



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 29, 2020 – 04:50 PM BST

PDB ID : 6ISF  
Title : Structure of 9N-I DNA polymerase incorporation with dT in the active site  
Authors : Linwu, S.W.; Maestre-Reyna, M.; Tsai, M.D.; Tu, Y.H.; Chang, W.H.  
Deposited on : 2018-11-16  
Resolution : 2.80 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

<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  
Xtriage (Phenix) : 1.13  
EDS : 2.13  
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) : 2.13

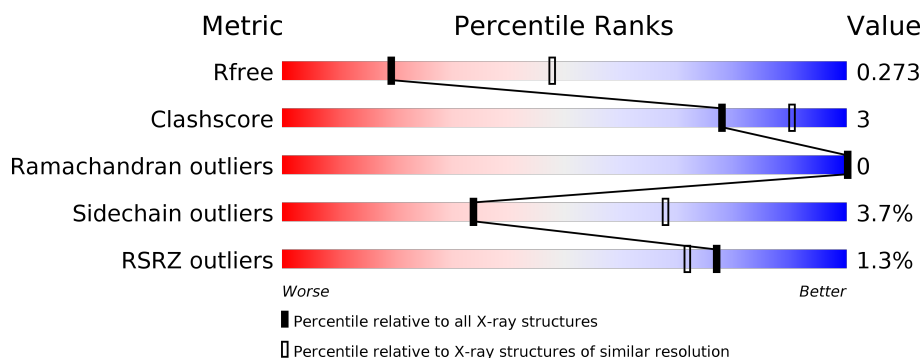
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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	783	<div> <div>11%</div> <div>86% 10% ..</div> </div>
1	B	783	<div> <div>89% 7% ..</div> </div>
2	C	15	<div> <div>87% 13%</div> </div>
2	E	15	<div> <div>100%</div> </div>
3	D	18	<div> <div>11% 100%</div> </div>
3	F	18	<div> <div>89% 11%</div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 13186 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	757	Total	C	N	O	S	0	1	0
			5933	3832	1011	1075	15			
1	B	756	Total	C	N	O	S	0	0	0
			5877	3793	996	1073	15			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	141	ALA	ASP	engineered mutation	UNP Q56366
A	143	ALA	GLU	engineered mutation	UNP Q56366
A	485	LEU	ALA	engineered mutation	UNP Q56366
A	776	LEU	-	expression tag	UNP Q56366
A	777	GLU	-	expression tag	UNP Q56366
A	778	HIS	-	expression tag	UNP Q56366
A	779	HIS	-	expression tag	UNP Q56366
A	780	HIS	-	expression tag	UNP Q56366
A	781	HIS	-	expression tag	UNP Q56366
A	782	HIS	-	expression tag	UNP Q56366
A	783	HIS	-	expression tag	UNP Q56366
B	141	ALA	ASP	engineered mutation	UNP Q56366
B	143	ALA	GLU	engineered mutation	UNP Q56366
B	485	LEU	ALA	engineered mutation	UNP Q56366
B	776	LEU	-	expression tag	UNP Q56366
B	777	GLU	-	expression tag	UNP Q56366
B	778	HIS	-	expression tag	UNP Q56366
B	779	HIS	-	expression tag	UNP Q56366
B	780	HIS	-	expression tag	UNP Q56366
B	781	HIS	-	expression tag	UNP Q56366
B	782	HIS	-	expression tag	UNP Q56366
B	783	HIS	-	expression tag	UNP Q56366

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	15	Total	C	N	O	P	0	0	0
			302	145	53	90	14			
2	E	15	Total	C	N	O	P	0	0	0
			301	145	53	89	14			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	18	Total	C	N	O	P	0	0	0
			359	170	69	103	17			
3	F	18	Total	C	N	O	P	0	0	0
			359	170	69	103	17			

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	5	Total	Ca	0	0
			5	5		
4	A	2	Total	Ca	0	0
			2	2		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	28	Total	O	0	0
			28	28		
5	B	19	Total	O	0	0
			19	19		
5	E	1	Total	O	0	0
			1	1		



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

Chain E:  100%


There are no outlier residues recorded for this chain.

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

Chain D:  11% 100%



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

Chain F:  89% 11%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	92.09Å 107.36Å 255.66Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.38 – 2.80 86.64 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.5 (46.38-2.80) 99.5 (86.64-2.80)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.58 (at 2.82Å)	Xtriage
Refinement program	REFMAC 5.8.0238	Depositor
R, $R_{free}$	0.232 , 0.272 0.235 , 0.273	Depositor DCC
$R_{free}$ test set	3128 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	62.5	Xtriage
Anisotropy	0.253	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 48.5	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	0.90	EDS
Total number of atoms	13186	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	63.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.54% 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: 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	0.75	1/6076 (0.0%)	0.79	0/8231
1	B	0.75	0/6013	0.80	0/8160
2	C	0.30	0/337	0.73	0/518
2	E	0.32	0/336	0.79	0/517
3	D	0.35	0/403	0.72	0/620
3	F	0.35	0/403	0.74	0/620
All	All	0.72	1/13568 (0.0%)	0.79	0/18666

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	189	GLU	CD-OE1	6.41	1.32	1.25

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	5933	0	5682	42	0
1	B	5877	0	5611	29	0
2	C	302	0	171	2	0
2	E	301	0	168	0	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	D	359	0	197	0	0
3	F	359	0	197	2	0
4	A	2	0	0	0	0
4	B	5	0	0	0	0
5	A	28	0	0	0	0
5	B	19	0	0	0	0
5	E	1	0	0	0	0
All	All	13186	0	12026	72	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 3.

The worst 5 of 72 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:B:146:TYR:CB	1:B:292:ALA:HB1	1.84	1.06
1:A:220:LYS:O	1:A:224:GLU:HG3	1.72	0.89
1:A:266[A]:ARG:NH1	1:A:347:SER:O	2.07	0.87
1:B:146:TYR:CB	1:B:292:ALA:CB	2.57	0.81
1:A:265:ARG:NH2	1:A:696:GLY:O	2.12	0.81

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	756/783 (97%)	717 (95%)	39 (5%)	0	100	100
1	B	754/783 (96%)	714 (95%)	40 (5%)	0	100	100
All	All	1510/1566 (96%)	1431 (95%)	79 (5%)	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	563/675 (83%)	541 (96%)	22 (4%)	32	66
1	B	562/675 (83%)	541 (96%)	21 (4%)	34	68
All	All	1125/1350 (83%)	1082 (96%)	43 (4%)	34	67

5 of 43 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	627	LEU
1	B	140	PHE
1	B	627	LEU
1	A	671	ARG
1	A	745	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	91	GLN
1	B	491	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](#)

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

## 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry

Of 7 ligands modelled in this entry, 7 are monoatomic - leaving 0 for Mogul analysis.

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.

## 5.7 Other polymers

There are no such residues in this entry.

## 5.8 Polymer linkage issues

There are no chain breaks in this entry.

## 6 Fit of model and data [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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	757/783 (96%)	0.23	9 (1%) 79 73	34, 62, 96, 119	3 (0%)
1	B	756/783 (96%)	0.27	10 (1%) 77 72	36, 62, 104, 141	1 (0%)
2	C	15/15 (100%)	-0.04	0 100 100	33, 47, 60, 66	0
2	E	15/15 (100%)	0.18	0 100 100	34, 40, 58, 61	0
3	D	18/18 (100%)	0.46	2 (11%) 5 3	34, 46, 80, 110	2 (11%)
3	F	18/18 (100%)	0.11	0 100 100	34, 48, 67, 89	0
All	All	1579/1632 (96%)	0.25	21 (1%) 77 72	33, 62, 101, 141	6 (0%)

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	292	ALA	3.4
1	B	299	TRP	2.9
1	A	51	VAL	2.8
3	D	1	DA	2.7
1	A	736	GLN	2.7

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

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

### 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
4	CA	B	801	1/1	0.72	0.15	85,85,85,85	0
4	CA	B	802	1/1	0.77	0.14	104,104,104,104	0
4	CA	A	802	1/1	0.85	0.19	101,101,101,101	0
4	CA	B	804	1/1	0.86	0.08	90,90,90,90	0
4	CA	B	805	1/1	0.92	0.17	75,75,75,75	0
4	CA	A	801	1/1	0.92	0.09	80,80,80,80	0
4	CA	B	803	1/1	0.97	0.09	65,65,65,65	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.