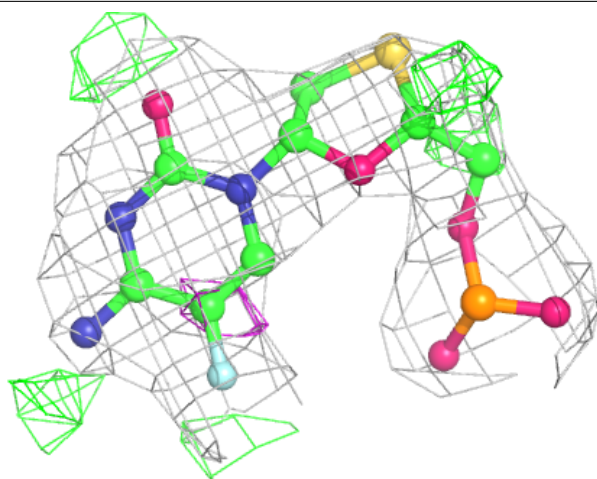
**Electron density around 43X P 901 (B):**

$2mF_o-DF_c$  (at 0.7 rmsd) in gray  
 $mF_o-DF_c$  (at 3 rmsd) in purple (negative)  
and green (positive)



**Electron density around 43X P 901 (A):**

$2mF_o-DF_c$  (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.