



# Full wwPDB X-ray Structure Validation Report ⓘ

May 22, 2020 – 06:53 am BST

PDB ID : 4EZ9  
Title : Bacillus DNA Polymerase I Large Fragment Complex 2  
Authors : Wang, W.; Beese, L.S.  
Deposited on : 2012-05-02  
Resolution : 1.64 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 : 2.11  
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) : 2.11

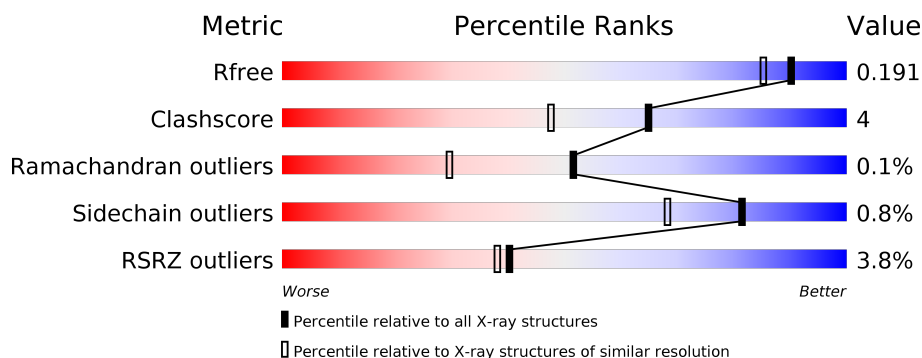
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.64 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	3122 (1.66-1.62)
Clashscore	141614	3268 (1.66-1.62)
Ramachandran outliers	138981	3215 (1.66-1.62)
Sidechain outliers	138945	3215 (1.66-1.62)
RSRZ outliers	127900	3079 (1.66-1.62)

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 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	592	<div> <div>7%</div> <div>88%</div> <div>10%</div> <div>.</div> </div>
1	D	592	<div> <div>%</div> <div>90%</div> <div>7%</div> <div>.</div> </div>
2	B	9	<div> <div>44%</div> <div>44%</div> <div>11%</div> </div>
2	E	9	<div> <div>67%</div> <div>22%</div> <div>11%</div> </div>
3	C	13	<div> <div>38%</div> <div>46%</div> <div>15%</div> </div>
3	F	13	<div> <div>31%</div> <div>62%</div> <div>8%</div> </div>

## 2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 21531 atoms, of which 9964 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	578	Total	C	H	N	O	S	0	2	0
			9362	2955	4718	805	867	17			
1	D	579	Total	C	H	N	O	S	0	9	0
			9457	2986	4771	809	873	18			

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(\*CP\*AP\*TP\*TP\*AP\*GP\*AP\*GP\*TP\*CP\*AP\*GP\*G)-3').

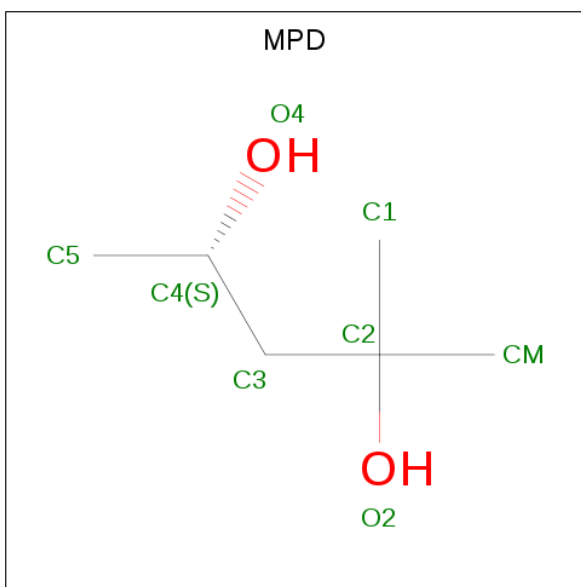
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	C	11	Total	C	H	N	O	P	0	0	0
			354	109	124	44	66	11			
3	F	12	Total	C	H	N	O	P	0	0	0
			383	119	135	49	69	11			

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



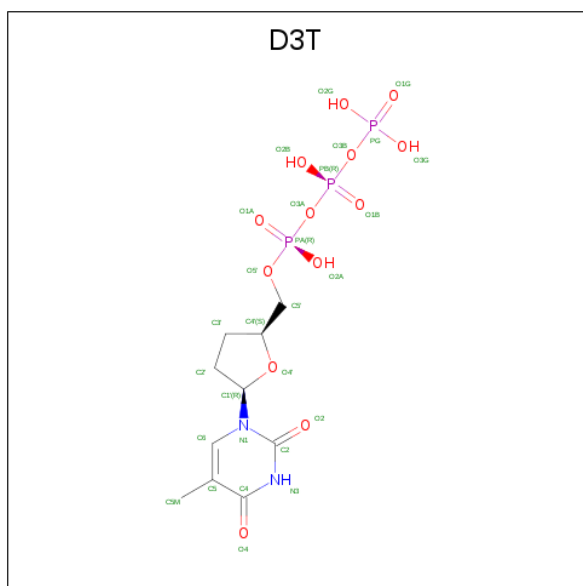
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	O	S	0	0
			5	4	1		
4	D	1	Total	O	S	0	0
			5	4	1		

- 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 2',3'-DIDEOXY-THYMIDINE-5'-TRIPHOSPHATE (three-letter code: D3T) (formula:  $\text{C}_{10}\text{H}_{17}\text{N}_2\text{O}_{13}\text{P}_3$ ).



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

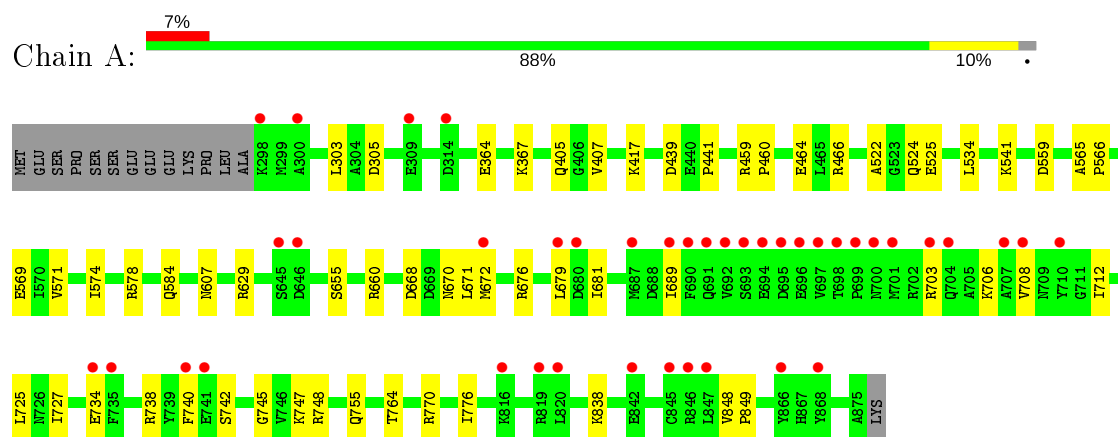
- Molecule 7 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	442	Total O 442 442	0	0
7	D	710	Total O 710 710	0	0
7	B	39	Total O 39 39	0	0
7	C	69	Total O 69 69	0	0
7	E	34	Total O 34 34	0	0
7	F	69	Total O 69 69	0	0

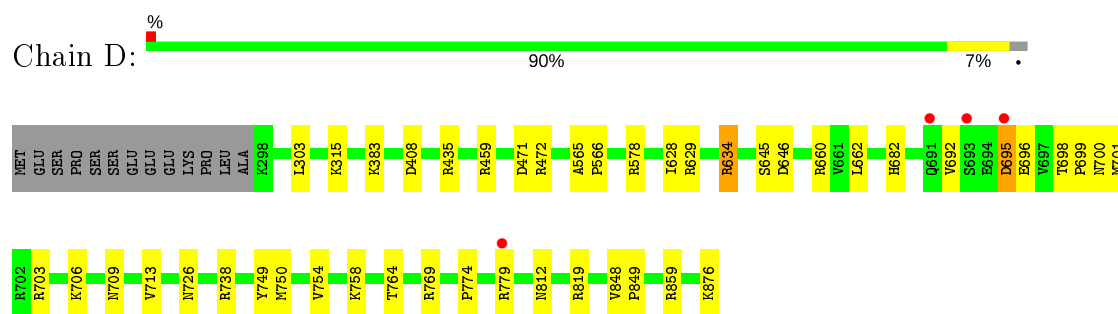
### 3 Residue-property plots [i](#)

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

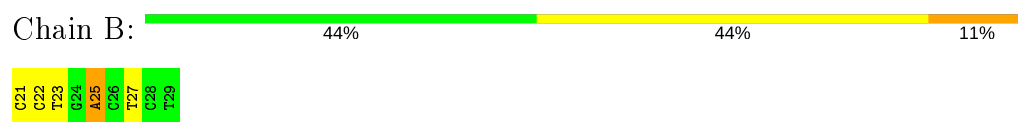
#### • Molecule 1: DNA polymerase



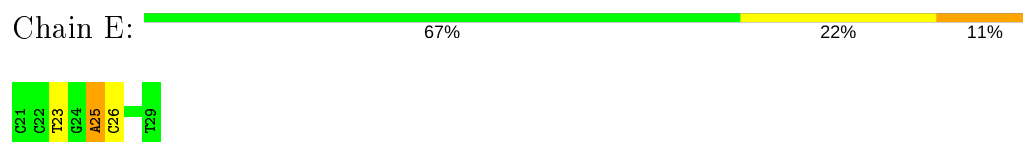
#### • Molecule 1: DNA polymerase



#### • 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')



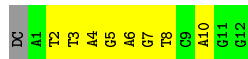
- Molecule 3: DNA (5'-D(\*CP\*AP\*TP\*TP\*AP\*GP\*AP\*GP\*TP\*CP\*AP\*GP\*G)-3')

Chain C:  38% 46% 15%



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

Chain F:  31% 62% 8%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	93.92Å 108.96Å 149.49Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.12 – 1.64 43.13 – 1.64	Depositor EDS
% Data completeness (in resolution range)	86.2 (43.12-1.64) 86.2 (43.13-1.64)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.58 (at 1.64Å)	Xtriage
Refinement program	PHENIX dev_1026	Depositor
R, $R_{free}$	0.168 , 0.191 0.168 , 0.191	Depositor DCC
$R_{free}$ test set	7359 reflections (4.55%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	18.6	Xtriage
Anisotropy	0.062	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.41 , 52.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	21531	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.99% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.



## 5 Model quality [i](#)

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

Bond lengths and bond angles in the following residue types are not validated in this section: 2DT, D3T, SO4, MPD

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	0.42	0/4734	0.56	0/6400
1	D	0.59	0/4801	0.71	4/6489 (0.1%)
2	B	1.12	0/173	1.95	7/264 (2.7%)
2	E	1.10	0/173	2.02	7/264 (2.7%)
3	C	1.07	0/258	1.76	4/397 (1.0%)
3	F	1.11	1/279 (0.4%)	1.72	8/430 (1.9%)
All	All	0.58	1/10418 (0.0%)	0.83	30/14244 (0.2%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	F	7	DG	N7-C5	6.06	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	-9.98	101.01	108.00
3	F	8	DT	O4'-C1'-N1	-9.80	101.14	108.00
2	E	23	DT	O4'-C1'-N1	-9.63	101.26	108.00
3	C	8	DT	O4'-C1'-N1	-8.29	102.20	108.00
2	E	25	DA	O5'-P-OP2	-7.87	98.62	105.70
1	D	634	ARG	NE-CZ-NH1	7.78	124.19	120.30
3	C	2	DT	O4'-C1'-N1	7.37	113.16	108.00
1	D	634	ARG	NE-CZ-NH2	-7.17	116.72	120.30
2	E	25	DA	C5'-C4'-C3'	7.08	126.84	114.10
2	E	26	DC	O4'-C4'-C3'	6.97	110.19	106.00
2	B	23	DT	N3-C4-O4	6.93	124.06	119.90
3	F	10	DA	O4'-C1'-N9	-6.59	103.39	108.00
2	E	25	DA	O4'-C4'-C3'	-6.38	101.95	104.50
3	C	5	DG	O4'-C4'-C3'	-6.38	101.95	104.50

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	F	2	DT	O4'-C1'-N1	6.34	112.44	108.00
2	B	25	DA	C5'-C4'-C3'	6.16	125.19	114.10
1	D	859	ARG	NE-CZ-NH1	6.10	123.35	120.30
2	B	25	DA	O4'-C4'-C3'	-5.94	102.12	104.50
2	E	26	DC	O4'-C1'-N1	5.73	112.01	108.00
3	F	6	DA	C5-C6-N1	-5.52	114.94	117.70
2	B	21	DC	O4'-C1'-C2'	5.48	110.28	105.90
2	B	25	DA	C1'-O4'-C4'	-5.48	104.62	110.10
3	F	6	DA	C2-N3-C4	-5.47	107.86	110.60
2	B	23	DT	C6-C5-C7	-5.42	119.64	122.90
3	C	5	DG	N3-C4-N9	-5.42	122.75	126.00
2	E	25	DA	O5'-P-OP1	5.41	117.19	110.70
1	D	660	ARG	NE-CZ-NH1	5.40	123.00	120.30
3	F	5	DG	O4'-C4'-C3'	-5.39	102.34	104.50
3	F	4	DA	O4'-C1'-C2'	5.19	110.05	105.90
3	F	3	DT	O4'-C1'-C2'	5.08	109.97	105.90

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	4644	4718	4706	37	0
1	D	4686	4771	4745	42	0
2	B	175	101	104	3	0
2	E	175	101	104	1	0
3	C	230	124	125	2	0
3	F	248	135	137	0	0
4	A	5	0	0	0	0
4	D	5	0	0	0	0
5	A	8	14	14	1	0
6	D	28	0	13	0	0
7	A	442	0	0	13	0
7	B	39	0	0	1	0
7	C	69	0	0	1	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	D	710	0	0	16	1
7	E	34	0	0	0	0
7	F	69	0	0	0	0
All	All	11567	9964	9948	83	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 4.

All (83) 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:749[A]:TYR:CE2	1:D:750[A]:MET:SD	2.57	0.97
1:D:459:ARG:NH1	7:D:1617:HOH:O	2.08	0.86
1:D:812[A]:ASN:OD1	7:D:1422:HOH:O	1.94	0.84
1:D:629:ARG:NH1	7:D:1218:HOH:O	2.11	0.84
1:D:472:ARG:NH2	7:D:1167:HOH:O	2.12	0.83
1:D:749[A]:TYR:CD2	1:D:750[A]:MET:SD	2.71	0.83
1:D:726:ASN:HB3	7:D:1614:HOH:O	1.82	0.79
5:A:902:MPD:HM2	7:A:1013:HOH:O	1.84	0.78
1:D:471:ASP:OD2	7:D:1475:HOH:O	2.02	0.76
1:D:459:ARG:NH2	7:D:1657:HOH:O	2.19	0.75
1:D:749[A]:TYR:HE2	1:D:750[A]:MET:SD	2.09	0.74
1:D:662:LEU:HD22	1:D:713[B]:VAL:HG21	1.70	0.73
1:D:408:ASP:OD1	7:D:1552:HOH:O	2.07	0.72
1:A:607:ASN:OD1	7:A:1330:HOH:O	2.13	0.66
1:D:749[A]:TYR:CE2	1:D:750[A]:MET:HE1	2.32	0.65
1:A:770:ARG:NH2	7:A:1422:HOH:O	2.29	0.64
1:A:584:GLN:HG3	7:A:1353:HOH:O	1.97	0.64
1:A:367:LYS:NZ	7:A:1439:HOH:O	2.31	0.62
1:D:383:LYS:HE3	7:D:1416:HOH:O	2.00	0.62
1:A:703:ARG:NH2	7:A:1200:HOH:O	2.26	0.61
1:D:738:ARG:HD3	7:D:1411:HOH:O	2.00	0.61
1:D:749[A]:TYR:CE2	1:D:750[A]:MET:CE	2.83	0.61
1:D:749[A]:TYR:HE2	1:D:750[A]:MET:CE	2.15	0.60
1:D:754:VAL:HG12	1:D:758:LYS:HE2	1.84	0.59
1:A:655:SER:O	1:A:660:ARG:NH1	2.35	0.59
1:A:838:LYS:NZ	7:A:1349:HOH:O	2.33	0.58
1:A:672:MET:CE	1:A:676:ARG:NH2	2.68	0.57
1:D:709:ASN:O	1:D:713[B]:VAL:HG12	2.07	0.55
1:D:315:LYS:HD2	7:D:1405:HOH:O	2.07	0.55
1:D:819:ARG:NH1	7:D:1611:HOH:O	2.39	0.55

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:726:ASN:CB	7:D:1614:HOH:O	2.47	0.54
1:D:692:VAL:HB	1:D:696:GLU:HB2	1.92	0.52
1:D:578:ARG:NH1	2:E:25:DA:H5''	2.25	0.52
1:A:364:GLU:HG2	7:A:1439:HOH:O	2.09	0.51
1:D:682:HIS:CG	1:D:706:LYS:HG2	2.44	0.51
1:A:708:VAL:O	1:A:712:ILE:HG12	2.12	0.50
1:A:740:PHE:CG	1:A:747:LYS:HD3	2.46	0.50
1:A:671:LEU:HD12	1:A:681:ILE:HD11	1.94	0.50
1:A:689:ILE:O	1:A:738:ARG:NH1	2.44	0.49
1:A:848:VAL:HB	1:A:849:PRO:HD3	1.94	0.49
1:D:695:ASP:N	1:D:695:ASP:OD2	2.45	0.49
1:A:522:ALA:O	1:A:541:LYS:HE2	2.14	0.48
1:A:565:ALA:HA	1:A:571:VAL:CG2	2.43	0.48
1:A:464:GLU:OE1	7:A:1189:HOH:O	2.20	0.48
3:C:12:DG:H1'	7:C:141:HOH:O	2.13	0.48
1:D:698:THR:HG22	1:D:700:ASN:H	1.77	0.48
1:D:698:THR:HB	1:D:701:MET:SD	2.53	0.48
1:D:754:VAL:CG1	1:D:758:LYS:HE2	2.42	0.48
1:D:435:ARG:NH2	7:D:1495:HOH:O	2.43	0.47
1:A:405:GLN:HB3	1:A:407:VAL:HG23	1.96	0.47
1:D:698:THR:HG23	1:D:699:PRO:HD2	1.97	0.47
1:A:670:ASN:ND2	1:A:742:SER:O	2.48	0.47
1:D:774:PRO:HA	7:D:1446:HOH:O	2.14	0.47
1:A:459:ARG:HB3	1:A:460:PRO:HD3	1.97	0.46
3:C:9:DC:H2'	3:C:10:DA:C8	2.51	0.46
1:A:569:GLU:HA	7:A:1298:HOH:O	2.15	0.46
1:D:565:ALA:N	1:D:566:PRO:CD	2.79	0.46
1:D:749[A]:TYR:CD2	1:D:750[A]:MET:HE1	2.50	0.46
1:D:749[A]:TYR:HD2	1:D:750[A]:MET:SD	2.37	0.45
1:A:668:ASP:OD2	1:A:745:GLY:N	2.50	0.45
1:D:848:VAL:HB	1:D:849:PRO:HD3	1.99	0.44
1:A:565:ALA:N	1:A:566:PRO:CD	2.82	0.43
1:D:645:SER:O	1:D:646:ASP:HB2	2.19	0.43
1:A:776:ILE:HG13	1:A:776:ILE:O	2.19	0.43
1:A:748:ARG:NE	7:A:1417:HOH:O	2.47	0.42
1:A:755:GLN:HG2	7:A:1112:HOH:O	2.20	0.42
2:B:22:DC:P	7:B:137:HOH:O	2.77	0.42
1:A:734:GLU:O	1:A:738:ARG:HG3	2.19	0.42
1:A:466:ARG:NH1	7:A:1424:HOH:O	2.51	0.41
1:D:634:ARG:HH22	1:D:876:LYS:HA	1.86	0.41
1:A:439:ASP:OD2	1:A:441:PRO:HD2	2.20	0.41

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:681:ILE:HD12	1:A:681:ILE:HA	1.97	0.41
1:D:749[A]:TYR:CD2	1:D:750[A]:MET:CE	3.03	0.41
1:A:764:THR:HG22	1:A:770:ARG:HG2	2.03	0.41
1:D:764:THR:HA	1:D:769:ARG:O	2.21	0.41
1:A:725:LEU:HB2	1:A:727:ILE:HG12	2.03	0.40
1:D:700:ASN:OD1	1:D:703:ARG:NH2	2.54	0.40
1:A:578:ARG:NH1	2:B:25:DA:H5"	2.36	0.40
1:A:629:ARG:HH22	2:B:27:DT:P	2.42	0.40
1:A:747:LYS:HD2	1:A:747:LYS:HA	1.80	0.40
1:A:524:GLN:HG2	1:A:525:GLU:O	2.21	0.40
1:A:534:LEU:HD11	1:A:574:ILE:HD13	2.04	0.40
1:D:779:ARG:NH1	7:D:1636:HOH:O	2.53	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)
7:D:1649:HOH:O	7:D:1659:HOH:O[3_854]	2.18	0.02

## 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	
1	A	578/592 (98%)	564 (98%)	14 (2%)	0	100	100
1	D	586/592 (99%)	572 (98%)	13 (2%)	1 (0%)	47	26
All	All	1164/1184 (98%)	1136 (98%)	27 (2%)	1 (0%)	51	29

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	628	ILE

### 5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	496/507 (98%)	490 (99%)	6 (1%)	71	51
1	D	503/507 (99%)	501 (100%)	2 (0%)	91	84
All	All	999/1014 (98%)	991 (99%)	8 (1%)	81	68

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	303	LEU
1	A	305	ASP
1	A	417	LYS
1	A	559	ASP
1	A	679	LEU
1	A	706	LYS
1	D	303	LEU
1	D	695	ASP

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

Mol	Chain	Res	Type
1	A	709	ASN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

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

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The

Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	2DT	B	29	3,2	14,20,21	1.32	1 (7%)	12,28,31	4.40	2 (16%)
2	2DT	E	29	3,2	14,20,21	1.16	2 (14%)	12,28,31	4.31	2 (16%)

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	-	0/4/18/19	0/2/2/2
2	2DT	E	29	3,2	-	0/4/18/19	0/2/2/2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	29	2DT	C4-N3	3.04	1.38	1.33
2	E	29	2DT	C6-C5	-2.15	1.34	1.40
2	E	29	2DT	C4-N3	2.13	1.36	1.33

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	29	2DT	C4-N3-C2	14.65	127.51	115.14
2	E	29	2DT	C4-N3-C2	14.40	127.30	115.14
2	B	29	2DT	C5M-C5-C6	2.48	123.92	118.68
2	E	29	2DT	C3'-C2'-C1'	2.14	105.25	102.78

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 carbohydrates in this entry.

## 5.6 Ligand geometry

4 ligands 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$
5	MPD	A	902	-	7,7,7	0.34	0	9,10,10	0.69	0
4	SO4	D	902	-	4,4,4	0.17	0	6,6,6	0.54	0
6	D3T	D	901	-	22,29,29	1.38	2 (9%)	25,45,45	1.84	2 (8%)
4	SO4	A	901	-	4,4,4	0.22	0	6,6,6	0.30	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
5	MPD	A	902	-	-	3/5/5/5	-
6	D3T	D	901	-	-	2/19/31/31	0/2/2/2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	D	901	D3T	O4-C4	3.91	1.34	1.24
6	D	901	D3T	C3'-C2'	-2.88	1.46	1.54

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	D	901	D3T	C4-N3-C2	7.03	121.08	115.14
6	D	901	D3T	C5-C6-N1	-4.13	117.74	122.19

There are no chirality outliers.

All (5) torsion outliers are listed below:



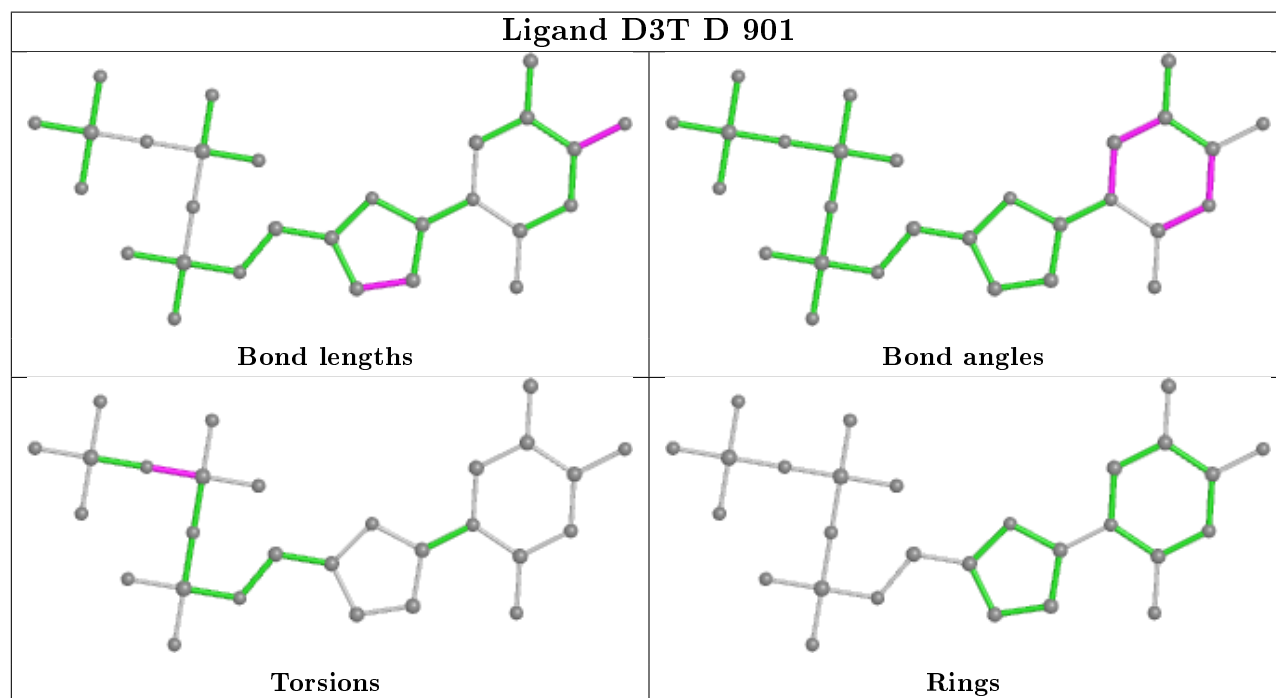
Mol	Chain	Res	Type	Atoms
6	D	901	D3T	PG-O3B-PB-O2B
5	A	902	MPD	C2-C3-C4-O4
5	A	902	MPD	CM-C2-C3-C4
6	D	901	D3T	PG-O3B-PB-O1B
5	A	902	MPD	C2-C3-C4-C5

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	902	MPD	1	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.



## 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	578/592 (97%)	0.24	41 (7%) 16 14	14, 34, 65, 108	0
1	D	579/592 (97%)	-0.12	4 (0%) 87 88	9, 20, 38, 54	0
2	B	8/9 (88%)	-0.17	0 100 100	19, 23, 40, 48	0
2	E	8/9 (88%)	-0.36	0 100 100	13, 20, 37, 48	0
3	C	11/13 (84%)	-0.14	0 100 100	16, 20, 49, 59	0
3	F	12/13 (92%)	-0.27	0 100 100	12, 17, 37, 48	0
All	All	1196/1228 (97%)	0.05	45 (3%) 40 38	9, 26, 56, 108	0

All (45) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	699	PRO	9.8
1	A	679	LEU	8.0
1	A	703	ARG	7.5
1	A	690	PHE	7.1
1	A	689	ILE	6.9
1	A	697	VAL	6.2
1	A	298	LYS	5.3
1	A	692	VAL	5.1
1	A	816	LYS	5.0
1	A	698	THR	4.8
1	A	819	ARG	4.7
1	A	691	GLN	4.5
1	A	846	ARG	4.4
1	A	842	GLU	4.0
1	A	694	GLU	3.6
1	A	695	ASP	3.3
1	A	868	TYR	3.3
1	A	740	PHE	3.3
1	A	735	PHE	3.2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	845	CYS	3.2
1	A	680	ASP	3.2
1	A	700	ASN	3.1
1	A	866	TYR	3.1
1	A	687	MET	3.0
1	A	704	GLN	3.0
1	D	691	GLN	2.8
1	D	779	ARG	2.8
1	A	672	MET	2.7
1	A	701	MET	2.6
1	A	847	LEU	2.6
1	A	741	GLU	2.6
1	A	693	SER	2.6
1	A	696	GLU	2.6
1	A	710	TYR	2.5
1	A	820	LEU	2.4
1	D	695	ASP	2.3
1	A	645	SER	2.2
1	A	300	ALA	2.2
1	A	708	VAL	2.1
1	A	314	ASP	2.1
1	A	646	ASP	2.1
1	A	734	GLU	2.1
1	A	707	ALA	2.1
1	D	693	SER	2.1
1	A	309	GLU	2.1

## 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	2DT	B	29	19/20	0.97	0.09	17,23,29,34	0
2	2DT	E	29	19/20	0.99	0.11	12,15,19,19	0

## 6.3 Carbohydrates [i](#)

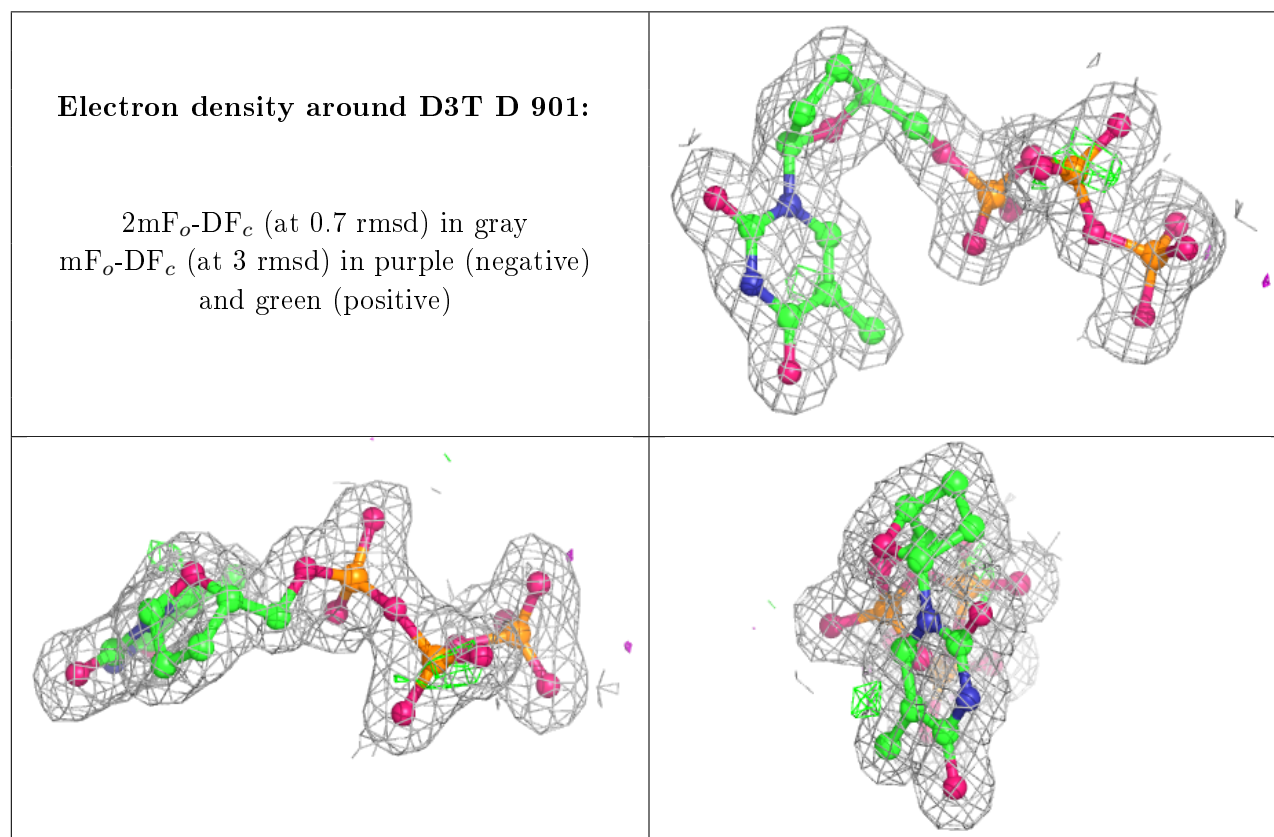
There are no carbohydrates in this entry.

## 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. 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( $\text{\AA}^2$ )	Q<0.9
5	MPD	A	902	8/8	0.94	0.10	16,26,38,50	0
4	SO4	A	901	5/5	0.96	0.09	30,33,35,35	0
6	D3T	D	901	28/28	0.98	0.11	12,16,22,24	0
4	SO4	D	902	5/5	0.99	0.07	25,26,35,40	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.



## 6.5 Other polymers [i](#)

There are no such residues in this entry.