



# wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 11, 2025 – 06:06 PM EST

PDB ID : 9MYE / pdb\_00009mye  
Title : C28 Polymerase Incorporating RNA, n+1  
Authors : Hajjar, M.; Ko, G.K.; Ho, E.J.; Chim, N.; Chaput, J.C.  
Deposited on : 2025-01-21  
Resolution : 2.22 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
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
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)	:	2.47

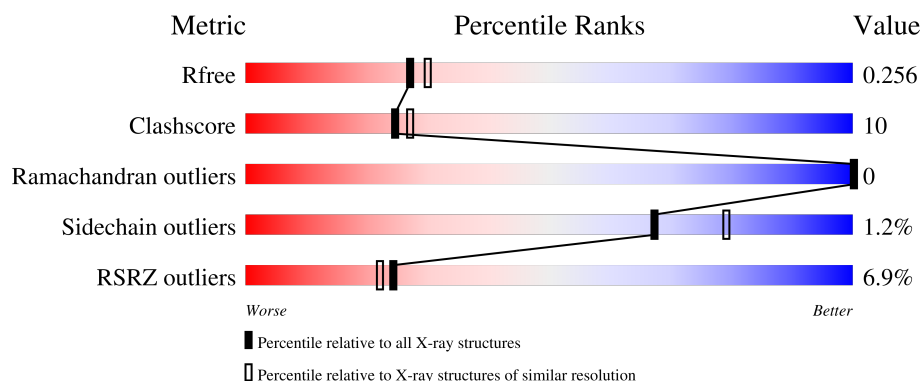
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*




The reported resolution of this entry is 2.22 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	7167 (2.24-2.20)
Clashscore	180529	8096 (2.24-2.20)
Ramachandran outliers	177936	8010 (2.24-2.20)
Sidechain outliers	177891	8011 (2.24-2.20)
RSRZ outliers	164620	7166 (2.24-2.20)

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	774	
2	T	18	
3	P	12	

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 6715 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	747	Total	C	N	O	S	0	0	0
			6117	3940	1038	1125	14			

There are 45 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	5	ALA	THR	conflict	UNP P56689
A	15	ILE	VAL	conflict	UNP P56689
A	28	VAL	ILE	conflict	UNP P56689
A	29	GLU	ASP	conflict	UNP P56689
A	35	ARG	GLU	conflict	UNP P56689
A	88	ASN	THR	conflict	UNP P56689
A	93	GLN	VAL	conflict	UNP P56689
A	99	ARG	LYS	conflict	UNP P56689
A	101	ARG	LYS	conflict	UNP P56689
A	102	ALA	GLU	conflict	UNP P56689
A	136	THR	LYS	conflict	UNP P56689
A	141	ALA	ASP	conflict	UNP P56689
A	143	ALA	GLU	conflict	UNP P56689
A	153	GLY	ALA	conflict	UNP P56689
A	154	THR	GLU	conflict	UNP P56689
A	165	GLY	GLU	conflict	UNP P56689
A	166	SER	GLU	conflict	UNP P56689
A	167	GLU	GLY	conflict	UNP P56689
A	175	LYS	ASN	conflict	UNP P56689
A	196	ARG	LYS	conflict	UNP P56689
A	199	ARG	LYS	conflict	UNP P56689
A	223	CYS	SER	conflict	UNP P56689
A	225	GLU	LYS	conflict	UNP P56689
A	228	ILE	VAL	conflict	UNP P56689
A	231	THR	ILE	conflict	UNP P56689
A	235	ASP	GLU	conflict	UNP P56689
A	327	LEU	PHE	conflict	UNP P56689

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	337	ILE	VAL	conflict	UNP P56689
A	377	TYR	LEU	conflict	UNP P56689
A	378	GLU	ALA	conflict	UNP P56689
A	381	LEU	-	insertion	UNP P56689
A	395	LYS	ARG	conflict	UNP P56689
A	400	GLY	ASN	conflict	UNP P56689
A	401	LEU	ILE	conflict	UNP P56689
A	403	SER	TYR	conflict	UNP P56689
A	410	GLY	TYR	conflict	UNP P56689
A	430	ARG	GLU	conflict	UNP P56689
A	437	GLU	GLN	conflict	UNP P56689
A	486	LEU	ALA	conflict	UNP P56689
A	665	LYS	GLU	conflict	UNP P56689
A	692	VAL	ILE	conflict	UNP P56689
A	724	THR	ALA	conflict	UNP P56689
A	749	SER	PHE	conflict	UNP P56689
A	750	HIS	GLY	conflict	UNP P56689
A	767	SER	GLY	conflict	UNP P56689

- Molecule 2 is a DNA chain called Template.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	T	13	Total	C	N	O	P	0	0	0
			266	126	48	79	13			

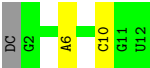
- Molecule 3 is DNA/RNA hybrid called Primer.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	P	11	Total	C	N	O	P	0	0	0
			227	106	43	67	11			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	101	Total	O	0	0
			101	101		
4	T	3	Total	O	0	0
			3	3		
4	P	1	Total	O	0	0
			1	1		





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	77.97Å 102.34Å 112.23Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	64.03 – 2.22 64.03 – 2.22	Depositor EDS
% Data completeness (in resolution range)	99.2 (64.03-2.22) 85.7 (64.03-2.22)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.33 (at 2.22Å)	Xtriage
Refinement program	PHENIX 1.20.1-4487-000	Depositor
R, $R_{free}$	0.241 , 0.272 0.240 , 0.256	Depositor DCC
$R_{free}$ test set	1986 reflections (4.42%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	37.8	Xtriage
Anisotropy	0.072	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 48.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	6715	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	55.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.61% 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

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	0.16	0/6253	0.35	1/8434 (0.0%)
2	T	0.22	0/297	0.41	0/456
3	P	0.22	0/254	0.45	0/390
All	All	0.17	0/6804	0.36	1/9280 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	543	ASP	N-CA-C	6.43	120.81	113.15

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	6117	0	6172	125	0
2	T	266	0	147	2	0
3	P	227	0	122	2	0
4	A	101	0	0	0	0
4	P	1	0	0	0	0
4	T	3	0	0	0	0
All	All	6715	0	6441	126	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 10.

The worst 5 of 126 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:468:MET:HE1	1:A:484:GLN:HG2	1.51	0.92
1:A:693:LYS:HD2	1:A:693:LYS:H	1.48	0.79
1:A:186:THR:HG23	1:A:189:GLU:H	1.50	0.77
1:A:101:ARG:HD3	1:A:109:ILE:HG12	1.67	0.75
1:A:137:MET:HE3	1:A:324:LYS:HE2	1.70	0.74

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	745/774 (96%)	716 (96%)	29 (4%)	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	646/670 (96%)	638 (99%)	8 (1%)	67	79

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	661	LEU
1	A	605	THR
1	A	495	TYR
1	A	394	GLU
1	A	517	TRP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	91	GLN
1	A	93	GLN
1	A	213	ASN
1	A	339	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](#)

There are no ligands in this entry.

## 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	747/774 (96%)	0.54	53 (7%) 23 21	25, 52, 99, 123	0
2	T	13/18 (72%)	-0.02	0 100 100	36, 50, 99, 112	0
3	P	11/12 (91%)	-0.30	0 100 100	44, 48, 71, 78	0
All	All	771/804 (95%)	0.52	53 (6%) 24 22	25, 52, 99, 123	0

The worst 5 of 53 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	472	ILE	4.4
1	A	299	TRP	4.0
1	A	495	TYR	3.7
1	A	290	VAL	3.5
1	A	588	PHE	3.4

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.