



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 29, 2020 – 03:13 PM BST

PDB ID : 5C4J  
Title : Crystal structure of a transcribing RNA Polymerase II complex reveals a complete transcription bubble  
Authors : Barnes, C.O.; Calero, M.; Malik, I.; Spahr, H.; Zhang, Q.; Pullara, F.; Kaplan, C.D.; Calero, G.  
Deposited on : 2015-06-18  
Resolution : 4.00 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](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 : 2.13  
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) : 2.13

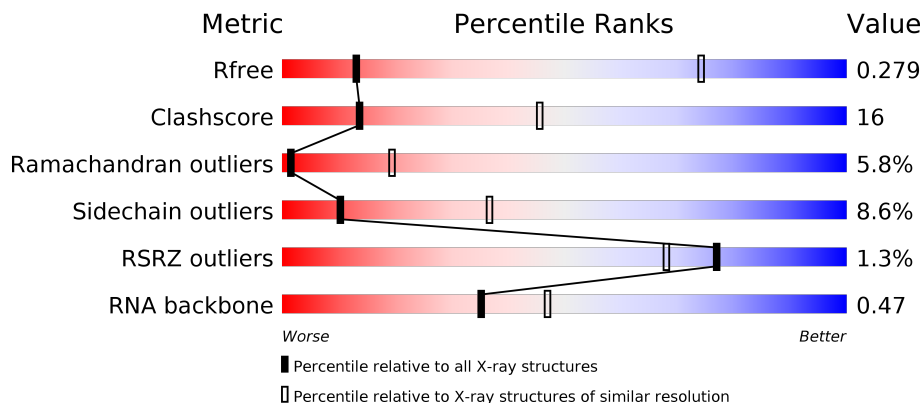
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 4.00 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.








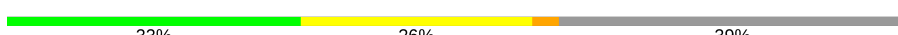

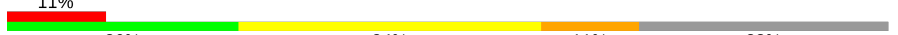

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1087 (4.30-3.70)
Clashscore	141614	1148 (4.30-3.70)
Ramachandran outliers	138981	1108 (4.30-3.70)
Sidechain outliers	138945	1099 (4.30-3.70)
RSRZ outliers	127900	1028 (4.34-3.66)
RNA backbone	3102	1048 (5.00-3.00)

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	1733	 52% 25% 5% 17%
2	B	1224	 60% 30% 5% 5%
3	C	318	 53% 28% 17%
4	E	215	 67% 29% 4% 2%

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	F	155	
6	H	146	
7	I	122	
8	J	70	
9	K	120	
10	L	70	
11	R	9	
12	S	53	
13	U	53	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
14	ZN	C	402	-	-	X	-

## 2 Entry composition [i](#)

There are 15 unique types of molecules in this entry. The entry contains 30496 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 II subunit RPB1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1432	11240	7079	1964	2136	61	0	0	0

- Molecule 2 is a protein called DNA-directed RNA polymerase II subunit RPB2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	1157	9145	5776	1599	1714	56	0	0	0

- Molecule 3 is a protein called DNA-directed RNA polymerase II subunit RPB3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	265	2074	1304	345	412	13	0	0	0

- Molecule 4 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	E	212	1735	1102	306	316	11	0	0	0

- Molecule 5 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	F	87	705	451	119	132	3	0	0	0

- Molecule 6 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
6	H	143	1102	689	189	220	4	0	0	0

- Molecule 7 is a protein called DNA-directed RNA polymerase II subunit RPB9.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
7	I	114	927	571	168	178	10	0	0	0

- Molecule 8 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
8	J	66	540	345	94	95	6	0	0	0

- Molecule 9 is a protein called DNA-directed RNA polymerase II subunit RPB11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
9	K	115	924	593	157	172	2	0	0	0

- Molecule 10 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
10	L	43	344	211	69	60	4	0	0	0

- Molecule 11 is a RNA chain called RNA (5'-R(P\*UP\*CP\*GP\*AP\*GP\*AP\*GP\*GP\*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
11	R	9	197	88	40	60	9	0	0	0

- Molecule 12 is a DNA chain called NON-TEMPLATE STRAND DNA (38-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
12	S	38	782	371	142	231	38	0	0	0

- Molecule 13 is a DNA chain called TEMPLATE STRAND DNA (38-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
13	U	38	771	366	144	223	38	0	0	0

- Molecule 14 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
14	J	1	Total 1	Zn 1	0	0
14	B	1	Total 1	Zn 1	0	0
14	I	2	Total 2	Zn 2	0	0
14	C	2	Total 2	Zn 2	0	0
14	A	2	Total 2	Zn 2	0	0
14	L	1	Total 1	Zn 1	0	0

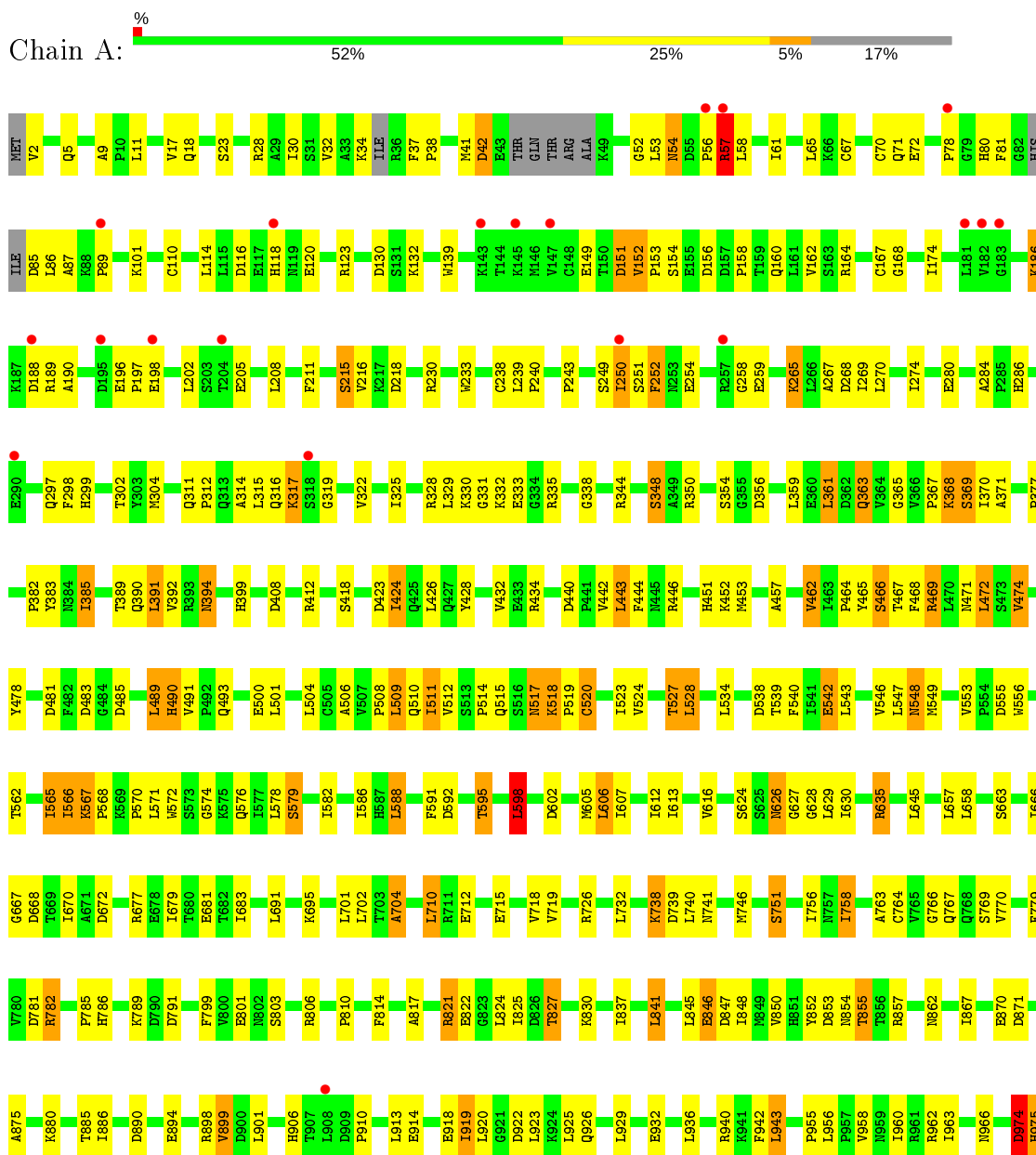
- Molecule 15 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

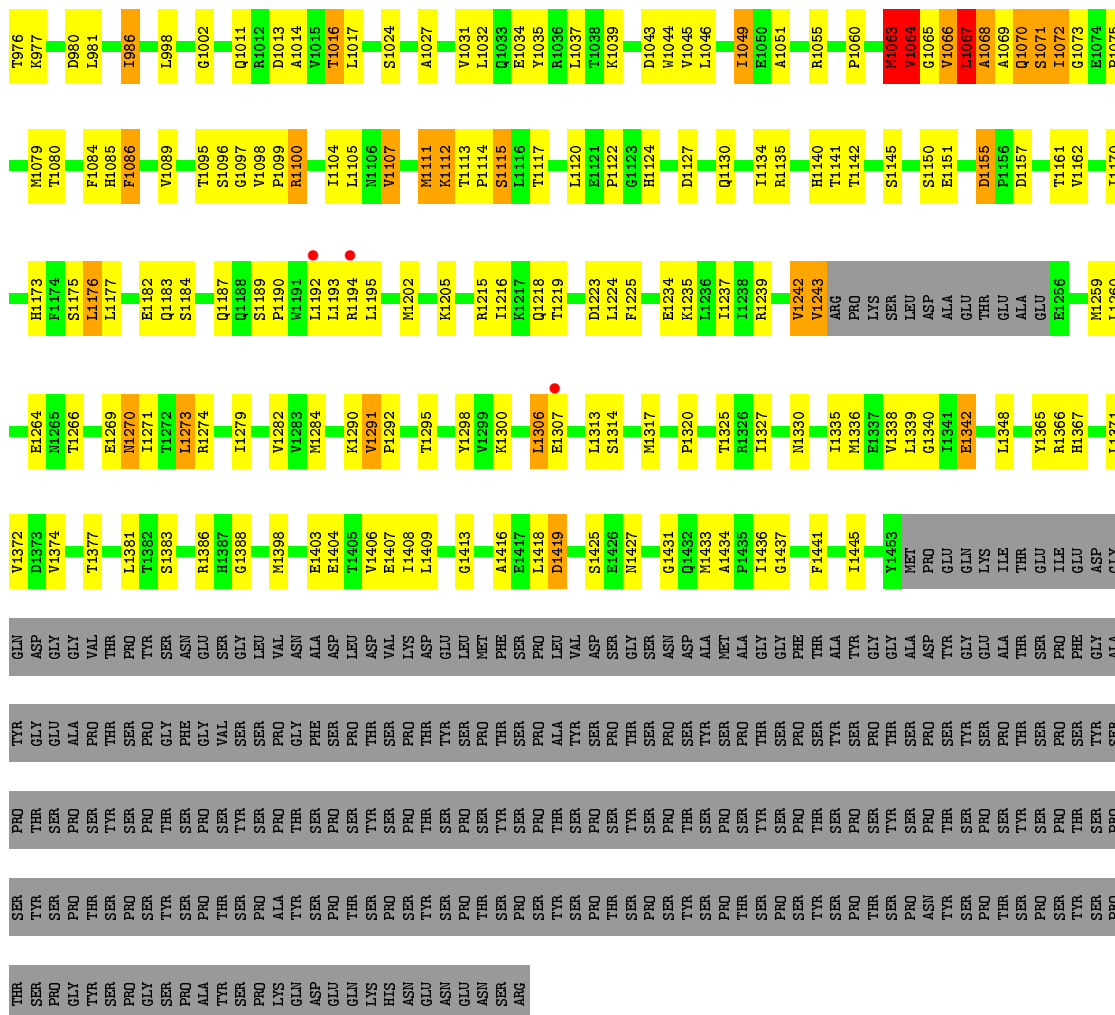
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
15	R	1	Total 1	Mg 1	0	0

### 3 Residue-property plots [i](#)

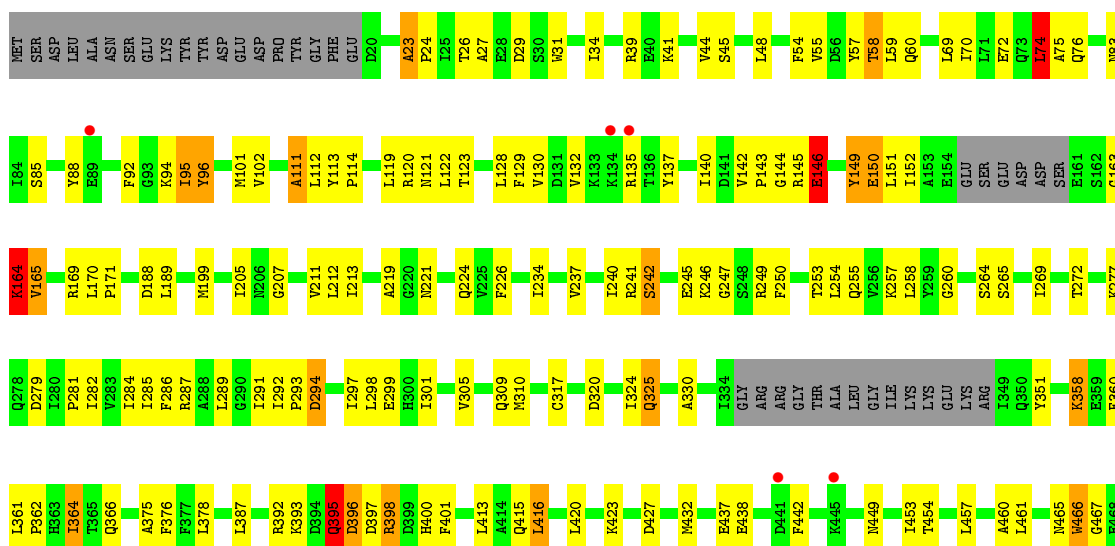
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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 II subunit RPB1

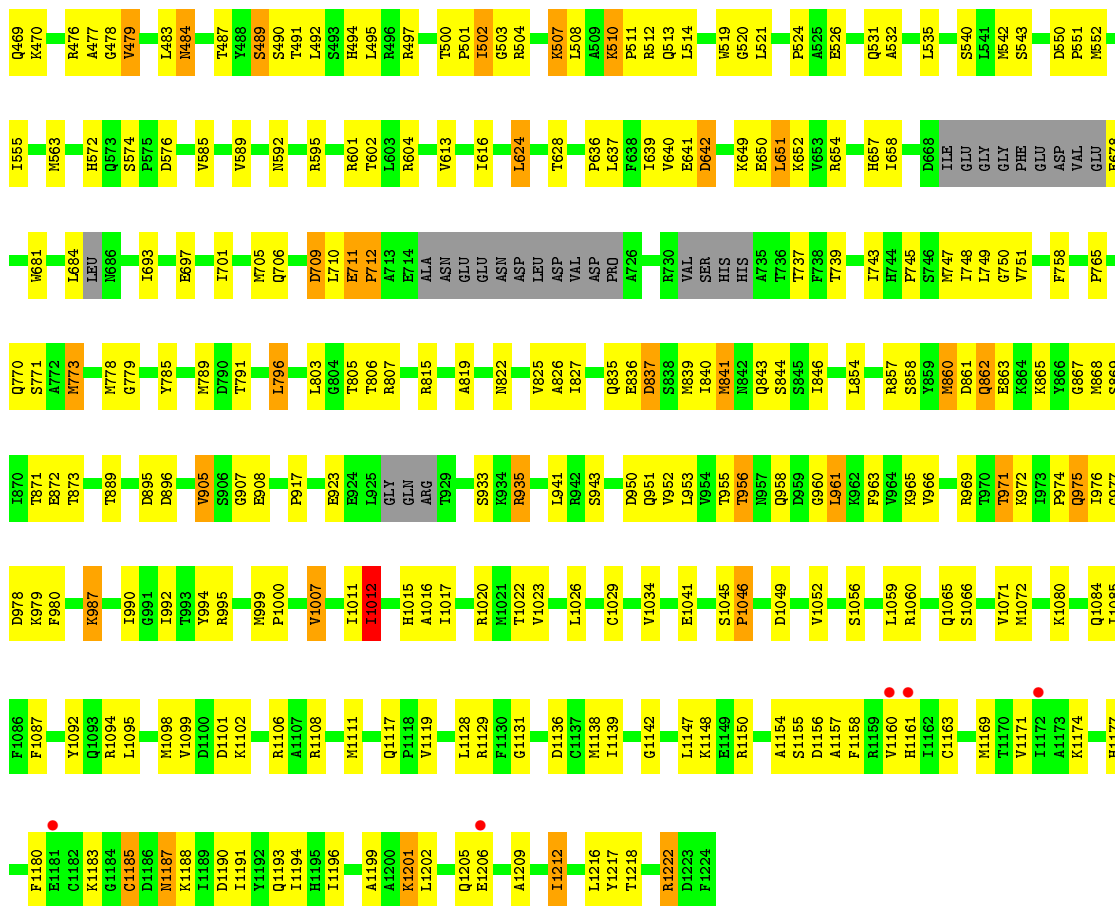




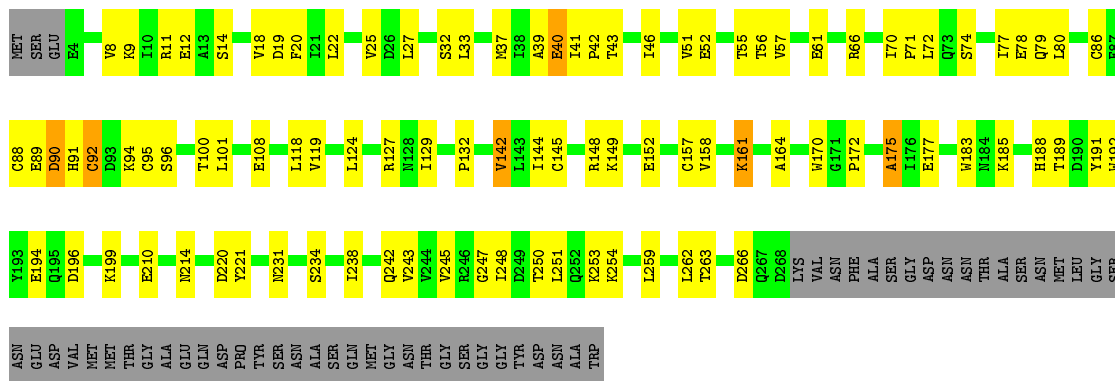
● Molecule 2: DNA-directed RNA polymerase II subunit RPB2



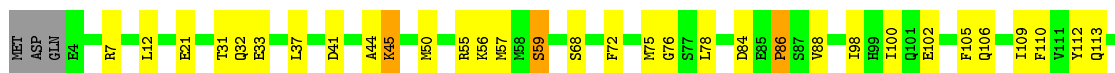




● Molecule 3: DNA-directed RNA polymerase II subunit RPB3



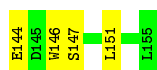
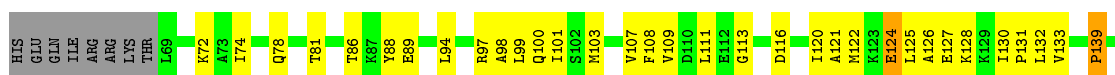
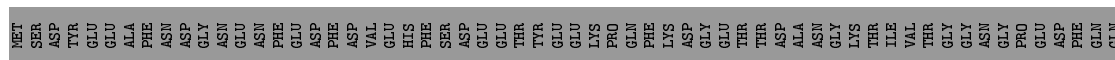
● Molecule 4: DNA-directed RNA polymerases I, II, and III subunit RPABC1





- Molecule 5: DNA-directed RNA polymerases I, II, and III subunit RPABC2

Chain F:



- Molecule 6: DNA-directed RNA polymerases I, II, and III subunit RPABC3

Chain H:



- Molecule 7: DNA-directed RNA polymerase II subunit RPB9

Chain I:



- Molecule 8: DNA-directed RNA polymerases I, II, and III subunit RPABC5

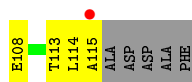
Chain J:



- Molecule 9: DNA-directed RNA polymerase II subunit RPB11

Chain K:





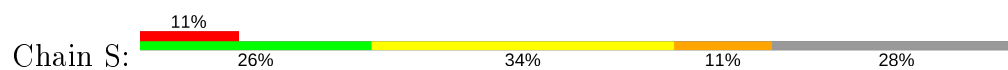
- Molecule 10: DNA-directed RNA polymerases I, II, and III subunit RPABC4



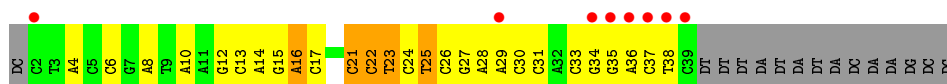
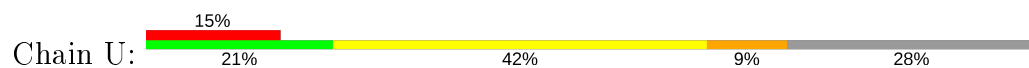
- Molecule 11: RNA (5'-R(P\*UP\*CP\*GP\*AP\*GP\*AP\*GP\*GP\*A)-3')



- Molecule 12: NON-TEMPLATE STRAND DNA (38-MER)



- Molecule 13: TEMPLATE STRAND DNA (38-MER)



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	280.71Å 223.38Å 156.42Å 90.00° 98.14° 90.00°	Depositor
Resolution (Å)	174.08 – 4.00 48.89 – 4.00	Depositor EDS
% Data completeness (in resolution range)	97.7 (174.08-4.00) 97.7 (48.89-4.00)	Depositor EDS
$R_{merge}$	0.15	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.51 (at 4.00Å)	Xtrriage
Refinement program	REFMAC 5.8.0049	Depositor
R, $R_{free}$	0.212 , 0.271 0.229 , 0.279	Depositor DCC
$R_{free}$ test set	2689 reflections (3.42%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	155.3	Xtrriage
Anisotropy	0.326	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 186.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.26$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	30496	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	231.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.93% 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 [i](#)

### 5.1 Standard geometry [i](#)

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	0.46	0/11441	0.77	5/15470 (0.0%)
2	B	0.45	0/9320	0.74	3/12568 (0.0%)
3	C	0.43	0/2112	0.68	0/2866
4	E	0.45	0/1771	0.74	1/2383 (0.0%)
5	F	0.43	0/717	0.77	1/967 (0.1%)
6	H	0.43	0/1120	0.81	1/1513 (0.1%)
7	I	0.42	0/945	0.78	1/1273 (0.1%)
8	J	0.45	0/549	0.73	0/738
9	K	0.42	0/942	0.69	0/1272
10	L	0.40	0/346	0.75	0/457
11	R	0.46	0/221	0.72	0/343
12	S	1.05	0/876	1.88	31/1351 (2.3%)
13	U	1.02	0/864	1.83	34/1328 (2.6%)
All	All	0.50	0/31224	0.86	77/42529 (0.2%)

There are no bond length outliers.

The worst 5 of 77 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
12	S	15	DC	O4'-C1'-N1	12.27	116.59	108.00
13	U	25	DT	P-O3'-C3'	12.22	134.36	119.70
12	S	10	DG	P-O3'-C3'	11.81	133.88	119.70
13	U	27	DG	P-O3'-C3'	11.61	133.64	119.70
13	U	35	DG	P-O3'-C3'	11.16	133.09	119.70

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	11240	0	11270	290	0
2	B	9145	0	9117	267	0
3	C	2074	0	2020	159	0
4	E	1735	0	1762	83	0
5	F	705	0	731	52	0
6	H	1102	0	1035	58	0
7	I	927	0	886	34	0
8	J	540	0	554	25	0
9	K	924	0	934	73	0
10	L	344	0	366	26	0
11	R	197	0	96	6	0
12	S	782	0	429	7	0
13	U	771	0	425	6	0
14	A	2	0	0	0	0
14	B	1	0	0	0	0
14	C	2	0	0	2	0
14	I	2	0	0	0	0
14	J	1	0	0	0	0
14	L	1	0	0	1	0
15	R	1	0	0	0	0
All	All	30496	0	29625	957	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 16.

The worst 5 of 957 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)
4:E:151:PRO:HD2	4:E:153:HIS:CE1	1.30	1.65
4:E:151:PRO:CD	4:E:153:HIS:HE1	1.24	1.50
7:I:56:ALA:HB2	7:I:89:GLN:CD	1.35	1.47
7:I:56:ALA:CB	7:I:89:GLN:OE1	1.64	1.46
3:C:221:TYR:CB	6:H:46:LEU:HD22	1.44	1.43

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	1422/1733 (82%)	1129 (79%)	199 (14%)	94 (7%)	1	17
2	B	1141/1224 (93%)	919 (80%)	158 (14%)	64 (6%)	2	20
3	C	263/318 (83%)	226 (86%)	30 (11%)	7 (3%)	5	34
4	E	210/215 (98%)	186 (89%)	17 (8%)	7 (3%)	4	30
5	F	85/155 (55%)	72 (85%)	9 (11%)	4 (5%)	2	23
6	H	139/146 (95%)	97 (70%)	26 (19%)	16 (12%)	0	6
7	I	112/122 (92%)	88 (79%)	16 (14%)	8 (7%)	1	16
8	J	64/70 (91%)	49 (77%)	11 (17%)	4 (6%)	1	18
9	K	113/120 (94%)	99 (88%)	12 (11%)	2 (2%)	8	41
10	L	41/70 (59%)	31 (76%)	6 (15%)	4 (10%)	0	10
All	All	3590/4173 (86%)	2896 (81%)	484 (14%)	210 (6%)	1	19

5 of 210 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	57	ARG
1	A	65	LEU
1	A	152	VAL
1	A	153	PRO
1	A	162	VAL

### 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	1244/1520 (82%)	1081 (87%)	163 (13%)	4	21
2	B	991/1061 (93%)	889 (90%)	102 (10%)	7	28
3	C	230/274 (84%)	229 (100%)	1 (0%)	91	94
4	E	194/197 (98%)	193 (100%)	1 (0%)	88	93
5	F	77/137 (56%)	76 (99%)	1 (1%)	69	82
6	H	115/128 (90%)	115 (100%)	0	100	100
7	I	108/116 (93%)	106 (98%)	2 (2%)	57	75
8	J	61/65 (94%)	60 (98%)	1 (2%)	62	79
9	K	99/102 (97%)	99 (100%)	0	100	100
10	L	38/57 (67%)	37 (97%)	1 (3%)	46	67
All	All	3157/3657 (86%)	2885 (91%)	272 (9%)	10	37

5 of 272 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1096	SER
1	A	1291	VAL
2	B	1072	MET
1	A	1112	LYS
1	A	1202	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 46 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1354	ASN
2	B	224	GLN
9	K	2	ASN
1	A	1387	HIS
2	B	46	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
11	R	7/9 (77%)	2 (28%)	0

All (2) RNA backbone outliers are listed below:



Mol	Chain	Res	Type
11	R	6	G
11	R	8	G

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 monosaccharides in this entry.

#### 5.6 Ligand geometry [i](#)

Of 10 ligands modelled in this entry, 10 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	1432/1733 (82%)	-0.21	23 (1%) 72 62	28, 216, 298, 382	0
2	B	1157/1224 (94%)	-0.22	10 (0%) 84 77	115, 221, 304, 366	0
3	C	265/318 (83%)	-0.50	0 100 100	134, 213, 270, 315	0
4	E	212/215 (98%)	-0.35	0 100 100	153, 232, 298, 322	0
5	F	87/155 (56%)	-0.36	0 100 100	134, 193, 260, 341	0
6	H	143/146 (97%)	-0.33	0 100 100	176, 258, 312, 334	0
7	I	114/122 (93%)	-0.35	0 100 100	177, 240, 306, 334	0
8	J	66/70 (94%)	-0.40	0 100 100	125, 223, 286, 318	0
9	K	115/120 (95%)	-0.31	1 (0%) 84 77	131, 202, 271, 293	0
10	L	43/70 (61%)	-0.30	0 100 100	170, 228, 283, 364	0
11	R	9/9 (100%)	0.18	0 100 100	285, 306, 337, 338	0
12	S	38/53 (71%)	0.98	6 (15%) 2 2	338, 360, 380, 381	0
13	U	38/53 (71%)	0.96	8 (21%) 1 1	275, 348, 378, 380	0
All	All	3719/4288 (86%)	-0.24	48 (1%) 77 68	28, 221, 309, 382	0

The worst 5 of 48 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	204	THR	6.9
1	A	183	GLY	4.3
13	U	35	DG	3.8
1	A	1194	ARG	3.7
2	B	1181	GLU	3.7

### 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
15	MG	R	101	1/1	0.56	0.16	201,201,201,201	0
14	ZN	I	202	1/1	0.87	0.03	213,213,213,213	0
14	ZN	J	101	1/1	0.92	0.32	207,207,207,207	0
14	ZN	A	1801	1/1	0.96	0.04	205,205,205,205	0
14	ZN	L	101	1/1	0.97	0.11	202,202,202,202	0
14	ZN	C	401	1/1	0.97	0.22	226,226,226,226	0
14	ZN	A	1802	1/1	0.97	0.10	200,200,200,200	0
14	ZN	B	1301	1/1	0.98	0.11	214,214,214,214	0
14	ZN	I	201	1/1	0.99	0.13	201,201,201,201	0
14	ZN	C	402	1/1	1.00	0.14	186,186,186,186	0

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

There are no such residues in this entry.