



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 18, 2024 – 09:34 AM EST

PDB ID : 4FBU  
Title : Dpo4 polymerase pre-insertion binary complex with the N-(deoxyguanosin-8-yl)-1-aminopyrene lesion  
Authors : Kirouac, K.; Basu, A.; Ling, H.  
Deposited on : 2012-05-23  
Resolution : 2.60 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.36  
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) : 2.36

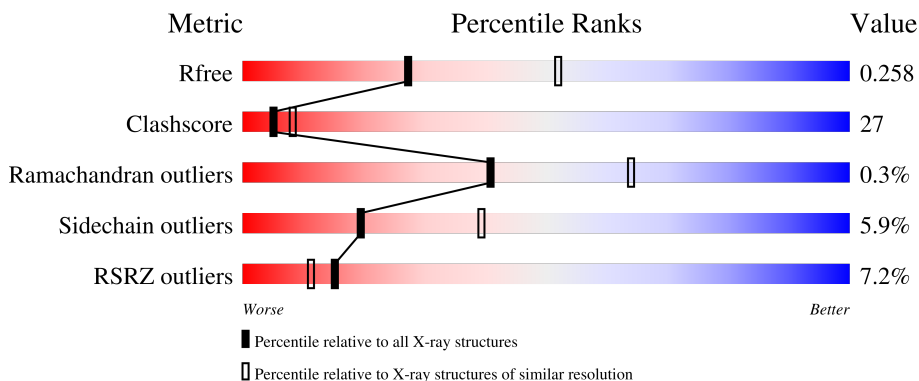
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.60 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	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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  $\geq 3$ , 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	341	<div style="display: flex; align-items: center;"> <div style="width: 7%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 65%; height: 10px; background-color: green; margin-right: 5px;"></div> <div style="width: 32%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 1%; height: 10px; background-color: orange; margin-right: 5px;"></div> <div style="width: 1%; height: 10px; background-color: grey;"></div> </div> <p style="margin-left: 20px;">7%      65%      32%      •</p>
1	B	341	<div style="display: flex; align-items: center;"> <div style="width: 5%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 65%; height: 10px; background-color: green; margin-right: 5px;"></div> <div style="width: 31%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 1%; height: 10px; background-color: orange; margin-right: 5px;"></div> <div style="width: 1%; height: 10px; background-color: grey;"></div> </div> <p style="margin-left: 20px;">5%      65%      31%      •</p>
2	D	13	<div style="display: flex; align-items: center;"> <div style="width: 23%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 8%; height: 10px; background-color: orange; margin-right: 5px;"></div> <div style="width: 77%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 15%; height: 10px; background-color: orange; margin-right: 5px;"></div> </div> <p style="margin-left: 20px;">23%      8%      77%      15%</p>
2	P	13	<div style="display: flex; align-items: center;"> <div style="width: 23%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 15%; height: 10px; background-color: orange; margin-right: 5px;"></div> <div style="width: 77%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 8%; height: 10px; background-color: orange; margin-right: 5px;"></div> </div> <p style="margin-left: 20px;">23%      15%      77%      8%</p>
3	C	14	<div style="display: flex; align-items: center;"> <div style="width: 21%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 21%; height: 10px; background-color: orange; margin-right: 5px;"></div> <div style="width: 57%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 21%; height: 10px; background-color: orange; margin-right: 5px;"></div> </div> <p style="margin-left: 20px;">21%      21%      57%      21%</p>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	T	14	

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	DG8	C	3	-	-	X	-
3	DG8	T	3	-	-	X	-

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 6855 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 IV.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	341	Total	C	N	O	S	0	0	0
			2738	1757	469	505	7			
1	B	341	Total	C	N	O	S	0	0	0
			2729	1751	466	505	7			

- Molecule 2 is a DNA chain called DNA primer.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	P	13	Total	C	N	O	P	0	0	0
			256	124	47	73	12			
2	D	13	Total	C	N	O	P	0	0	0
			256	124	47	73	12			

- Molecule 3 is a DNA chain called DNA template.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	T	14	Total	C	N	O	P	0	0	0
			312	155	57	86	14			
3	C	14	Total	C	N	O	P	0	0	0
			312	155	57	86	14			

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	
4	A	2	Total	Ca	0	0
			2	2		
4	B	2	Total	Ca	0	0
			2	2		

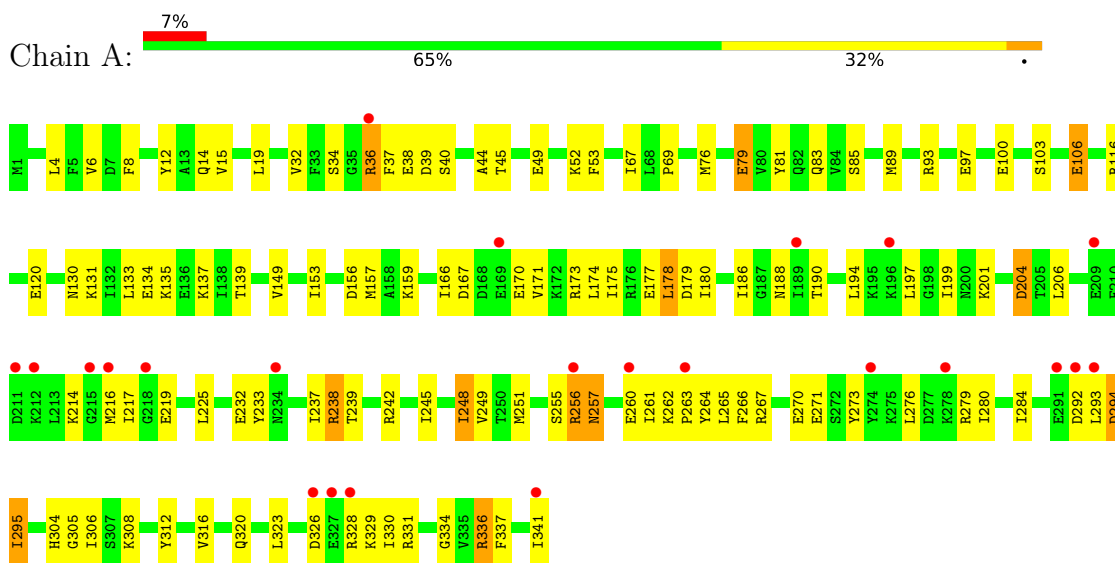
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	108	Total 108	O 108	0	0
5	B	97	Total 97	O 97	0	0
5	P	7	Total 7	O 7	0	0
5	T	12	Total 12	O 12	0	0
5	D	11	Total 11	O 11	0	0
5	C	13	Total 13	O 13	0	0

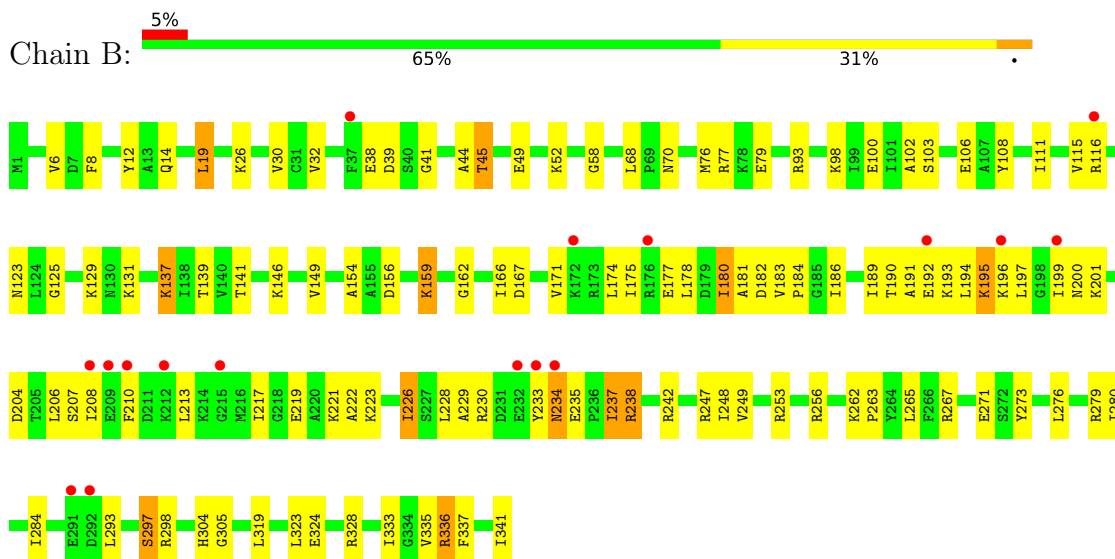
### 3 Residue-property plots

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

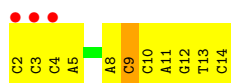
- Molecule 1: DNA polymerase IV



- Molecule 1: DNA polymerase IV



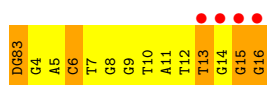
- Molecule 2: DNA primer



- Molecule 2: DNA primer



- Molecule 3: DNA template



- Molecule 3: DNA template



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	52.77Å 181.46Å 52.74Å 90.00° 109.49° 90.00°	Depositor
Resolution (Å)	31.24 – 2.60 31.24 – 2.60	Depositor EDS
% Data completeness (in resolution range)	90.0 (31.24-2.60) 90.0 (31.24-2.60)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.30 (at 2.61Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.6.4_486)	Depositor
R, $R_{free}$	0.217 , 0.261 0.211 , 0.258	Depositor DCC
$R_{free}$ test set	1324 reflections (5.13%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	50.6	Xtrriage
Anisotropy	0.326	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 70.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.098 for l,-k,h	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	6855	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	68.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.76% 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: CA, DG8

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	0/2777	0.57	2/3729 (0.1%)
1	B	0.41	0/2768	0.58	0/3718
2	D	0.97	0/286	1.44	6/438 (1.4%)
2	P	0.86	0/286	1.31	1/438 (0.2%)
3	C	0.74	0/306	1.56	7/472 (1.5%)
3	T	0.72	0/306	1.61	6/472 (1.3%)
All	All	0.50	0/6729	0.84	22/9267 (0.2%)

There are no bond length outliers.

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	13	DT	O4'-C1'-N1	-8.80	101.84	108.00
3	T	15	DG	P-O3'-C3'	-8.46	109.55	119.70
3	T	16	DG	O4'-C4'-C3'	-8.29	101.03	106.00
2	P	9	DC	O4'-C1'-N1	7.71	113.40	108.00
2	D	13	DT	O4'-C4'-C3'	-7.19	101.62	104.50

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	2738	0	2878	113	1
1	B	2729	0	2858	119	0
2	D	256	0	144	21	0
2	P	256	0	144	23	0
3	C	312	0	169	49	1
3	T	312	0	168	57	0
4	A	2	0	0	0	0
4	B	2	0	0	0	0
5	A	108	0	0	6	0
5	B	97	0	0	11	0
5	C	13	0	0	0	0
5	D	11	0	0	0	0
5	P	7	0	0	0	0
5	T	12	0	0	0	0
All	All	6855	0	6361	348	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 27.

The worst 5 of 348 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:C:15:DG:C8	3:C:15:DG:H5''	1.43	1.50
3:T:3:DG8:C5'	3:T:6:DC:H41	1.45	1.28
3:C:3:DG8:O1P	3:C:3:DG8:H8	1.36	1.21
3:C:15:DG:C8	3:C:15:DG:C5'	2.30	1.14
3:C:15:DG:H4'	3:C:15:DG:OP1	1.41	1.11

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:206:LEU:O	3:C:16:DG:C2'[2_556]	2.06	0.14

## 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	339/341 (99%)	305 (90%)	33 (10%)	1 (0%)	41	64
1	B	339/341 (99%)	309 (91%)	29 (9%)	1 (0%)	41	64
All	All	678/682 (99%)	614 (91%)	62 (9%)	2 (0%)	41	64

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	199	ILE
1	B	180	ILE

### 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	299/300 (100%)	283 (95%)	16 (5%)	22	44
1	B	297/300 (99%)	278 (94%)	19 (6%)	17	35
All	All	596/600 (99%)	561 (94%)	35 (6%)	19	39

5 of 35 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	237	ILE
1	B	238	ARG
1	B	297	SER
1	A	323	LEU
1	A	295	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	188	ASN
1	B	304	HIS
1	A	304	HIS
1	B	14	GLN
1	B	70	ASN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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	DG8	C	3	-	38,45,46	1.42	8 (21%)	49,68,71	1.03	4 (8%)
3	DG8	T	3	-	38,45,46	1.48	10 (26%)	49,68,71	1.02	2 (4%)

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	DG8	C	3	-	-	5/5/25/26	0/7/7/7
3	DG8	T	3	-	-	3/5/25/26	0/7/7/7

The worst 5 of 18 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	3	DG8	C25-C30	3.38	1.50	1.43
3	T	3	DG8	C25-C30	3.07	1.49	1.43
3	T	3	DG8	C5-C6	-2.64	1.42	1.47

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	T	3	DG8	C35-C28	2.61	1.48	1.42
3	C	3	DG8	C35-C28	2.60	1.48	1.42

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	T	3	DG8	C25-N24-C8	-2.94	119.92	128.05
3	T	3	DG8	C2'-C1'-N9	2.85	119.13	116.01
3	C	3	DG8	C25-C30-C35	-2.60	115.85	118.98
3	C	3	DG8	C2'-C3'-C4'	-2.50	97.55	102.76
3	C	3	DG8	O6-C6-C5	2.49	129.24	124.37

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	T	3	DG8	C3'-C4'-C5'-O5'
3	T	3	DG8	O4'-C4'-C5'-O5'
3	C	3	DG8	C30-C25-N24-C8
3	C	3	DG8	C26-C25-N24-C8
3	C	3	DG8	O4'-C4'-C5'-O5'

There are no ring outliers.

2 monomers are involved in 50 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	3	DG8	21	0
3	T	3	DG8	29	0

## 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

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.

No monomer is involved in short contacts.

## 5.7 Other polymers

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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, 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	341/341 (100%)	0.30	23 (6%) 17 13	22, 59, 122, 176	0
1	B	341/341 (100%)	0.23	17 (4%) 28 23	21, 53, 110, 145	0
2	D	13/13 (100%)	0.98	3 (23%) 0 0	46, 71, 132, 137	0
2	P	13/13 (100%)	1.14	3 (23%) 0 0	58, 78, 124, 157	0
3	C	13/14 (92%)	0.94	3 (23%) 0 0	42, 67, 169, 173	0
3	T	13/14 (92%)	1.00	4 (30%) 0 0	58, 77, 166, 178	0
All	All	734/736 (99%)	0.32	53 (7%) 15 11	21, 58, 123, 178	0

The worst 5 of 53 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	P	2	DC	5.4
1	B	196	LYS	4.8
1	B	212	LYS	4.6
3	C	16	DG	4.0
1	B	291	GLU	4.0

### 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	DG8	C	3	39/40	0.71	0.37	74,91,103,113	0
3	DG8	T	3	39/40	0.73	0.30	68,81,88,90	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
4	CA	B	402	1/1	0.60	0.33	67,67,67,67	0
4	CA	A	402	1/1	0.86	0.31	62,62,62,62	0
4	CA	B	401	1/1	0.92	0.13	53,53,53,53	0
4	CA	A	401	1/1	0.94	0.16	58,58,58,58	0

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

There are no such residues in this entry.