



# Full wwPDB X-ray Structure Validation Report ⓘ

Oct 3, 2021 – 12:18 PM EDT

PDB ID : 3KHG  
Title : Dpo4 extension ternary complex with misinserted A opposite the 2-aminofluorene-guanine [AF]G lesion  
Authors : Rechkoblit, O.; Malinina, L.; Patel, D.J.  
Deposited on : 2009-10-30  
Resolution : 2.96 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.23.2  
buster-report : 1.1.7 (2018)  
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) : 2.23.2

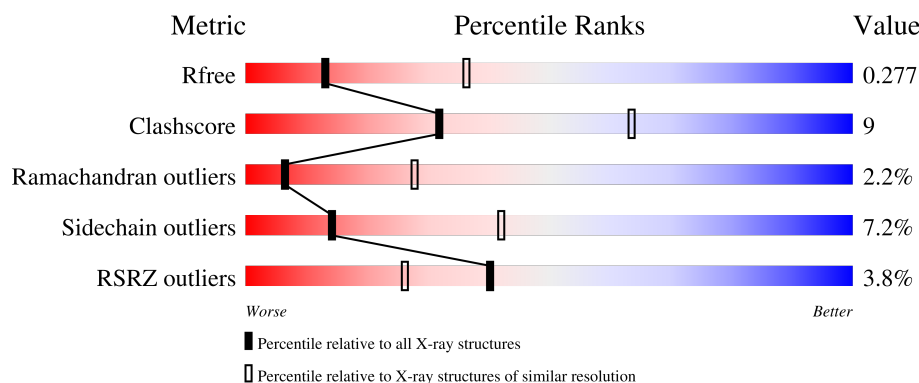
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.96 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	3104 (3.00-2.92)
Clashscore	141614	3462 (3.00-2.92)
Ramachandran outliers	138981	3340 (3.00-2.92)
Sidechain outliers	138945	3343 (3.00-2.92)
RSRZ outliers	127900	2986 (3.00-2.92)

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> <div>6%</div> <div>77%</div> <div>19%</div> <div>.</div> </div>
1	B	341	<div> <div>6%</div> <div>72%</div> <div>25%</div> <div>.</div> </div>
2	D	13	<div> <div>8%</div> <div>31%</div> <div>69%</div> </div>
2	H	13	<div> <div>8%</div> <div>38%</div> <div>23%</div> <div>38%</div> </div>
3	E	19	<div> <div>16%</div> <div>37%</div> <div>32%</div> <div>32%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	J	19	 <p>11% 32% 16% 5% 47%</p>

## 2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 6640 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
1	A	341	Total	C	N	O	S	0	0	0
			2740	1757	472	505	6			
1	B	341	Total	C	N	O	S	0	0	0
			2740	1757	472	505	6			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP Q97W02
B	1001	GLY	-	expression tag	UNP Q97W02

- Molecule 2 is a DNA chain called 5'-D(\*GP\*TP\*TP\*GP\*GP\*AP\*TP\*GP\*GP\*TP\*AP\*GP\*(2DA))-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	13	Total	C	N	O	P	0	0	0
			254	120	48	74	12			
2	H	8	Total	C	N	O	P	0	0	0
			170	80	34	48	8			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	814	2DA	A	engineered mutation	PDB 3KHG
H	1814	2DA	A	engineered mutation	PDB 3KHG

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	19	Total	C	N	O	P	0	0	0
			376	181	68	109	18			

*Continued on next page...*

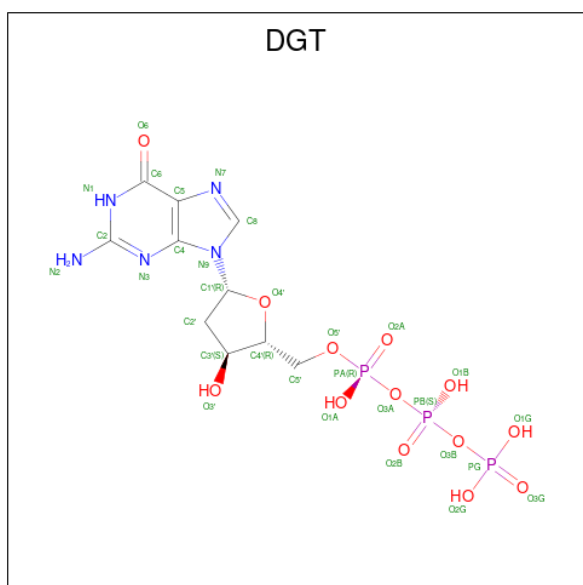
Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	J	10	Total	C	N	O	P	0	0	0
			199	95	34	60	10			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	906	DG	C	engineered mutation	PDB 3KHG
J	1906	DG	C	engineered mutation	PDB 3KHG

- Molecule 4 is 2'-DEOXYGUANOSINE-5'-TRIPHOSPHATE (three-letter code: DGT) (formula:  $C_{10}H_{16}N_5O_{13}P_3$ ).

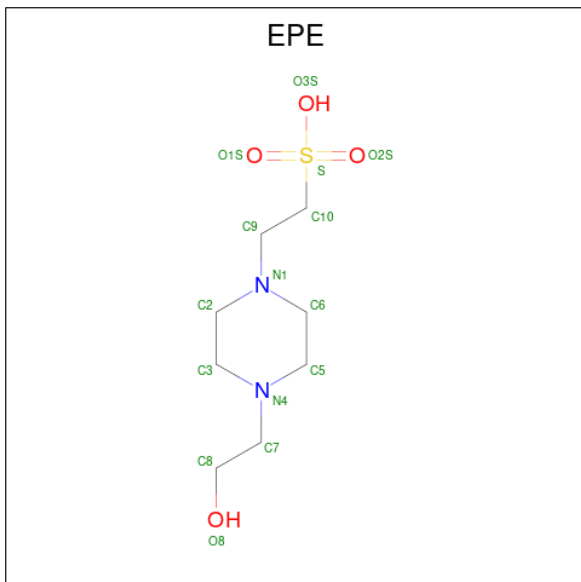


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	A	1	Total	C	N	O	P	0	0
			31	10	5	13	3		
4	B	1	Total	C	N	O	P	0	0
			31	10	5	13	3		

- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

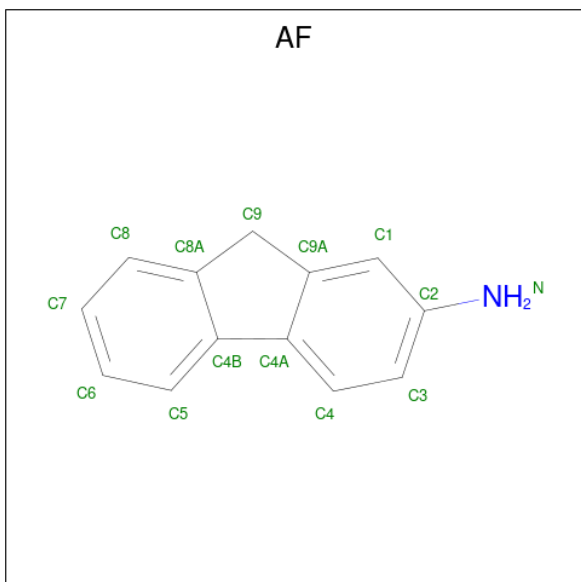
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	3	Total Ca 3 3	0	0
5	B	3	Total Ca 3 3	0	0

- Molecule 6 is 4-(2-HYDROXYETHYL)-1-PIPERAZINE ETHANESULFONIC ACID (three-letter code: EPE) (formula:  $C_8H_{18}N_2O_4S$ ).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
6	A	1	Total	C	N	O	S	0	0
			15	8	2	4	1		

- Molecule 7 is 2-AMINOFLUORENE (three-letter code: AF) (formula:  $C_{13}H_{11}N$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	E	1	Total	C	N	0	0
			14	13	1		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	J	1	Total	C	N	0	0
			14	13	1		

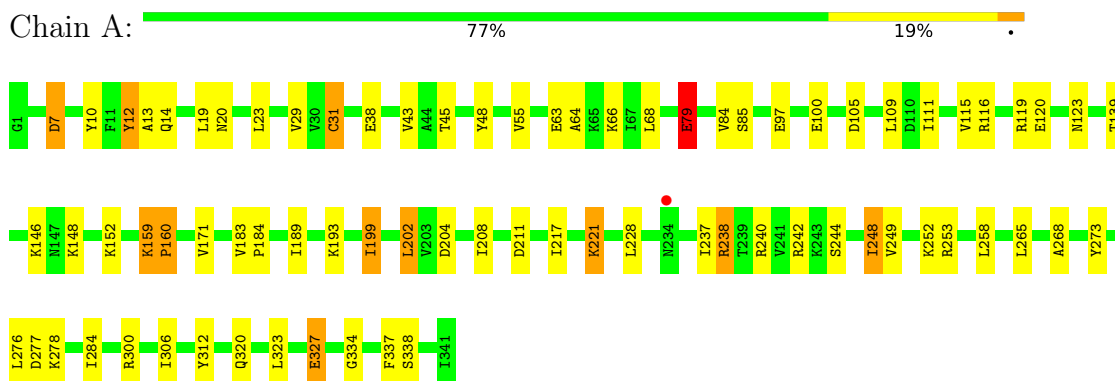
- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	25	Total	O	0	0
			25	25		
8	B	17	Total	O	0	0
			17	17		
8	D	4	Total	O	0	0
			4	4		
8	E	3	Total	O	0	0
			3	3		
8	J	1	Total	O	0	0
			1	1		

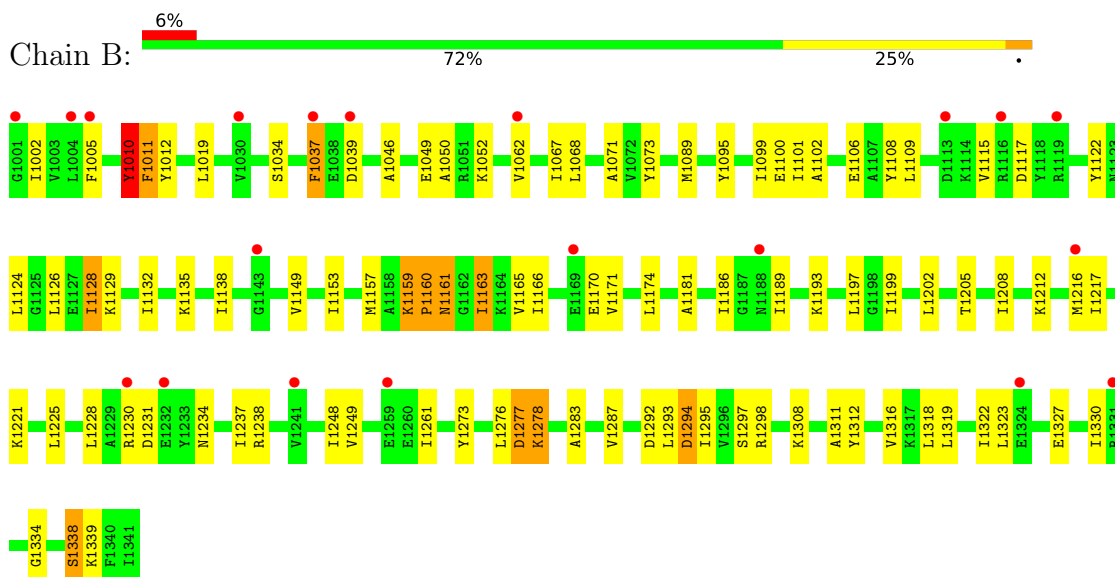
### 3 Residue-property plots [i](#)

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



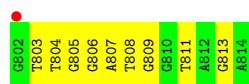
- Molecule 1: DNA polymerase IV



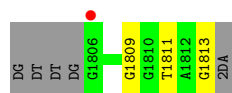
- Molecule 2: 5'-D(\*GP\*TP\*TP\*GP\*GP\*AP\*TP\*GP\*GP\*TP\*AP\*GP\*(2DA))-3'







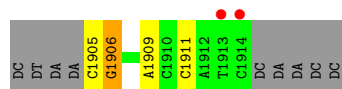
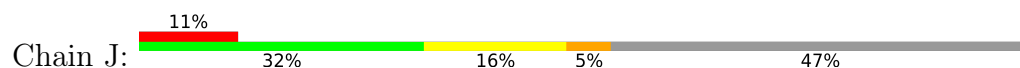
- Molecule 2: 5'-D(\*GP\*TP\*TP\*GP\*GP\*AP\*TP\*GP\*GP\*TP\*AP\*GP\*(2DA))-3'



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



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



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	52.81Å 110.95Å 100.77Å 90.00° 101.38° 90.00°	Depositor
Resolution (Å)	20.00 – 2.96 19.86 – 2.96	Depositor EDS
% Data completeness (in resolution range)	96.2 (20.00-2.96) 96.2 (19.86-2.96)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.85 (at 2.98Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, $R_{free}$	0.205 , 0.282 0.206 , 0.277	Depositor DCC
$R_{free}$ test set	1224 reflections (5.36%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	65.6	Xtriage
Anisotropy	0.445	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 64.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	6640	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.41% 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: AF, DGT, EPE, 2DA, CA

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	0.65	2/2779 (0.1%)	0.70	1/3731 (0.0%)
1	B	0.49	0/2779	0.60	0/3731
2	D	0.93	0/282	1.76	11/436 (2.5%)
2	H	0.80	0/191	1.47	2/294 (0.7%)
3	E	1.08	0/420	1.94	17/643 (2.6%)
3	J	0.84	0/221	1.73	6/337 (1.8%)
All	All	0.66	2/6672 (0.0%)	0.97	37/9172 (0.4%)

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
1	A	0	1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	79	GLU	CG-CD	5.88	1.60	1.51
1	A	12	TYR	CD2-CE2	-5.11	1.31	1.39

All (37) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	809	DG	P-O3'-C3'	9.27	130.82	119.70
3	E	906	DG	O4'-C1'-N9	9.20	114.44	108.00
3	E	901	DC	O4'-C4'-C3'	-8.13	101.12	106.00
3	E	901	DC	N1-C2-O2	8.07	123.75	118.90
3	E	910	DC	O4'-C1'-N1	-7.76	102.57	108.00

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	E	917	DA	O4'-C1'-N9	7.70	113.39	108.00
2	D	805	DG	O4'-C1'-N9	7.34	113.14	108.00
3	E	902	DT	O5'-P-OP2	-7.22	99.20	105.70
3	E	918	DC	O4'-C1'-N1	6.97	112.88	108.00
3	J	1906	DG	O4'-C1'-N9	6.82	112.77	108.00
1	A	159	LYS	C-N-CD	-6.79	105.66	120.60
2	D	803	DT	O4'-C1'-N1	6.51	112.56	108.00
2	D	811	DT	O5'-P-OP2	-6.43	99.91	105.70
3	E	916	DA	O4'-C1'-N9	6.26	112.38	108.00
3	E	901	DC	C1'-O4'-C4'	-6.12	103.98	110.10
3	J	1906	DG	N3-C4-C5	-6.10	125.55	128.60
3	J	1911	DC	O4'-C1'-N1	5.93	112.15	108.00
3	E	919	DC	O4'-C1'-N1	5.89	112.12	108.00
3	E	909	DA	O5'-P-OP2	-5.85	100.44	105.70
2	D	808	DT	O4'-C1'-C2'	-5.73	101.32	105.90
3	E	902	DT	O4'-C1'-N1	-5.72	103.99	108.00
2	D	807	DA	O4'-C1'-N9	-5.64	104.05	108.00
2	D	809	DG	O4'-C1'-N9	5.53	111.87	108.00
3	E	912	DA	O4'-C1'-N9	5.51	111.86	108.00
2	D	808	DT	C1'-O4'-C4'	-5.48	104.62	110.10
3	E	905	DC	O4'-C1'-N1	5.47	111.83	108.00
3	J	1911	DC	O4'-C1'-C2'	-5.46	101.53	105.90
3	E	909	DA	O4'-C4'-C3'	5.45	109.27	106.00
2	D	803	DT	P-O3'-C3'	5.44	126.23	119.70
2	H	1809	DG	P-O3'-C3'	5.37	126.14	119.70
2	D	804	DT	O4'-C1'-N1	5.22	111.65	108.00
2	D	806	DG	P-O3'-C3'	5.21	125.95	119.70
3	J	1909	DA	C1'-O4'-C4'	-5.17	104.93	110.10
2	H	1809	DG	C1'-O4'-C4'	-5.10	105.00	110.10
3	E	918	DC	O4'-C1'-C2'	-5.08	101.83	105.90
3	E	913	DT	P-O3'-C3'	5.05	125.76	119.70
3	J	1906	DG	N3-C4-N9	5.03	129.02	126.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	159	LYS	Peptide

## 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	A	2740	0	2883	42	0
1	B	2740	0	2880	65	0
2	D	254	0	137	2	0
2	H	170	0	91	2	0
3	E	376	0	214	3	0
3	J	199	0	113	2	0
4	A	31	0	12	2	0
4	B	31	0	12	2	0
5	A	3	0	0	0	0
5	B	3	0	0	0	0
6	A	15	0	17	1	0
7	E	14	0	10	0	0
7	J	14	0	10	2	0
8	A	25	0	0	0	0
8	B	17	0	0	2	0
8	D	4	0	0	0	0
8	E	3	0	0	0	0
8	J	1	0	0	0	0
All	All	6640	0	6379	115	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 9.

All (115) 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:A:14:GLN:HE22	1:A:139:THR:H	1.15	0.90
1:B:1323:LEU:HD11	1:B:1330:ILE:HD11	1.62	0.79
1:A:238:ARG:HD3	1:A:238:ARG:C	2.02	0.78
1:B:1159:LYS:HB3	1:B:1160:PRO:HD3	1.67	0.76
1:B:1199:ILE:HD13	1:B:1205:THR:HG22	1.69	0.75
1:B:1261:ILE:HG21	1:B:1319:LEU:HD21	1.74	0.69
1:B:1046:ALA:HB1	1:B:1050:ALA:HB3	1.76	0.66
1:A:14:GLN:NE2	1:A:139:THR:H	1.92	0.65
1:B:1312:TYR:O	1:B:1316:VAL:HG23	1.98	0.64

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:217:ILE:HD12	1:A:221:LYS:HB3	1.80	0.63
6:A:342:EPE:H62	2:D:813:DG:H21	1.62	0.63
1:A:100:GLU:HB2	1:A:237:ILE:HG23	1.82	0.62
1:B:1248:ILE:HA	1:B:1334:GLY:HA3	1.82	0.61
1:A:199:ILE:HG23	1:A:204:ASP:HB2	1.83	0.60
1:A:248:ILE:HA	1:A:334:GLY:HA3	1.83	0.59
1:A:189:ILE:HG22	1:A:193:LYS:HE3	1.85	0.59
1:B:1249:VAL:HA	8:B:38:HOH:O	2.02	0.59
1:B:1287:VAL:HG22	1:B:1297:SER:HB3	1.85	0.59
1:B:1273:TYR:HA	1:B:1276:LEU:HD12	1.85	0.58
1:B:1010:TYR:HA	4:B:1414:DGT:O3B	2.03	0.58
1:B:1039:ASP:O	1:B:1062:VAL:HG13	2.03	0.57
1:A:277:ASP:O	1:A:278:LYS:HB2	2.04	0.57
1:B:1166:ILE:HG22	1:B:1171:VAL:HG23	1.86	0.57
1:A:55:VAL:HG21	1:A:68:LEU:HD12	1.87	0.56
3:J:1906:DG:N7	7:J:1926:AF:H3	2.21	0.56
1:B:1308:LYS:O	1:B:1311:ALA:HB3	2.05	0.56
1:B:1166:ILE:HG22	1:B:1171:VAL:CG2	2.36	0.56
1:B:1129:LYS:HZ1	1:B:1161:ASN:ND2	2.04	0.56
1:B:1129:LYS:NZ	1:B:1161:ASN:ND2	2.54	0.55
1:A:119:ARG:O	1:A:123:ASN:ND2	2.40	0.55
1:B:1159:LYS:CB	1:B:1160:PRO:HD3	2.36	0.55
1:B:1166:ILE:HG23	1:B:1170:GLU:HG2	1.88	0.54
1:B:1287:VAL:HG22	1:B:1297:SER:CB	2.38	0.54
1:A:12:TYR:CZ	4:A:414:DGT:H2'	2.43	0.54
1:A:29:VAL:CG1	1:A:43:VAL:HG22	2.37	0.54
1:B:1197:LEU:HD21	1:B:1216:MET:HE3	1.89	0.54
1:B:1181:ALA:HB1	8:B:1515:HOH:O	2.08	0.53
1:B:1248:ILE:HD12	1:B:1287:VAL:HG12	1.91	0.52
1:B:1277:ASP:O	1:B:1278:LYS:HB2	2.10	0.52
1:B:1283:ALA:HB3	1:B:1338:SER:HB2	1.91	0.52
1:B:1261:ILE:HG21	1:B:1319:LEU:CD2	2.39	0.52
1:A:202:LEU:HD21	1:A:228:LEU:HB2	1.91	0.52
1:A:111:ILE:O	1:A:115:VAL:HG22	2.10	0.52
2:H:1813:DG:H2''	7:J:1926:AF:H7	1.92	0.52
1:B:1002:ILE:HD12	1:B:1115:VAL:HG23	1.91	0.51
1:B:1102:ALA:HB3	1:B:1106:GLU:HB2	1.92	0.51
1:A:12:TYR:CE2	4:A:414:DGT:H2'	2.47	0.49
1:B:1189:ILE:HG21	2:H:1811:DT:OP1	2.13	0.49
1:B:1149:VAL:O	1:B:1153:ILE:HG13	2.12	0.49
1:B:1010:TYR:O	1:B:1011:PHE:CB	2.60	0.48

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:1132:ILE:HG22	1:B:1138:ILE:O	2.13	0.48
1:B:1159:LYS:HB3	1:B:1160:PRO:CD	2.39	0.48
1:B:1318:LEU:O	1:B:1322:ILE:HG13	2.12	0.48
1:B:1100:GLU:HB2	1:B:1237:ILE:HG23	1.95	0.48
1:B:1277:ASP:O	1:B:1278:LYS:CB	2.61	0.48
1:A:238:ARG:HD3	1:A:238:ARG:O	2.14	0.48
1:A:10:TYR:HD1	1:A:13:ALA:HB3	1.79	0.48
1:A:152:LYS:NZ	2:D:813:DG:O3'	2.46	0.48
1:B:1089:MET:HB3	1:B:1101:ILE:HD11	1.96	0.47
1:A:284:ILE:HD12	1:A:337:PHE:CE1	2.49	0.47
1:B:1005:PHE:CD1	1:B:1108:TYR:CD1	3.03	0.47
1:B:1217:ILE:HD12	1:B:1221:LYS:HB3	1.96	0.47
1:B:1089:MET:SD	1:B:1101:ILE:CD1	3.03	0.47
3:E:905:DC:C5	3:E:906:DG:C6	3.03	0.47
1:A:273:TYR:HA	1:A:276:LEU:HD12	1.96	0.47
1:B:1294:ASP:C	1:B:1295:ILE:HD12	2.36	0.47
1:B:1005:PHE:CD1	1:B:1108:TYR:CE1	3.02	0.46
1:B:1126:LEU:HG	1:B:1163:ILE:HD13	1.97	0.46
1:A:100:GLU:OE2	1:A:148:LYS:NZ	2.39	0.46
1:B:1126:LEU:HD21	1:B:1163:ILE:HG21	1.98	0.46
1:B:1149:VAL:HG21	1:B:1228:LEU:HD11	1.97	0.46
1:B:1011:PHE:O	1:B:1012:TYR:C	2.53	0.46
1:B:1153:ILE:HG22	1:B:1157:MET:CE	2.45	0.45
3:E:918:DC:H2''	3:E:919:DC:H5''	1.97	0.45
1:A:55:VAL:CG2	1:A:68:LEU:HD12	2.46	0.45
3:J:1905:DC:H2'	3:J:1906:DG:N3	2.32	0.45
1:A:7:ASP:OD1	1:A:7:ASP:C	2.56	0.44
1:B:1049:GLU:OE1	1:B:1052:LYS:NZ	2.44	0.44
1:B:1170:GLU:O	1:B:1174:LEU:N	2.45	0.44
1:A:79:GLU:H	1:A:79:GLU:CD	2.21	0.44
1:B:1166:ILE:HG23	1:B:1170:GLU:CG	2.47	0.44
1:A:273:TYR:OH	1:A:306:ILE:O	2.31	0.44
1:B:1067:ILE:HG22	1:B:1068:LEU:HG	1.99	0.43
1:B:1197:LEU:HD21	1:B:1216:MET:CE	2.48	0.43
1:A:109:LEU:HD12	1:A:109:LEU:N	2.34	0.43
1:A:120:GLU:HA	1:A:123:ASN:HD22	1.82	0.43
1:A:265:LEU:O	1:A:268:ALA:HB3	2.18	0.43
1:B:1186:ILE:HD11	1:B:1225:LEU:HD11	2.00	0.43
1:B:1124:LEU:O	1:B:1128:ILE:N	2.50	0.43
1:A:84:VAL:O	1:A:85:SER:C	2.57	0.43
1:A:327:GLU:H	1:A:327:GLU:CD	2.22	0.43

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:1102:ALA:HB3	1:B:1106:GLU:CB	2.48	0.42
1:B:1099:ILE:HG23	1:B:1099:ILE:O	2.18	0.42
1:A:63:GLU:O	1:A:64:ALA:C	2.57	0.42
1:A:29:VAL:HG13	1:A:43:VAL:HG22	2.02	0.42
1:B:1298:ARG:C	1:B:1318:LEU:HD22	2.40	0.42
1:B:1129:LYS:HZ1	1:B:1161:ASN:HD22	1.68	0.42
3:E:901:DC:H2'	3:E:902:DT:C6	2.55	0.41
1:B:1153:ILE:HG22	1:B:1157:MET:HE2	2.02	0.41
1:A:238:ARG:NH2	1:A:240:ARG:HA	2.35	0.41
1:B:1071:ALA:HB3	1:B:1073:TYR:CE2	2.55	0.41
1:A:183:VAL:HG11	1:A:228:LEU:HD13	2.03	0.41
1:B:1202:LEU:HD23	1:B:1202:LEU:O	2.21	0.41
1:A:20:ASN:HB3	1:A:23:LEU:HD22	2.02	0.41
1:A:248:ILE:HD12	1:A:249:VAL:N	2.36	0.41
1:B:1095:TYR:HD1	1:B:1124:LEU:HD11	1.86	0.41
1:A:63:GLU:O	1:A:66:LYS:HB2	2.20	0.41
1:A:146:LYS:HD3	1:A:171:VAL:HG11	2.02	0.41
1:A:48:TYR:CZ	1:A:160:PRO:HD3	2.55	0.41
1:B:1109:LEU:HD12	1:B:1109:LEU:N	2.37	0.40
1:B:1323:LEU:HD11	1:B:1330:ILE:CD1	2.44	0.40
1:A:258:LEU:CD2	1:A:320:GLN:HE21	2.35	0.40
1:B:1166:ILE:HG22	1:B:1166:ILE:O	2.21	0.40
4:B:1414:DGT:H3'	4:B:1414:DGT:O2B	2.22	0.40
1:A:31:CYS:HB3	1:A:43:VAL:HA	2.03	0.40

There are no symmetry-related clashes.

## 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%)	314 (93%)	22 (6%)	3 (1%)	17 51

Continued on next page...



*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	339/341 (99%)	299 (88%)	28 (8%)	12 (4%)	3	17
All	All	678/682 (99%)	613 (90%)	50 (7%)	15 (2%)	6	28

All (15) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	160	PRO
1	B	1010	TYR
1	B	1011	PHE
1	B	1159	LYS
1	B	1231	ASP
1	B	1277	ASP
1	B	1034	SER
1	B	1037	PHE
1	B	1122	TYR
1	B	1160	PRO
1	B	1278	LYS
1	B	1339	LYS
1	A	312	TYR
1	A	38	GLU
1	B	1163	ILE

### 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	299/299 (100%)	275 (92%)	24 (8%)	12	37
1	B	299/299 (100%)	280 (94%)	19 (6%)	17	47
All	All	598/598 (100%)	555 (93%)	43 (7%)	14	42

All (43) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	7	ASP

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	19	LEU
1	A	31	CYS
1	A	45	THR
1	A	79	GLU
1	A	97	GLU
1	A	105	ASP
1	A	116	ARG
1	A	184	PRO
1	A	199	ILE
1	A	202	LEU
1	A	208	ILE
1	A	211	ASP
1	A	221	LYS
1	A	238	ARG
1	A	242	ARG
1	A	244	SER
1	A	248	ILE
1	A	252	LYS
1	A	253	ARG
1	A	300	ARG
1	A	323	LEU
1	A	327	GLU
1	A	338	SER
1	B	1010	TYR
1	B	1019	LEU
1	B	1037	PHE
1	B	1117	ASP
1	B	1128	ILE
1	B	1135	LYS
1	B	1161	ASN
1	B	1165	VAL
1	B	1193	LYS
1	B	1208	ILE
1	B	1212	LYS
1	B	1230	ARG
1	B	1234	ASN
1	B	1238	ARG
1	B	1292	ASP
1	B	1293	LEU
1	B	1294	ASP
1	B	1327	GLU
1	B	1338	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (7) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	14	GLN
1	A	123	ASN
1	A	320	GLN
1	B	1161	ASN
1	B	1188	ASN
1	B	1304	HIS
1	B	1320	GLN

### 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	$\# Z  > 2$
2	2DA	D	814	2	0,2,23	-	-	0,1,34	-	-

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.

## 5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

## 5.6 Ligand geometry

Of 11 ligands modelled in this entry, 6 are monoatomic - leaving 5 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
4	DGT	B	1414	5	26,33,33	1.08	2 (7%)	32,52,52	3.50	13 (40%)
7	AF	J	1926	3	16,16,16	0.54	0	23,23,23	0.86	2 (8%)
6	EPE	A	342	-	15,15,15	0.84	1 (6%)	18,20,20	1.53	3 (16%)
7	AF	E	926	3	16,16,16	0.56	0	23,23,23	1.02	2 (8%)
4	DGT	A	414	5	26,33,33	1.02	1 (3%)	32,52,52	3.63	13 (40%)

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
4	DGT	B	1414	5	-	7/18/34/34	0/3/3/3
7	AF	J	1926	3	-	-	0/3/3/3
6	EPE	A	342	-	-	5/9/19/19	0/1/1/1
7	AF	E	926	3	-	-	0/3/3/3
4	DGT	A	414	5	-	4/18/34/34	0/3/3/3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	1414	DGT	C6-N1	3.99	1.40	1.33
4	A	414	DGT	C6-N1	3.40	1.39	1.33
6	A	342	EPE	C10-S	2.71	1.81	1.77
4	B	1414	DGT	C8-N7	-2.11	1.30	1.34

All (33) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	414	DGT	C5-C6-N1	-8.37	111.98	123.43
4	A	414	DGT	O1G-PG-O3G	-8.31	78.16	110.68
4	B	1414	DGT	C5-C6-N1	-7.99	112.50	123.43
4	A	414	DGT	PB-O3B-PG	-7.54	106.95	132.83
4	B	1414	DGT	O2G-PG-O3G	-7.52	81.25	110.68
4	B	1414	DGT	O1G-PG-O3G	-6.59	84.90	110.68
4	A	414	DGT	C6-N1-C2	6.52	126.29	115.93
4	B	1414	DGT	C6-N1-C2	6.48	126.23	115.93
4	A	414	DGT	PA-O3A-PB	-6.36	110.99	132.83
4	B	1414	DGT	PB-O3B-PG	-6.25	111.37	132.83
4	B	1414	DGT	PA-O3A-PB	-5.87	112.70	132.83
4	A	414	DGT	O1G-PG-O3B	5.09	121.72	104.64
4	B	1414	DGT	N3-C2-N1	-5.06	120.47	127.22
4	B	1414	DGT	O3B-PG-O3G	-4.67	85.28	111.19
4	A	414	DGT	O3B-PG-O3G	-4.63	85.51	111.19
4	A	414	DGT	N3-C2-N1	-4.62	121.06	127.22
4	A	414	DGT	O2G-PG-O3G	-4.25	94.05	110.68
4	B	1414	DGT	O2G-PG-O1G	3.86	122.40	107.64
4	B	1414	DGT	O2G-PG-O3B	3.73	117.16	104.64
4	B	1414	DGT	O1G-PG-O3B	3.67	116.96	104.64
6	A	342	EPE	O3S-S-C10	3.67	111.71	105.77
4	A	414	DGT	O2G-PG-O1G	3.40	120.62	107.64
4	A	414	DGT	O2G-PG-O3B	3.39	116.02	104.64
7	E	926	AF	C1-C2-N	-3.04	115.33	120.57
4	A	414	DGT	C2'-C1'-N9	-3.03	107.27	114.27
6	A	342	EPE	O1S-S-C10	2.73	110.20	106.92
7	E	926	AF	C3-C2-N	2.70	125.94	120.91
4	A	414	DGT	C6-C5-C4	-2.66	118.26	120.80
6	A	342	EPE	C6-N1-C2	2.63	114.74	108.83
4	B	1414	DGT	C6-C5-C4	-2.58	118.34	120.80
7	J	1926	AF	C1-C2-N	-2.28	116.63	120.57
4	B	1414	DGT	N2-C2-N3	2.09	121.20	117.79
7	J	1926	AF	C3-C2-N	2.03	124.68	120.91

There are no chirality outliers.

All (16) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	414	DGT	PB-O3A-PA-O5'
4	A	414	DGT	C5'-O5'-PA-O1A
4	A	414	DGT	C5'-O5'-PA-O2A
4	B	1414	DGT	C5'-O5'-PA-O1A
4	B	1414	DGT	C5'-O5'-PA-O2A

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms
6	A	342	EPE	S-C10-C9-N1
6	A	342	EPE	C9-C10-S-O1S
6	A	342	EPE	C9-C10-S-O3S
6	A	342	EPE	N4-C7-C8-O8
4	B	1414	DGT	PB-O3B-PG-O3G
4	B	1414	DGT	PG-O3B-PB-O2B
6	A	342	EPE	C9-C10-S-O2S
4	B	1414	DGT	PB-O3B-PG-O2G
4	A	414	DGT	C5'-O5'-PA-O3A
4	B	1414	DGT	C5'-O5'-PA-O3A
4	B	1414	DGT	PA-O3A-PB-O1B

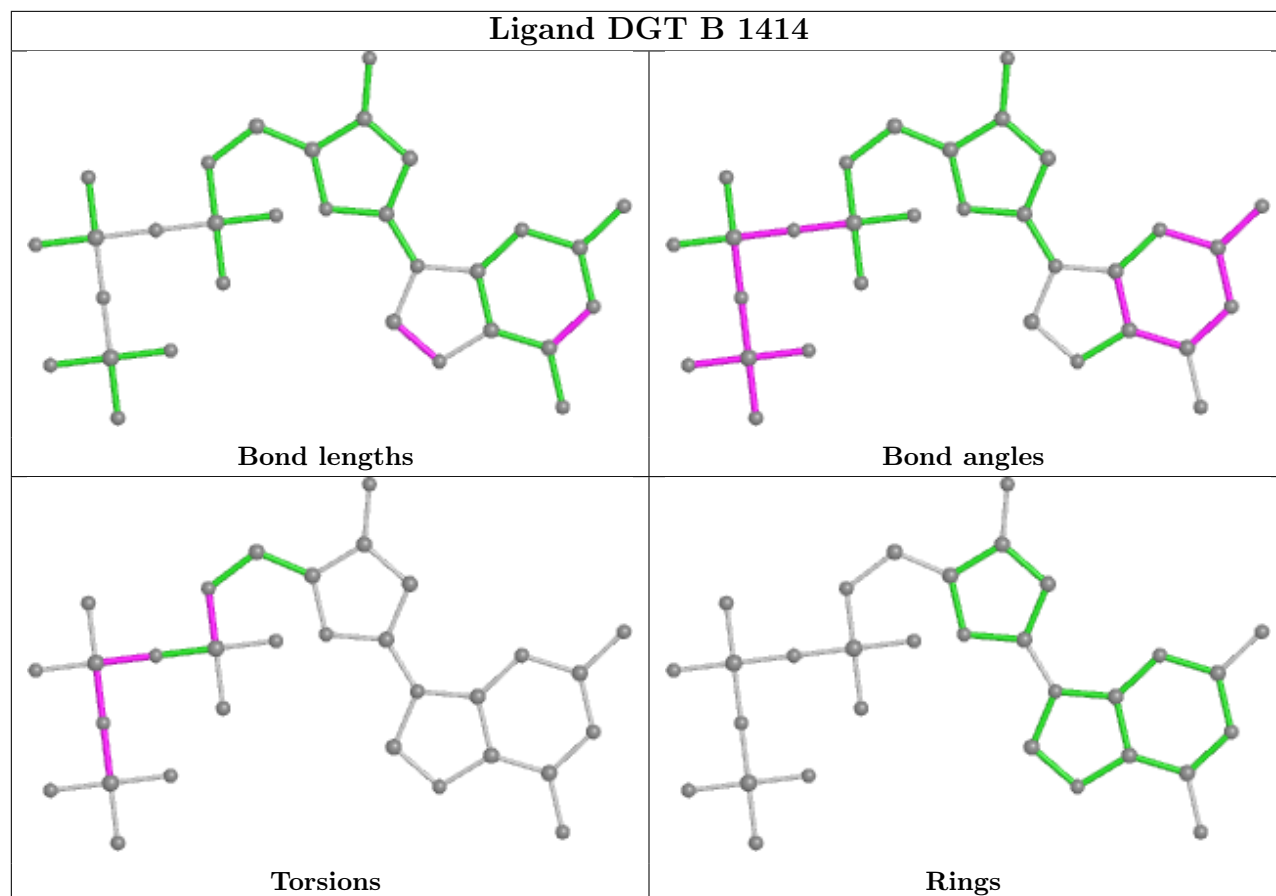
There are no ring outliers.

4 monomers are involved in 7 short contacts:

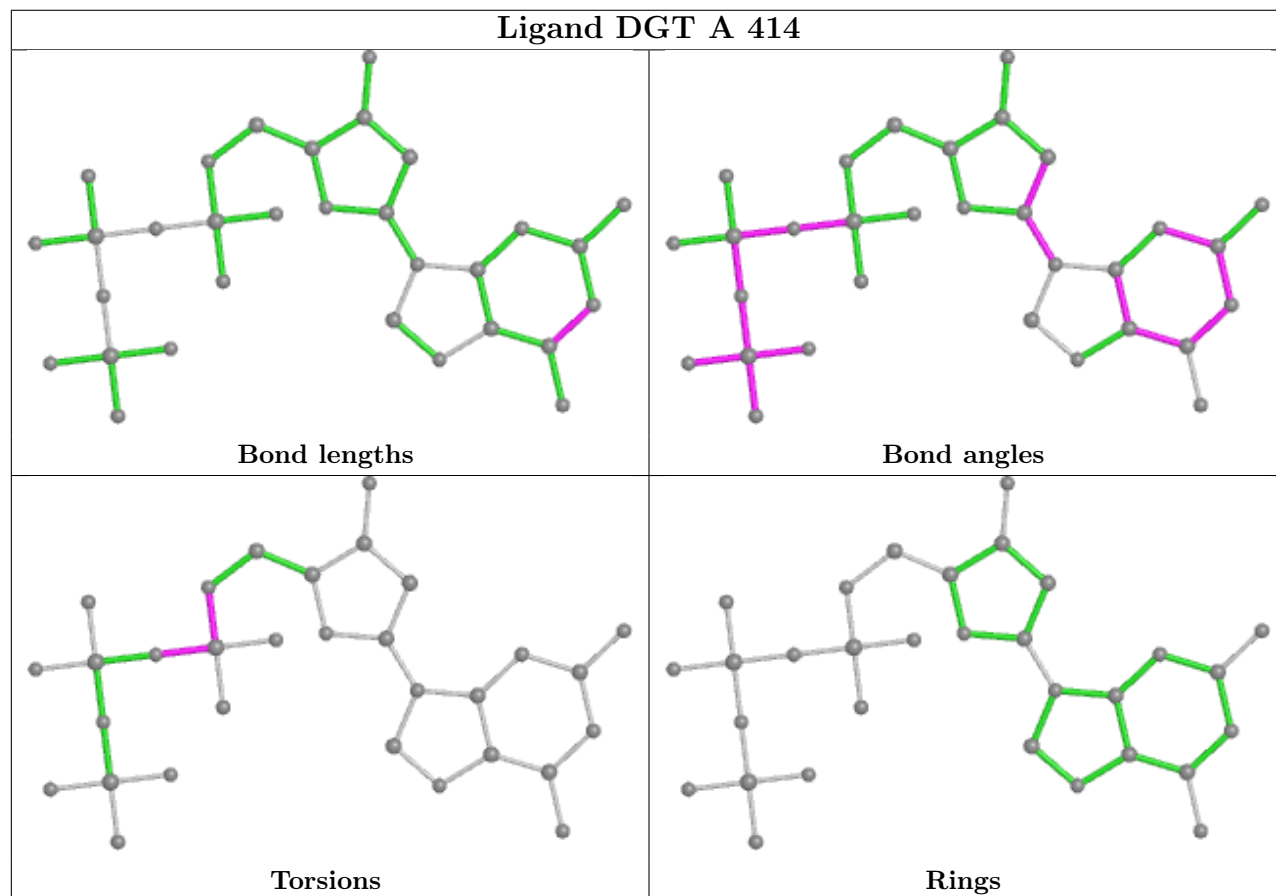
Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	1414	DGT	2	0
7	J	1926	AF	2	0
6	A	342	EPE	1	0
4	A	414	DGT	2	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less than 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

## Ligand DGT B 1414



## Ligand DGT A 414



## 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	341/341 (100%)	-0.25	1 (0%) 94 87	2, 18, 29, 35	0
1	B	341/341 (100%)	0.26	20 (5%) 22 13	32, 50, 60, 68	0
2	D	12/13 (92%)	0.15	1 (8%) 11 6	20, 32, 72, 77	0
2	H	8/13 (61%)	0.77	1 (12%) 3 2	49, 69, 89, 89	0
3	E	19/19 (100%)	0.42	3 (15%) 2 1	5, 26, 67, 68	0
3	J	10/19 (52%)	0.46	2 (20%) 1 0	44, 65, 93, 128	0
All	All	731/746 (97%)	0.03	28 (3%) 40 26	2, 34, 60, 128	0

All (28) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	E	919	DC	5.6
1	B	1143	GLY	4.6
1	B	1116	ARG	4.5
3	J	1914	DC	3.8
1	B	1004	LEU	3.8
1	B	1119	ARG	3.8
3	E	918	DC	3.6
1	B	1001	GLY	3.2
1	B	1259	GLU	3.0
1	A	234	ASN	2.9
1	B	1216	MET	2.8
1	B	1113	ASP	2.8
1	B	1232	GLU	2.6
1	B	1169	GLU	2.6
1	B	1331	ARG	2.6
1	B	1230	ARG	2.5
1	B	1030	VAL	2.5
3	E	917	DA	2.5
1	B	1005	PHE	2.4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	D	802	DG	2.4
1	B	1039	ASP	2.3
1	B	1037	PHE	2.2
1	B	1241	VAL	2.1
1	B	1062	VAL	2.1
2	H	1806	DG	2.1
1	B	1324	GLU	2.0
1	B	1188	ASN	2.0
3	J	1913	DT	2.0

## 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
2	2DA	D	814	3/21	0.84	0.55	67,67,68,71	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
5	CA	B	1415	1/1	0.80	0.15	63,63,63,63	0
5	CA	B	1417	1/1	0.87	0.17	63,63,63,63	0
5	CA	B	1416	1/1	0.88	0.05	57,57,57,57	0
7	AF	E	926	14/14	0.89	0.28	47,50,53,53	0
4	DGT	B	1414	31/31	0.90	0.20	52,56,60,61	0
7	AF	J	1926	14/14	0.92	0.33	70,70,71,72	0
5	CA	A	415	1/1	0.94	0.06	38,38,38,38	0
5	CA	A	417	1/1	0.96	0.05	38,38,38,38	0

*Continued on next page...*

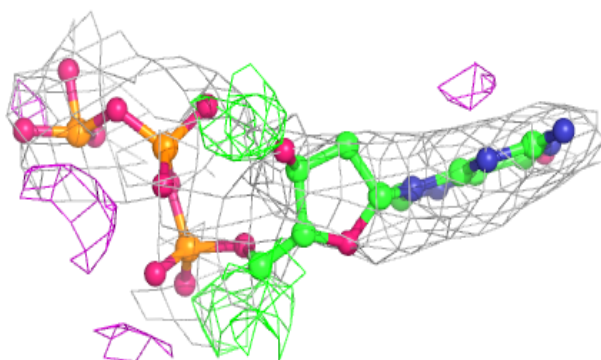
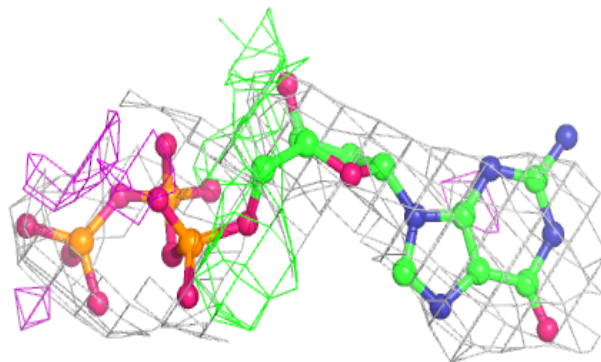
*Continued from previous page...*

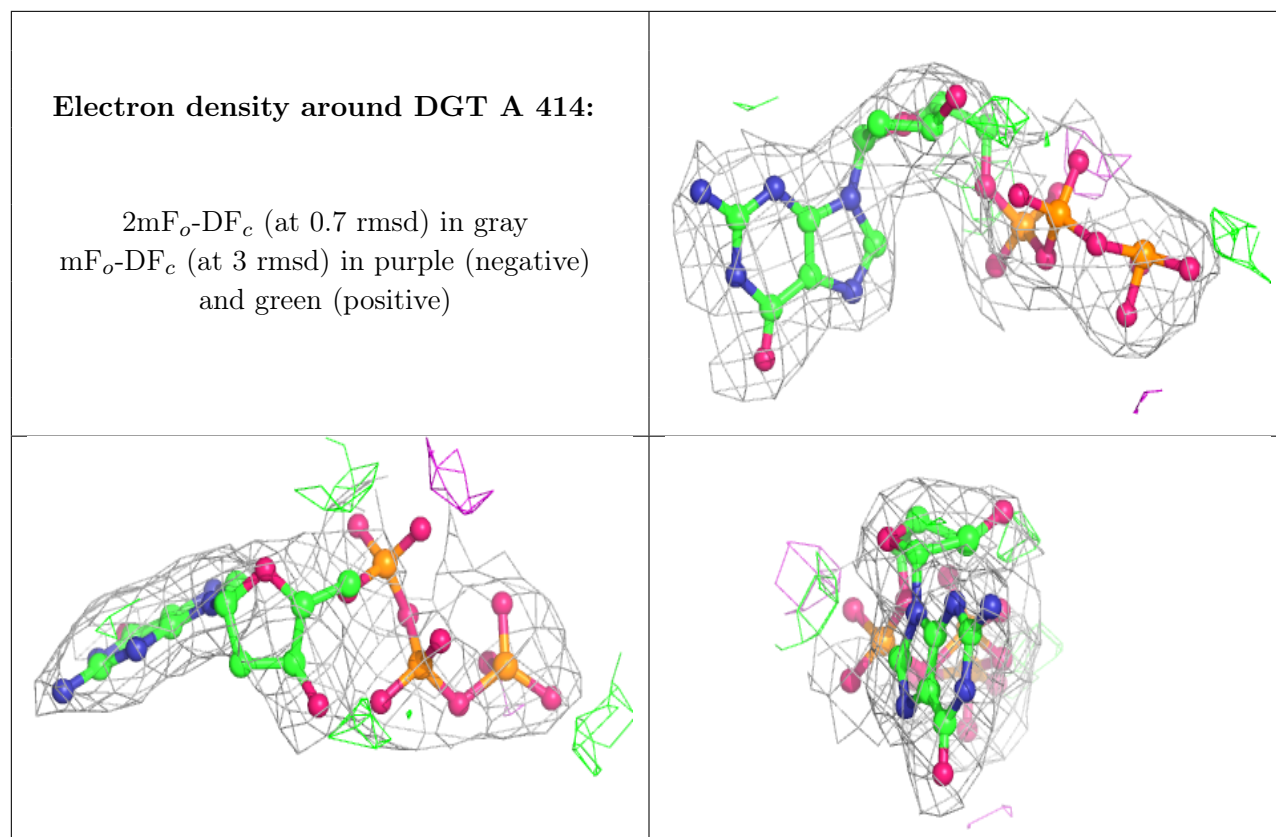
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
6	EPE	A	342	15/15	0.96	0.19	71,74,77,77	0
4	DGT	A	414	31/31	0.97	0.14	2,19,25,28	0
5	CA	A	416	1/1	0.98	0.06	29,29,29,29	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

**Electron density around DGT B 1414:**

2mF<sub>o</sub>-DF<sub>c</sub> (at 0.7 rmsd) in gray  
mF<sub>o</sub>-DF<sub>c</sub> (at 3 rmsd) in purple (negative)  
and green (positive)





## 6.5 Other polymers [i](#)

There are no such residues in this entry.