



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 27, 2023 – 04:36 AM EDT

PDB ID : 3GTK
Title : Backtracked RNA polymerase II complex with 18mer RNA
Authors : Wang, D.; Bushnell, D.A.; Huang, X.; Westover, K.D.; Levitt, M.; Kornberg, R.D.
Deposited on : 2009-03-27
Resolution : 3.80 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 : 4.02b-467
Xtriage (Phenix) : 1.13
EDS : 2.35
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) : 2.35

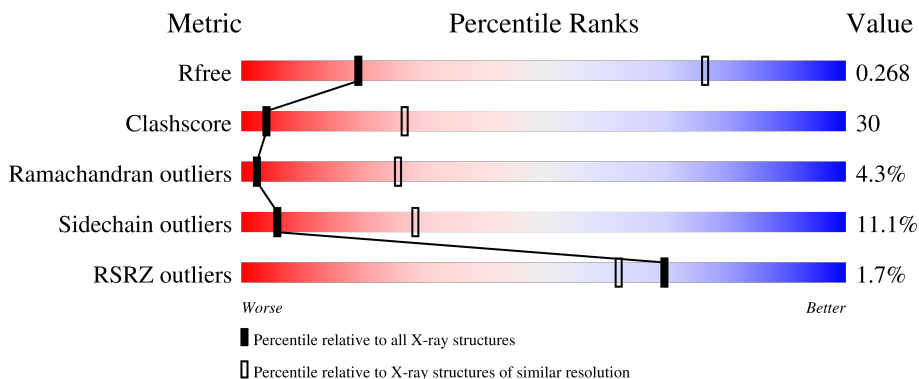
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.80 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.




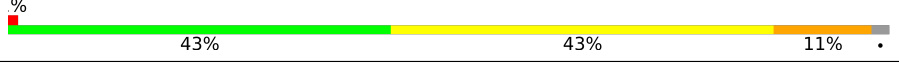
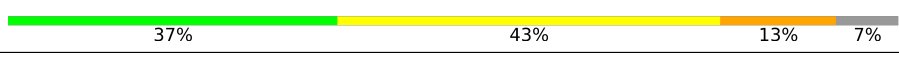

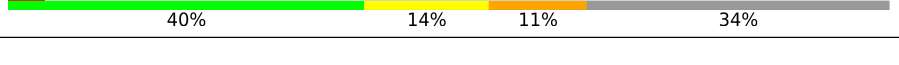
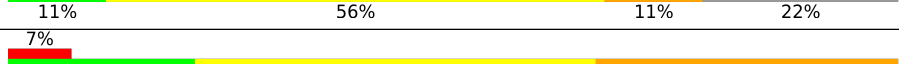
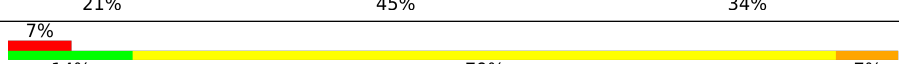
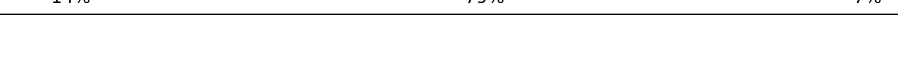
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1212 (4.00-3.60)
Clashscore	141614	1288 (4.00-3.60)
Ramachandran outliers	138981	1243 (4.00-3.60)
Sidechain outliers	138945	1237 (4.00-3.60)
RSRZ outliers	127900	1121 (4.00-3.60)

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	1733	 2% 41% 34% 7% 17%
2	B	1224	 2% 44% 42% 7% 6%
3	C	318	 % 45% 36% 5% 15%
4	E	215	 68% 30% 2%
5	F	155	 33% 21% 45%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	H	146	
7	I	122	
8	J	70	
9	K	120	
10	L	70	
11	R	18	
12	T	29	
13	N	14	

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	A	1734	-	-	X	-
14	ZN	C	319	-	-	X	-
14	ZN	J	101	-	-	X	-

2 Entry composition [i](#)

There are 14 unique types of molecules in this entry. The entry contains 30111 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	1442	11332	7133	1982	2156	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	1153	9168	5795	1604	1713	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	271	2135	1344	355	423	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	215	1757	1114	310	322	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	85	684	437	116	128	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	136	1087	684	183	215	5	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	119	971	596	179	186	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	65	532	339	93	94	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	114	919	590	156	171	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	46	363	224	72	63	4	0	0	0

- Molecule 11 is DNA/RNA hybrid called DNA/RNA (5'-R(*AP*UP*CP*GP*AP*GP*AP*GP*GP*AP*UP*GP*C)-D(P*AP*GP*AP*CP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
11	R	14	284	126	55	90	13	0	0	0

- Molecule 12 is a DNA chain called DNA (29-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
12	T	29	587	281	109	169	28	0	0	0

- Molecule 13 is a DNA chain called DNA (5'-D(*CP*TP*GP*CP*TP*TP*AP*TP*CP*GP*GP*TP*AP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
13	N	14	284	137	49	85	13	0	0	0

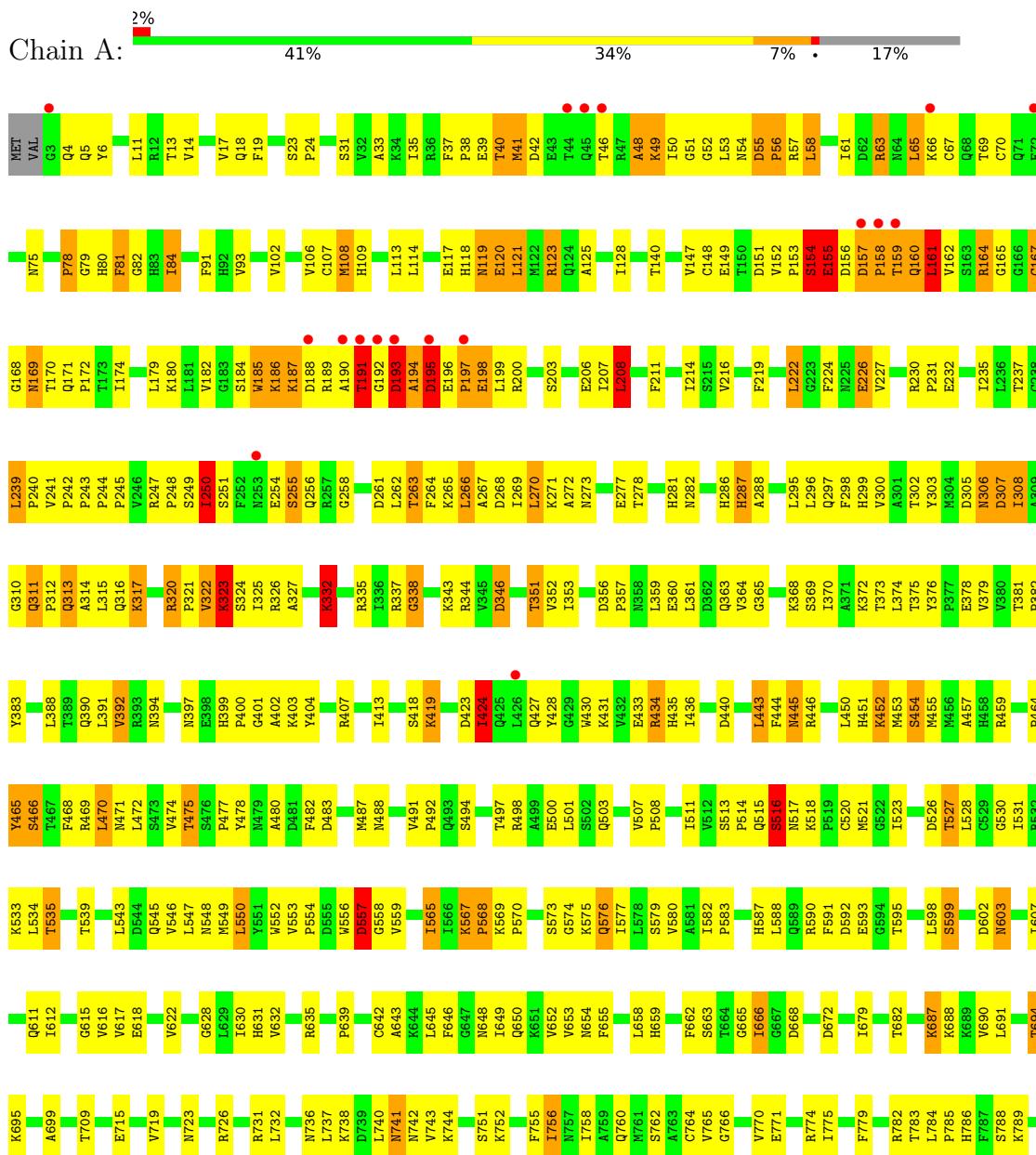
- Molecule 14 is ZINC ION (three-letter code: ZN) (formula: Zn).

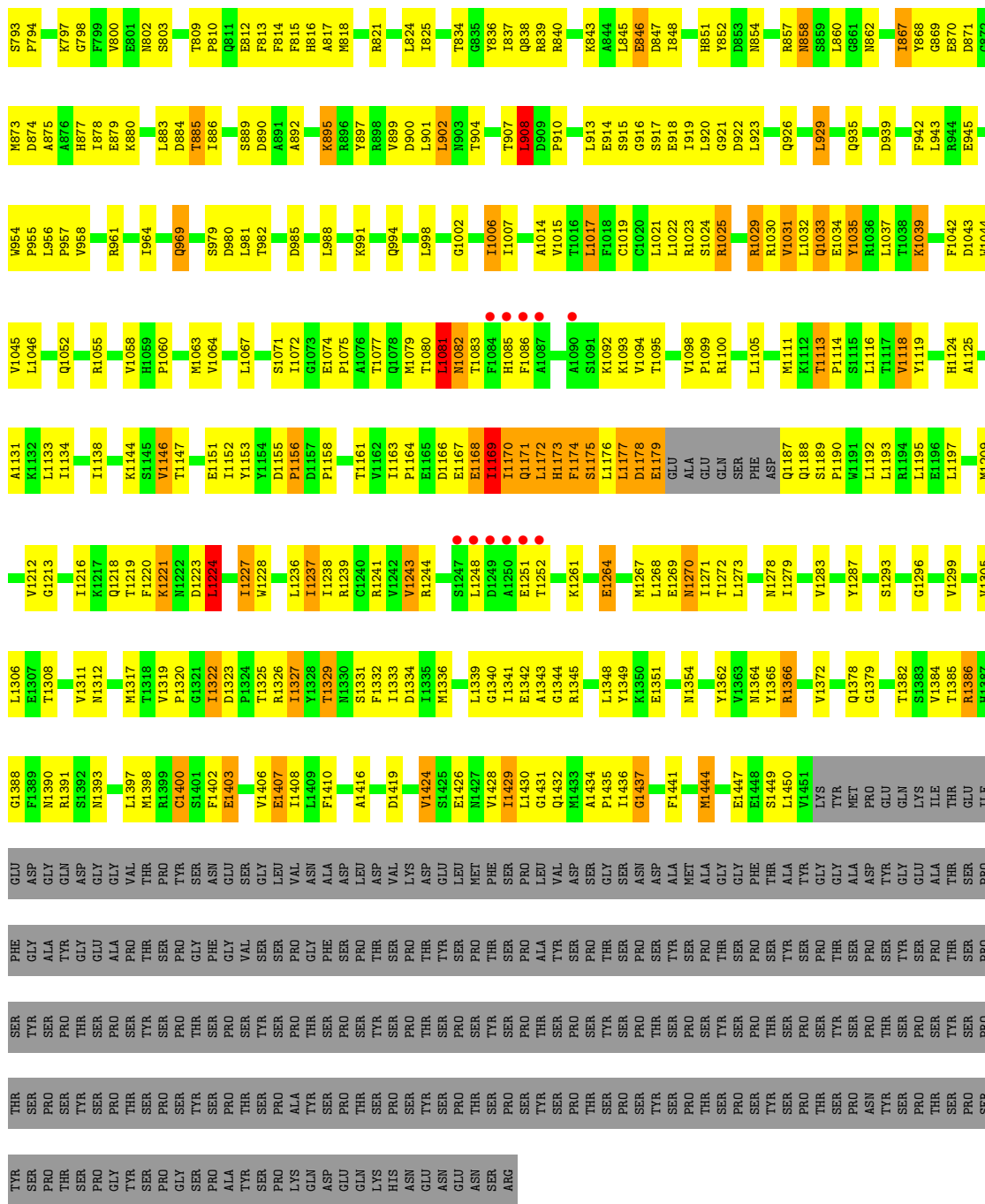
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
14	A	2	Total 2	Zn 2	0	0
14	B	1	Total 1	Zn 1	0	0
14	C	1	Total 1	Zn 1	0	0
14	I	2	Total 2	Zn 2	0	0
14	J	1	Total 1	Zn 1	0	0
14	L	1	Total 1	Zn 1	0	0

3 Residue-property plots [i](#)

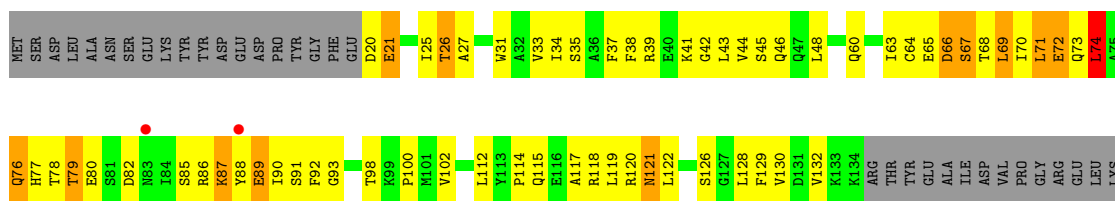
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density.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

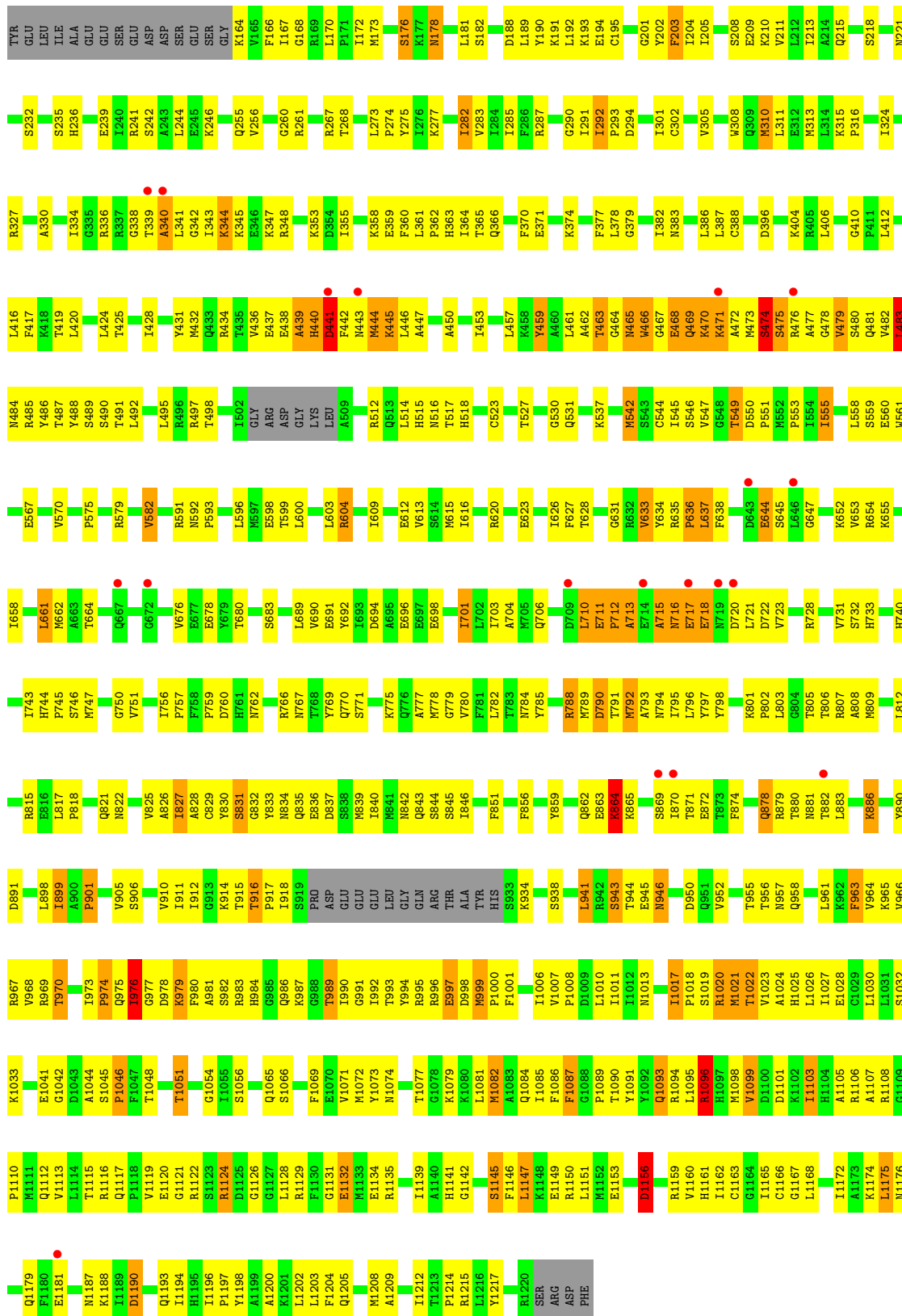
- Molecule 1: DNA-directed RNA polymerase II subunit RPB1





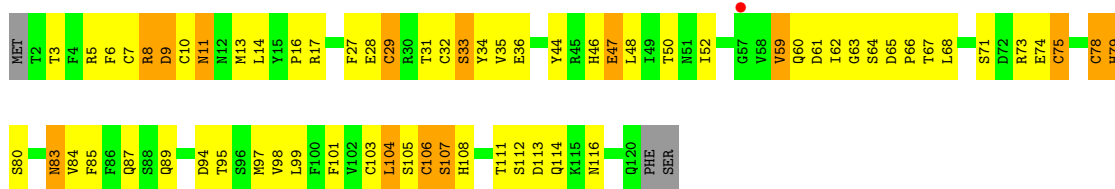
● Molecule 2: DNA-directed RNA polymerase II subunit RPB2





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





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

Chain J: 37% 43% 13% 7%



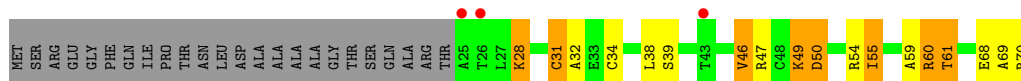
- Molecule 9: DNA-directed RNA polymerase II subunit RPB11

Chain K: 56% 33% 6% 5%



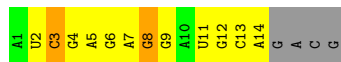
- Molecule 10: DNA-directed RNA polymerases I, II, and III subunit RPABC4

Chain L: 4% 40% 14% 11% 34%



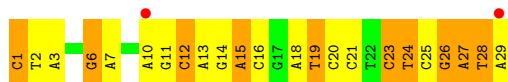
- Molecule 11: DNA/RNA (5'-R(*AP*UP*CP*GP*AP*GP*AP*GP*GP*AP*UP*GP*C)-D(P*AP*GP*AP*CP*G)-3')

Chain R: 11% 56% 11% 22%




- Molecule 12: DNA (29-MER)

Chain T: 7% 21% 45% 34%



- Molecule 13: DNA (5'-D(*CP*TP*GP*CP*TP*TP*AP*TP*CP*GP*GP*TP*AP*G)-3')

Chain N: 7% 14% 79% 7%



C1
T2
G3
C4
T5
T6
A7
T8
C9
G10
G11
T12
A13
G14

4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	170.45Å 222.29Å 195.09Å 90.00° 101.72° 90.00°	Depositor
Resolution (Å)	50.00 – 3.80 39.90 – 3.80	Depositor EDS
% Data completeness (in resolution range)	97.4 (50.00-3.80) 97.5 (39.90-3.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.96 (at 3.76Å)	Xtrriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.247 , 0.278 0.240 , 0.268	Depositor DCC
R_{free} test set	3444 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å ²)	116.7	Xtrriage
Anisotropy	0.063	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.26 , 76.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	30111	wwPDB-VP
Average B, all atoms (Å ²)	142.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.11% 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

Bond lengths and bond angles in the following residue types are not validated in this section:
ZN

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	0.58	1/11536 (0.0%)	0.72	14/15605 (0.1%)
2	B	0.64	1/9348 (0.0%)	0.76	12/12611 (0.1%)
3	C	0.65	0/2174	0.76	1/2946 (0.0%)
4	E	0.53	0/1793	0.60	0/2413
5	F	0.53	0/696	0.67	0/940
6	H	0.54	0/1105	0.68	0/1495
7	I	0.59	0/989	0.65	0/1331
8	J	0.66	0/541	0.75	0/727
9	K	0.63	0/937	0.71	1/1265 (0.1%)
10	L	0.68	0/365	0.84	1/485 (0.2%)
11	R	1.15	1/318 (0.3%)	1.79	5/496 (1.0%)
12	T	1.07	2/658 (0.3%)	1.88	27/1012 (2.7%)
13	N	1.54	0/317	2.03	19/488 (3.9%)
All	All	0.64	5/30777 (0.0%)	0.82	80/41814 (0.2%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	206	GLU	CD-OE2	5.85	1.32	1.25
12	T	28	DT	C1'-N1	5.59	1.56	1.49
2	B	544	CYS	CB-SG	-5.45	1.73	1.81
12	T	23	DC	C1'-N1	5.20	1.56	1.49
11	R	11	U	C1'-N1	5.17	1.56	1.48

The worst 5 of 80 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
12	T	25	DC	O4'-C1'-N1	10.24	115.17	108.00
12	T	28	DT	O4'-C4'-C3'	-10.00	100.00	106.00
12	T	27	DA	O4'-C4'-C3'	-9.05	100.57	106.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	715	ALA	CB-CA-C	8.60	122.99	110.10
12	T	28	DT	O4'-C1'-N1	8.54	113.98	108.00

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	11332	0	11391	806	0
2	B	9168	0	9179	616	0
3	C	2135	0	2091	139	0
4	E	1757	0	1781	48	0
5	F	684	0	703	30	0
6	H	1087	0	1062	59	0
7	I	971	0	930	64	0
8	J	532	0	544	49	0
9	K	919	0	929	55	0
10	L	363	0	389	13	0
11	R	284	0	142	19	0
12	T	587	0	327	24	0
13	N	284	0	161	3	0
14	A	2	0	0	2	0
14	B	1	0	0	0	0
14	C	1	0	0	2	0
14	I	2	0	0	0	0
14	J	1	0	0	2	0
14	L	1	0	0	0	0
All	All	30111	0	29629	1766	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 30.

The worst 5 of 1766 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:438:GLU:HB3	2:B:440:HIS:CD2	1.53	1.43
1:A:1168:GLU:CA	1:A:1171:GLN:HB2	1.51	1.41
1:A:1172:LEU:O	1:A:1174:PHE:N	1.56	1.34
2:B:439:ALA:CB	2:B:440:HIS:HA	1.56	1.29
1:A:1169:ILE:CB	1:A:1170:ILE:HB	1.60	1.29

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	1438/1733 (83%)	1083 (75%)	289 (20%)	66 (5%)	2	24
2	B	1145/1224 (94%)	877 (77%)	219 (19%)	49 (4%)	2	26
3	C	269/318 (85%)	205 (76%)	57 (21%)	7 (3%)	5	36
4	E	213/215 (99%)	180 (84%)	28 (13%)	5 (2%)	6	38
5	F	83/155 (54%)	66 (80%)	13 (16%)	4 (5%)	2	24
6	H	132/146 (90%)	100 (76%)	26 (20%)	6 (4%)	2	25
7	I	117/122 (96%)	82 (70%)	26 (22%)	9 (8%)	1	15
8	J	63/70 (90%)	53 (84%)	6 (10%)	4 (6%)	1	20
9	K	112/120 (93%)	97 (87%)	13 (12%)	2 (2%)	8	42
10	L	44/70 (63%)	29 (66%)	13 (30%)	2 (4%)	2	25
All	All	3616/4173 (87%)	2772 (77%)	690 (19%)	154 (4%)	2	26

5 of 154 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	55	ASP
1	A	117	GLU
1	A	120	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	169	ASN
1	A	186	LYS

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	1258/1520 (83%)	1103 (88%)	155 (12%)	4 24
2	B	1000/1061 (94%)	882 (88%)	118 (12%)	5 26
3	C	238/274 (87%)	212 (89%)	26 (11%)	6 29
4	E	196/197 (100%)	184 (94%)	12 (6%)	18 50
5	F	74/137 (54%)	72 (97%)	2 (3%)	44 69
6	H	119/128 (93%)	112 (94%)	7 (6%)	19 51
7	I	113/116 (97%)	101 (89%)	12 (11%)	6 30
8	J	60/65 (92%)	52 (87%)	8 (13%)	4 22
9	K	99/102 (97%)	90 (91%)	9 (9%)	9 36
10	L	40/57 (70%)	33 (82%)	7 (18%)	2 13
All	All	3197/3657 (87%)	2841 (89%)	356 (11%)	6 29

5 of 356 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	864	LYS
3	C	129	ILE
2	B	916	THR
2	B	1103	ILE
4	E	33	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 91 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	762	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	C	31	ASN
2	B	794	ASN
2	B	984	HIS
3	C	112	ASN

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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 8 ligands modelled in this entry, 8 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, 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	1442/1733 (83%)	-0.27	29 (2%) 65 58	77, 132, 243, 401	0
2	B	1153/1224 (94%)	-0.23	21 (1%) 68 61	77, 116, 250, 399	0
3	C	271/318 (85%)	-0.37	2 (0%) 87 83	86, 111, 177, 330	0
4	E	215/215 (100%)	-0.33	0 100 100	107, 153, 272, 354	0
5	F	85/155 (54%)	-0.33	0 100 100	107, 129, 169, 217	0
6	H	136/146 (93%)	-0.18	4 (2%) 51 42	120, 167, 307, 372	0
7	I	119/122 (97%)	-0.28	1 (0%) 86 81	106, 148, 191, 275	0
8	J	65/70 (92%)	-0.45	0 100 100	83, 101, 151, 176	0
9	K	114/120 (95%)	-0.35	0 100 100	83, 115, 150, 171	0
10	L	46/70 (65%)	0.11	3 (6%) 18 14	102, 161, 254, 269	0
11	R	14/18 (77%)	-0.29	0 100 100	106, 130, 215, 216	0
12	T	29/29 (100%)	0.21	2 (6%) 16 13	105, 216, 417, 435	0
13	N	14/14 (100%)	0.13	1 (7%) 16 12	263, 310, 353, 397	0
All	All	3703/4234 (87%)	-0.26	63 (1%) 70 62	77, 128, 255, 435	0

The worst 5 of 63 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	882	THR	6.7
1	A	44	THR	5.6
10	L	25	ALA	4.8
1	A	1250	ALA	4.8
6	H	85	GLY	4.7

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, 95th 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(\AA^2)	Q<0.9
14	ZN	A	1734	1/1	0.70	0.06	228,228,228,228	0
14	ZN	I	204	1/1	0.96	0.13	310,310,310,310	0
14	ZN	A	1735	1/1	0.98	0.12	149,149,149,149	0
14	ZN	C	319	1/1	0.99	0.06	136,136,136,136	0
14	ZN	I	203	1/1	0.99	0.11	132,132,132,132	0
14	ZN	B	1307	1/1	0.99	0.08	157,157,157,157	0
14	ZN	J	101	1/1	0.99	0.17	193,193,193,193	0
14	ZN	L	105	1/1	0.99	0.02	131,131,131,131	0

6.5 Other polymers [i](#)

There are no such residues in this entry.