



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 06:10 PM GMT

PDB ID : 4KPY  
Title : DNA binding protein and DNA complex structure  
Authors : Sheng, G.; Zhao, H.; Wang, J.; Rao, Y.; Wang, Y.  
Deposited on : 2013-05-14  
Resolution : 2.41 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](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

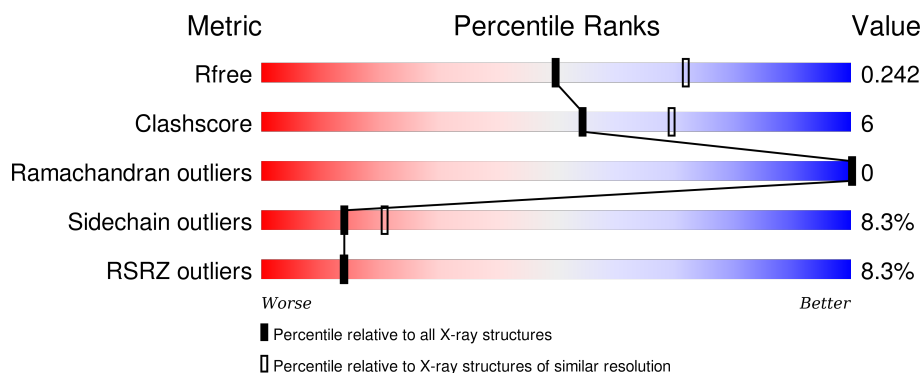
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.41 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	2919 (2.40-2.40)
Clashscore	102246	3407 (2.40-2.40)
Ramachandran outliers	100387	3351 (2.40-2.40)
Sidechain outliers	100360	3352 (2.40-2.40)
RSRZ outliers	91569	2928 (2.40-2.40)

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	685	<div> <div>8%</div> <div>80%</div> <div>16%</div> <div>..</div> </div>
1	B	685	<div> <div>9%</div> <div>76%</div> <div>20%</div> <div>..</div> </div>
2	C	21	<div> <div>33%</div> <div>38%</div> <div>5%</div> <div>24%</div> </div>
2	E	21	<div> <div>43%</div> <div>29%</div> <div>5%</div> <div>24%</div> </div>
3	D	10	<div> <div>60%</div> <div>40%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	F	10	 70% 20% 10%
4	M	9	 44% 22% 33%
4	N	9	 33% 22% 11% 33%

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 12268 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 Uncharacterized protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	676	Total	C	N	O	S	0	0	0
			5278	3372	992	907	7			
1	B	675	Total	C	N	O	S	0	1	0
			5272	3371	991	904	6			

- Molecule 2 is a DNA chain called DNA (5'-D(P\*TP\*GP\*AP\*GP\*GP\*TP\*AP\*GP\*TP\*AP\*GP\*GP\*TP\*TP\*GP\*TP\*AP\*TP\*AP\*GP\*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	16	Total	C	N	O	P	0	0	0
			338	160	62	100	16			
2	E	16	Total	C	N	O	P	0	0	0
			338	160	62	100	16			

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

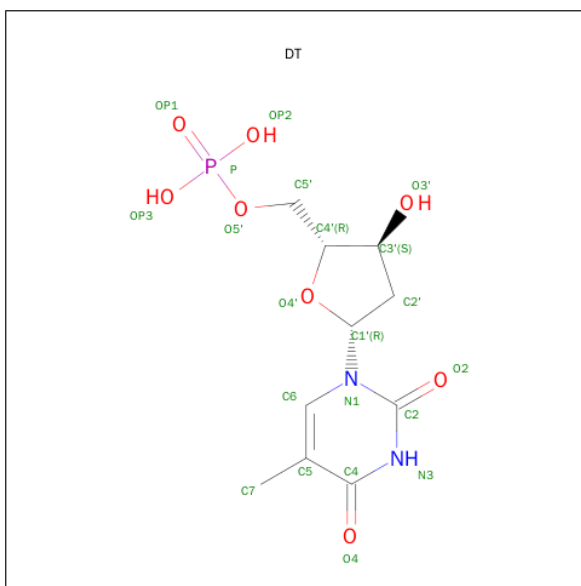
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	10	Total	C	N	O	P	0	0	0
			201	96	33	62	10			
3	F	10	Total	C	N	O	P	0	0	0
			201	96	33	62	10			

- Molecule 4 is a DNA chain called DNA (5'-D(\*TP\*AP\*TP\*AP\*CP\*AP\*AP\*CP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	M	6	Total	C	N	O	P	0	0	0
			120	57	24	33	6			
4	N	6	Total	C	N	O	P	0	0	0
			116	57	24	30	5			

- Molecule 5 is THYMIDINE-5'-MONOPHOSPHATE (three-letter code: DT) (formula:

C<sub>10</sub>H<sub>15</sub>N<sub>2</sub>O<sub>8</sub>P).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	A	1	Total	C	N	O	P	0	0
			21	10	2	8	1		
5	B	1	Total	C	N	O	P	0	0
			21	10	2	8	1		

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	B	2	Total	Mn	0	0
			2	2		
6	A	2	Total	Mn	0	0
			2	2		
6	C	1	Total	Mn	0	0
			1	1		
6	E	1	Total	Mn	0	0
			1	1		

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	138	Total	O	0	0
			138	138		
7	B	161	Total	O	0	0
			161	161		
7	C	17	Total	O	0	0
			17	17		

*Continued on next page...*

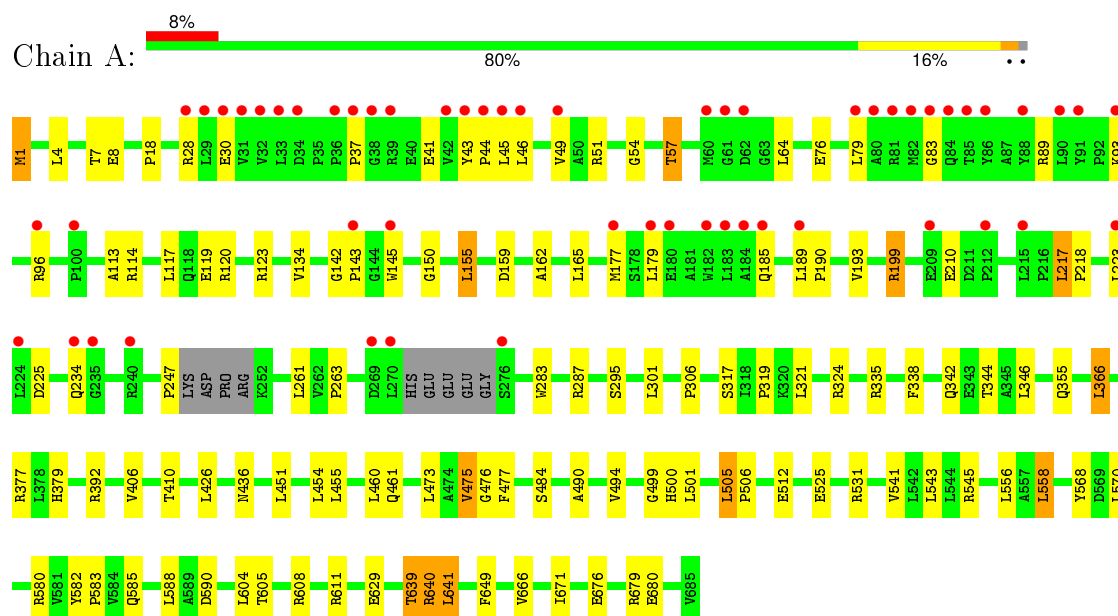
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	D	9	Total 9	O 9	0	0
7	E	20	Total 20	O 20	0	0
7	F	8	Total 8	O 8	0	0
7	M	3	Total 3	O 3	0	0

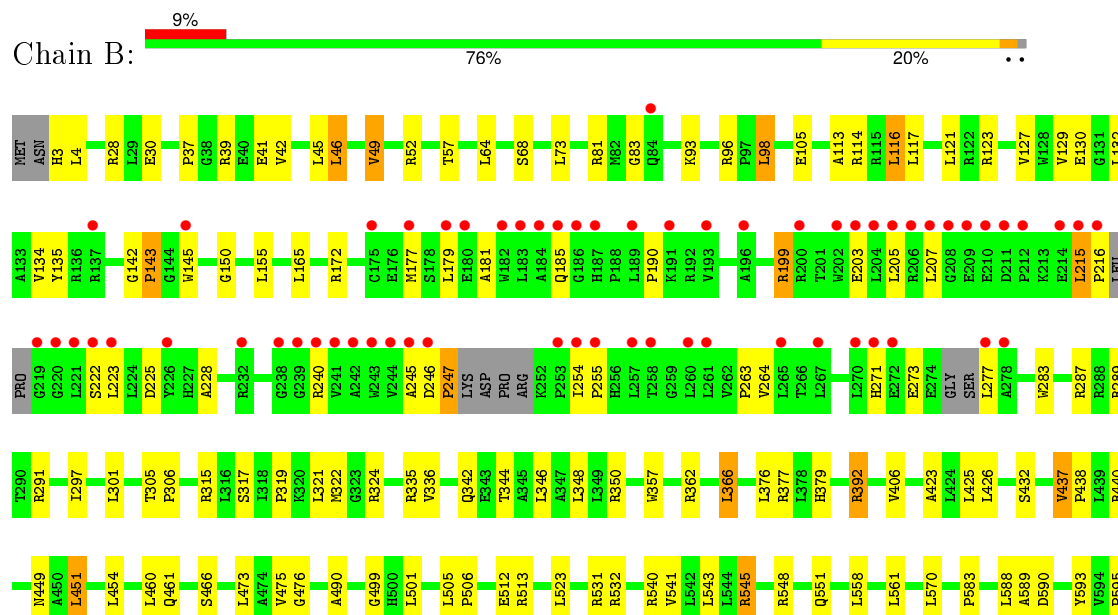
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $\text{RSRZ} > 2$ ). 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: Uncharacterized protein

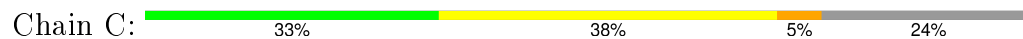


#### • Molecule 1: Uncharacterized protein





- Molecule 2: DNA (5'-D(P\*TP\*GP\*AP\*GP\*GP\*TP\*AP\*GP\*TP\*AP\*GP\*GP\*TP\*TP\*GP\*T P\*AP\*TP\*AP\*GP\*T)-3')



- Molecule 2: DNA (5'-D(P\*TP\*GP\*AP\*GP\*GP\*TP\*AP\*GP\*TP\*AP\*GP\*GP\*TP\*TP\*GP\*T P\*AP\*TP\*AP\*GP\*T)-3')



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



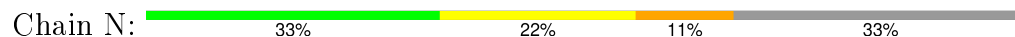
- Molecule 3: DNA (5'-D(P\*TP\*AP\*CP\*TP\*AP\*CP\*CP\*TP\*CP\*G)-3')



- Molecule 4: DNA (5'-D(\*TP\*AP\*TP\*AP\*CP\*AP\*AP\*CP\*C)-3')



- Molecule 4: DNA (5'-D(\*TP\*AP\*TP\*AP\*CP\*AP\*AP\*CP\*C)-3')





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	111.33Å 118.36Å 160.87Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.31 – 2.41 48.31 – 2.41	Depositor EDS
% Data completeness (in resolution range)	97.1 (48.31-2.41) 97.0 (48.31-2.41)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.13 (at 2.39Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, $R_{free}$	0.190 , 0.235 0.201 , 0.242	Depositor DCC
$R_{free}$ test set	4036 reflections (5.28%)	DCC
Wilson B-factor (Å <sup>2</sup> )	44.7	Xtriage
Anisotropy	0.025	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 48.2	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Outliers	3 of 80738 reflections (0.004%)	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	12268	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.03% 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: MN

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	0.33	0/5402	0.55	2/7333 (0.0%)
1	B	0.34	0/5399	0.56	3/7329 (0.0%)
2	C	0.87	1/379 (0.3%)	1.35	6/584 (1.0%)
2	E	0.81	1/379 (0.3%)	1.34	4/584 (0.7%)
3	D	0.92	1/223 (0.4%)	1.26	2/339 (0.6%)
3	F	0.89	1/223 (0.4%)	1.29	2/339 (0.6%)
4	M	0.68	0/134	1.31	1/203 (0.5%)
4	N	0.54	0/130	1.23	2/198 (1.0%)
All	All	0.43	4/12269 (0.0%)	0.70	22/16909 (0.1%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	1	DT	OP3-P	-11.03	1.48	1.61
2	C	1	DT	OP3-P	-10.51	1.48	1.61
3	D	10	DT	OP3-P	-10.45	1.48	1.61
3	F	10	DT	OP3-P	-9.81	1.49	1.61

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	F	11	DA	O4'-C1'-N9	8.26	113.78	108.00
2	E	1	DT	OP1-P-OP2	-8.13	107.41	119.60
4	N	7	DA	O4'-C1'-N9	7.36	113.15	108.00
2	C	1	DT	OP1-P-OP2	-7.33	108.60	119.60
2	C	15	DG	O4'-C1'-N9	6.23	112.36	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	5278	0	5338	65	0
1	B	5272	0	5330	82	0
2	C	338	0	183	5	0
2	E	338	0	183	4	0
3	D	201	0	114	1	0
3	F	201	0	114	1	0
4	M	120	0	66	2	0
4	N	116	0	64	2	0
5	A	21	0	13	1	0
5	B	21	0	13	0	0
6	A	2	0	0	0	0
6	B	2	0	0	0	0
6	C	1	0	0	0	0
6	E	1	0	0	0	0
7	A	138	0	0	2	0
7	B	161	0	0	6	0
7	C	17	0	0	0	0
7	D	9	0	0	0	0
7	E	20	0	0	1	0
7	F	8	0	0	0	0
7	M	3	0	0	0	0
All	All	12268	0	11418	150	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 6.

The worst 5 of 150 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:319:PRO:HG2	1:A:640:ARG:HD2	1.63	0.81
1:B:344:THR:HG21	1:B:460:LEU:HD11	1.66	0.76
1:B:513:ARG:NH2	1:B:551:GLN:O	2.17	0.75
1:B:597:GLU:O	7:B:900:HOH:O	2.05	0.73
1:A:512:GLU:OE2	1:A:545:ARG:NH2	2.23	0.71

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	670/685 (98%)	657 (98%)	13 (2%)	0	100	100
1	B	668/685 (98%)	656 (98%)	12 (2%)	0	100	100
All	All	1338/1370 (98%)	1313 (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	521/549 (95%)	481 (92%)	40 (8%)	16	24
1	B	520/549 (95%)	474 (91%)	46 (9%)	12	18
All	All	1041/1098 (95%)	955 (92%)	86 (8%)	14	21

5 of 86 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	676	GLU
1	B	165	LEU
1	B	604	LEU
1	B	3	HIS
1	B	52	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

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

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 8 ligands modelled in this entry, 6 are monoatomic - leaving 2 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
5	DT	A	701	-	17,22,22	1.25	2 (11%)	22,33,33	3.38	3 (13%)
5	DT	B	701	-	17,22,22	1.27	3 (17%)	22,33,33	3.54	3 (13%)

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	DT	A	701	-	-	0/6/22/22	0/2/2/2
5	DT	B	701	-	-	0/6/22/22	0/2/2/2

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	701	DT	C6-C5	-2.33	1.33	1.40
5	B	701	DT	C6-C5	-2.15	1.34	1.40
5	B	701	DT	C6-N1	2.29	1.38	1.35
5	A	701	DT	C4-N3	2.76	1.38	1.33
5	B	701	DT	C4-N3	3.01	1.38	1.33

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	B	701	DT	C5-C4-N3	-8.89	115.24	125.14
5	A	701	DT	C5-C4-N3	-8.64	115.51	125.14
5	A	701	DT	OP2-P-OP1	2.21	117.69	110.58
5	B	701	DT	OP2-P-OP1	2.44	118.44	110.58
5	A	701	DT	C4-N3-C2	12.69	126.21	115.25

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	701	DT	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	676/685 (98%)	0.22	55 (8%) 15 14	19, 45, 95, 141	0
1	B	675/685 (98%)	0.24	62 (9%) 11 11	20, 40, 111, 140	0
2	C	16/21 (76%)	-0.62	0 100 100	27, 41, 70, 92	0
2	E	16/21 (76%)	-0.57	0 100 100	32, 38, 60, 62	0
3	D	10/10 (100%)	-0.78	0 100 100	36, 46, 52, 59	0
3	F	10/10 (100%)	-0.59	0 100 100	37, 56, 59, 62	0
4	M	6/9 (66%)	-0.23	0 100 100	39, 53, 76, 79	0
4	N	6/9 (66%)	-0.07	0 100 100	38, 60, 89, 94	0
All	All	1415/1450 (97%)	0.19	117 (8%) 14 14	19, 42, 101, 141	0

The worst 5 of 117 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	205	LEU	8.1
1	B	207	LEU	8.0
1	B	189	LEU	7.9
1	A	79	LEU	7.7
1	B	223	LEU	7.4

### 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 [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	A	703	1/1	0.99	0.12	0.39	32,32,32,32	0
5	DT	B	701	21/21	0.93	0.37	0.22	82,102,118,122	0
6	MN	C	101	1/1	0.99	0.13	0.09	28,28,28,28	0
5	DT	A	701	21/21	0.97	0.20	-0.02	43,53,84,93	0
6	MN	E	101	1/1	1.00	0.13	-0.54	33,33,33,33	0
6	MN	A	702	1/1	1.00	0.11	-0.62	32,32,32,32	0
6	MN	B	702	1/1	0.99	0.11	-1.40	32,32,32,32	0
6	MN	B	703	1/1	0.99	0.09	-2.52	34,34,34,34	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.