



wwPDB EM Validation Summary Report i

Jul 15, 2025 – 10:58 AM JST

PDB ID : 8Z85 / pdb_00008z85
EMDB ID : EMD-39838
Title : Cryo-EM structure of Thogoto virus polymerase in transcription pre-initiation conformation 1
Authors : Xue, L.; Chang, T.; Li, Z.; Zhao, H.; Li, M.; He, J.; Chen, X.; Xiong, X.
Deposited on : 2024-04-21
Resolution : 2.30 Å(reported)

This is a wwPDB EM Validation Summary Report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/EMValidationReportHelp>

with specific help available everywhere you see the i 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](#) i) were used in the production of this report:

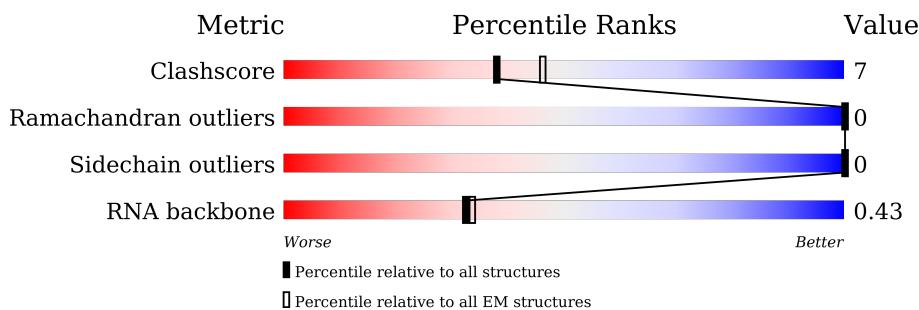
EMDB validation analysis	: FAILED
MolProbit	: 4-5-2 with Phenix2.0rc1
Percentile statistics	: 20231227.v01 (using entries in the PDB archive December 27th 2023)
MapQ	: FAILED
Ideal geometry (proteins)	: Engh & Huber (2001)
Ideal geometry (DNA, RNA)	: Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	: 2.44

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

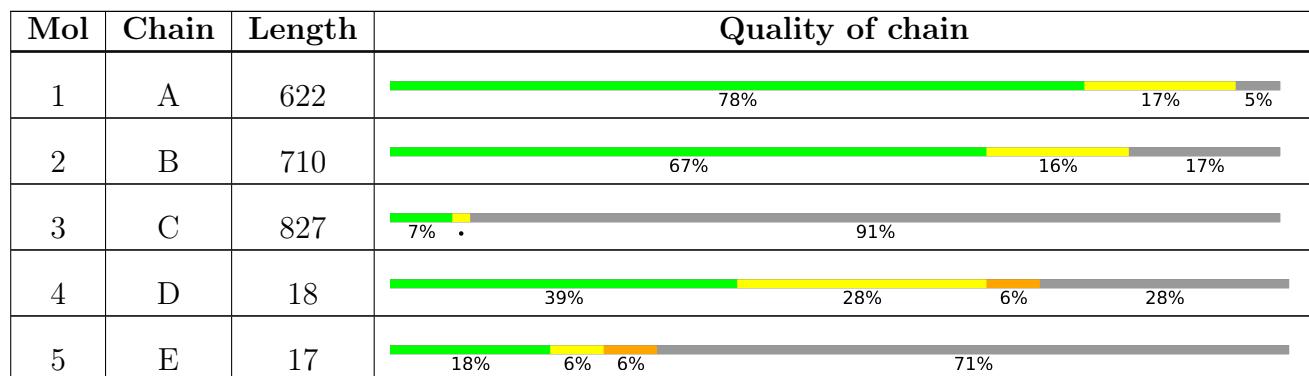
The reported resolution of this entry is 2.30 Å.

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)	EM structures (#Entries)
Clashscore	210492	15764
Ramachandran outliers	207382	16835
Sidechain outliers	206894	16415
RNA backbone	6643	2191

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the bar indicate the fraction of residues that contain outliers for ≥ 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 $\leq 5\%$



2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 10467 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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 Polymerase acidic protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	591	Total	C	N	O	S	0	0

4771 3019 825 900 27

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	471	GLU	GLY	conflict	UNP P27194

- Molecule 2 is a protein called RNA-directed RNA polymerase catalytic subunit.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	590	Total	C	N	O	S	0	0

4702 2996 804 868 34

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	7	LEU	ARG	conflict	UNP O41353
B	230	TRP	CYS	conflict	UNP O41353

- Molecule 3 is a protein called Polymerase basic protein 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	75	Total	C	N	O	S	0	0

608 396 101 109 2

There are 58 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	770	LEU	-	expression tag	UNP Q9YNA4
C	771	GLU	-	expression tag	UNP Q9YNA4
C	772	VAL	-	expression tag	UNP Q9YNA4

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	773	LEU	-	expression tag	UNP Q9YNA4
C	774	PHE	-	expression tag	UNP Q9YNA4
C	775	GLN	-	expression tag	UNP Q9YNA4
C	776	GLY	-	expression tag	UNP Q9YNA4
C	777	PRO	-	expression tag	UNP Q9YNA4
C	778	GLY	-	expression tag	UNP Q9YNA4
C	779	HIS	-	expression tag	UNP Q9YNA4
C	780	HIS	-	expression tag	UNP Q9YNA4
C	781	HIS	-	expression tag	UNP Q9YNA4
C	782	HIS	-	expression tag	UNP Q9YNA4
C	783	HIS	-	expression tag	UNP Q9YNA4
C	784	HIS	-	expression tag	UNP Q9YNA4
C	785	HIS	-	expression tag	UNP Q9YNA4
C	786	HIS	-	expression tag	UNP Q9YNA4
C	787	SER	-	expression tag	UNP Q9YNA4
C	788	ALA	-	expression tag	UNP Q9YNA4
C	789	ASP	-	expression tag	UNP Q9YNA4
C	790	TYR	-	expression tag	UNP Q9YNA4
C	791	LYS	-	expression tag	UNP Q9YNA4
C	792	ASP	-	expression tag	UNP Q9YNA4
C	793	ASP	-	expression tag	UNP Q9YNA4
C	794	ASP	-	expression tag	UNP Q9YNA4
C	795	ASP	-	expression tag	UNP Q9YNA4
C	796	LYS	-	expression tag	UNP Q9YNA4
C	797	GLY	-	expression tag	UNP Q9YNA4
C	798	GLY	-	expression tag	UNP Q9YNA4
C	799	TRP	-	expression tag	UNP Q9YNA4
C	800	SER	-	expression tag	UNP Q9YNA4
C	801	HIS	-	expression tag	UNP Q9YNA4
C	802	PRO	-	expression tag	UNP Q9YNA4
C	803	GLN	-	expression tag	UNP Q9YNA4
C	804	PHE	-	expression tag	UNP Q9YNA4
C	805	GLU	-	expression tag	UNP Q9YNA4
C	806	LYS	-	expression tag	UNP Q9YNA4
C	807	GLY	-	expression tag	UNP Q9YNA4
C	808	GLY	-	expression tag	UNP Q9YNA4
C	809	GLY	-	expression tag	UNP Q9YNA4
C	810	SER	-	expression tag	UNP Q9YNA4
C	811	GLY	-	expression tag	UNP Q9YNA4
C	812	GLY	-	expression tag	UNP Q9YNA4
C	813	GLY	-	expression tag	UNP Q9YNA4
C	814	GLY	-	expression tag	UNP Q9YNA4

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	815	SER	-	expression tag	UNP Q9YNA4
C	816	GLY	-	expression tag	UNP Q9YNA4
C	817	GLY	-	expression tag	UNP Q9YNA4
C	818	SER	-	expression tag	UNP Q9YNA4
C	819	ALA	-	expression tag	UNP Q9YNA4
C	820	TRP	-	expression tag	UNP Q9YNA4
C	821	SER	-	expression tag	UNP Q9YNA4
C	822	HIS	-	expression tag	UNP Q9YNA4
C	823	PRO	-	expression tag	UNP Q9YNA4
C	824	GLN	-	expression tag	UNP Q9YNA4
C	825	PHE	-	expression tag	UNP Q9YNA4
C	826	GLU	-	expression tag	UNP Q9YNA4
C	827	LYS	-	expression tag	UNP Q9YNA4

- Molecule 4 is a RNA chain called RNA (5'-R(*AP*GP*AP*GP*AP*AP*AP*UP*CP*AP*AP*GP*GP*CP*AP*GP*UP*U)-3').

Mol	Chain	Residues	Atoms					AltConf	Trace
4	D	13	Total	C	N	O	P	0	0
			283	128	60	83	12		

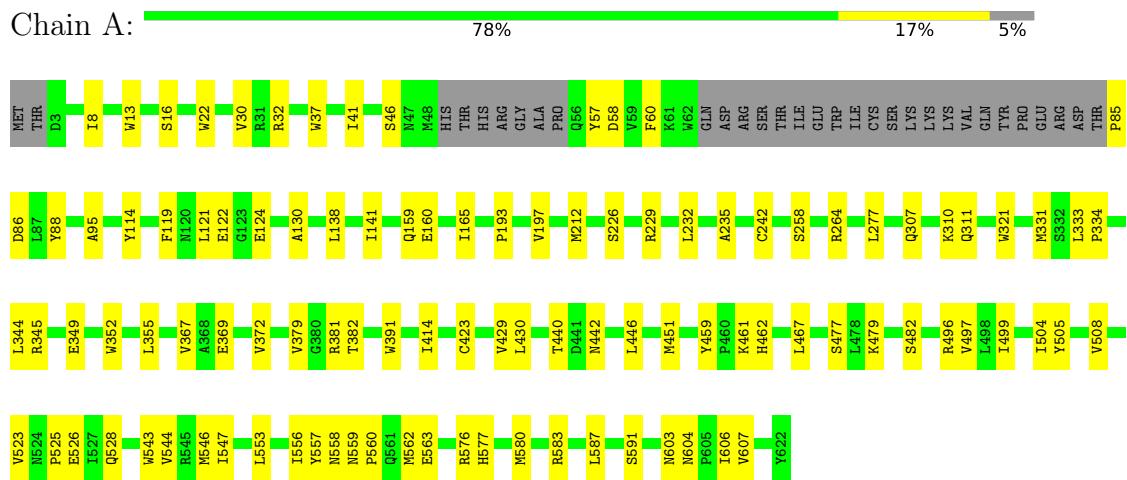
- Molecule 5 is a RNA chain called RNA (5'-R(*GP*AP*CP*UP*GP*CP*CP*UP*GP*UP*UP*UP*UP*GP*CP*U)-3').

Mol	Chain	Residues	Atoms					AltConf	Trace
5	E	5	Total	C	N	O	P	0	0
			103	46	14	38	5		

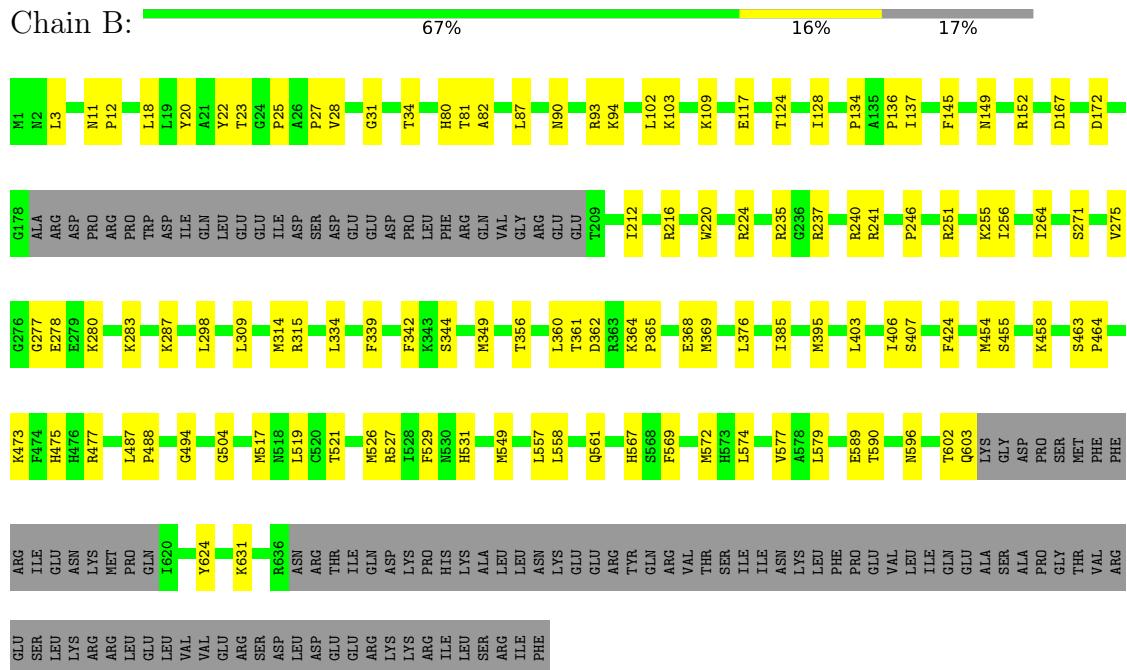
3 Residue-property plots [i](#)

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. 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. 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: Polymerase acidic protein



- Molecule 2: RNA-directed RNA polymerase catalytic subunit



- Molecule 3: Polymerase basic protein 2

Chain C: 7% .



- Molecule 4: RNA (5'-R(*AP*GP*AP*GP*AP*AP*AP*UP*CP*AP*AP*GP*GP*CP*AP*GP*UP*U)-3')

Chain D:



A 4x4 grid of colored cells. The cells are arranged in a 4x4 pattern. The colors of the cells are: Row 1: A1 (green), G2 (green), A3 (yellow), G4 (yellow). Row 2: A7 (green), A11 (orange), G12 (orange), G13 (orange). Row 3: C (grey), A (grey), G (grey), U (grey). Row 4: U (grey), U (grey), U (grey), U (grey).

- Molecule 5: RNA (5'-R(*GP*AP*CP*UP*GP*CP*CP*UP*GP*UP*UP*UP*UP*UP*GP*CP*U)-3')

Chain E:





4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	350712	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	TFS KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	50	Depositor
Minimum defocus (nm)	600	Depositor
Maximum defocus (nm)	2400	Depositor
Magnification	Not provided	
Image detector	TFS FALCON 4i (4k x 4k)	Depositor

5 Model quality [i](#)

5.1 Standard geometry [i](#)

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	0.09	0/4877	0.25	0/6612
2	B	0.10	0/4802	0.28	0/6493
3	C	0.08	0/623	0.26	0/846
4	D	0.07	0/319	0.16	0/497
5	E	0.08	0/113	0.13	0/173
All	All	0.09	0/10734	0.26	0/14621

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	4771	0	4716	78	0
2	B	4702	0	4735	89	0
3	C	608	0	621	10	0
4	D	283	0	144	4	0
5	E	103	0	53	3	0
All	All	10467	0	10269	149	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 7.

The worst 5 of 149 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)
2:B:487:LEU:HD12	2:B:488:PRO:HD2	1.65	0.78
3:C:66:PRO:HD2	3:C:100:ASN:HA	1.68	0.75
2:B:424:PHE:HZ	2:B:454:MET:HE3	1.51	0.75
1:A:451:MET:SD	1:A:477:SER:OG	2.45	0.75
1:A:546:MET:HE3	2:B:3:LEU:HD11	1.68	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 PDB entries followed by that with respect to all EM entries.

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	585/622 (94%)	571 (98%)	14 (2%)	0	100 100
2	B	584/710 (82%)	569 (97%)	15 (3%)	0	100 100
3	C	71/827 (9%)	70 (99%)	1 (1%)	0	100 100
All	All	1240/2159 (57%)	1210 (98%)	30 (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 PDB entries followed by that with respect to all EM entries.

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	532/561 (95%)	532 (100%)	0	100 100
2	B	520/633 (82%)	520 (100%)	0	100 100
3	C	69/736 (9%)	69 (100%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1121/1930 (58%)	1121 (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. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	222	GLN
2	B	227	ASN
3	C	131	GLN
2	B	476	HIS
2	B	530	ASN

5.3.3 RNA [\(i\)](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
4	D	12/18 (66%)	2 (16%)	0
5	E	4/17 (23%)	1 (25%)	0
All	All	16/35 (45%)	3 (18%)	0

All (3) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
4	D	11	A
4	D	13	G
5	E	17	U

There are no RNA pucker outliers to report.

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.