



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 04:28 PM GMT

PDB ID : 4F3O  
Title : DNA Polymerase I Large Fragment Complex 5  
Authors : Wang, W.; Beese, L.S.  
Deposited on : 2012-05-09  
Resolution : 1.57 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  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : 1.7 (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

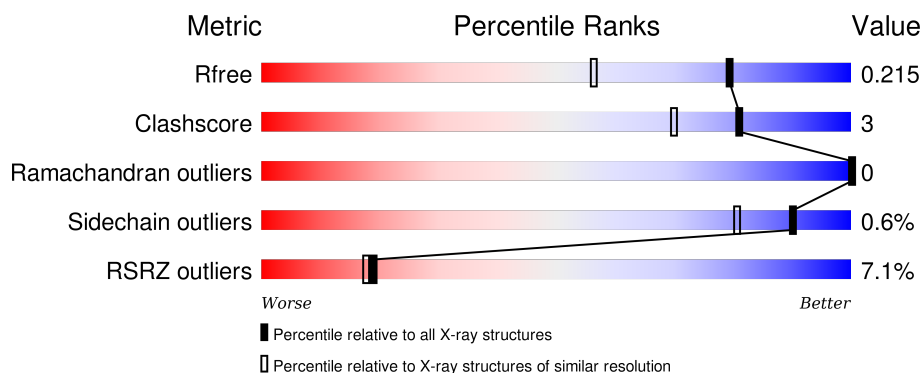
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.57 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	3815 (1.60-1.56)
Clashscore	102246	4131 (1.60-1.56)
Ramachandran outliers	100387	4021 (1.60-1.56)
Sidechain outliers	100360	4018 (1.60-1.56)
RSRZ outliers	91569	3823 (1.60-1.56)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	592	<div> <div>12%</div> <div>88%</div> <div>7%</div> <div>•</div> </div>
1	D	592	<div> <div>2%</div> <div>91%</div> <div>7%</div> <div>•</div> </div>
2	B	9	<div> <div>56%</div> <div>44%</div> </div>
2	E	9	<div> <div>44%</div> <div>56%</div> </div>
3	C	13	<div> <div>8%</div> <div>69%</div> <div>23%</div> <div>8%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	F	13	

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
5	MPD	A	903	-	-	-	X
6	MN	D	904	-	-	-	X

## 2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 21393 atoms, of which 9882 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.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	566	Total	C	H	N	O	S	0	2	0
			9177	2894	4628	794	845	16			
1	D	579	Total	C	H	N	O	S	0	9	0
			9459	2985	4770	812	875	17			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	598	ALA	ASP	ENGINEERED MUTATION	UNP Q5KWC1
A	710	TYR	PHE	ENGINEERED MUTATION	UNP Q5KWC1
A	823	HIS	ARG	SEE REMARK 999	UNP Q5KWC1
D	598	ALA	ASP	ENGINEERED MUTATION	UNP Q5KWC1
D	710	TYR	PHE	ENGINEERED MUTATION	UNP Q5KWC1
D	823	HIS	ARG	SEE REMARK 999	UNP Q5KWC1

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	9	Total	C	H	N	O	P	0	0	0
			276	86	101	28	53	8			
2	E	9	Total	C	H	N	O	P	0	0	0
			276	86	101	28	53	8			

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	C	12	Total	C	H	N	O	P	0	0	0
			384	119	134	52	68	11			
3	F	12	Total	C	H	N	O	P	0	0	0
			384	119	134	52	68	11			

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	O	S	0	0
			5	4	1		
4	A	1	Total	O	S	0	0
			5	4	1		
4	D	1	Total	O	S	0	0
			5	4	1		
4	D	1	Total	O	S	0	0
			5	4	1		

- Molecule 5 is (4S)-2-METHYL-2,4-PENTANEDIOL (three-letter code: MPD) (formula: C<sub>6</sub>H<sub>14</sub>O<sub>2</sub>).

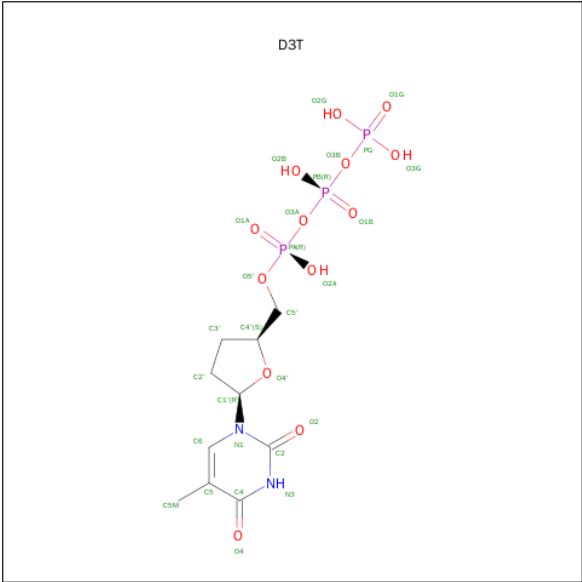


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	C	H	O	0	0
			22	6	14	2		

- Molecule 6 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	1	Total	Mn	0	0
			1	1		
6	D	1	Total	Mn	0	0
			1	1		

- Molecule 7 is 2',3'-DIDEOXY-THYMIDINE-5'-TRIPHOSPHATE (three-letter code: D3T) (formula: C<sub>10</sub>H<sub>17</sub>N<sub>2</sub>O<sub>13</sub>P<sub>3</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
7	D	1	Total	C	N	O	P	0	0
			28	10	2	13	3		

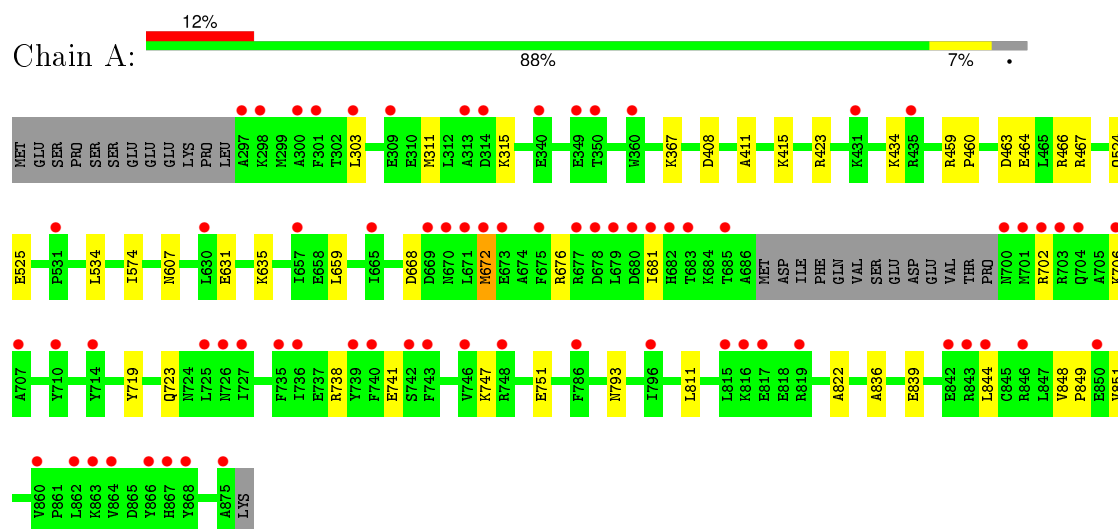
- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	406	Total	O	0	0
			406	406		
8	C	64	Total	O	0	0
			64	64		
8	D	762	Total	O	0	0
			762	762		
8	B	31	Total	O	0	0
			31	31		
8	E	32	Total	O	0	0
			32	32		
8	F	70	Total	O	0	0
			70	70		

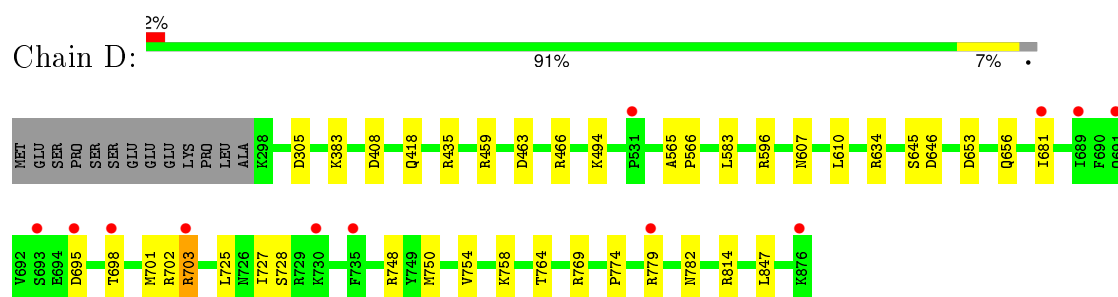
### 3 Residue-property plots [i](#)

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: DNA polymerase



- Molecule 1: DNA polymerase



- Molecule 2: DNA (5'-D(\*CP\*CP\*TP\*GP\*AP\*CP\*TP\*CP\*(2DT))-3')



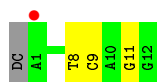
- Molecule 2: DNA (5'-D(\*CP\*CP\*TP\*GP\*AP\*CP\*TP\*CP\*(2DT))-3')



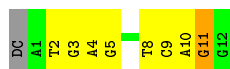




- Molecule 3: DNA (5'-D(\*C\*AP\*TP\*GP\*AP\*GP\*AP\*GP\*TP\*CP\*AP\*GP\*G)-3')



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



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	93.83Å 109.33Å 150.11Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	79.56 – 1.57 79.57 – 1.57	Depositor EDS
% Data completeness (in resolution range)	89.3 (79.56-1.57) 89.3 (79.57-1.57)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.98 (at 1.57Å)	Xtriage
Refinement program	PHENIX (phenix.refine: dev_1026)	Depositor
R, $R_{free}$	0.179 , 0.204 0.193 , 0.215	Depositor DCC
$R_{free}$ test set	8792 reflections (4.80%)	DCC
Wilson B-factor (Å <sup>2</sup> )	18.3	Xtriage
Anisotropy	0.100	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.39 , 50.2	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 191803 reflections	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	21393	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.71% 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, 0.375 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: 2DT, MPD, MN, D3T, SO4

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	0.41	0/4636	0.56	0/6264
1	D	0.60	0/4800	0.69	3/6487 (0.0%)
2	B	1.03	0/173	1.89	7/264 (2.7%)
2	E	1.08	0/173	2.09	8/264 (3.0%)
3	C	0.88	0/282	1.50	4/435 (0.9%)
3	F	1.14	1/282 (0.4%)	1.77	8/435 (1.8%)
All	All	0.58	1/10346 (0.0%)	0.82	30/14149 (0.2%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	F	3	DG	N7-C5	5.09	1.42	1.39

All (30) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	23	DT	O4'-C1'-N1	-10.09	100.94	108.00
3	F	8	DT	O4'-C1'-N1	-9.60	101.28	108.00
3	C	8	DT	O4'-C1'-N1	-9.43	101.40	108.00
2	E	27	DT	N3-C4-O4	8.90	125.24	119.90
2	E	25	DA	O5'-P-OP2	-8.60	97.96	105.70
2	E	26	DC	O4'-C1'-N1	8.21	113.75	108.00
3	F	2	DT	O4'-C1'-N1	7.51	113.25	108.00
1	D	634	ARG	NE-CZ-NH2	-7.32	116.64	120.30
2	B	27	DT	O4'-C4'-C3'	-7.22	101.61	104.50
3	F	11	DG	O4'-C1'-N9	-7.00	103.10	108.00
1	D	634	ARG	NE-CZ-NH1	6.68	123.64	120.30
2	E	23	DT	O4'-C1'-N1	-6.55	103.41	108.00
3	F	4	DA	O4'-C4'-C3'	6.37	109.82	106.00
2	E	22	DC	O4'-C1'-N1	-6.35	103.55	108.00

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	27	DT	C4'-C3'-C2'	-6.19	97.53	103.10
2	E	27	DT	C5-C4-O4	-6.00	120.70	124.90
3	F	9	DC	C2-N3-C4	5.57	122.69	119.90
2	E	25	DA	O5'-P-OP1	5.55	117.36	110.70
2	B	25	DA	O4'-C4'-C3'	-5.55	102.28	104.50
2	B	26	DC	O4'-C4'-C3'	5.53	109.32	106.00
3	F	5	DG	N9-C4-C5	5.40	107.56	105.40
3	F	10	DA	O4'-C1'-N9	-5.38	104.24	108.00
1	D	653	ASP	CB-CG-OD1	5.33	123.10	118.30
2	B	23	DT	N3-C4-O4	5.31	123.09	119.90
3	C	9	DC	C1'-O4'-C4'	-5.28	104.82	110.10
3	C	11	DG	O4'-C1'-N9	-5.23	104.34	108.00
3	F	2	DT	C4-C5-C7	5.20	122.12	119.00
2	B	25	DA	O5'-P-OP1	5.20	116.94	110.70
2	B	23	DT	C5-C4-O4	-5.12	121.31	124.90
3	C	9	DC	O4'-C1'-N1	-5.01	104.50	108.00

There are no chirality outliers.

There are no planarity outliers.

## 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	4549	4628	4616	32	0
1	D	4689	4770	4771	31	0
2	B	175	101	104	0	0
2	E	175	101	104	0	0
3	C	250	134	136	0	0
3	F	250	134	136	1	0
4	A	10	0	0	0	0
4	D	10	0	0	1	0
5	A	8	14	14	1	0
6	A	1	0	0	0	0
6	D	1	0	0	0	0
7	D	28	0	13	2	0
8	A	406	0	0	17	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
8	B	31	0	0	0	0
8	C	64	0	0	0	0
8	D	762	0	0	17	1
8	E	32	0	0	1	0
8	F	70	0	0	1	0
All	All	11511	9882	9894	66	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 3.

All (66) 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:D:702:ARG:NH2	7:D:901:D3T:O2G	1.96	0.97
3:F:11:DG:OP1	8:F:149:HOH:O	1.84	0.95
1:D:750:MET:HG3	8:D:1355:HOH:O	1.66	0.94
1:A:607[A]:ASN:OD1	8:A:1172:HOH:O	1.99	0.78
1:A:676:ARG:NH1	8:A:1141:HOH:O	2.19	0.75
1:A:459:ARG:NH2	8:A:1070:HOH:O	2.19	0.75
1:A:839:GLU:HB2	8:A:1385:HOH:O	1.92	0.69
1:A:408:ASP:OD1	8:A:1116:HOH:O	2.11	0.69
5:A:903:MPD:HM2	8:A:1011:HOH:O	1.95	0.67
1:D:408:ASP:OD1	8:D:1174:HOH:O	2.13	0.65
1:A:315:LYS:HB2	8:A:1264:HOH:O	1.97	0.62
4:D:903:SO4:O1	8:D:1610:HOH:O	2.15	0.61
1:A:463:ASP:HB2	8:A:1250:HOH:O	1.99	0.61
1:D:466:ARG:NE	8:D:1334:HOH:O	2.33	0.60
1:A:311:MET:O	8:A:1134:HOH:O	2.16	0.60
1:D:459:ARG:NH2	8:D:1540:HOH:O	2.35	0.60
1:A:747:LYS:HE3	1:A:751:GLU:OE1	2.02	0.59
1:D:698:THR:OG1	1:D:701:MET:HG3	2.04	0.57
1:A:848:VAL:HB	1:A:849:PRO:HD3	1.89	0.55
1:A:631:GLU:OE2	1:A:635:LYS:HE3	2.07	0.55
1:D:305:ASP:OD1	8:D:1564:HOH:O	2.18	0.55
1:A:466:ARG:NH1	8:A:1324:HOH:O	2.39	0.55
1:A:668:ASP:O	1:A:672:MET:HB2	2.07	0.54
1:A:659:LEU:HD13	1:A:681:ILE:HG21	1.88	0.54
1:D:748:ARG:HD3	8:D:1501:HOH:O	2.08	0.53
1:A:467:ARG:HD2	8:A:1350:HOH:O	2.11	0.51
1:D:779:ARG:NH1	8:D:1552:HOH:O	2.44	0.51
1:D:418:GLN:HA	8:D:1227:HOH:O	2.11	0.51

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:702:ARG:HH22	7:D:901:D3T:PG	2.30	0.50
1:D:703:ARG:NH2	8:D:1199:HOH:O	2.44	0.50
1:A:459:ARG:HB2	1:A:460:PRO:HD3	1.92	0.50
1:A:793:ASN:HB2	8:A:1008:HOH:O	2.13	0.48
1:D:814:ARG:NH2	1:D:847[A]:LEU:HD13	2.29	0.48
1:D:754:VAL:CG1	1:D:758:LYS:HE2	2.44	0.47
1:D:754:VAL:HG12	1:D:758:LYS:HE2	1.96	0.47
1:D:435:ARG:NH2	8:D:1385:HOH:O	2.45	0.47
1:D:596[B]:ARG:HD3	8:D:1028:HOH:O	2.15	0.46
1:A:464:GLU:OE1	8:A:1287:HOH:O	2.21	0.46
1:D:607[A]:ASN:ND2	1:D:610:LEU:HD12	2.31	0.46
1:D:774:PRO:HA	8:D:1545:HOH:O	2.15	0.46
1:D:814:ARG:CZ	1:D:847[A]:LEU:HD13	2.46	0.46
1:D:656:GLN:OE1	1:D:681:ILE:HG22	2.16	0.46
1:A:738:ARG:O	1:A:741:GLU:HB3	2.16	0.46
1:D:583:LEU:HG	8:E:219:HOH:O	2.16	0.45
1:D:725:LEU:O	1:D:727:ILE:HG23	2.17	0.45
1:D:703:ARG:NH2	8:D:1237:HOH:O	2.50	0.45
1:D:494:LYS:HE3	8:D:1649:HOH:O	2.16	0.45
1:D:782:ASN:ND2	8:D:1754:HOH:O	2.51	0.44
1:D:463:ASP:OD2	1:D:466:ARG:NH2	2.51	0.44
1:A:631:GLU:O	1:A:635:LYS:HG2	2.18	0.43
1:A:467:ARG:NH2	8:A:1189:HOH:O	2.51	0.43
1:A:534:LEU:HD11	1:A:574:ILE:HD13	2.01	0.43
1:A:702:ARG:CZ	1:A:706:LYS:HE3	2.48	0.43
1:A:524:GLN:HG2	1:A:525:GLU:O	2.18	0.43
1:D:565:ALA:N	1:D:566:PRO:CD	2.82	0.43
1:A:367:LYS:NZ	8:A:1371:HOH:O	2.51	0.43
1:A:822:ALA:CB	1:A:836:ALA:HB2	2.49	0.43
1:A:811:LEU:HA	1:A:851:VAL:HG11	2.02	0.42
1:A:434:LYS:NZ	8:A:1369:HOH:O	2.34	0.42
1:A:719:TYR:O	1:A:723:GLN:HG2	2.20	0.42
1:A:411:ALA:O	1:A:415:LYS:HG2	2.20	0.42
1:A:423:ARG:NH1	8:A:1401:HOH:O	2.53	0.41
1:D:383:LYS:HE3	8:D:1494:HOH:O	2.20	0.41
1:D:645:SER:O	1:D:646:ASP:HB2	2.21	0.41
1:A:463:ASP:O	1:A:467:ARG:HG3	2.21	0.41
1:D:764:THR:HA	1:D:769:ARG:O	2.22	0.40

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)
8:D:1161:HOH:O	8:D:1195:HOH:O[4_445]	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	564/592 (95%)	553 (98%)	11 (2%)	0	100	100
1	D	586/592 (99%)	572 (98%)	14 (2%)	0	100	100
All	All	1150/1184 (97%)	1125 (98%)	25 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	483/507 (95%)	480 (99%)	3 (1%)	90	81
1	D	504/507 (99%)	501 (99%)	3 (1%)	90	81
All	All	987/1014 (97%)	981 (99%)	6 (1%)	90	81

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	303	LEU
1	A	672	MET
1	A	844	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	D	695	ASP
1	D	703	ARG
1	D	728	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	D	867	HIS

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

2 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
2	2DT	B	29	3,2	11,20,21	1.16	2 (18%)	11,28,31	1.37	2 (18%)
2	2DT	E	29	3,2	11,20,21	1.40	2 (18%)	11,28,31	1.03	2 (18%)

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
2	2DT	B	29	3,2	-	0/3/18/19	0/2/2/2
2	2DT	E	29	3,2	-	0/3/18/19	0/2/2/2

All (4) bond length outliers are listed below:



Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	29	2DT	C6-C5	-2.69	1.32	1.40
2	B	29	2DT	C6-C5	-2.17	1.34	1.40
2	B	29	2DT	C6-N1	2.62	1.38	1.35
2	E	29	2DT	C6-N1	2.82	1.39	1.35

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	29	2DT	C5M-C5-C4	-3.00	117.83	120.97
2	E	29	2DT	O4'-C4'-C3'	-2.35	100.76	104.69
2	E	29	2DT	O4'-C1'-N1	2.14	111.42	107.72
2	B	29	2DT	C5M-C5-C6	2.72	124.10	118.62

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 [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 8 ligands modelled in this entry, 2 are monoatomic - leaving 6 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	SO4	A	901	-	4,4,4	0.10	0	6,6,6	0.06	0
4	SO4	A	902	-	4,4,4	0.28	0	6,6,6	0.30	0
5	MPD	A	903	-	6,7,7	0.40	0	7,10,10	0.65	0
7	D3T	D	901	6	20,29,29	1.62	2 (10%)	27,45,45	2.19	10 (37%)
4	SO4	D	902	-	4,4,4	0.33	0	6,6,6	0.57	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
4	SO4	D	903	-	4,4,4	0.23	0	6,6,6	0.09	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
4	SO4	A	901	-	-	0/0/0/0	0/0/0/0
4	SO4	A	902	-	-	0/0/0/0	0/0/0/0
5	MPD	A	903	-	-	0/5/5/5	0/0/0/0
7	D3T	D	901	6	-	0/18/31/31	0/2/2/2
4	SO4	D	902	-	-	0/0/0/0	0/0/0/0
4	SO4	D	903	-	-	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	D	901	D3T	C3'-C2'	-2.75	1.46	1.54
7	D	901	D3T	O4-C4	5.34	1.37	1.24

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	D	901	D3T	O4'-C1'-C2'	-4.61	101.67	106.67
7	D	901	D3T	O4'-C4'-C3'	-3.15	99.41	104.69
7	D	901	D3T	C5-C4-N3	-2.96	121.84	125.14
7	D	901	D3T	C3'-C4'-C5'	-2.05	107.70	116.05
7	D	901	D3T	O2A-PA-O3A	2.02	114.26	105.09
7	D	901	D3T	O2B-PB-O3A	2.27	115.37	105.09
7	D	901	D3T	O4'-C1'-N1	2.80	112.57	107.72
7	D	901	D3T	C3'-C2'-C1'	3.29	106.38	102.71
7	D	901	D3T	C4-N3-C2	3.93	118.65	115.25
7	D	901	D3T	O4'-C4'-C5'	4.48	116.15	109.54

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	903	MPD	1	0
7	D	901	D3T	2	0
4	D	903	SO4	1	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<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	566/592 (95%)	0.67	71 (12%) 5 5	16, 37, 62, 81	0
1	D	579/592 (97%)	0.27	12 (2%) 67 67	9, 19, 36, 49	0
2	B	8/9 (88%)	-0.03	0 100 100	22, 27, 46, 59	0
2	E	8/9 (88%)	-0.13	0 100 100	13, 24, 42, 51	0
3	C	12/13 (92%)	-0.10	1 (8%) 14 13	18, 27, 66, 75	0
3	F	12/13 (92%)	0.08	0 100 100	12, 20, 44, 51	0
All	All	1185/1228 (96%)	0.45	84 (7%) 19 18	9, 28, 57, 81	0

All (84) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	297	ALA	7.5
1	A	679	LEU	6.1
1	A	700	ASN	5.8
1	A	866	TYR	5.7
1	A	298	LYS	5.6
1	A	681	ILE	5.6
1	A	735	PHE	5.2
1	A	675	PHE	5.2
1	A	844	LEU	4.8
1	A	703	ARG	4.4
1	A	842	GLU	4.4
1	A	743	PHE	4.2
1	A	740	PHE	4.2
1	A	727	ILE	4.1
1	A	683	THR	4.1
1	A	680	ASP	4.0
1	A	819	ARG	3.9
1	A	677	ARG	3.8
1	A	704	GLN	3.8

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	678	ASP	3.7
1	A	531	PRO	3.7
1	A	682	HIS	3.7
1	D	698	THR	3.5
1	A	725	LEU	3.5
1	A	746	VAL	3.3
1	A	816	LYS	3.3
1	A	671	LEU	3.3
1	A	706	LYS	3.1
1	A	817	GLU	3.1
1	A	673	GLU	3.1
1	A	867	HIS	3.1
1	A	868	TYR	3.1
1	A	714	TYR	3.0
1	A	726	ASN	3.0
1	A	300	ALA	3.0
1	A	850	GLU	2.9
1	A	360	TRP	2.9
1	A	846	ARG	2.9
1	A	864	VAL	2.8
1	A	786	PHE	2.8
1	A	815	LEU	2.8
1	A	301	PHE	2.8
1	A	701	MET	2.8
1	A	875	ALA	2.7
1	A	313	ALA	2.6
1	D	693	SER	2.6
1	A	657	ILE	2.6
1	A	314	ASP	2.6
1	D	779	ARG	2.6
1	A	863	LYS	2.6
1	A	739	TYR	2.6
1	A	349	GLU	2.5
1	D	735	PHE	2.5
1	A	669	ASP	2.5
1	D	531	PRO	2.5
1	A	702	ARG	2.5
1	A	309	GLU	2.5
1	A	843	ARG	2.4
1	A	860	VAL	2.4
1	A	435	ARG	2.4
1	A	665	ILE	2.4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	303	LEU	2.4
1	A	862	LEU	2.4
1	A	672	MET	2.4
1	A	748	ARG	2.3
1	A	685	THR	2.3
1	A	707	ALA	2.2
1	D	876	LYS	2.2
1	A	340	GLU	2.2
1	A	742	SER	2.2
1	A	710	TYR	2.2
1	A	670	ASN	2.1
1	A	350	THR	2.1
1	D	681	ILE	2.1
1	D	689	ILE	2.1
1	A	431	LYS	2.1
1	A	736	ILE	2.1
1	A	796	ILE	2.1
1	D	695	ASP	2.1
1	A	630	LEU	2.1
3	C	1	DA	2.1
1	D	691	GLN	2.0
1	D	703	ARG	2.0
1	D	730	LYS	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. 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<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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	2DT	E	29	19/20	0.99	0.10	-	11,13,16,19	0
2	2DT	B	29	19/20	0.96	0.08	-	22,26,32,32	0

## 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. 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<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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
6	MN	D	904	1/1	0.97	0.21	10.01	8,8,8,8	1
5	MPD	A	903	8/8	0.89	0.18	2.66	20,27,40,50	0
4	SO4	D	903	5/5	0.95	0.21	1.39	66,69,71,71	0
4	SO4	A	902	5/5	0.83	0.14	1.18	42,42,44,46	0
7	D3T	D	901	28/28	0.93	0.13	0.83	11,16,21,21	28
6	MN	A	904	1/1	0.96	0.11	-0.85	16,16,16,16	1
4	SO4	A	901	5/5	0.89	0.20	-	46,65,74,76	0
4	SO4	D	902	5/5	0.98	0.07	-	25,33,37,43	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.