



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 9, 2016 – 06:20 PM EDT

PDB ID : 5C51  
Title : Probing the Structural and Molecular Basis of Nucleotide Selectivity by Human Mitochondrial DNA Polymerase gamma  
Authors : Sohl, C.D.; Szymanski, M.R.; Mislak, A.C.; Shumate, C.K.; Amiralaie, S.; Schinazi, R.F.; Anderson, K.S.; Whitney, Y.Y.  
Deposited on : 2015-06-19  
Resolution : 3.43 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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/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 : unknown  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027457  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027457

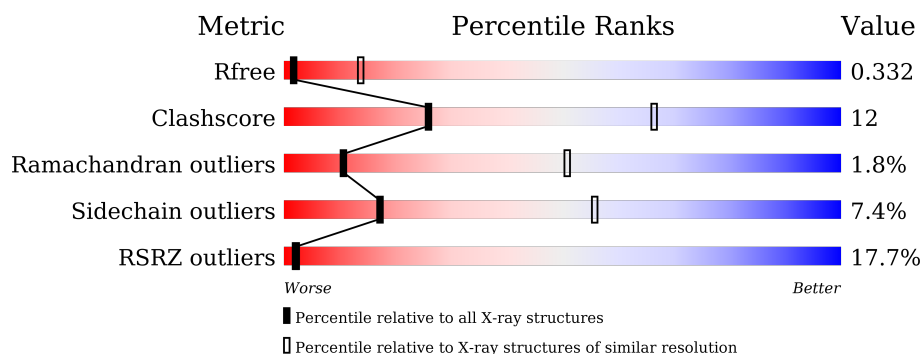
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.43 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))
$R_{free}$	91344	1049 (3.52-3.32)
Clashscore	102246	1032 (3.50-3.34)
Ramachandran outliers	100387	1002 (3.50-3.34)
Sidechain outliers	100360	1003 (3.50-3.34)
RSRZ outliers	91569	1054 (3.52-3.32)

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. 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	1205	<div> <div>15%</div> <div>55%</div> <div>23%</div> <div>18%</div> </div>
2	B	485	<div> <div>9%</div> <div>56%</div> <div>17%</div> <div>25%</div> </div>
2	C	485	<div> <div>13%</div> <div>55%</div> <div>17%</div> <div>26%</div> </div>
3	P	22	<div> <div>45%</div> <div>41%</div> <div>59%</div> </div>
4	T	25	<div> <div>28%</div> <div>32%</div> <div>64%</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	DOC	P	24	-	-	X	-
6	1RY	A	4003	-	-	X	-

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 14620 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called DNA polymerase subunit gamma-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	983	Total	C	N	O	S	0	0	0
			7802	4966	1374	1413	49			

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	GLN	deletion	UNP P54098
A	?	-	GLN	deletion	UNP P54098
A	?	-	GLN	deletion	UNP P54098
A	?	-	GLN	deletion	UNP P54098
A	?	-	GLN	deletion	UNP P54098
A	?	-	GLN	deletion	UNP P54098
A	?	-	GLN	deletion	UNP P54098
A	?	-	GLN	deletion	UNP P54098
A	?	-	GLN	deletion	UNP P54098
A	?	-	GLN	deletion	UNP P54098
A	948	ARG	ILE	conflict	UNP P54098

- Molecule 2 is a protein called DNA polymerase subunit gamma-2, mitochondrial.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	363	Total	C	N	O	S	0	0	0
			2943	1885	520	522	16			
2	C	358	Total	C	N	O	S	0	0	0
			2888	1852	506	514	16			

- Molecule 3 is a DNA chain called DNA (5'-D(\*(AD)P\*AP\*AP\*AP\*CP\*GP\*AP\*GP\*GP\*GP\*CP\*CP\*AP\*GP\*TP\*GP\*CP\*CP\*GP\*TP\*AP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	P	22	Total	C	N	O	P	0	0	0
			451	214	92	124	21			

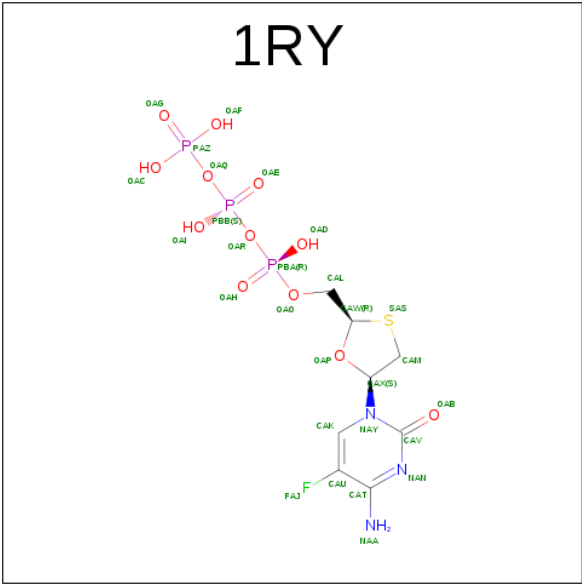
- Molecule 4 is a DNA chain called DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	T	25	Total	C	N	O	P	0	0	0
			506	241	87	154	24			

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

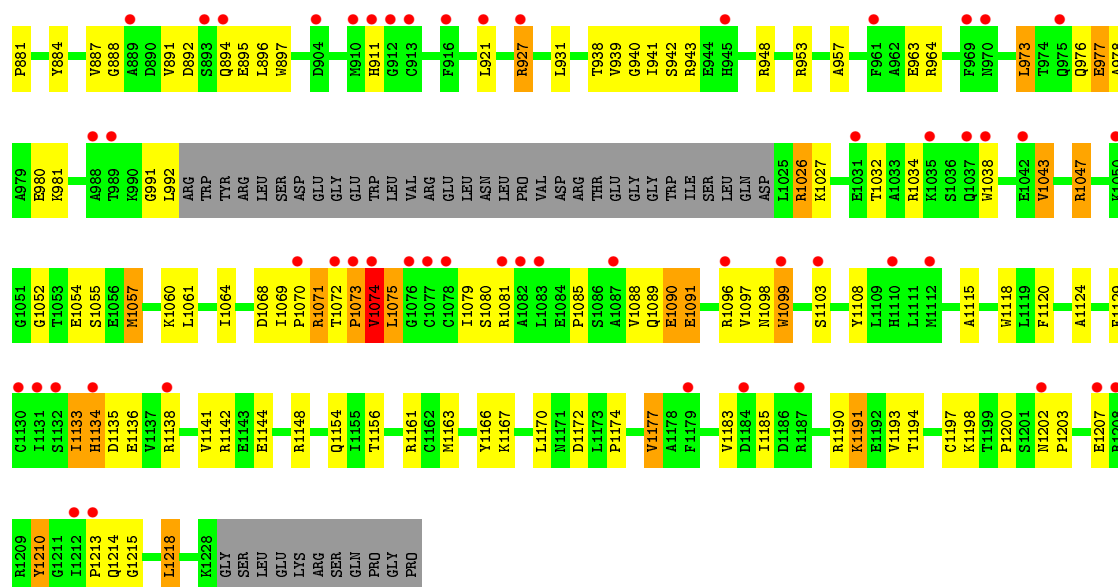
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	2	Total	Mg	0	0
			2	2		

- Molecule 6 is [[(2R,5S)-5-(4-azanyl-5-fluoranyl-2-oxidanylidene-pyrimidin-1-yl)-1,3-oxathiolan-2-yl]methoxy-oxidanyl-phosphoryl] phosphono hydrogen phosphate (three-letter code: 1RY) (formula: C<sub>8</sub>H<sub>13</sub>FN<sub>3</sub>O<sub>12</sub>P<sub>3</sub>S).

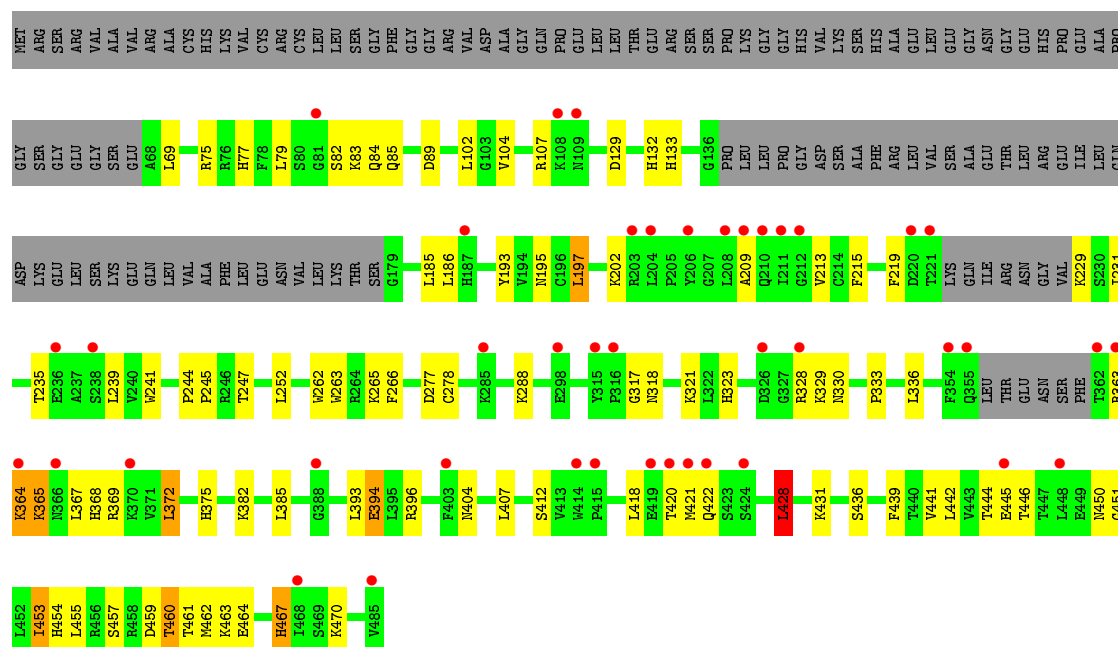


Mol	Chain	Residues	Atoms							ZeroOcc	AltConf
6	A	1	Total	C	F	N	O	P	S	0	0
			28	8	1	3	12	3	1		

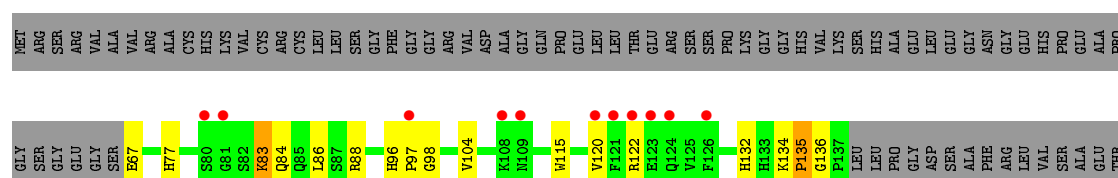


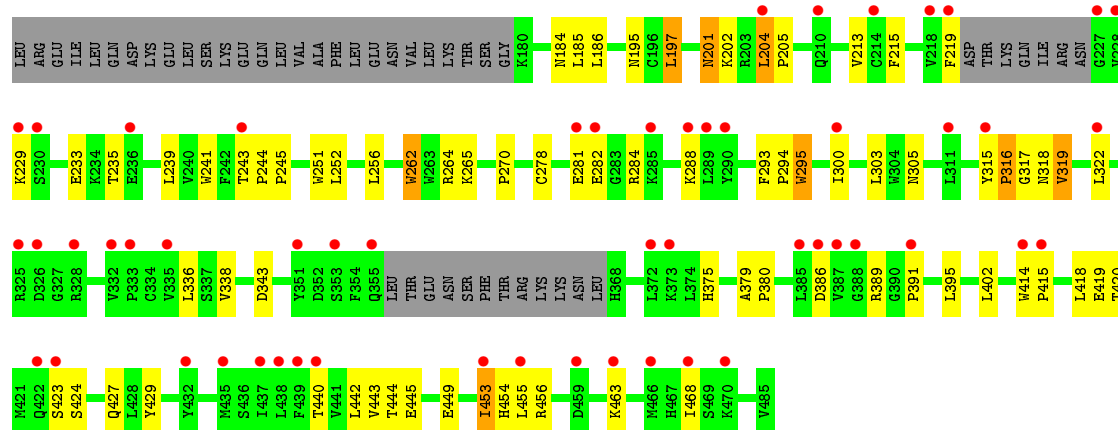


• Molecule 2: DNA polymerase subunit gamma-2, mitochondrial

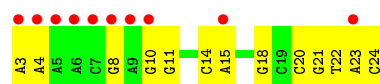


• Molecule 2: DNA polymerase subunit gamma-2, mitochondrial

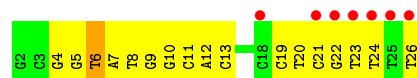




- Molecule 3: DNA (5'-D(\*((AD)P\*AP\*AP\*AP\*CP\*GP\*AP\*GP\*GP\*GP\*CP\*CP\*AP\*GP\*TP\*GP\*CP\*CP\*GP\*TP\*AP\*C)-3')



- Molecule 4: DNA





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	215.05Å 215.05Å 161.69Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.02 – 3.43 49.64 – 3.42	Depositor EDS
% Data completeness (in resolution range)	98.2 (47.02-3.43) 84.8 (49.64-3.42)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.32 (at 3.40Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, $R_{free}$	0.315 , 0.345 0.322 , 0.332	Depositor DCC
$R_{free}$ test set	1694 reflections (3.89%)	DCC
Wilson B-factor (Å <sup>2</sup> )	139.0	Xtriage
Anisotropy	0.414	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 8.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.87	EDS
Total number of atoms	14620	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.07% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: 1RY, MG, DOC

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	0.22	0/8002	0.43	2/10855 (0.0%)
2	B	0.22	0/3016	0.39	1/4074 (0.0%)
2	C	0.23	0/2961	0.42	2/4002 (0.0%)
3	P	0.48	0/488	0.75	0/752
4	T	0.44	0/565	1.13	1/870 (0.1%)
All	All	0.25	0/15032	0.49	6/20553 (0.0%)

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
2	C	135	PRO	CA-N-CD	-8.68	99.35	111.50
2	C	96	HIS	C-N-CD	-7.00	105.19	120.60
1	A	752	LEU	C-N-CD	-6.60	106.07	120.60
4	T	6	DT	N3-C4-O4	5.48	123.19	119.90
1	A	752	LEU	C-N-CA	5.08	143.34	122.00

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	A	7802	0	7694	212	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	2943	0	2939	64	0
2	C	2888	0	2862	65	1
3	P	451	0	245	31	0
4	T	506	0	279	37	0
5	A	2	0	0	0	0
6	A	28	0	11	15	0
All	All	14620	0	14030	350	1

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 12.

The worst 5 of 350 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)
6:A:4003:1RY:CAX	6:A:4003:1RY:OAP	1.63	1.38
1:A:948:ARG:NE	6:A:4003:1RY:NAA	1.86	1.23
1:A:948:ARG:NH2	3:P:24:DOC:N4	1.88	1.21
1:A:948:ARG:NH2	3:P:24:DOC:HN42	1.42	1.18
2:C:135:PRO:HD2	2:C:136:GLY:H	1.13	1.12

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:136:ASP:OD2	2:C:318:ASN:ND2[5_545]	2.19	0.01

## 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	969/1205 (80%)	837 (86%)	111 (12%)	21 (2%)	8 47

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	355/485 (73%)	326 (92%)	27 (8%)	2 (1%)	30	73
2	C	350/485 (72%)	324 (93%)	19 (5%)	7 (2%)	9	49
All	All	1674/2175 (77%)	1487 (89%)	157 (9%)	30 (2%)	11	51

5 of 30 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	752	LEU
1	A	1070	PRO
2	C	97	PRO
1	A	749	PHE
1	A	767	ALA

### 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	823/1017 (81%)	752 (91%)	71 (9%)	13	49
2	B	325/426 (76%)	306 (94%)	19 (6%)	25	65
2	C	317/426 (74%)	298 (94%)	19 (6%)	24	64
All	All	1465/1869 (78%)	1356 (93%)	109 (7%)	17	56

5 of 109 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1027	LYS
1	A	1120	PHE
2	C	262	TRP
1	A	1038	TRP
1	A	1074	VAL

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	1134	HIS

### 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	DOC	P	24	3,4	12,19,20	0.82	0	14,26,29	0.49	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
3	DOC	P	24	3,4	-	0/3/18/19	0/2/2/2

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	P	24	DOC	7	0

## 5.5 Carbohydrates

There are no carbohydrates in this entry.

## 5.6 Ligand geometry

Of 3 ligands modelled in this entry, 2 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	$\# Z  > 2$
6	1RY	A	4003	5	21,29,29	3.20	6 (28%)	24,45,45	1.81	3 (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
6	1RY	A	4003	5	-	0/18/31/31	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)
6	A	4003	1RY	CAW-SAS	-7.09	1.56	1.81
6	A	4003	1RY	CAM-SAS	-5.70	1.64	1.81
6	A	4003	1RY	CAV-NAN	-2.21	1.33	1.38
6	A	4003	1RY	CAL-CAW	3.27	1.60	1.50
6	A	4003	1RY	CAT-NAA	3.67	1.43	1.34

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	A	4003	1RY	CAW-OAP-CAX	-2.78	106.18	112.31
6	A	4003	1RY	NAA-CAT-NAN	2.73	120.92	116.92
6	A	4003	1RY	CAM-SAS-CAW	6.73	105.46	88.23

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 15 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	A	4003	1RY	15	0

## 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	983/1205 (81%)	1.13	185 (18%) 2 2	38, 67, 106, 401	0
2	B	363/485 (74%)	0.80	42 (11%) 6 7	42, 59, 96, 263	0
2	C	358/485 (73%)	1.00	65 (18%) 2 2	41, 65, 97, 154	0
3	P	21/22 (95%)	2.16	10 (47%) 0 0	101, 105, 106, 111	0
4	T	25/25 (100%)	1.68	7 (28%) 1 1	65, 82, 104, 115	0
All	All	1750/2222 (78%)	1.06	309 (17%) 2 2	38, 66, 105, 401	0

The worst 5 of 309 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	746	GLY	29.8
1	A	745	PRO	19.7
1	A	747	CYS	12.6
2	C	228	VAL	12.3
1	A	510	ALA	10.9

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
3	DOC	P	24	18/19	0.71	0.36	-	99,102,106,106	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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
6	1RY	A	4003	28/28	0.78	0.34	-0.29	63,66,78,79	0
5	MG	A	4001	1/1	0.77	0.40	-0.65	88,88,88,88	0
5	MG	A	4002	1/1	0.76	0.33	-	87,87,87,87	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.