



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 08:35 am BST

PDB ID : 5UAJ  
Title : Escherichia coli RNA polymerase RpoB S531L mutant  
Authors : Molodtsov, V.; Scharf, N.T.; Stefan, M.A.; Garcia, G.A.; Murakami, K.S.  
Deposited on : 2016-12-19  
Resolution : 3.92 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

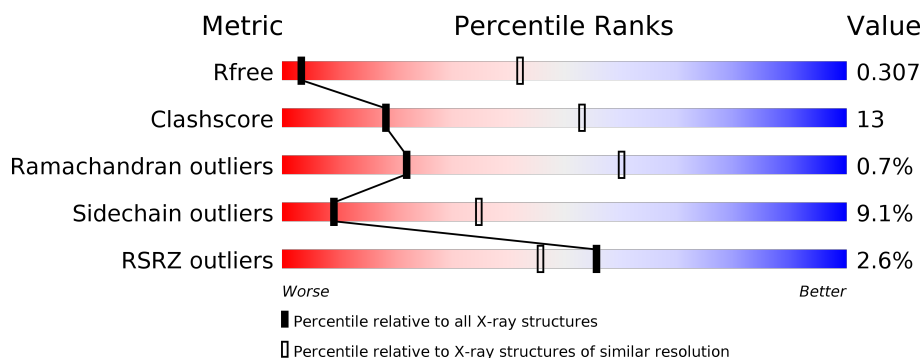
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.92 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))
$R_{free}$	130704	1019 (4.18-3.66)
Clashscore	141614	1016 (4.16-3.68)
Ramachandran outliers	138981	1039 (4.18-3.66)
Sidechain outliers	138945	1032 (4.18-3.66)
RSRZ outliers	127900	1002 (4.20-3.64)

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 on 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	329	<div> <div>41%</div> <div>25%</div> <div>•</div> <div>31%</div> </div>
1	B	329	<div>3%</div> <div>36%</div> <div>27%</div> <div>•</div> <div>35%</div>
1	G	329	<div>2%</div> <div>43%</div> <div>21%</div> <div>•</div> <div>32%</div>
1	H	329	<div>2%</div> <div>36%</div> <div>27%</div> <div>•</div> <div>35%</div>
2	C	1342	<div>3%</div> <div>63%</div> <div>33%</div> <div>•</div>
2	I	1342	<div>3%</div> <div>67%</div> <div>30%</div> <div>•</div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	D	1407	<div><div><div></div><div></div><div></div><div></div></div><div>%</div><div><div></div><div></div><div></div><div></div></div><div>54%25%•17%</div></div>
3	J	1407	<div><div><div></div><div></div><div></div><div></div></div><div>2%</div><div><div></div><div></div><div></div><div></div></div><div>53%25%•18%</div></div>
4	E	91	<div><div><div></div><div></div><div></div><div></div></div><div>75%20%••</div></div>
4	K	91	<div><div><div></div><div></div><div></div><div></div></div><div>13%</div><div><div></div><div></div><div></div><div></div></div><div>56%29%•13%</div></div>
5	F	613	<div><div><div></div><div></div><div></div><div></div></div><div>2%</div><div><div></div><div></div><div></div><div></div></div><div>50%24%•24%</div></div>
5	L	613	<div><div><div></div><div></div><div></div><div></div></div><div>2%</div><div><div></div><div></div><div></div><div></div></div><div>51%23%•23%</div></div>

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 54994 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	227	Total	C	N	O	S	0	0	0
			1753	1091	311	345	6			
1	B	214	Total	C	N	O	S	0	0	0
			1649	1029	290	324	6			
1	G	224	Total	C	N	O	S	0	0	0
			1730	1076	308	340	6			
1	H	215	Total	C	N	O	S	0	0	0
			1659	1037	291	325	6			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1342	Total	C	N	O	S	0	0	0
			10587	6644	1843	2056	44			
2	I	1340	Total	C	N	O	S	0	0	0
			10568	6632	1840	2053	43			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	531	LEU	SER	engineered mutation	UNP P0A8V2
I	531	LEU	SER	engineered mutation	UNP P0A8V2

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1166	Total	C	N	O	S	0	0	0
			9089	5714	1627	1702	46			
3	J	1155	Total	C	N	O	S	0	0	0
			9001	5659	1612	1684	46			

- Molecule 4 is a protein called DNA-directed RNA polymerase subunit omega.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	89	Total	C	N	O	S	0	0	0
			691	421	129	140	1			
4	K	79	Total	C	N	O	S	0	0	0
			627	382	118	126	1			

- Molecule 5 is a protein called RNA polymerase sigma factor RpoD.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	468	Total	C	N	O	S	0	0	0
			3813	2389	678	723	23			
5	L	469	Total	C	N	O	S	0	0	0
			3821	2393	679	726	23			

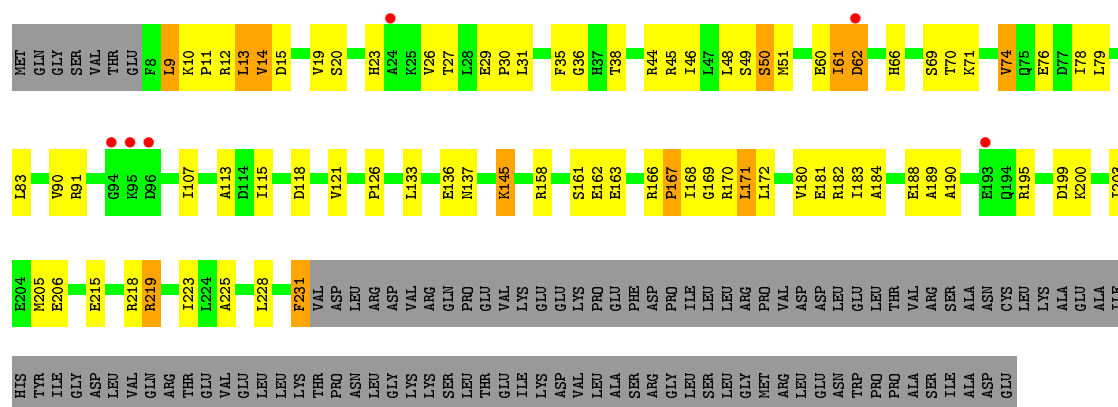
- Molecule 6 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	J	1	Total	Mg	0	0
			1	1		
6	D	1	Total	Mg	0	0
			1	1		

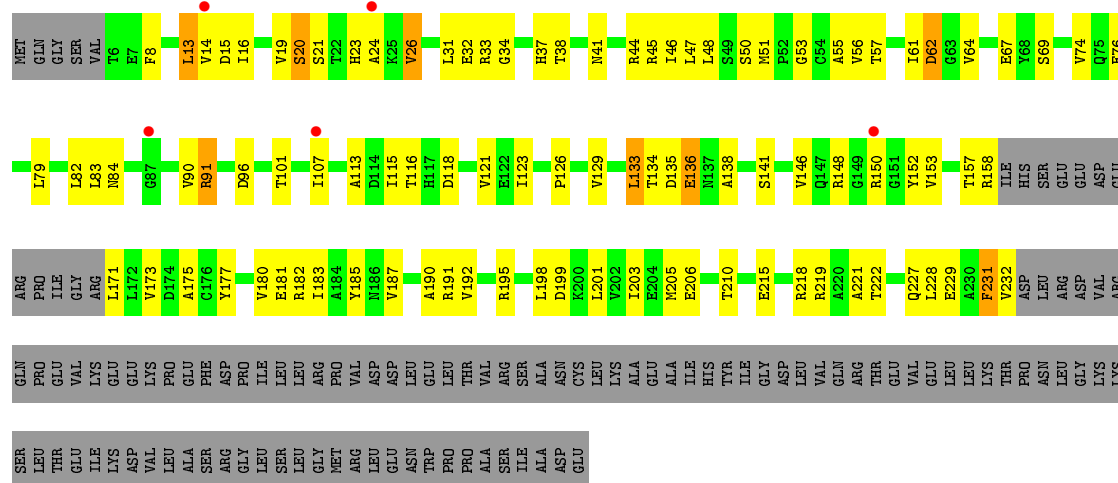
- Molecule 7 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	J	2	Total	Zn	0	0
			2	2		
7	D	2	Total	Zn	0	0
			2	2		

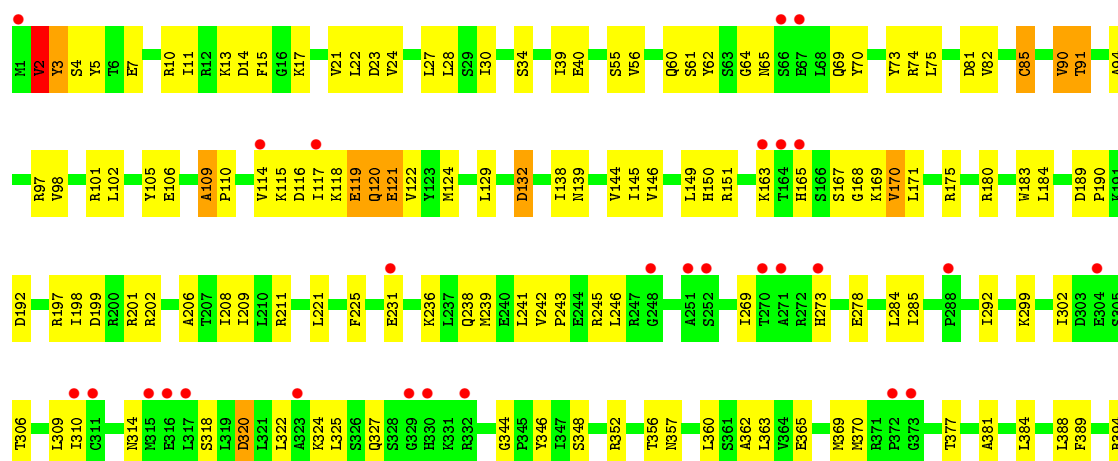


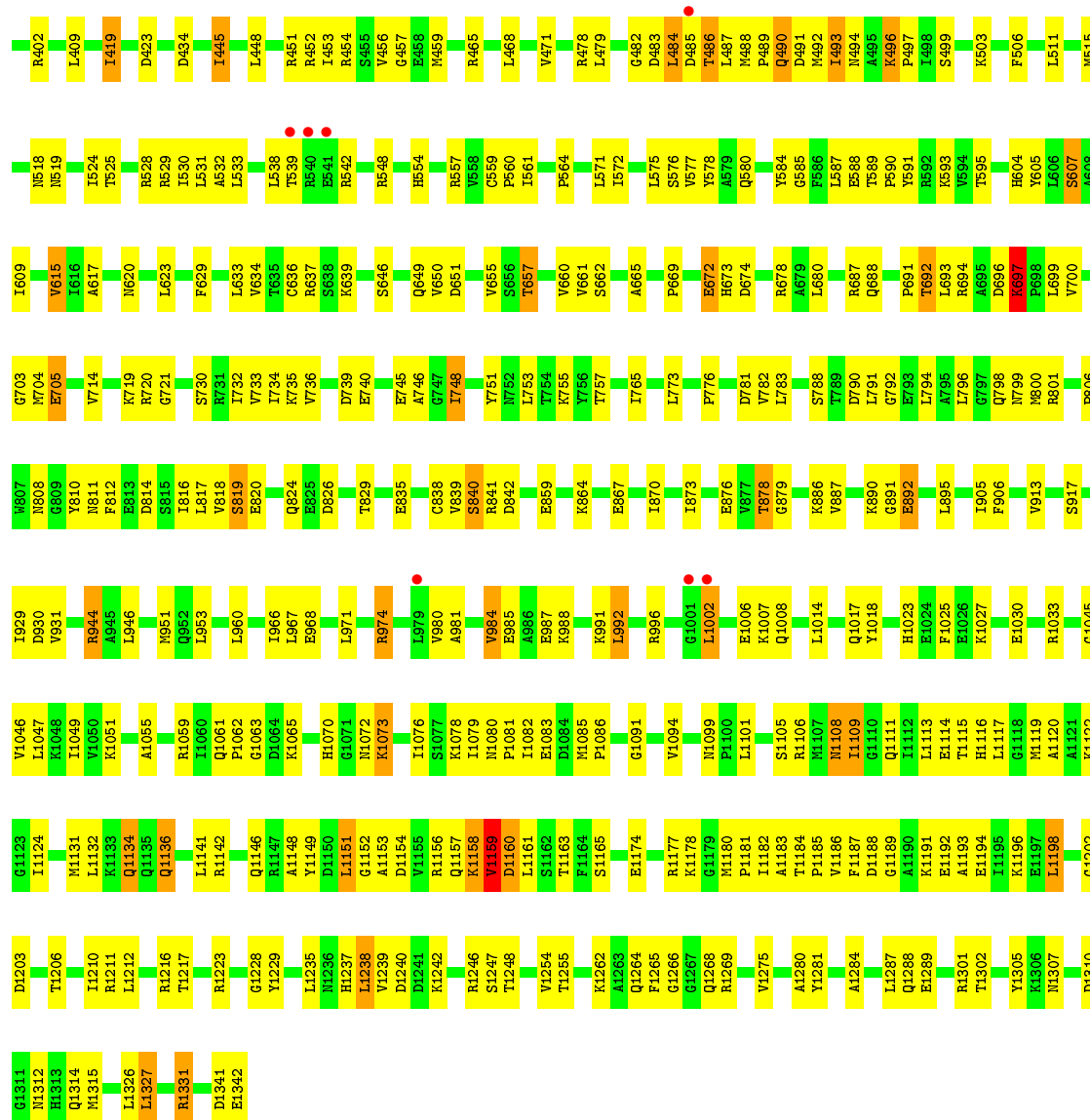


• Molecule 1: DNA-directed RNA polymerase subunit alpha

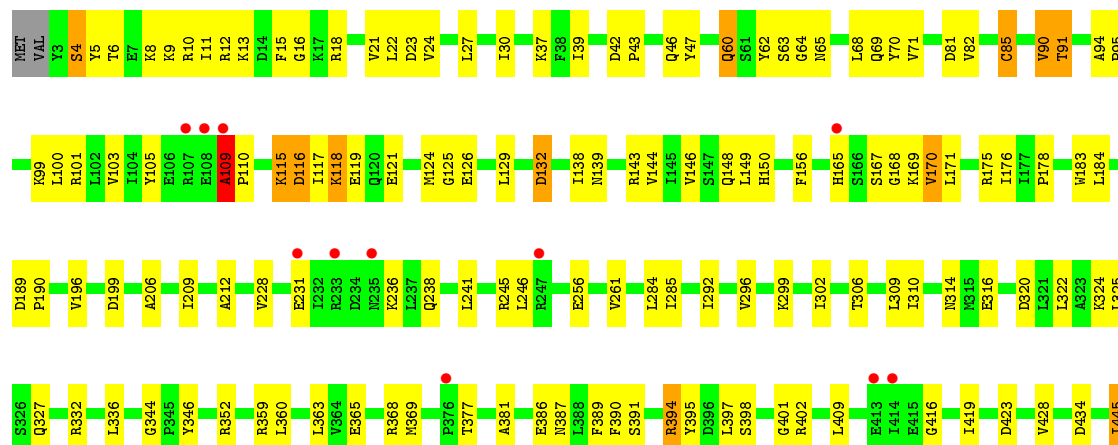


• Molecule 2: DNA-directed RNA polymerase subunit beta

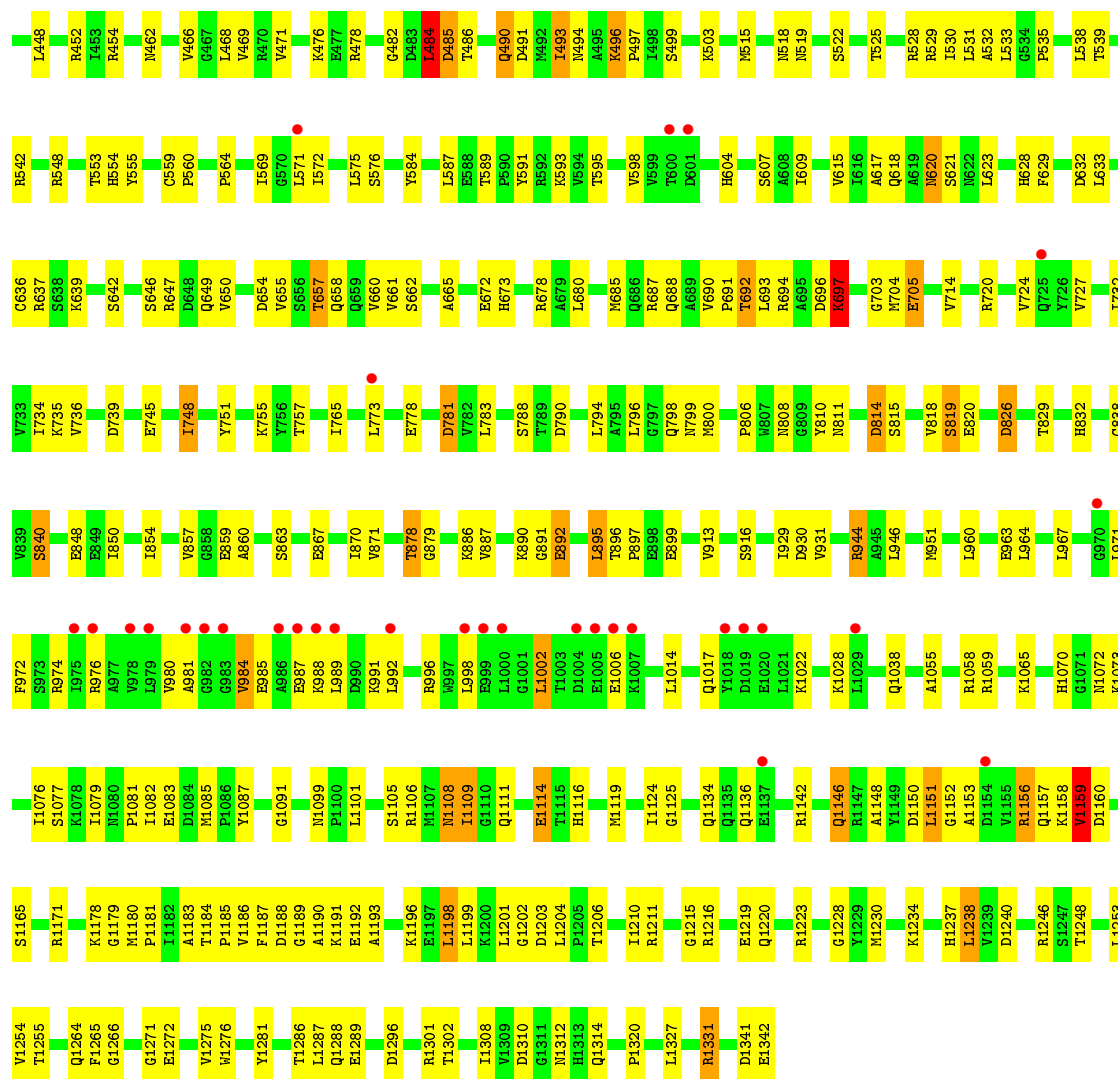




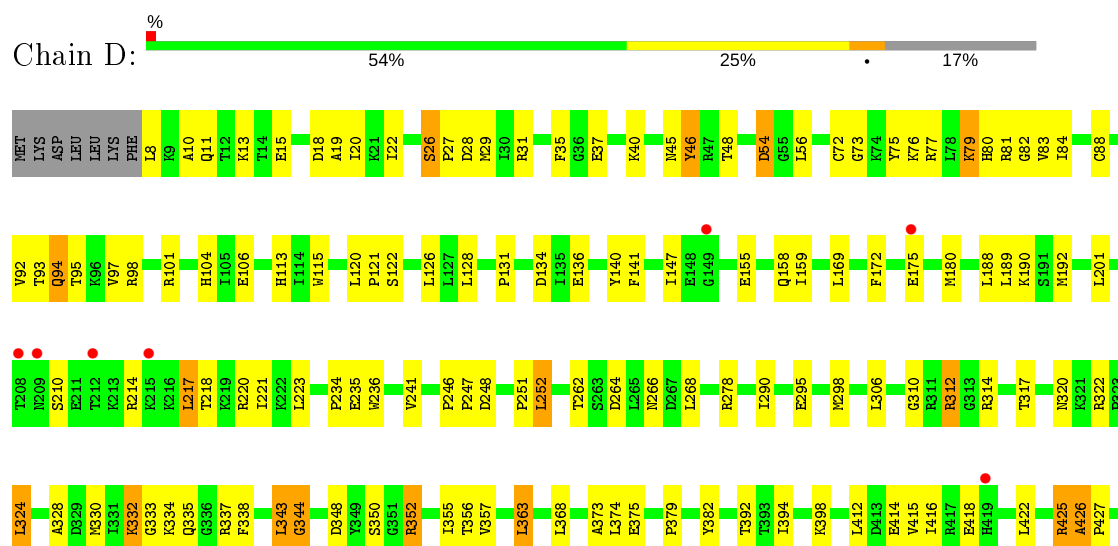
• Molecule 2: DNA-directed RNA polymerase subunit beta





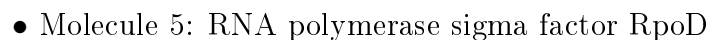


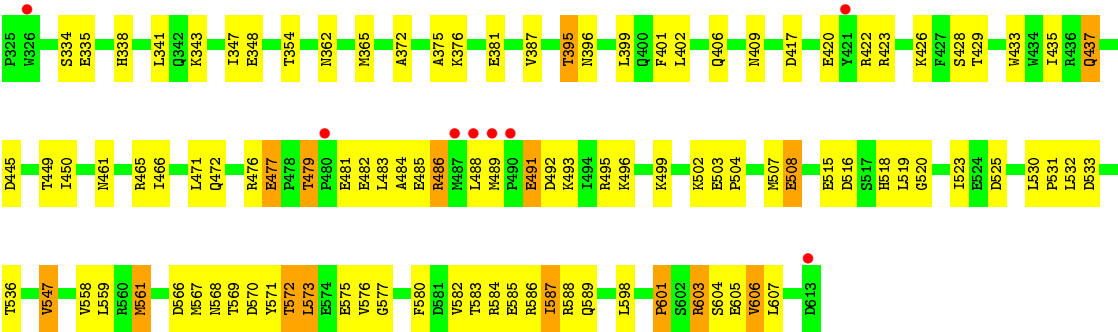
• Molecule 3: DNA-directed RNA polymerase subunit beta'











## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	187.22Å 204.43Å 310.59Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.96 – 3.92 29.96 – 3.91	Depositor EDS
% Data completeness (in resolution range)	79.1 (29.96-3.92) 79.1 (29.96-3.91)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.37 (at 3.86Å)	Xtriage
Refinement program	PHENIX (1.10.1 _2155: ???)	Depositor
R, $R_{free}$	0.260 , 0.307 0.260 , 0.307	Depositor DCC
$R_{free}$ test set	1978 reflections (2.33%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	167.5	Xtriage
Anisotropy	0.166	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 113.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.53$ , $\langle L^2 \rangle = 0.37$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	54994	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	195.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.00% 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: ZN, MG

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	1/1774 (0.1%)	0.54	0/2405
1	B	0.32	0/1668	0.70	2/2260 (0.1%)
1	G	0.33	1/1751 (0.1%)	0.54	0/2373
1	H	0.28	0/1678	0.62	0/2274
2	C	0.30	0/10756	0.54	1/14512 (0.0%)
2	I	0.27	0/10737	0.52	1/14487 (0.0%)
3	D	0.32	1/9229 (0.0%)	0.56	2/12459 (0.0%)
3	J	0.29	0/9140	0.54	1/12341 (0.0%)
4	E	0.25	0/693	0.48	0/935
4	K	0.24	0/629	0.47	0/847
5	F	0.26	0/3864	0.51	1/5194 (0.0%)
5	L	0.26	0/3872	0.48	0/5205
All	All	0.30	3/55791 (0.0%)	0.54	8/75292 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
2	C	0	3
2	I	0	2
3	D	0	2
3	J	0	2
5	F	0	1
5	L	0	1
All	All	0	11

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	29	GLU	C-N	11.10	1.55	1.34
3	D	426	ALA	C-N	9.32	1.51	1.34
1	G	29	GLU	C-N	7.73	1.49	1.34

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	9	LEU	C-N-CA	9.97	146.62	121.70
2	I	109	ALA	C-N-CD	-5.66	108.15	120.60
1	B	13	LEU	CA-CB-CG	5.43	127.80	115.30
3	D	857	LEU	CA-CB-CG	5.22	127.30	115.30
3	J	857	LEU	CA-CB-CG	5.20	127.26	115.30

There are no chirality outliers.

5 of 11 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	109	ALA	Peptide
2	C	2	VAL	Peptide
2	C	236	LYS	Peptide
3	D	1184	ASP	Peptide
3	D	1296	GLY	Peptide

## 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	1753	0	1780	72	0
1	B	1649	0	1673	72	0
1	G	1730	0	1756	61	0
1	H	1659	0	1692	72	0
2	C	10587	0	10609	319	0
2	I	10568	0	10582	269	0
3	D	9089	0	9263	265	0
3	J	9001	0	9167	278	0
4	E	691	0	695	13	0
4	K	627	0	634	21	0
5	F	3813	0	3880	96	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	L	3821	0	3884	96	0
6	D	1	0	0	0	0
6	J	1	0	0	0	0
7	D	2	0	0	0	0
7	J	2	0	0	0	0
All	All	54994	0	55615	1475	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 13.

The worst 5 of 1475 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:I:1272:GLU:H	3:J:343:LEU:HD12	1.37	0.90
1:H:32:GLU:HA	1:H:198:LEU:HD22	1.51	0.89
1:B:32:GLU:HA	1:B:198:LEU:HD22	1.54	0.89
3:D:660:GLU:HB3	3:D:685:ILE:HD12	1.55	0.88
2:I:1105:SER:HB2	3:J:731:ARG:HG2	1.57	0.87

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	225/329 (68%)	193 (86%)	28 (12%)	4 (2%)	8	41
1	B	210/329 (64%)	182 (87%)	23 (11%)	5 (2%)	6	36
1	G	222/329 (68%)	191 (86%)	27 (12%)	4 (2%)	8	41
1	H	211/329 (64%)	183 (87%)	22 (10%)	6 (3%)	5	34
2	C	1340/1342 (100%)	1230 (92%)	102 (8%)	8 (1%)	25	63

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	I	1338/1342 (100%)	1234 (92%)	98 (7%)	6 (0%)	34	71
3	D	1162/1407 (83%)	1069 (92%)	86 (7%)	7 (1%)	25	63
3	J	1151/1407 (82%)	1058 (92%)	84 (7%)	9 (1%)	19	57
4	E	87/91 (96%)	81 (93%)	6 (7%)	0	100	100
4	K	77/91 (85%)	74 (96%)	3 (4%)	0	100	100
5	F	462/613 (75%)	425 (92%)	36 (8%)	1 (0%)	47	79
5	L	463/613 (76%)	423 (91%)	39 (8%)	1 (0%)	47	79
All	All	6948/8222 (84%)	6343 (91%)	554 (8%)	51 (1%)	22	60

5 of 51 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	13	LEU
2	C	1159	VAL
3	D	332	LYS
1	H	20	SER
2	I	1159	VAL

### 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	194/286 (68%)	181 (93%)	13 (7%)	16	45
1	B	182/286 (64%)	170 (93%)	12 (7%)	16	46
1	G	191/286 (67%)	179 (94%)	12 (6%)	18	47
1	H	184/286 (64%)	174 (95%)	10 (5%)	22	51
2	C	1157/1157 (100%)	1054 (91%)	103 (9%)	9	36
2	I	1154/1157 (100%)	1051 (91%)	103 (9%)	9	36
3	D	970/1168 (83%)	873 (90%)	97 (10%)	7	30
3	J	960/1168 (82%)	863 (90%)	97 (10%)	7	30
4	E	72/75 (96%)	64 (89%)	8 (11%)	6	27

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
4	K	67/75 (89%)	63 (94%)	4 (6%)	19	48
5	F	417/540 (77%)	374 (90%)	43 (10%)	7	29
5	L	418/540 (77%)	376 (90%)	42 (10%)	7	30
All	All	5966/7024 (85%)	5422 (91%)	544 (9%)	9	35

5 of 544 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	F	401	PHE
2	I	82	VAL
5	L	98	VAL
5	F	450	ILE
1	G	9	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 63 such sidechains are listed below:

Mol	Chain	Res	Type
5	F	446	GLN
2	I	568	ASN
5	L	131	GLN
5	F	455	HIS
1	H	186	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 ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry

Of 6 ligands modelled in this entry, 6 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

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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	227/329 (68%)	-0.25	1 (0%) 92 87	133, 170, 230, 390	0
1	B	214/329 (65%)	0.15	10 (4%) 31 26	146, 213, 308, 463	0
1	G	224/329 (68%)	-0.08	6 (2%) 54 44	171, 216, 299, 368	0
1	H	215/329 (65%)	0.09	5 (2%) 60 51	173, 222, 293, 418	0
2	C	1342/1342 (100%)	-0.15	35 (2%) 56 46	107, 161, 300, 496	0
2	I	1340/1342 (99%)	-0.00	42 (3%) 49 38	146, 194, 325, 509	0
3	D	1166/1407 (82%)	-0.17	17 (1%) 73 64	108, 149, 239, 419	0
3	J	1155/1407 (82%)	-0.07	30 (2%) 56 46	127, 170, 264, 409	0
4	E	89/91 (97%)	-0.02	0 100 100	186, 210, 243, 257	0
4	K	79/91 (86%)	0.69	12 (15%) 2 2	242, 288, 354, 398	0
5	F	468/613 (76%)	-0.08	10 (2%) 63 54	151, 209, 338, 459	0
5	L	469/613 (76%)	-0.10	14 (2%) 50 39	162, 215, 360, 527	0
All	All	6988/8222 (84%)	-0.08	182 (2%) 56 46	107, 184, 306, 527	0

The worst 5 of 182 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	F	167	ASP	9.1
2	C	373	GLY	8.1
2	I	982	GLY	7.0
2	C	252	SER	6.2
2	C	251	ALA	5.9

### 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 carbohydrates 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
6	MG	J	1501	1/1	0.93	0.39	130,130,130,130	0
7	ZN	J	1502	1/1	0.94	0.06	213,213,213,213	0
6	MG	D	1501	1/1	0.95	0.27	104,104,104,104	0
7	ZN	J	1503	1/1	0.98	0.17	135,135,135,135	0
7	ZN	D	1503	1/1	0.99	0.17	117,117,117,117	0
7	ZN	D	1502	1/1	0.99	0.09	177,177,177,177	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.