



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 8, 2026 – 05:44 AM UTC

PDB ID : 2GHO / pdb_00002gho
Title : Recombinant *Thermus aquaticus* RNA polymerase for Structural Studies
Authors : Lamour, V.; Darst, S.A.
Deposited on : 2006-03-27
Resolution : 5.00 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

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	:	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

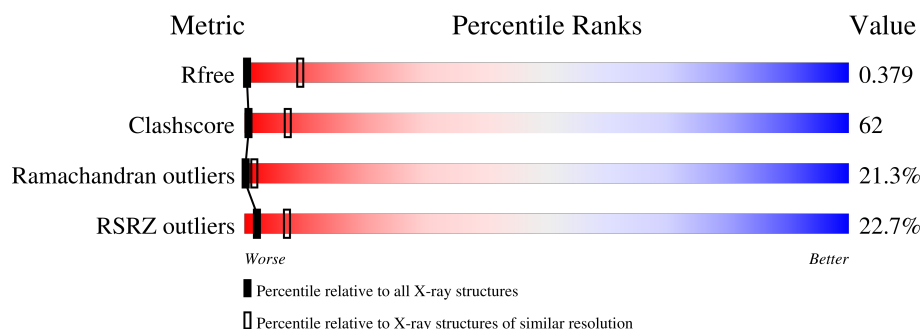
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 5.00 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	1009 (6.02-3.98)
Clashscore	190562	1040 (6.00-4.00)
Ramachandran outliers	187476	1015 (6.06-3.92)
RSRZ outliers	180081	1004 (6.02-3.98)

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 ≥ 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	314	15% 42% 24% 7% 27%
1	B	314	16% 45% 22% 5% 28%
2	C	1119	22% 54% 35% 10%
3	D	1233	23% 39% 38% 19% ••

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 11060 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 RNA polymerase subunit alpha.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	230	Total	C	N	O	0	0	0
			920	460	230	230			
1	B	225	Total	C	N	O	0	0	0
			900	450	225	225			

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	1114	Total	C	N	O	0	0	0
			4456	2228	1114	1114			

- Molecule 3 is a protein called DNA-directed RNA polymerase subunit beta',DNA-directed RNA polymerase subunit beta'.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	D	1196	Total	C	N	O	0	0	0
			4784	2392	1196	1196			

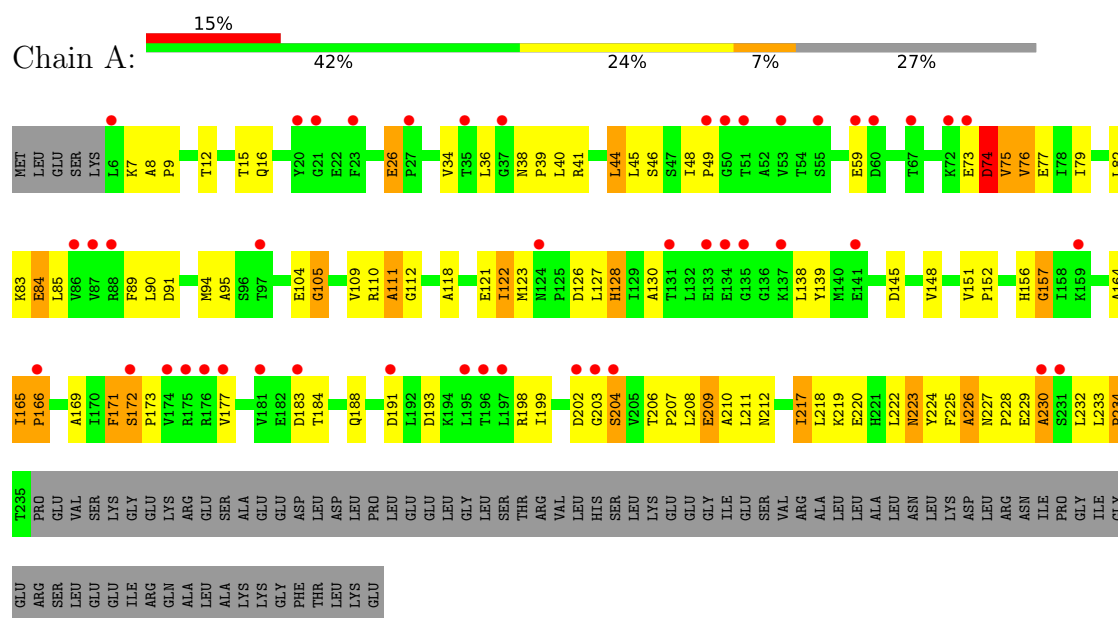
There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	159	GLY	-	linker	UNP Q9KWU6
D	160	GLY	-	linker	UNP Q9KWU6

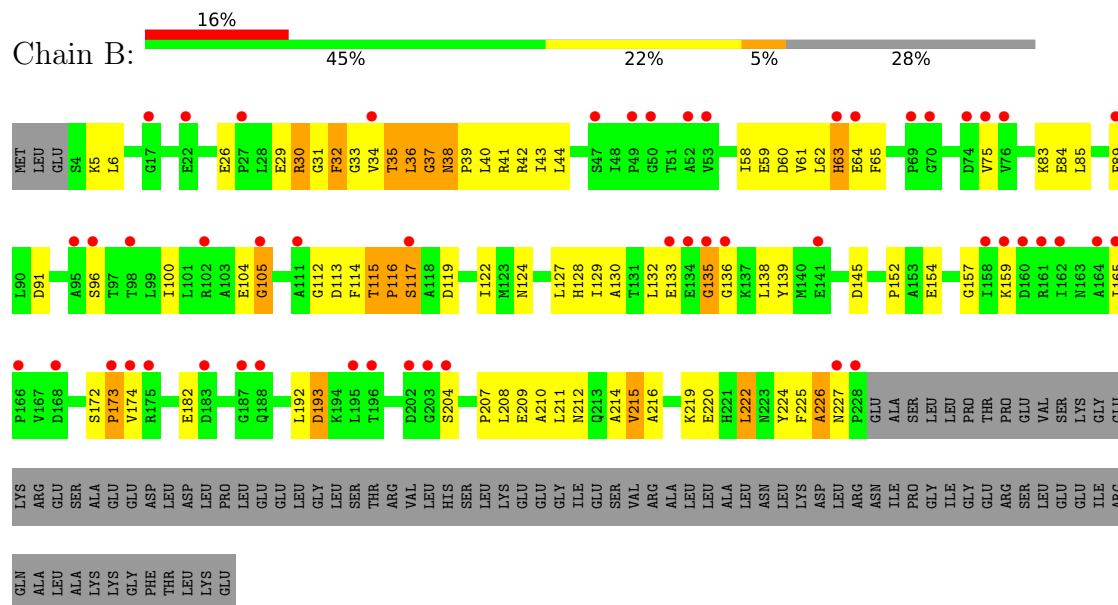
3 Residue-property plots [i](#)

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.

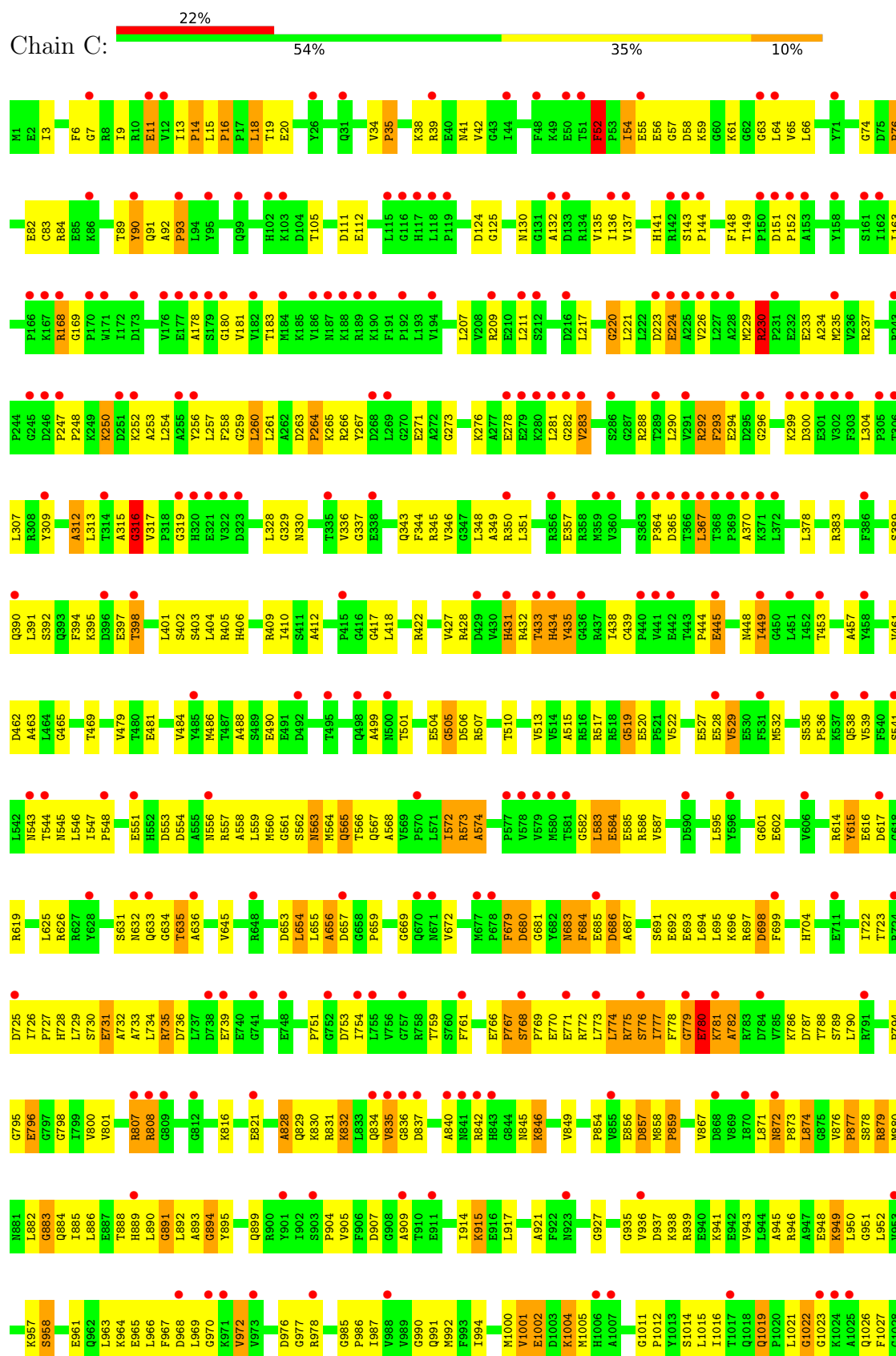
• Molecule 1: DNA-directed RNA polymerase subunit alpha

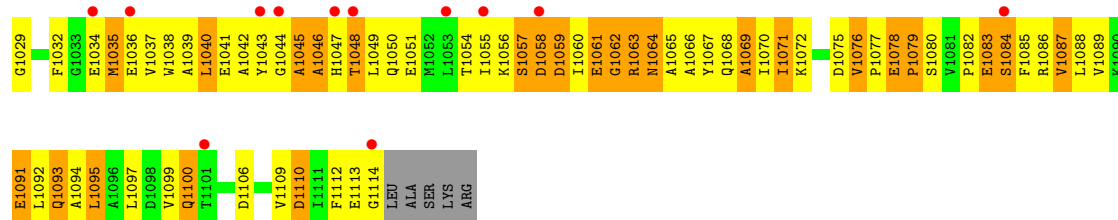


• Molecule 1: DNA-directed RNA polymerase subunit alpha

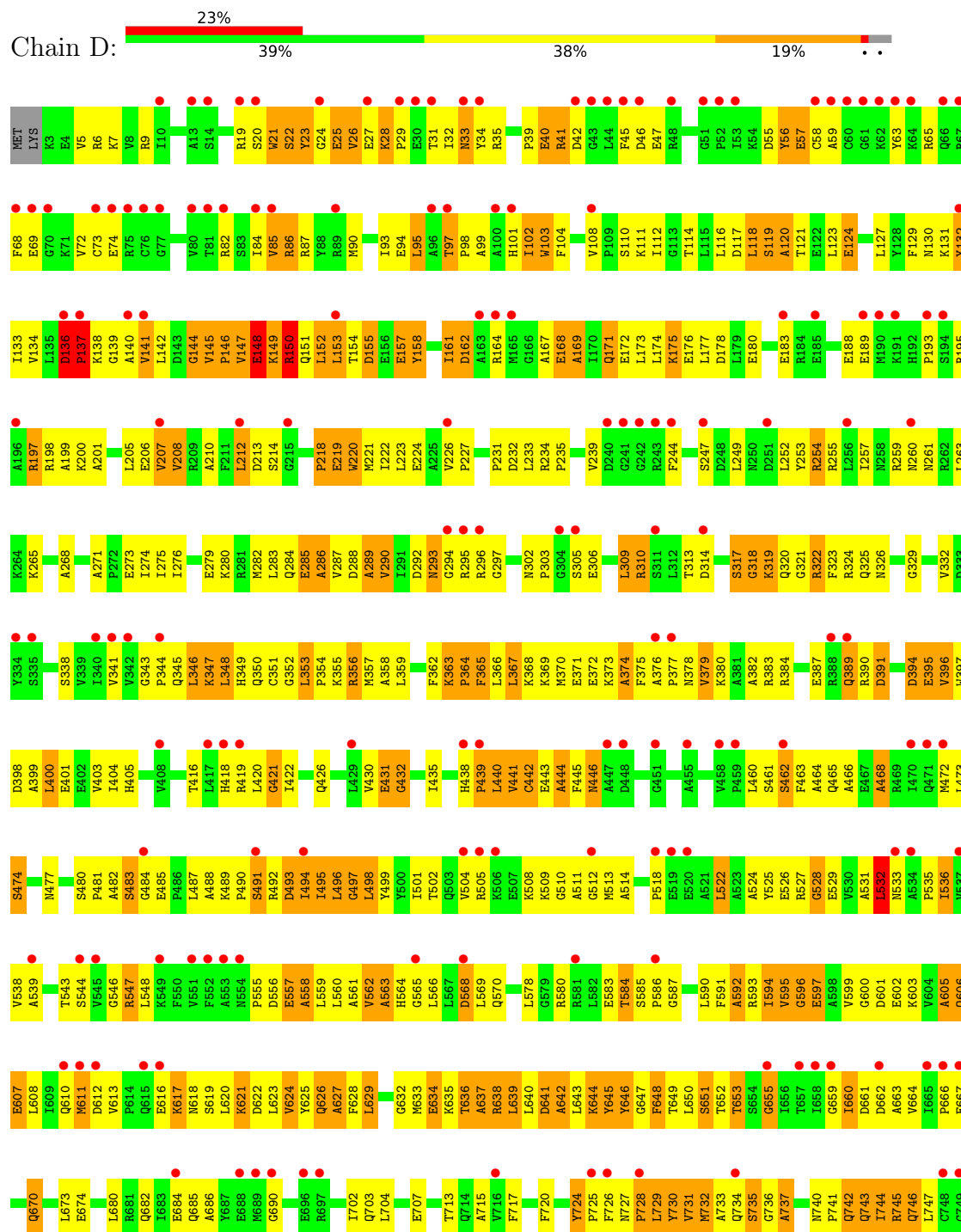


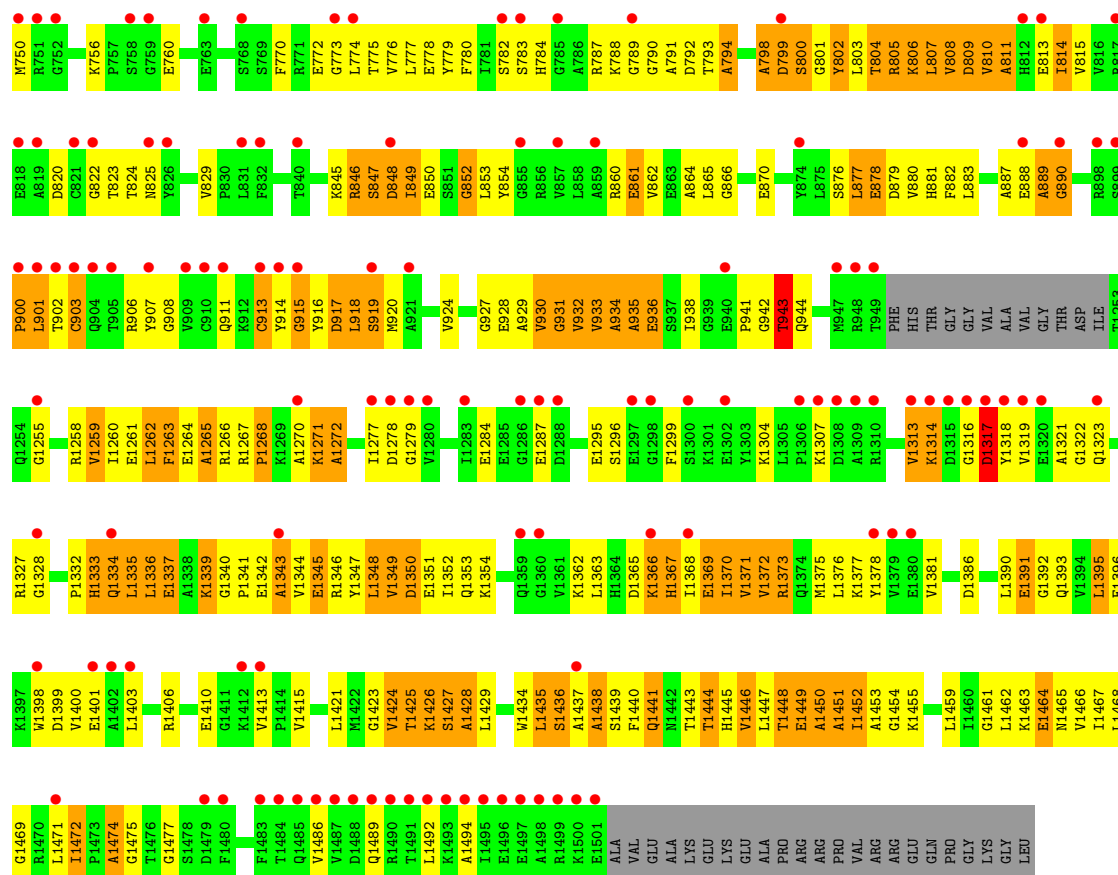
● Molecule 2: DNA-directed RNA polymerase subunit beta





● Molecule 3: DNA-directed RNA polymerase subunit beta', DNA-directed RNA polymerase subunit beta'





4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α , β , γ	202.80Å 202.80Å 326.80Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	25.00 – 5.00 25.00 – 5.00	Depositor EDS
% Data completeness (in resolution range)	87.1 (25.00-5.00) 86.3 (25.00-5.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.08	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.05 (at 4.10Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.336 , 0.337 0.385 , 0.379	Depositor DCC
R_{free} test set	3069 reflections (7.06%)	wwPDB-VP
Wilson B-factor (Å ²)	138.4	Xtriage
Anisotropy	0.324	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.17 , 999.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.81	EDS
Total number of atoms	11060	wwPDB-VP
Average B, all atoms (Å ²)	193.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.59% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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](#)

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	0.62	0/919	1.43	19/1147 (1.7%)
1	B	0.67	0/899	1.46	17/1122 (1.5%)
2	C	0.95	11/4455 (0.2%)	1.69	131/5567 (2.4%)
3	D	0.94	21/4782 (0.4%)	1.68	140/5974 (2.3%)
All	All	0.90	32/11055 (0.3%)	1.65	307/13810 (2.2%)

The worst 5 of 32 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	704	HIS	C-N	20.58	1.60	1.33
3	D	943	THR	C-N	-16.07	1.11	1.33
2	C	828	ALA	C-N	12.53	1.51	1.33
3	D	1435	LEU	C-N	-10.68	1.18	1.33
2	C	1057	SER	C-N	10.48	1.48	1.33

The worst 5 of 307 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	207	VAL	N-CA-C	-13.50	99.95	111.56
3	D	140	ALA	CA-C-N	12.71	144.84	121.97
3	D	140	ALA	C-N-CA	12.71	144.84	121.97
3	D	141	VAL	N-CA-C	12.54	135.41	109.34
3	D	918	LEU	N-CA-C	-12.32	99.04	113.21

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	920	0	246	47	0
1	B	900	0	242	41	0
2	C	4456	0	1247	271	0
3	D	4784	0	1312	517	0
All	All	11060	0	3047	873	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 62.

The worst 5 of 873 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)
3:D:145:VAL:O	3:D:146:PRO:O	1.56	1.19
2:C:775:ARG:O	2:C:778:PHE:N	1.88	1.05
3:D:1438:ALA:O	3:D:1440:PHE:N	1.98	0.94
2:C:775:ARG:O	2:C:779:GLY:N	2.00	0.94
3:D:917:ASP:C	3:D:919:SER:H	1.68	0.90

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	228/314 (73%)	138 (60%)	48 (21%)	42 (18%)	0	2
1	B	223/314 (71%)	133 (60%)	58 (26%)	32 (14%)	0	3
2	C	1112/1119 (99%)	640 (58%)	275 (25%)	197 (18%)	0	2
3	D	1192/1233 (97%)	536 (45%)	339 (28%)	317 (27%)	0	0
All	All	2755/2980 (92%)	1447 (52%)	720 (26%)	588 (21%)	0	1

5 of 588 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	7	LYS
1	A	45	LEU
1	A	59	GLU
1	A	74	ASP
1	A	75	VAL

5.3.2 Protein sidechains [i](#)

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

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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
3	D	2

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	D	1435:LEU	C	1436:SER	N	1.18
1	D	943:THR	C	944:GLN	N	1.11

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, 95th 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 ²)	Q<0.9
1	A	230/314 (73%)	1.16	46 (20%) 3 8	194, 194, 194, 194	0
1	B	225/314 (71%)	1.13	51 (22%) 2 7	194, 194, 194, 194	0
2	C	1114/1119 (99%)	1.28	247 (22%) 2 7	194, 194, 194, 194	0
3	D	1196/1233 (96%)	1.35	285 (23%) 2 6	194, 194, 194, 194	0
All	All	2765/2980 (92%)	1.28	629 (22%) 2 7	194, 194, 194, 194	0

The worst 5 of 629 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	D	1315	ASP	27.7
3	D	241	GLY	17.9
3	D	242	GLY	16.7
3	D	505	ARG	11.7
3	D	74	GLU	11.3

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.