



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 07:30 AM UTC

PDB ID : 4YD2 / pdb\_00004yd2  
Title : Nicked complex of human DNA Polymerase Mu with 2-nt gapped DNA substrate  
Authors : Moon, A.F.; Gosavi, R.A.; Kunkel, T.A.; Pedersen, L.C.; Bebenek, K.  
Deposited on : 2015-02-20  
Resolution : 2.47 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

MolProbity : 4-5-2 with Phenix2.0  
Mogul : 2022.3.0, CSD as543be (2022)  
Xtriage (Phenix) : 2.0  
EDS : 3.0  
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)  
CCP4 : 9.0.010 (Gargrove)  
Density-Fitness : 1.0.12  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.49

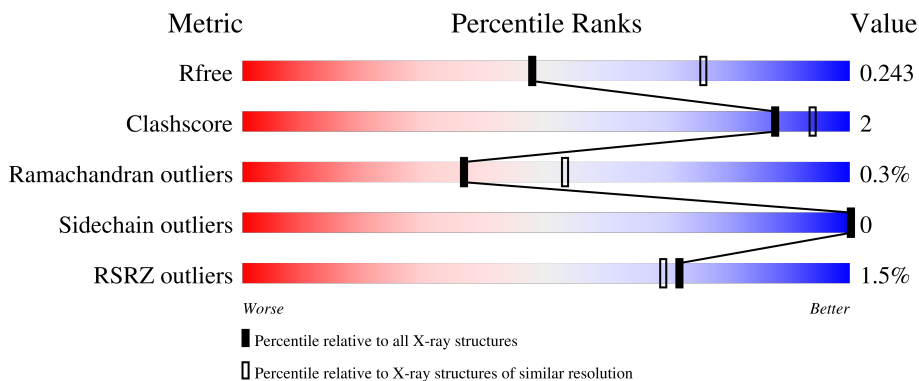
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.47 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))
$R_{free}$	180053	7589 (2.50-2.46)
Clashscore	190562	8295 (2.50-2.46)
Ramachandran outliers	187476	8164 (2.50-2.46)
Sidechain outliers	187428	8166 (2.50-2.46)
RSRZ outliers	180081	7593 (2.50-2.46)

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	354	
2	T	10	
3	P	5	
4	D	4	

## 2 Entry composition i

There are 8 unique types of molecules in this entry. The entry contains 3031 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-directed DNA/RNA polymerase mu.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	323	2551	1620	456	465	10	0	10	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	129	GLY	-	expression tag	UNP Q9NP87
A	130	SER	-	expression tag	UNP Q9NP87
A	131	ALA	-	expression tag	UNP Q9NP87
A	132	ALA	-	expression tag	UNP Q9NP87
A	133	ALA	-	expression tag	UNP Q9NP87
A	?	-	PRO	deletion	UNP Q9NP87
A	?	-	GLY	deletion	UNP Q9NP87
A	?	-	ALA	deletion	UNP Q9NP87
A	?	-	ALA	deletion	UNP Q9NP87
A	?	-	VAL	deletion	UNP Q9NP87
A	?	-	GLY	deletion	UNP Q9NP87
A	?	-	GLY	deletion	UNP Q9NP87
A	?	-	SER	deletion	UNP Q9NP87
A	?	-	THR	deletion	UNP Q9NP87
A	?	-	ARG	deletion	UNP Q9NP87
A	?	-	PRO	deletion	UNP Q9NP87
A	?	-	CYS	deletion	UNP Q9NP87
A	410	GLY	-	insertion	UNP Q9NP87

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	T	7	141	68	28	39	6	0	0	0

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace	
			Total	C	N	O				P
3	P	5	99	49	17	29	4	0	0	0

- Molecule 4 is a DNA chain called DNA (5'-D(P\*GP\*CP\*CP\*G)-3').

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace	
			Total	C	N	O				P
4	D	4	83	38	16	25	4	0	0	0

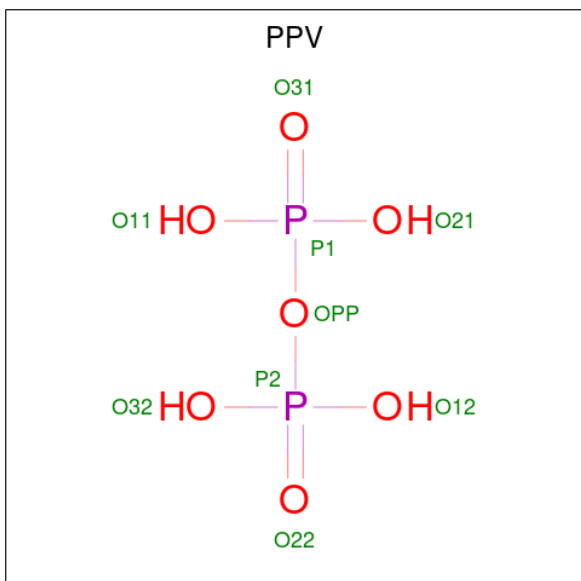
- Molecule 5 is SODIUM ION (CCD ID: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Na		
5	A	3	3	3	0	0

- Molecule 6 is CHLORIDE ION (CCD ID: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Cl		
6	A	3	3	3	0	0

- Molecule 7 is PYROPHOSPHATE (CCD ID: PPV) (formula: H<sub>4</sub>O<sub>7</sub>P<sub>2</sub>).



Mol	Chain	Residues	Atoms		ZeroOcc	AltConf	
			Total	O			P
7	A	1	5	4	1	0	0

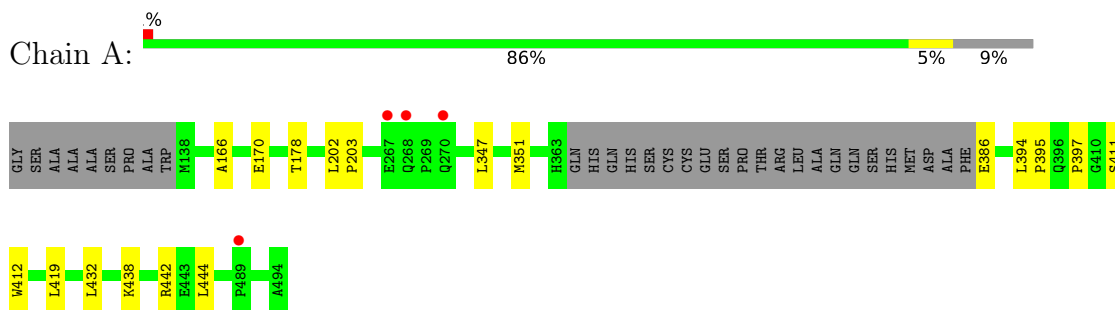
- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	135	Total 139	O 139	0	5
8	T	3	Total 3	O 3	0	0
8	P	3	Total 3	O 3	0	0
8	D	1	Total 1	O 1	0	0

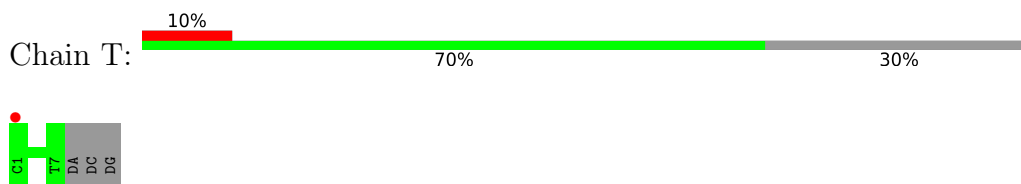
### 3 Residue-property plots [i](#)

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-directed DNA/RNA polymerase mu



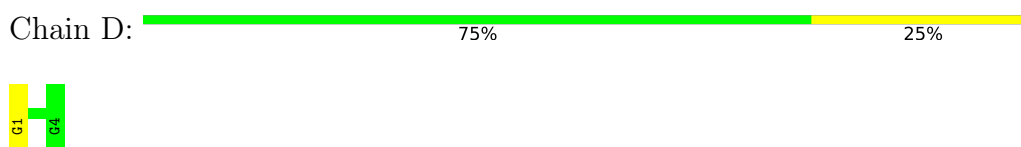
- Molecule 2: DNA (5'-D(\*CP\*GP\*GP\*CP\*AP\*AP\*T)-3')



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



- Molecule 4: DNA (5'-D(P\*GP\*CP\*CP\*G)-3')



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	60.08Å 68.70Å 114.98Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	22.61 – 2.47 22.61 – 2.47	Depositor EDS
% Data completeness (in resolution range)	97.0 (22.61-2.47) 96.8 (22.61-2.47)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.04 (at 2.47Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.8.4_1496)	Depositor
R, $R_{free}$	0.212 , 0.242 0.215 , 0.243	Depositor DCC
$R_{free}$ test set	858 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	29.0	Xtrriage
Anisotropy	0.076	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 51.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	3031	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	34.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.29% 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: PPV, NA, CL

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	0.27	0/2630	0.68	0/3563
2	T	0.18	0/158	0.32	0/242
3	P	0.15	0/110	0.34	0/168
4	D	0.16	0/92	0.27	0/138
All	All	0.26	0/2990	0.64	0/4111

There are no bond length outliers.

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	2551	0	2482	10	0
2	T	141	0	80	0	0
3	P	99	0	59	1	0
4	D	83	0	45	1	0
5	A	3	0	0	0	0
6	A	3	0	0	0	0
7	A	5	0	0	0	0
8	A	139	0	0	0	0
8	D	1	0	0	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
8	P	3	0	0	0	0
8	T	3	0	0	0	0
All	All	3031	0	2666	11	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 2.

The worst 5 of 11 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:347:LEU:HD12	1:A:386:GLU:HB3	1.87	0.54
1:A:178:THR:OG1	4:D:1:DG:N3	2.41	0.53
1:A:411:SER:OG	1:A:412:TRP:N	2.43	0.50
1:A:351[A]:MET:HE2	1:A:351[A]:MET:HA	2.00	0.43
3:P:4:DA:H2'	3:P:5:DT:C6	2.53	0.42

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	329/354 (93%)	322 (98%)	6 (2%)	1 (0%)	36 53

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	397	PRO

### 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	263/299 (88%)	263 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

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

Mol	Chain	Res	Type
1	A	355	GLN

### 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 oligosaccharides in this entry.

### 5.6 Ligand geometry [i](#)

Of 7 ligands modelled in this entry, 6 are monoatomic - leaving 1 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
7	PPV	A	507	5	4,4,8	1.56	1 (25%)	6,6,13	0.41	0

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	A	507	PPV	P2-O22	2.46	1.56	1.50

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 [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 [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	323/354 (91%)	0.03	4 (1%) 76 74	8, 31, 52, 81	10 (3%)
2	T	7/10 (70%)	1.09	1 (14%) 6 5	35, 48, 69, 96	0
3	P	5/5 (100%)	0.71	0 100 100	35, 42, 55, 64	0
4	D	4/4 (100%)	-0.07	0 100 100	45, 46, 54, 58	0
All	All	339/373 (90%)	0.06	5 (1%) 72 69	8, 31, 54, 96	10 (2%)

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	489[A]	PRO	6.4
2	T	1	DC	5.3
1	A	270	GLN	3.3
1	A	268	GLN	2.4
1	A	267	GLU	2.1

### 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 oligosaccharides 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	NA	A	501	1/1	0.82	0.09	18,18,18,18	0
5	NA	A	504	1/1	0.90	0.15	30,30,30,30	0
5	NA	A	505	1/1	0.92	0.07	37,37,37,37	0
6	CL	A	506	1/1	0.94	0.07	52,52,52,52	0
6	CL	A	503	1/1	0.95	0.06	40,40,40,40	0
6	CL	A	502	1/1	0.95	0.06	35,35,35,35	1
7	PPV	A	507	5/9	0.95	0.10	28,29,33,37	5

## 6.5 Other polymers [i](#)

There are no such residues in this entry.