



# wwPDB X-ray Structure Validation Summary Report ⓘ

Apr 22, 2025 – 03:03 AM EDT

PDB ID : 5J0U / pdb\_00005j0u  
Title : Binary complex crystal structure of DNA polymerase Beta with G:G mismatch at the primer terminus  
Authors : Batra, V.K.; Wilson, S.H.  
Deposited on : 2016-03-28  
Resolution : 2.10 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	:	4.02b-467
Xtrriage (Phenix)	:	2.0rc1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.006 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.42

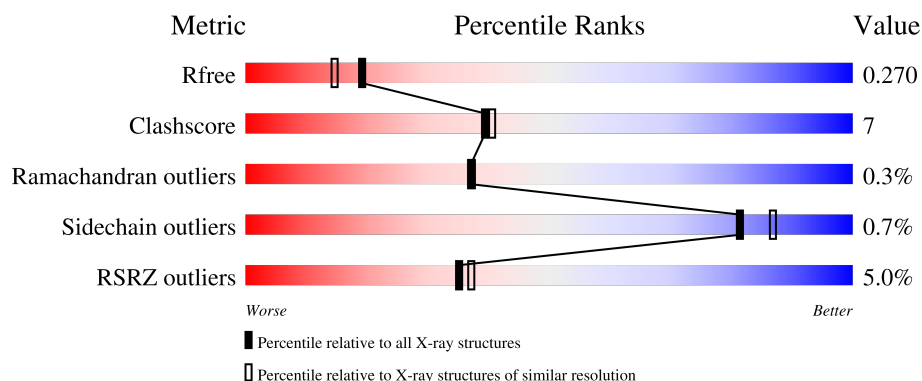
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.10 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))
$R_{free}$	164625	6234 (2.10-2.10)
Clashscore	180529	6893 (2.10-2.10)
Ramachandran outliers	177936	6839 (2.10-2.10)
Sidechain outliers	177891	6840 (2.10-2.10)
RSRZ outliers	164620	6234 (2.10-2.10)

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	335	<div> <div>5%</div> <div> <div></div> <div>85%</div> <div>11%</div> <div>..</div> </div> </div>
2	T	16	<div> <div>88%</div> <div>6%</div> <div>6%</div> </div>
3	P	10	<div> <div>60%</div> <div>30%</div> <div>10%</div> </div>
4	D	5	<div> <div>80%</div> <div>20%</div> </div>

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 3636 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 beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	326	Total	C	N	O	S	0	0	0
			2598	1644	454	491	9			

- Molecule 2 is a DNA chain called Template Strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	T	16	Total	C	N	O	P	0	0	0
			322	153	63	91	15			

- Molecule 3 is a DNA chain called Primer Strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	P	10	Total	C	N	O	P	0	0	0
			206	98	40	59	9			

- Molecule 4 is a DNA chain called Downstream Primer Strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	5	Total	C	N	O	P	0	0	0
			106	49	20	32	5			

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	4	Total	Na	0	0
			4	4		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	287	Total	O	0	0
			287	287		

*Continued on next page...*

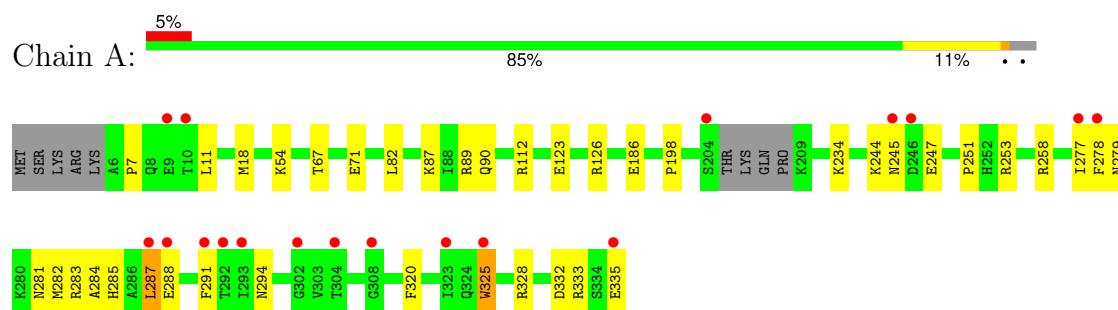
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	T	55	Total 55	O 55	0	0
6	P	32	Total 32	O 32	0	0
6	D	26	Total 26	O 26	0	0

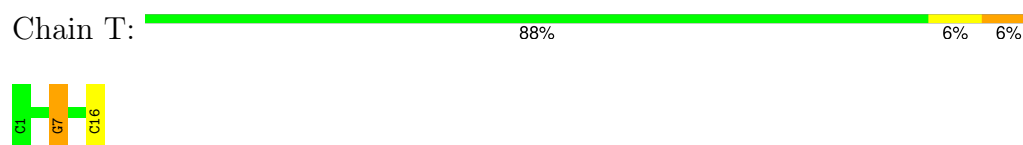
### 3 Residue-property plots

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 polymerase beta



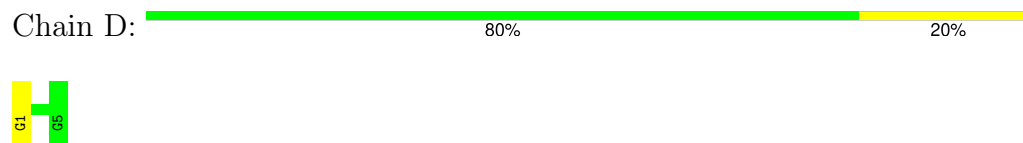
- Molecule 2: Template Strand



- Molecule 3: Primer Strand



- Molecule 4: Downstream Primer Strand



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	54.51Å 79.47Å 54.95Å 90.00° 105.66° 90.00°	Depositor
Resolution (Å)	20.68 – 2.10 20.68 – 2.10	Depositor EDS
% Data completeness (in resolution range)	97.9 (20.68-2.10) 85.5 (20.68-2.10)	Depositor EDS
$R_{merge}$	0.14	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.75 (at 2.09Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, $R_{free}$	0.205 , 0.281 0.196 , 0.270	Depositor DCC
$R_{free}$ test set	2545 reflections (9.82%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	19.9	Xtriage
Anisotropy	0.117	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 44.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	0.059 for l,-k,h	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	3636	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.30% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section:  
NA

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	0.40	0/2646	0.56	1/3555 (0.0%)
2	T	1.00	1/361 (0.3%)	0.98	0/554
3	P	1.20	2/231 (0.9%)	1.21	2/356 (0.6%)
4	D	1.30	1/118 (0.8%)	1.06	0/179
All	All	0.62	4/3356 (0.1%)	0.71	3/4644 (0.1%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	1	DG	OP3-P	-8.87	1.50	1.61
3	P	10	DG	C2-N3	-6.25	1.27	1.32
3	P	10	DG	N9-C4	-5.94	1.33	1.38
2	T	7	DG	C3'-O3'	-5.22	1.37	1.44

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	P	10	DG	N3-C4-N9	-6.43	122.14	126.00
3	P	10	DG	N9-C4-C5	5.57	107.63	105.40
1	A	287	LEU	CA-CB-CG	5.19	127.23	115.30

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	2598	0	2608	31	0
2	T	322	0	179	8	0
3	P	206	0	114	8	0
4	D	106	0	57	0	0
5	A	4	0	0	0	0
6	A	287	0	0	10	2
6	D	26	0	0	0	1
6	P	32	0	0	1	0
6	T	55	0	0	2	0
All	All	3636	0	2958	41	2

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 7.

The worst 5 of 41 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:234:LYS:NZ	6:A:501:HOH:O	1.79	1.12
2:T:7:DG:N2	3:P:10:DG:H22	1.55	1.04
2:T:7:DG:H22	3:P:10:DG:H22	1.01	0.93
2:T:7:DG:H22	3:P:10:DG:N2	1.70	0.88
1:A:283:ARG:HH21	1:A:294:ASN:HA	1.45	0.79

All (2) 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)
6:A:755:HOH:O	6:D:111:HOH:O[1_556]	2.15	0.05
6:A:664:HOH:O	6:A:665:HOH:O[1_455]	2.17	0.03

## 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	322/335 (96%)	310 (96%)	11 (3%)	1 (0%)	37 37

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	245	ASN

### 5.3.2 Protein sidechains ⓘ

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	282/295 (96%)	280 (99%)	2 (1%)	81 87

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	258	ARG
1	A	325	TRP

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	281	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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 4 ligands modelled in this entry, 4 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 [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	326/335 (97%)	0.01	18 (5%) 32 34	13, 24, 58, 68	0
2	T	16/16 (100%)	-0.51	0 100 100	17, 25, 35, 38	0
3	P	10/10 (100%)	-0.43	0 100 100	16, 26, 35, 35	0
4	D	5/5 (100%)	-0.83	0 100 100	17, 18, 27, 32	0
All	All	357/366 (97%)	-0.04	18 (5%) 35 37	13, 24, 57, 68	0

The worst 5 of 18 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	245	ASN	3.5
1	A	246	ASP	3.3
1	A	325	TRP	2.9
1	A	287	LEU	2.8
1	A	293	ILE	2.8

### 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( $\text{\AA}^2$ )	Q<0.9
5	NA	A	402	1/1	0.91	0.06	24,24,24,24	0
5	NA	A	404	1/1	0.94	0.07	25,25,25,25	0
5	NA	A	403	1/1	0.95	0.14	24,24,24,24	0
5	NA	A	401	1/1	0.96	0.03	13,13,13,13	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.