



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 8, 2018 – 07:07 pm GMT

PDB ID : 2DPJ  
Title : structure of hPoli with DNA and dTTP  
Authors : Nair, D.T.; Johnson, R.E.; Prakash, L.; Prakash, S.; Aggarwal, A.K.  
Deposited on : 2006-05-12  
Resolution : 2.30 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

<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 : 1.7.3 (157068), CSD as539be (2018)  
Xtriage (Phenix) : 1.13  
EDS : trunk30967  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Refmac : 5.8.0158  
CCP4 : 7.0 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk30967

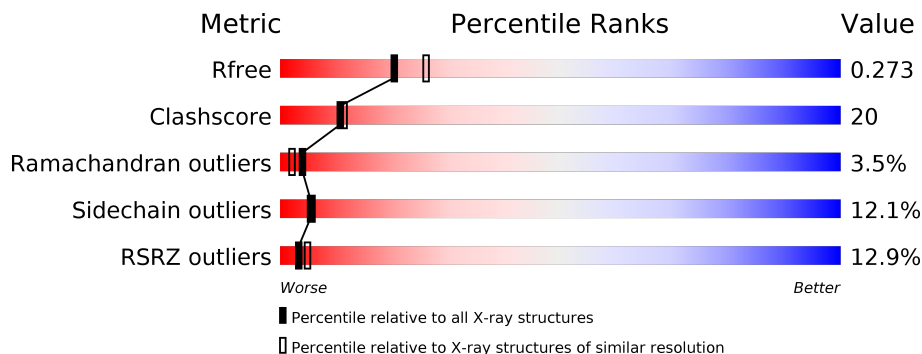
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

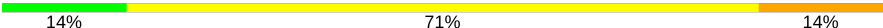

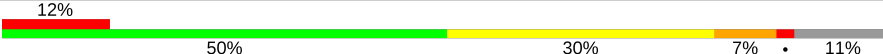
The reported resolution of this entry is 2.30 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	111664	4477 (2.30-2.30)
Clashscore	122126	5072 (2.30-2.30)
Ramachandran outliers	120053	5022 (2.30-2.30)
Sidechain outliers	120020	5021 (2.30-2.30)
RSRZ outliers	108989	4374 (2.30-2.30)

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	P	7	
2	T	9	
3	A	420	

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 3357 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 5'-D(\*AP\*GP\*GP\*AP\*CP\*CP\*(DOC))-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	P	7	Total	C	N	O	P	0	0	0
			139	67	29	37	6			

- Molecule 2 is a DNA chain called 5'-D(\*TP\*(EDA)P\*GP\*GP\*GP\*TP\*CP\*CP\*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	T	9	Total	C	N	O	P	0	0	0
			184	90	32	54	8			

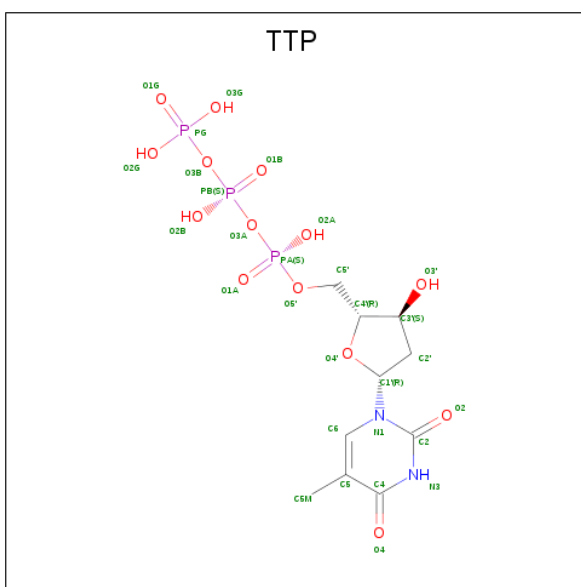
- Molecule 3 is a protein called DNA polymerase iota.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	373	Total	C	N	O	S	0	0	0
			2876	1809	504	542	21			

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

- Molecule 5 is THYMIDINE-5'-TRIPHOSPHATE (three-letter code: TTP) (formula: C<sub>10</sub>H<sub>17</sub>N<sub>2</sub>O<sub>14</sub>P<sub>3</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	
5	A	1	Total 29	C 10	N 2	O 14	P 3	0	0

- Molecule 6 is water.

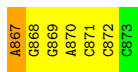
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	P	3	Total O 3 3	0	0
6	T	14	Total O 14 14	0	0
6	A	110	Total O 110 110	0	0

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

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: 5'-D(\*AP\*GP\*GP\*AP\*CP\*CP\*(DOC))-3'

Chain P: 



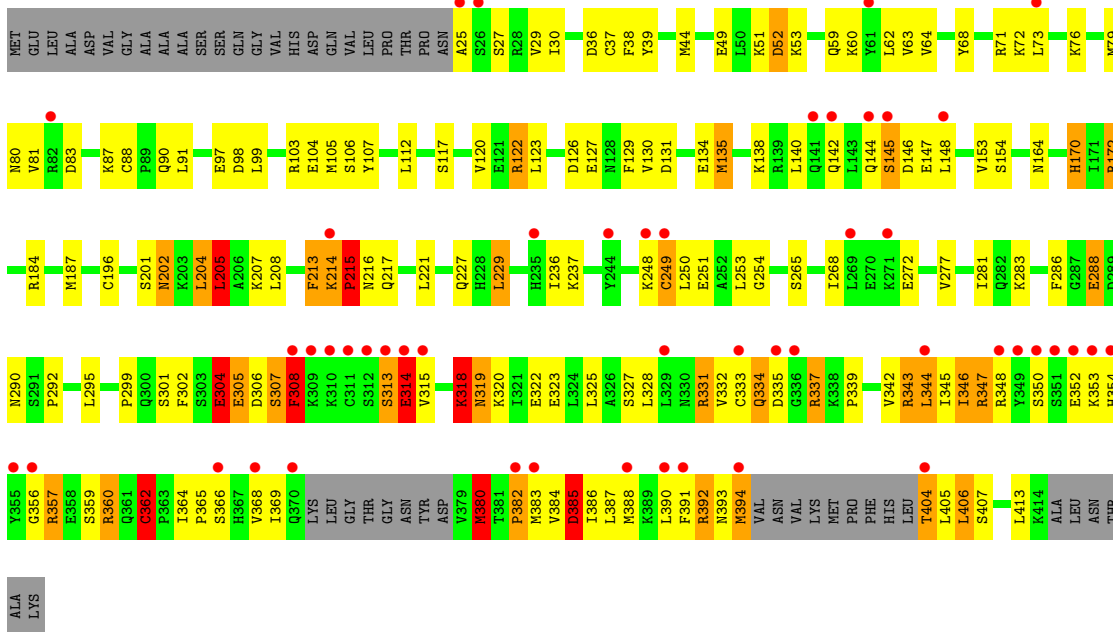
- Molecule 2: 5'-D(\*TP\*(EDA)P\*GP\*GP\*GP\*TP\*CP\*CP\*T)-3'

Chain T: 



- Molecule 3: DNA polymerase iota

Chain A: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	98.06Å 98.06Å 203.10Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	36.76 – 2.30 36.64 – 2.30	Depositor EDS
% Data completeness (in resolution range)	94.0 (36.76-2.30) 94.0 (36.64-2.30)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	8.14 (at 2.29Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, $R_{free}$	0.238 , 0.277 0.232 , 0.273	Depositor DCC
$R_{free}$ test set	2085 reflections (7.89%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	37.3	Xtriage
Anisotropy	0.236	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 51.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	3357	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.67% 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: DOC, MG, EDA, TTP

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	P	2.98	11/136 (8.1%)	3.74	28/208 (13.5%)
2	T	3.66	32/178 (18.0%)	3.93	48/271 (17.7%)
3	A	1.75	44/2914 (1.5%)	1.34	27/3934 (0.7%)
All	All	1.97	87/3228 (2.7%)	1.79	103/4413 (2.3%)

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	T	0	1
3	A	0	3
All	All	0	4

The worst 5 of 87 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	313	SER	CB-OG	30.32	1.81	1.42
3	A	318	LYS	CE-NZ	14.28	1.84	1.49
2	T	843	DG	P-O5'	12.34	1.72	1.59
2	T	841	DG	P-O5'	11.60	1.71	1.59
1	P	869	DG	N7-C5	10.56	1.45	1.39

The worst 5 of 103 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	P	870	DA	O4'-C1'-N9	15.09	118.57	108.00
2	T	843	DG	O4'-C1'-C2'	-14.57	94.24	105.90
1	P	870	DA	O4'-C4'-C3'	14.44	114.66	106.00

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	T	843	DG	C4'-C3'-C2'	-13.25	91.17	103.10
1	P	867	DA	O4'-C4'-C3'	-11.95	98.83	106.00

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	213	PHE	Peptide
3	A	346	ILE	Peptide
3	A	380	MET	Peptide
2	T	843	DG	Sidechain

## 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	P	139	0	79	0	1
2	T	184	0	104	10	1
3	A	2876	0	2899	121	0
4	A	2	0	0	0	0
5	A	29	0	13	0	0
6	A	110	0	0	4	0
6	P	3	0	0	0	0
6	T	14	0	0	1	0
All	All	3357	0	3095	124	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 20.

The worst 5 of 124 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)
3:A:318:LYS:CE	3:A:318:LYS:NZ	1.84	1.38
3:A:313:SER:OG	3:A:313:SER:CB	1.81	1.27
3:A:344:LEU:HD11	3:A:387:LEU:HD22	1.21	1.08

*Continued on next page...*



*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:T:842:DG:H2''	2:T:843:DG:H5'	1.44	0.97
3:A:249:CYS:O	3:A:253:LEU:HD12	1.69	0.93

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:P:867:DA:O5'	2:T:847:DT:O3'[10_665]	1.91	0.29

## 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
3	A	367/420 (87%)	329 (90%)	25 (7%)	13 (4%)	<b>4</b> <b>2</b>

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	308	PHE
3	A	314	GLU
3	A	315	VAL
3	A	333	CYS
3	A	353	LYS

### 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
3	A	321/376 (85%)	282 (88%)	39 (12%)	<b>5</b> <b>6</b>

5 of 39 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	229	LEU
3	A	308	PHE
3	A	404	THR
3	A	295	LEU
3	A	304	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	202	ASN
3	A	334	GLN
3	A	262	GLN
3	A	170	HIS
3	A	279	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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
1	DOC	P	873	1,2	13,19,20	<b>2.33</b>	<b>6 (46%)</b>	12,26,29	1.80	<b>4 (33%)</b>
2	EDA	T	840	2	18,26,27	<b>2.52</b>	<b>5 (27%)</b>	17,38,41	<b>2.17</b>	<b>5 (29%)</b>

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
1	DOC	P	873	1,2	-	0/3/18/19	0/2/2/2
2	EDA	T	840	2	-	0/3/21/22	0/4/4/4

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	T	840	EDA	O5'-C5'	-7.39	1.34	1.44
1	P	873	DOC	C3'-C2'	-3.80	1.43	1.54
2	T	840	EDA	C8-N9	-3.69	1.32	1.36
2	T	840	EDA	C5-C6	-3.45	1.35	1.41
1	P	873	DOC	O4'-C1'	-2.61	1.36	1.42

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	T	840	EDA	O4'-C1'-N9	-5.90	97.83	107.78
2	T	840	EDA	C4'-O4'-C1'	-3.05	102.02	109.43
1	P	873	DOC	O4'-C1'-N1	-2.32	103.87	107.78
1	P	873	DOC	C2'-C1'-N1	-2.17	108.39	112.48
2	T	840	EDA	C2-N3-C4	2.33	120.07	116.51

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	T	840	EDA	4	0

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

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	$\# Z  > 2$
5	TTP	A	875	4	22,30,30	1.50	5 (22%)	27,47,47	1.78	4 (14%)

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	TTP	A	875	4	-	0/18/34/34	0/2/2/2

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	875	TTP	PA-O1A	-2.59	1.41	1.50
5	A	875	TTP	C2'-C3'	-2.38	1.46	1.52
5	A	875	TTP	C5'-C4'	-2.03	1.45	1.51
5	A	875	TTP	C4-N3	2.82	1.38	1.33
5	A	875	TTP	O4'-C1'	3.60	1.50	1.42

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	875	TTP	PB-O3B-PG	-2.48	124.28	132.63
5	A	875	TTP	O3G-PG-O2G	2.30	116.71	107.59
5	A	875	TTP	C2'-C3'-C4'	2.54	108.15	102.76
5	A	875	TTP	C4-N3-C2	6.37	120.56	115.14

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	T	2

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	T	839:DT	O3'	840:EDA	P	1.35
1	T	840:EDA	O3'	841:DG	P	1.35

## 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, 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	P	6/7 (85%)	0.27	0 <a href="#">100</a> <a href="#">100</a>	26, 31, 37, 39	0
2	T	8/9 (88%)	0.53	1 (12%) <a href="#">4</a> <a href="#">5</a>	20, 22, 25, 88	0
3	A	373/420 (88%)	0.61	49 (13%) <a href="#">3</a> <a href="#">4</a>	8, 37, 79, 97	0
All	All	387/436 (88%)	0.60	50 (12%) <a href="#">3</a> <a href="#">5</a>	8, 36, 79, 97	0

The worst 5 of 50 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	354	HIS	14.0
3	A	355	TYR	13.5
3	A	25	ALA	11.8
3	A	353	LYS	6.9
3	A	350	SER	6.1

### 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
2	EDA	T	840	23/24	0.91	0.15	24,38,61,65	0
1	DOC	P	873	18/19	0.97	0.19	15,18,27,29	0

### 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 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
4	MG	A	872	1/1	0.93	0.24	78,78,78,78	0
5	TTP	A	875	29/29	0.97	0.14	13,18,22,24	0
4	MG	A	871	1/1	0.98	0.10	18,18,18,18	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.