wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 08:32 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 X-ray Structure 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/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

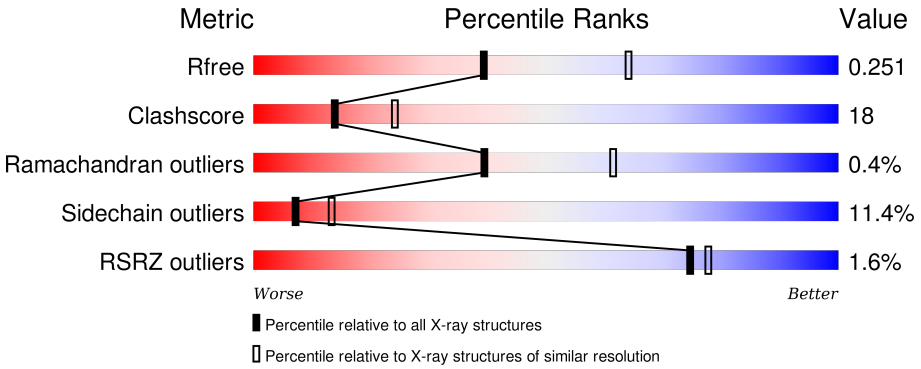
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 2.51 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}	91344	3553 (2.50-2.50)
Clashscore	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (2.50-2.50)
RSRZ outliers	91569	3562 (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. 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 <=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	17	6% 35% 18% 24% 24%
2	T	22	36% 23% 18% 5% 18%
3	A	1041	% 68% 23% • 5%

2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 8574 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(*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

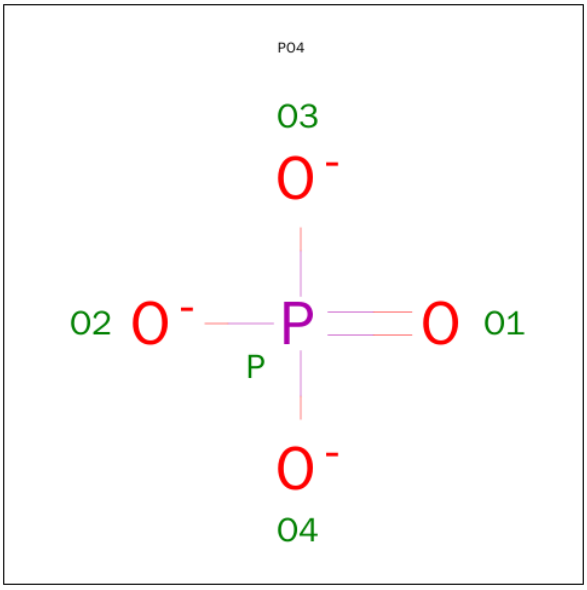
- Molecule 4 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

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

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

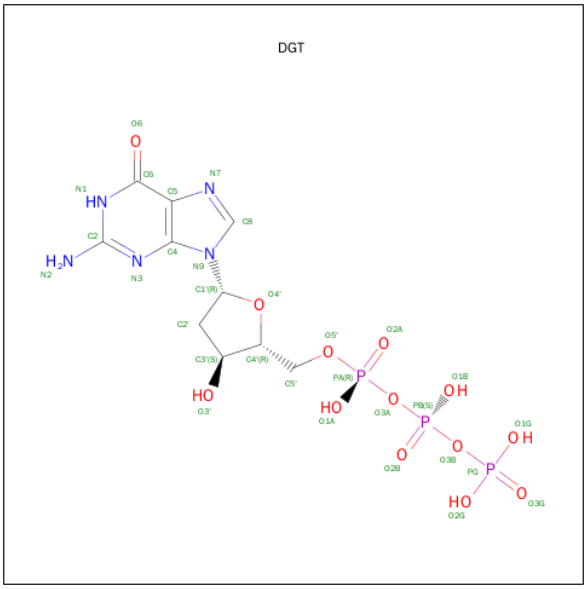
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	2	Total Zn 2 2	0	0

- Molecule 6 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



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

- Molecule 7 is 2'-DEOXYGUANOSINE-5'-TRIPHOSPHATE (three-letter code: DGT) (formula: $C_{10}H_{16}N_5O_{13}P_3$).



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.

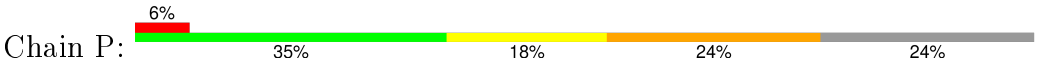
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	P	1	Total	O	0	0
			1	1		
8	T	8	Total	O	0	0
			8	8		
8	A	92	Total	O	0	0
			92	92		

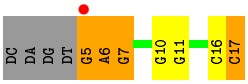
3 Residue-property plots (i)

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

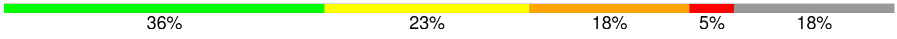
Chain P:

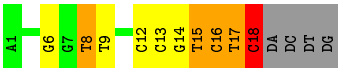




• Molecule 2: 5'-D(*DAP*DTP*DAP*DAP*DCP*DGP*DGP*DTP*DTP*DGP*DCP*DCP*DCP*DGP*DTP*DCP*DTP*DCP*DAP*DCP*DTP*DG)-3'

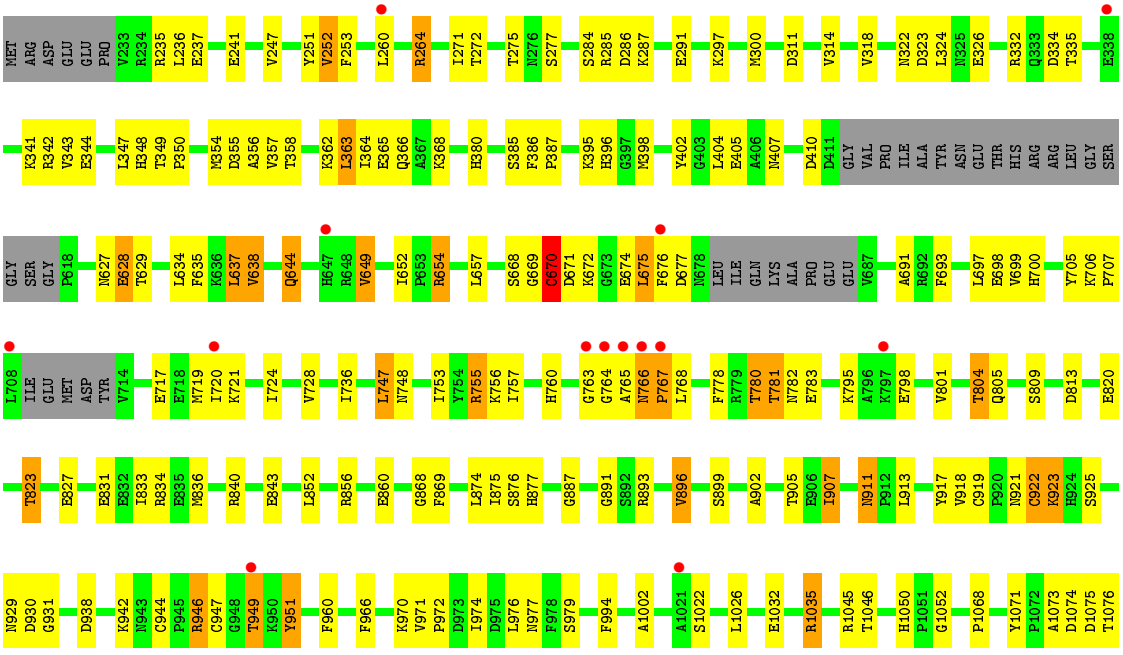
Chain T:

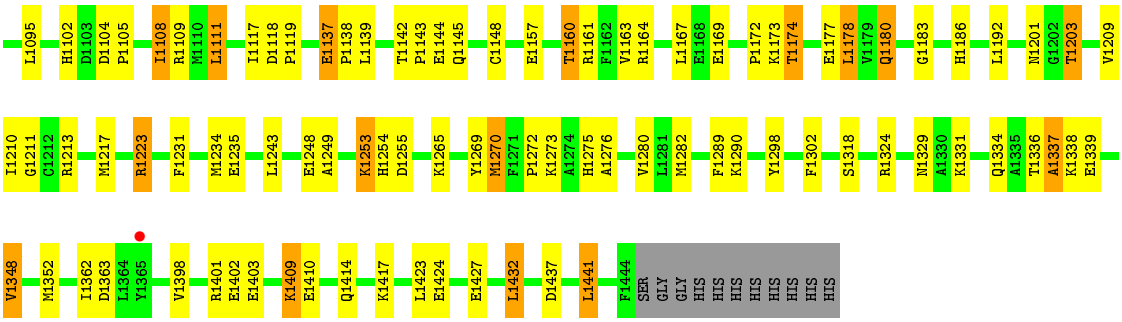




• Molecule 3: GEOBACILLUS KAUSTOPHILUS DNA POLC

Chain A:




L1095
H1102
D1103
D1104
P1105
I1108
R1109
M1110
L1111
I1117
D1118
P1119
E1137
P1138
L1139
T1142
P1143
E1144
Q1145
C1148
E1157
T1160
R1161
F1162
V1163
R1164
L1167
E1168
E1169
P1172
K1173
T1174
E1177
L1178
V1179
Q1180
G1183
H1186
L1192
N1201
G1202
T1203
V1209
I1210
G1211
C1212
R1213
M1217
R1223
F1231
M1234
E1235
L1243
E1248
A1249
K1253
H1254
D1255
K1265
Y1269
M1270
F1271
P1272
K1273
A1274
H1275
A1276
V1280
L1281
M1282
F1289
K1290
Y1298
F1302
S1318
R1324
N1329
A1330
K1331
Q1334
A1335
T1336
A1337
K1338
E1339
V1348
M1352
I1362
D1363
L1364
Y1365
V1398
R1401
E1402
E1403
K1409
E1410
Q1414
K1417
L1423
E1424
E1427
L1432
D1437
L1441
F1444
SER
GLY
GLY
HIS
HIS
HIS
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

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
$< I/\sigma(I) >$ [1]	2.24 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.4.0069	Depositor
R, R_{free}	0.216 , 0.254 0.214 , 0.251	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)	0.29 , 42.4	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Outliers	1 of 50748 reflections (0.002%)	Xtriage
F_o,F_c correlation	0.94	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.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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: 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(^{o})	Ideal(^{o})
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 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	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

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 18.

The worst 5 of 305 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)
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%)	39 61

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%)	7 13

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 molecules 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	25,33,33	1.18	2 (8%)	35,52,52	2.44	12 (34%)
6	PO4	A	6	5,4	4,4,4	0.65	0	6,6,6	0.29	0
6	PO4	A	7	-	4,4,4	0.42	0	6,6,6	0.27	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	-	0/18/34/34	0/3/3/3
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 (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	A	1456	DGT	C5-C4	2.70	1.46	1.40
7	A	1456	DGT	C6-C5	3.35	1.47	1.41

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
7	A	1456	DGT	C2'-C1'-N9	-5.73	100.23	114.16
7	A	1456	DGT	C5-C6-N1	-4.05	118.05	123.59
7	A	1456	DGT	C1'-N9-C4	-3.99	120.40	127.16
7	A	1456	DGT	C6-C5-C4	-3.97	116.16	120.90
7	A	1456	DGT	N3-C2-N1	-3.79	121.67	127.44

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
7	A	1456	DGT	7	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^{th} 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%)	0.28	1 (7%) 16 18	32, 47, 85, 90	0
2	T	18/22 (81%)	0.02	0 100 100	32, 46, 70, 81	0
3	A	993/1041 (95%)	-0.04	15 (1%) 76 79	32, 45, 59, 84	0
All	All	1024/1080 (94%)	-0.03	16 (1%) 74 78	32, 45, 60, 90	0

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	767	PRO	5.9
3	A	765	ALA	5.4
3	A	766	ASN	4.5
3	A	764	GLY	4.2
1	P	5	DG	4.1

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, 95^{th} 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^2)	Q<0.9
6	PO4	A	7	5/5	0.92	0.13	-0.65	80,81,82,83	0
6	PO4	A	6	5/5	0.95	0.13	-1.89	68,69,70,70	0
5	ZN	A	4	1/1	0.98	0.06	-1.91	79,79,79,79	0
4	MN	A	2	1/1	0.95	0.06	-3.62	63,63,63,63	0
7	DGT	A	1456	31/31	0.98	0.07	-3.66	25,28,30,30	0
4	MN	A	3	1/1	1.00	0.06	-4.37	36,36,36,36	0
5	ZN	A	5	1/1	0.98	0.04	-8.94	53,53,53,53	0
4	MN	A	1	1/1	0.99	0.05	-	30,30,30,30	0

6.5 Other polymers (i)

There are no such residues in this entry.