



wwPDB X-ray Structure Validation Summary Report

Feb 27, 2014 – 01:34 AM GMT

PDB ID : 3F2D
Title : DNA Polymerase PolC from *Geobacillus kaustophilus* complex with DNA, dGTP, Mn and Zn
Authors : Davies, D.R.; Evans, R.J.; Bullard, J.M.; Christensen, J.; Green, L.S.; Guiles, J.W.; Ribble, W.K.; Janjic, N.; Jarvis, T.C.
Deposited on : 2008-10-29
Resolution : 2.51 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

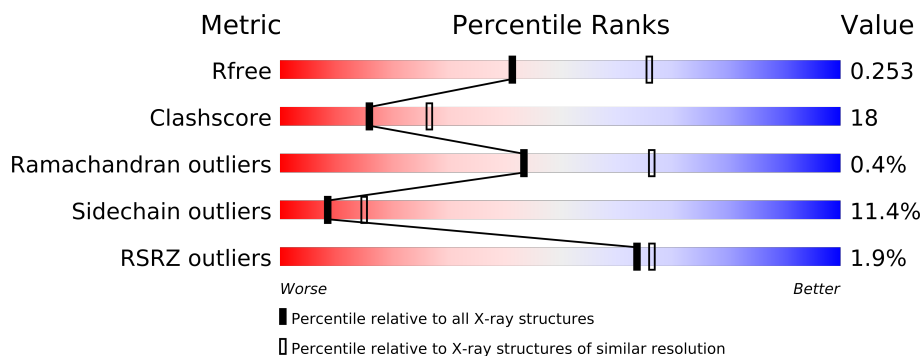
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.15 2013
Xtriage (Phenix) : dev-1323
EDS : stable22639
Percentile statistics : 21963
Refmac : 5.8.0049
CCP4 : 6.3.0 (Settle)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : stable22683

1 Overall quality at a glance

The reported resolution of this entry is 2.51 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}	66092	2784 (2.50-2.50)
Clashscore	79885	3562 (2.50-2.50)
Ramachandran outliers	78287	3480 (2.50-2.50)
Sidechain outliers	78261	3482 (2.50-2.50)
RSRZ outliers	66119	2785 (2.50-2.50)

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 ≥ 3 , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	P	17	
2	T	22	
3	A	1041	

2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 8574 atoms, of which 0 are hydrogen and 0 are deuterium.

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(*DCP*DAP*DGP*DTP*DGP*DAP*DGP*DAP*DCP*DGP*DGP*DGP*DCP*DAP*DAP*DCP*DC)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	P	13	Total	C	N	O	P	0	0	0
			269	126	57	73	13			

- Molecule 2 is a DNA chain called 5'-D(*DAP*DTP*DAP*DAP*DCP*DGP*DGP*DTP*DTP*DGP*DCP*DCP*DCP*DGP*DTP*DCP*DTP*DCP*DAP*DCP*DTP*DG)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	T	18	Total	C	N	O	P	0	0	0
			362	174	63	108	17			

- Molecule 3 is a protein called GEOBACILLUS KAUSTOPHILUS DNA POLC.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	993	Total	C	N	O	S	0	0	0
			7796	4961	1318	1481	36			

There are 17 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	227	MET	-	EXPRESSION TAG	UNP Q5L0J3
A	425	GLY	-	LINKER	UNP Q5L0J3
A	426	SER	-	LINKER	UNP Q5L0J3
A	427	GLY	-	LINKER	UNP Q5L0J3
A	428	SER	-	LINKER	UNP Q5L0J3
A	429	GLY	-	LINKER	UNP Q5L0J3
A	1445	SER	-	EXPRESSION TAG	UNP Q5L0J3
A	1446	GLY	-	EXPRESSION TAG	UNP Q5L0J3
A	1447	GLY	-	EXPRESSION TAG	UNP Q5L0J3
A	1448	HIS	-	EXPRESSION TAG	UNP Q5L0J3
A	1449	HIS	-	EXPRESSION TAG	UNP Q5L0J3
A	1450	HIS	-	EXPRESSION TAG	UNP Q5L0J3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	1451	HIS	-	EXPRESSION TAG	UNP Q5L0J3
A	1452	HIS	-	EXPRESSION TAG	UNP Q5L0J3
A	1453	HIS	-	EXPRESSION TAG	UNP Q5L0J3
A	1454	HIS	-	EXPRESSION TAG	UNP Q5L0J3
A	1455	HIS	-	EXPRESSION TAG	UNP Q5L0J3

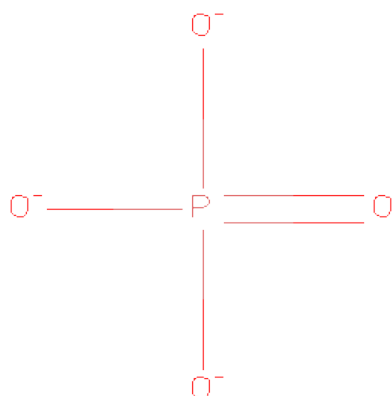
- Molecule 4 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	3	Total	Mn	0	0
			3	3		

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

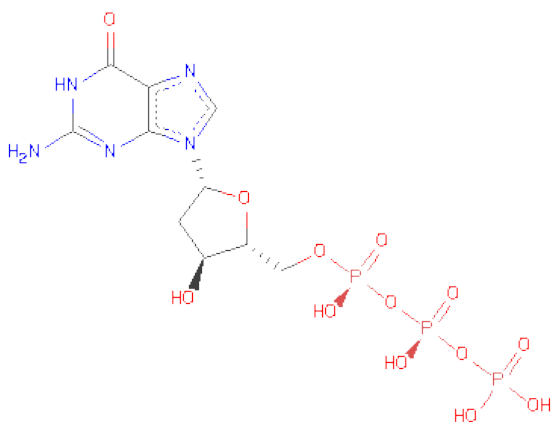
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	2	Total	Zn	0	0
			2	2		

- Molecule 6 is PHOSPHATE ION (three-letter code: PO4) (formula: O₄P).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	O	P	0	0
			5	4	1		
6	A	1	Total	O	P	0	0
			5	4	1		

- Molecule 7 is 2'-DEOXYGUANOSINE-5'-TRIPHOSPHATE (three-letter code: DGT) (formula: C₁₀H₁₆N₅O₁₃P₃).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
7	A	1	Total	C	N	O	P	0	0
			31	10	5	13	3		

- Molecule 8 is water.

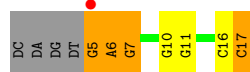
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	P	1	Total	O	0	0
			1	1		
8	T	8	Total	O	0	0
			8	8		
8	A	92	Total	O	0	0
			92	92		

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

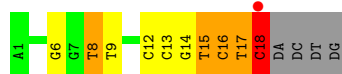
- Molecule 1: 5'-D(*DCP*DAP*DGP*DTP*DGP*DAP*DGP*DAP*DCP*DGP*DGP*DGP*DCP*DAP*DAP*DCP*DC)-3'

Chain P: 



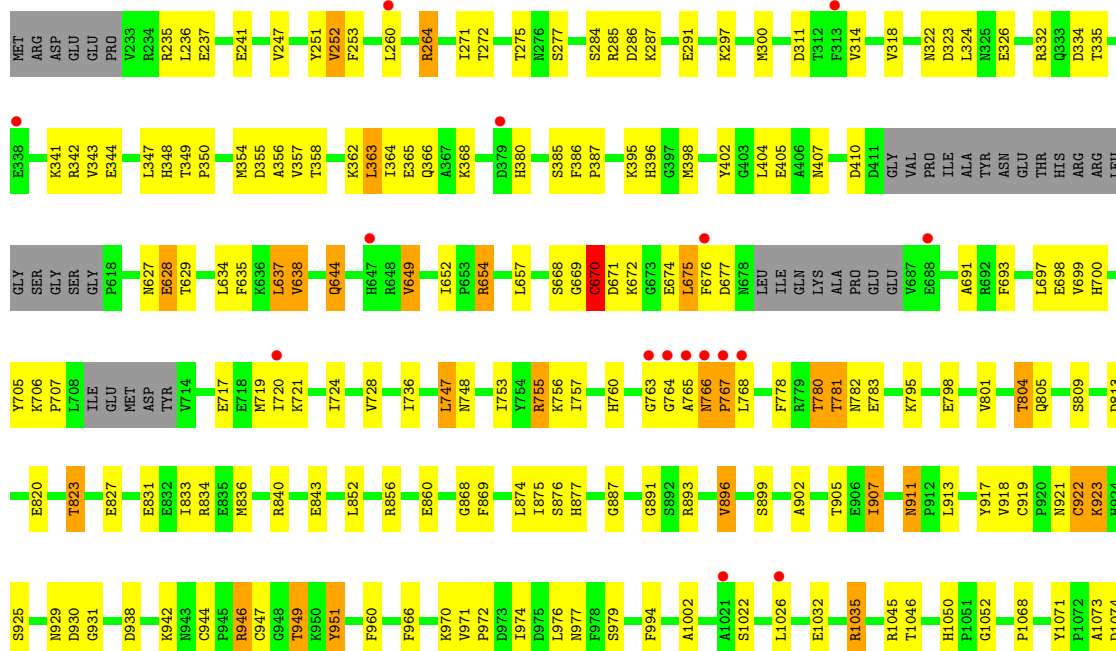
- Molecule 2: 5'-D(*DAP*DTP*DAP*DAP*DCP*DGP*DGP*DTP*DTP*DGP*DCP*DCP*DCP*DGP*DTP*DCP*DTP*DCP*DAP*DCP*DTP*DG)-3'

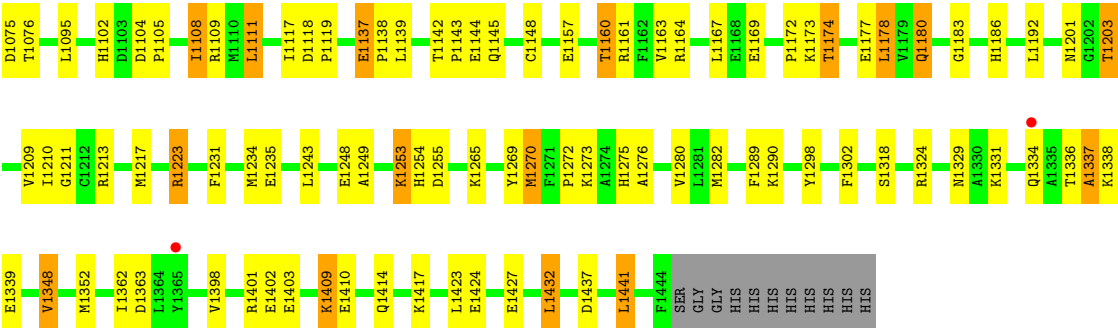
Chain T: 



- Molecule 3: GEOBACILLUS KAUSTOPHILUS DNA POLC

Chain A: 





4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	116.00Å 139.48Å 184.07Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.15 – 2.51 43.16 – 2.51	Depositor EDS
% Data completeness (in resolution range)	(Not available) (43.15-2.51) 99.2 (43.16-2.51)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.24 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.4.0069	Depositor
R, R_{free}	0.216 , 0.254 0.216 , 0.253	Depositor DCC
R_{free} test set	2573 reflections (5.34%)	DCC
Wilson B-factor (Å ²)	47.7	Xtriage
Anisotropy	0.060	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 35.2	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	1 of 50748 reflections (0.002%)	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	8574	wwPDB-VP
Average B, all atoms (Å ²)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.04% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, MN, PO4, DGT

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	2.60	12/303 (4.0%)	2.78	28/465 (6.0%)
2	T	2.09	13/404 (3.2%)	3.27	30/621 (4.8%)
3	A	0.58	1/7960 (0.0%)	0.71	1/10776 (0.0%)
All	All	0.87	26/8667 (0.3%)	1.15	59/11862 (0.5%)

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

The worst 5 of 26 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	P	7	DG	N1-C2	19.35	1.53	1.37
1	P	7	DG	C8-N7	16.06	1.40	1.30
1	P	6	DA	N9-C4	15.93	1.47	1.37
2	T	18	DC	C3'-C2'	-15.20	1.34	1.52
1	P	7	DG	C5-C4	13.93	1.48	1.38

The worst 5 of 59 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	T	18	DC	O4'-C1'-C2'	33.13	132.41	105.90
2	T	18	DC	C1'-O4'-C4'	-21.28	88.82	110.10
2	T	18	DC	C4'-C3'-C2'	21.07	122.06	103.10
2	T	18	DC	C2-N1-C1'	-20.70	96.03	118.80
2	T	18	DC	N1-C1'-C2'	-20.46	73.73	112.60

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	T	18	DC	Sidechain

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	P	269	0	142	24	0
2	T	362	0	205	19	0
3	A	7796	0	7707	266	0
4	A	3	0	0	0	0
5	A	2	0	0	0	0
6	A	10	0	0	0	0
7	A	31	0	12	7	0
8	A	92	0	0	7	0
8	P	1	0	0	0	0
8	T	8	0	0	2	0
All	All	8574	0	8066	305	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 18.

The worst 5 of 305 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:P:5:DG:H1'	1:P:6:DA:N7	1.35	1.38
1:P:5:DG:H1'	1:P:6:DA:C8	1.61	1.35
3:A:947:CYS:SG	3:A:949:THR:OG1	1.97	1.21
3:A:877:HIS:HA	3:A:907:ILE:HD11	1.25	1.17
2:T:18:DC:H6	2:T:18:DC:H5''	1.11	1.12

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
3	A	985/1041 (95%)	932 (95%)	49 (5%)	4 (0%)	43 66

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	670	CYS
3	A	767	PRO
3	A	1052	GLY
3	A	1337	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
3	A	850/904 (94%)	753 (89%)	97 (11%)	8 15

5 of 97 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	813	ASP
3	A	923	LYS
3	A	1403	GLU
3	A	827	GLU
3	A	896	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	911	ASN
3	A	943	ASN
3	A	1200	GLN
3	A	803	ASN
3	A	805	GLN

5.3.3 RNA ⓘ

There are no RNA chains in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

Of 8 ligands modelled in this entry, 5 are monoatomic - leaving 3 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	$\# Z > 2$
7	DGT	A	1456	4	33,33,33	1.42	4 (12%)	49,52,52	4.24	14 (28%)
6	PO4	A	6	5,4	4,4,4	0.46	0	6,6,6	0.32	0
6	PO4	A	7	-	4,4,4	0.26	0	6,6,6	0.31	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
7	DGT	A	1456	4	-	0/20/34/34	0/1/3/3
6	PO4	A	6	5,4	-	0/0/0/0	0/0/0/0
6	PO4	A	7	-	-	0/0/0/0	0/0/0/0

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	A	1456	DGT	C6-C5	4.12	1.47	1.41
7	A	1456	DGT	C2-N2	3.10	1.37	1.32
7	A	1456	DGT	C5-C4	2.70	1.46	1.40
7	A	1456	DGT	C4-N9	-2.68	1.33	1.37

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	A	1456	DGT	C6-C5-N7	26.13	137.66	134.14
7	A	1456	DGT	C2'-C1'-N9	-5.33	100.23	114.08
7	A	1456	DGT	O4'-C1'-N9	5.02	117.12	107.68
7	A	1456	DGT	C2-N3-C4	4.41	121.28	115.09
7	A	1456	DGT	C5-C4-N3	-4.00	120.14	125.94

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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 ⓘ

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, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q < 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ > 2	OWAB(Å ²)	Q < 0.9
1	P	13/17 (76%)	0.26	1 (7%) 13 13	32, 47, 85, 90	0
2	T	18/22 (81%)	0.00	1 (5%) 24 24	32, 46, 70, 81	0
3	A	993/1041 (95%)	0.05	18 (1%) 65 68	32, 45, 59, 84	0
All	All	1024/1080 (94%)	0.05	20 (1%) 64 64	32, 45, 60, 90	0

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	767	PRO	5.4
3	A	765	ALA	4.3
1	P	5	DG	3.8
3	A	766	ASN	3.7
3	A	764	GLY	3.2

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

6.4 Ligands ⓘ

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, 95th percentile and maximum values of B factors

of atoms in the group. The column labelled 'Q < 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors(\AA^2)	Q<0.9
6	PO4	A	6	5/5	0.21	0.25	68,69,70,70	0
6	PO4	A	7	5/5	0.14	-0.53	80,81,82,83	0
7	DGT	A	1456	31/31	0.14	-1.03	25,28,30,30	0
4	MN	A	3	1/1	0.18	-1.15	36,36,36,36	0
4	MN	A	2	1/1	0.14	-1.23	63,63,63,63	0
5	ZN	A	4	1/1	0.07	-1.85	79,79,79,79	0
5	ZN	A	5	1/1	0.14	-2.26	53,53,53,53	0
4	MN	A	1	1/1	0.14	-2.76	30,30,30,30	0

6.5 Other polymers ⓘ

There are no such residues in this entry.