wwPDB X-ray Structure Validation Summary Report (i)

Sep 3, 2023 – 01:44 PM EDT

PDB ID	:	3S1M
Title	:	RNA Polymerase II Initiation Complex with a 5-nt RNA (variant 1)
Authors	:	Liu, X.; Bushnell, D.A.; Silva, D.A.; Huang, X.; Kornberg, R.D.
Deposited on	:	2011-05-15
Resolution	:	3.13 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 (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 (1)) were used in the production of this report:

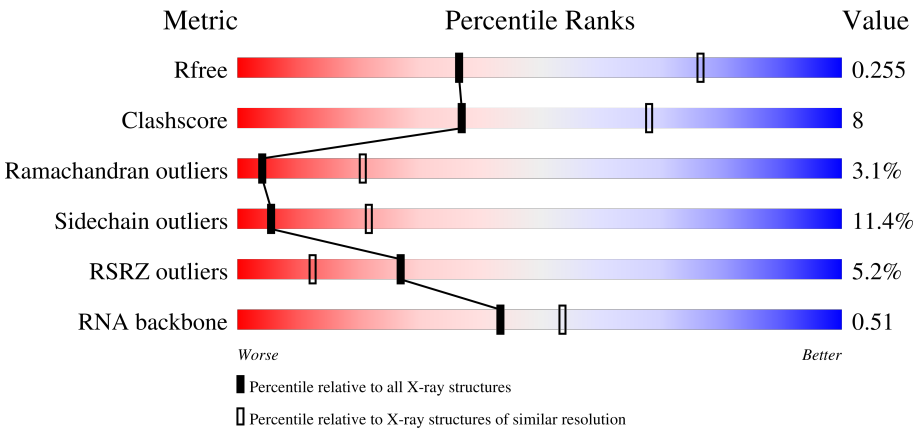
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.13 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	1626 (3.18-3.10)
Clashscore	141614	1735 (3.18-3.10)
Ramachandran outliers	138981	1677 (3.18-3.10)
Sidechain outliers	138945	1677 (3.18-3.10)
RSRZ outliers	127900	1588 (3.18-3.10)
RNA backbone	3102	1000 (3.46-2.82)

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 <=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	8% 58% 18% • 19%
2	B	1224	3% 65% 23% • 9%
3	C	318	60% 20% • 16%
4	E	215	% 81% 16% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	F	155	48% 7% 45%
6	H	146	3% 64% 21% 6% 9%
7	I	122	74% 21%
8	J	70	61% 24% 6% 7%
9	K	120	72% 18% 5%
10	L	70	3% 29% 30% 7% 34%
11	R	5	20% 80% 20%
12	T	29	10% 17% 72%

2 Entry composition (i)

There are 14 unique types of molecules in this entry. The entry contains 28570 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	1405	11043	6965	1936	2081	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	1114	8861	5610	1549	1647	55	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	266	2095	1317	348	417	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	214	1752	1111	309	321	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	688	439	116	130	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	133	1068	673	180	211	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	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 a RNA chain called RNA (5'-R(*AP*GP*AP*CP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
11	R	5	107	49	23	31	4	0	0	0

- Molecule 12 is a DNA chain called DNA (5'-D(*CP*TP*AP*CP*CP*GP*AP*TP*AP*AP *GP*CP*AP*GP*AP*CP*GP*AP*TP*CP*GP*TP*CP*TP*CP*GP*AP*TP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
12	T	8	162	77	28	49	8	0	0	0

- Molecule 13 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
13	A	2	2	2	0	0

Continued on next page...

Continued from previous page...

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

- Molecule 14 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

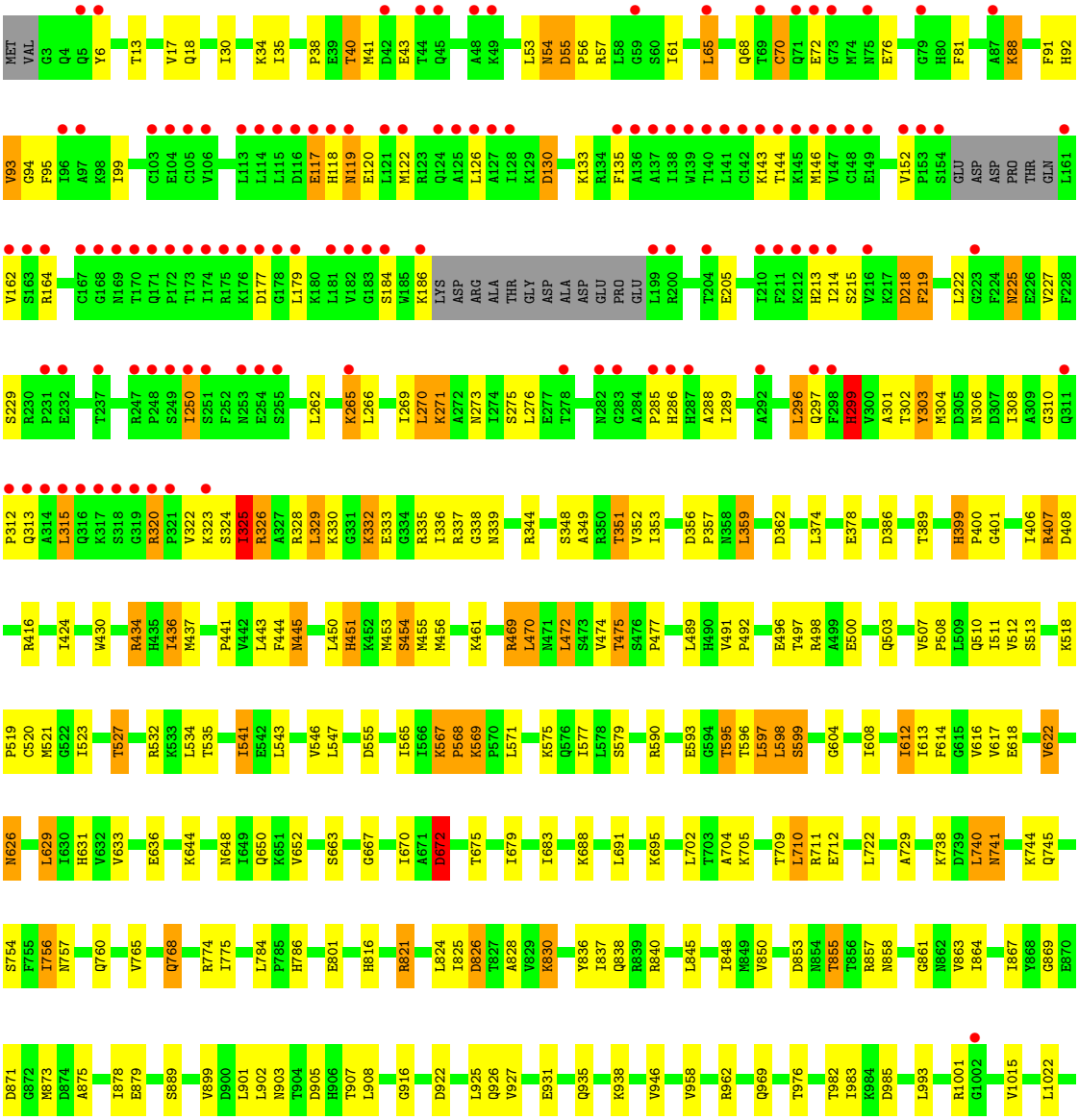
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mg		
14	A	1	1	1	0	0

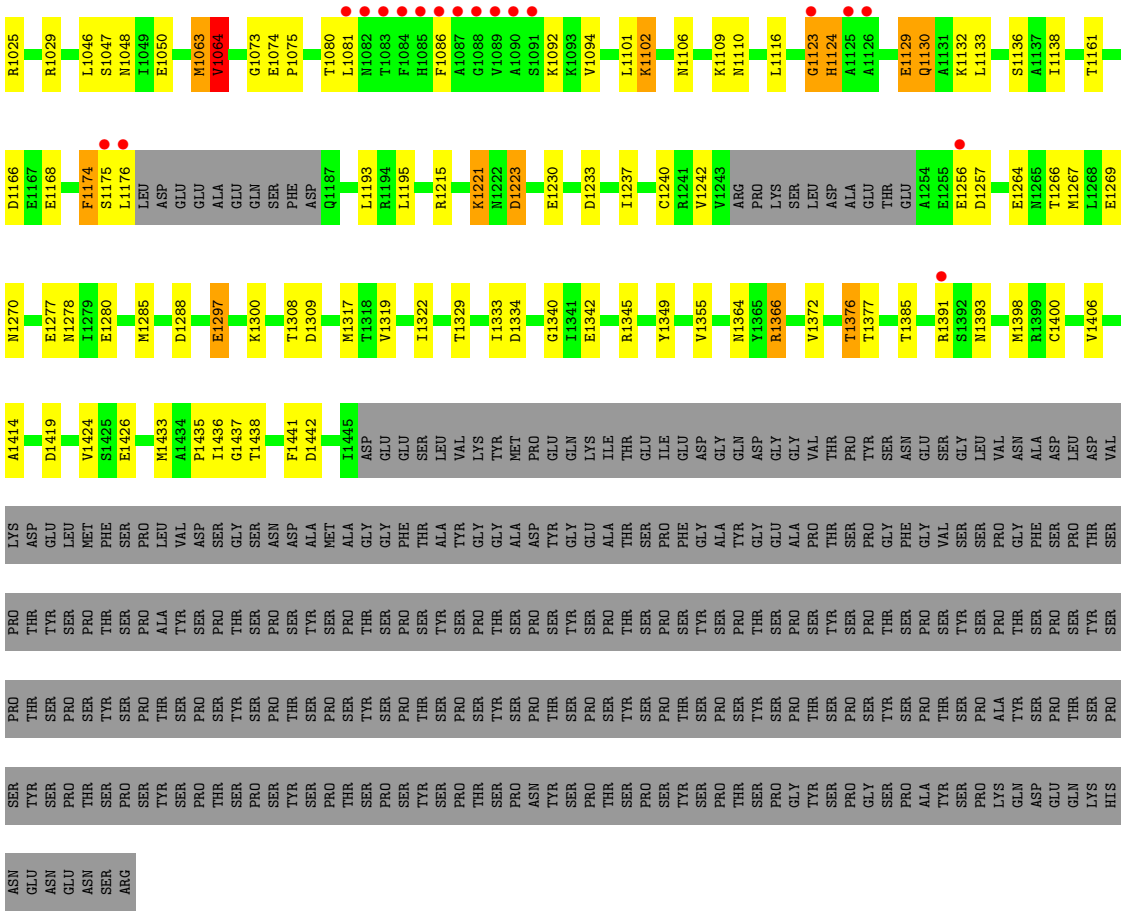
3 Residue-property plots (i)

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

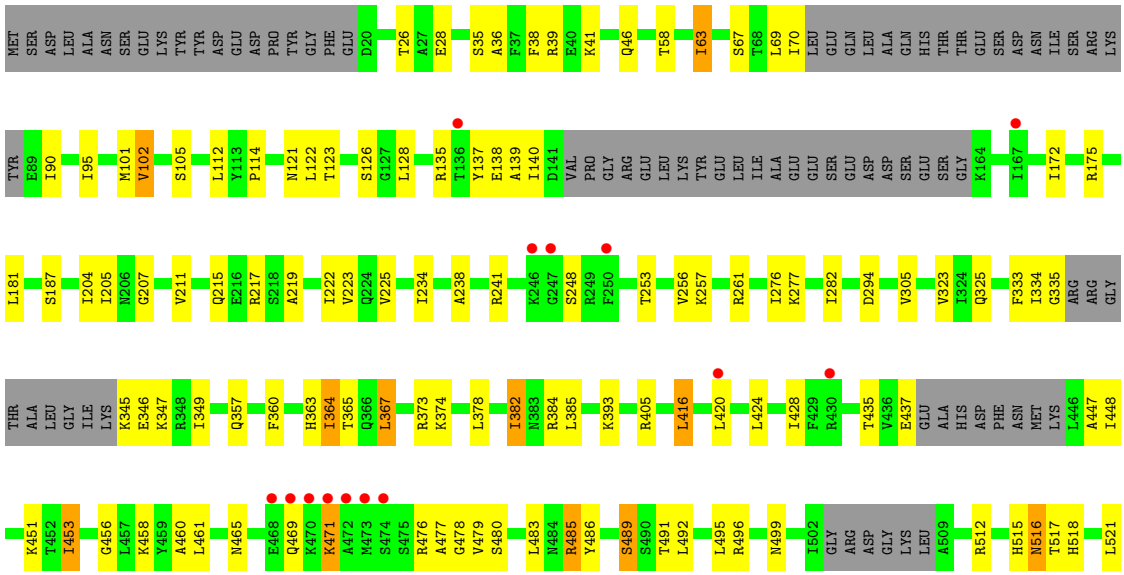
Chain A: 8% 58% 18% 19%

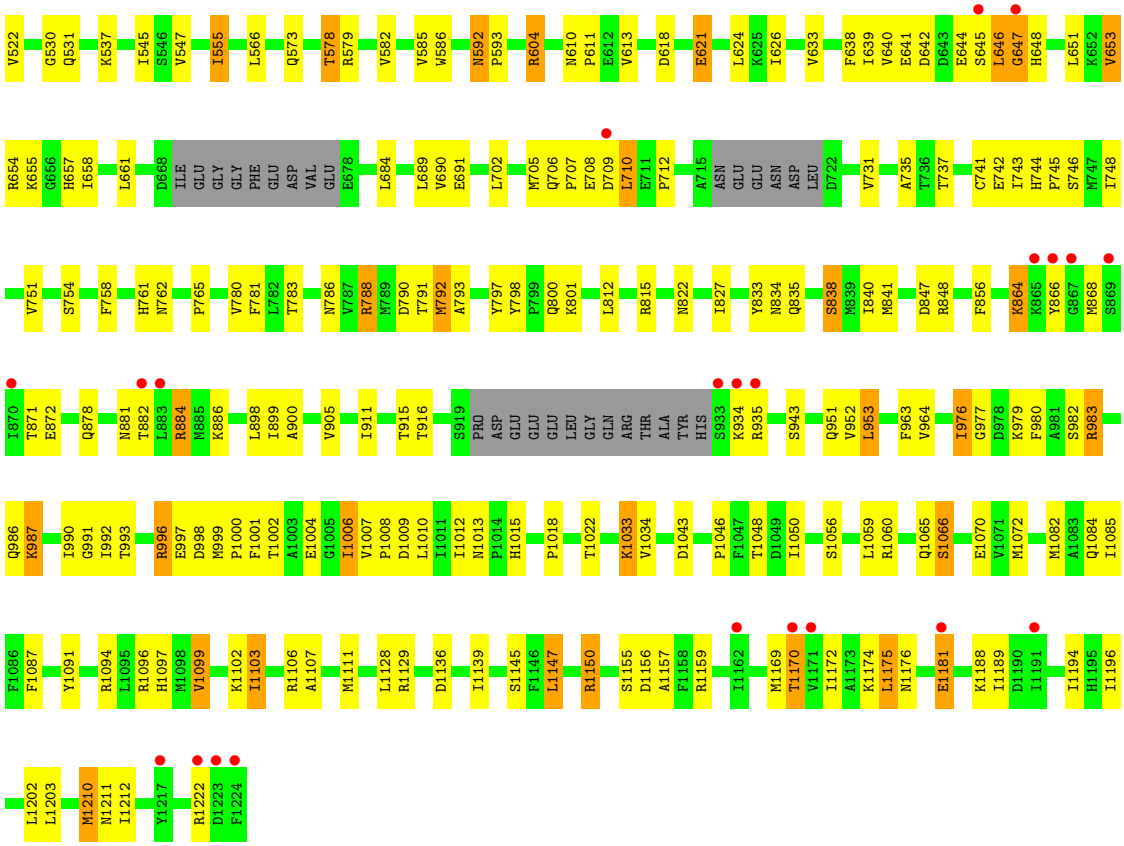




- Molecule 2: DNA-directed RNA polymerase II subunit RPB2

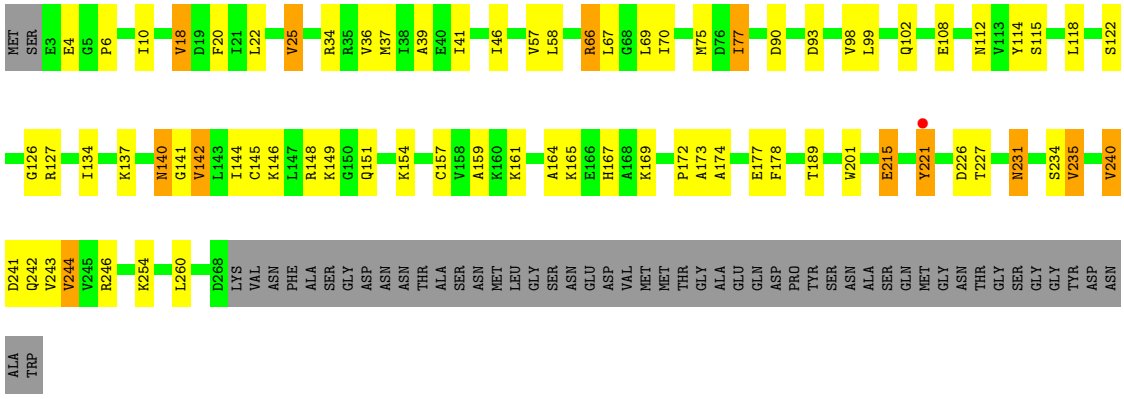
Chain B:





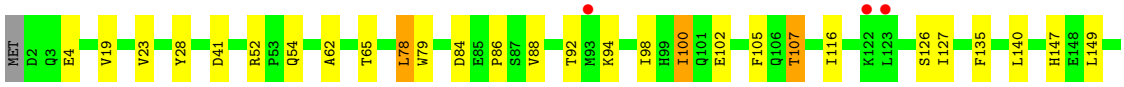
- Molecule 3: DNA-directed RNA polymerase II subunit RPB3

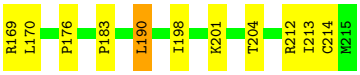
Chain C:



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

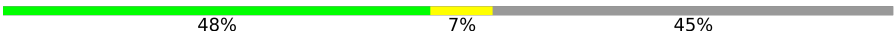
Chain E:

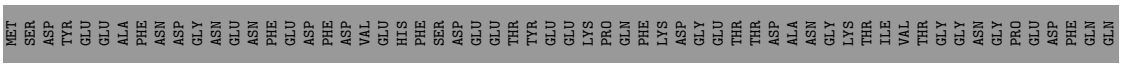


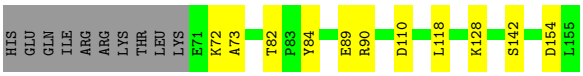


• Molecule 5: DNA-directed RNA polymerases I, II, and III subunit RPABC2

Chain F:

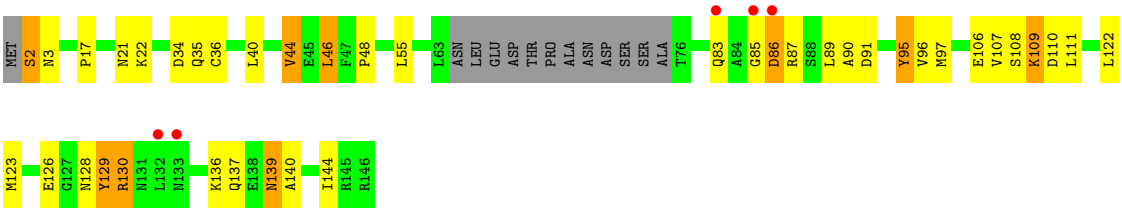






• Molecule 6: DNA-directed RNA polymerases I, II, and III subunit RPABC3

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

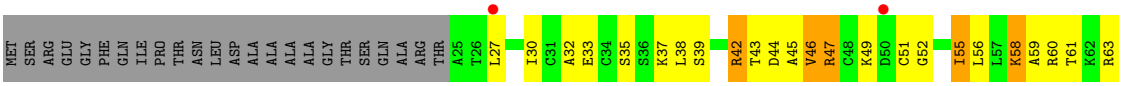
Chain J:

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

Chain K:

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

Chain L:

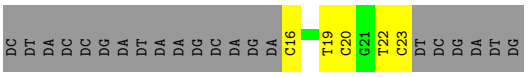


• Molecule 11: RNA (5'-R(*AP*GP*AP*CP*G)-3')

Chain R:

• Molecule 12: DNA (5'-D(*CP*TP*AP*CP*CP*GP*AP*TP*AP*AP*GP*CP*AP*GP*AP*CP*GP*AP*TP*CP*GP*TP*CP*TP*CP*GP*AP*TP*G)-3')

Chain T:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	156.92Å 220.71Å 191.77Å 90.00° 97.48° 90.00°	Depositor
Resolution (Å)	47.72 – 3.13 47.72 – 3.12	Depositor EDS
% Data completeness (in resolution range)	(Not available) (47.72-3.13) 98.7 (47.72-3.12)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.29 (at 3.12Å)	Xtriage
Refinement program	BUSTER-TNT BUSTER 2.8.0, BUSTER 2.8.0	Depositor
R, R_{free}	0.183 , 0.233 0.205 , 0.255	Depositor DCC
R_{free} test set	5668 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	79.3	Xtriage
Anisotropy	0.725	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 84.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	28570	wwPDB-VP
Average B, all atoms (Å^2)	118.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.37% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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	0.47	0/11241	0.75	2/15199 (0.0%)
2	B	0.48	0/9033	0.76	1/12181 (0.0%)
3	C	0.44	0/2133	0.77	0/2891
4	E	0.43	0/1788	0.67	0/2406
5	F	0.44	0/700	0.68	0/945
6	H	0.44	0/1086	0.78	1/1470 (0.1%)
7	I	0.47	0/989	0.81	0/1331
8	J	0.49	0/541	0.83	0/727
9	K	0.43	0/937	0.68	0/1265
10	L	0.51	0/365	0.90	0/485
11	R	0.95	0/120	1.47	0/186
12	T	1.24	0/180	1.92	7/275 (2.5%)
All	All	0.48	0/29113	0.77	11/39361 (0.0%)

There are no bond length outliers.

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
12	T	22	DT	O4'-C1'-N1	7.68	113.38	108.00
1	A	218	ASP	C-N-CA	6.95	139.08	121.70
12	T	23	DC	O4'-C1'-N1	6.54	112.58	108.00
12	T	20	DC	O4'-C1'-N1	6.33	112.43	108.00
12	T	19	DT	O4'-C1'-N1	5.92	112.14	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	11043	0	11133	196	0
2	B	8861	0	8884	152	0
3	C	2095	0	2051	46	0
4	E	1752	0	1776	23	0
5	F	688	0	707	6	0
6	H	1068	0	1040	19	0
7	I	971	0	927	14	0
8	J	532	0	542	22	0
9	K	919	0	929	20	0
10	L	363	0	386	11	0
11	R	107	0	56	0	0
12	T	162	0	91	0	0
13	A	2	0	0	0	0
13	B	1	0	0	0	0
13	C	1	0	0	0	0
13	I	2	0	0	0	0
13	J	1	0	0	0	0
13	L	1	0	0	0	0
14	A	1	0	0	0	0
All	All	28570	0	28522	447	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 8.

The worst 5 of 447 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)
1:A:756:ILE:CG1	1:A:756:ILE:CD1	1.74	1.57
2:B:801:LYS:O	8:J:52:THR:HG23	1.74	0.87
1:A:567:LYS:HB2	1:A:568:PRO:HD2	1.58	0.86
2:B:783:THR:HG22	8:J:63:TYR:HE1	1.41	0.85
1:A:567:LYS:HB3	6:H:96:VAL:H	1.44	0.82

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	1395/1733 (80%)	1242 (89%)	110 (8%)	43 (3%)	4	21
2	B	1096/1224 (90%)	949 (87%)	112 (10%)	35 (3%)	4	20
3	C	264/318 (83%)	239 (90%)	18 (7%)	7 (3%)	5	23
4	E	212/215 (99%)	202 (95%)	8 (4%)	2 (1%)	17	50
5	F	83/155 (54%)	74 (89%)	6 (7%)	3 (4%)	3	18
6	H	129/146 (88%)	107 (83%)	14 (11%)	8 (6%)	1	8
7	I	117/122 (96%)	101 (86%)	15 (13%)	1 (1%)	17	50
8	J	63/70 (90%)	57 (90%)	4 (6%)	2 (3%)	4	20
9	K	112/120 (93%)	105 (94%)	6 (5%)	1 (1%)	17	50
10	L	44/70 (63%)	26 (59%)	10 (23%)	8 (18%)	0	0
All	All	3515/4173 (84%)	3102 (88%)	303 (9%)	110 (3%)	4	21

5 of 110 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	54	ASN
1	A	55	ASP
1	A	325	ILE
1	A	399	HIS
1	A	543	LEU

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	1225/1520 (81%)	1067 (87%)	158 (13%)	4	17
2	B	967/1061 (91%)	863 (89%)	104 (11%)	6	24
3	C	234/274 (85%)	213 (91%)	21 (9%)	9	32
4	E	196/197 (100%)	181 (92%)	15 (8%)	13	39
5	F	75/137 (55%)	73 (97%)	2 (3%)	44	72
6	H	117/128 (91%)	101 (86%)	16 (14%)	3	15
7	I	113/116 (97%)	103 (91%)	10 (9%)	10	33
8	J	60/65 (92%)	48 (80%)	12 (20%)	1	5
9	K	99/102 (97%)	90 (91%)	9 (9%)	9	31
10	L	40/57 (70%)	31 (78%)	9 (22%)	1	3
All	All	3126/3657 (86%)	2770 (89%)	356 (11%)	5	22

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

Mol	Chain	Res	Type
2	B	987	LYS
4	E	54	GLN
2	B	1033	LYS
2	B	1211	ASN
6	H	21	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 52 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	515	HIS
2	B	957	ASN
4	E	147	HIS
2	B	516	ASN
2	B	657	HIS

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
11	R	4/5 (80%)	1 (25%)	0

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
11	R	7	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 9 ligands modelled in this entry, 9 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	1405/1733 (81%)	0.25	138 (9%) 7 3	62, 106, 212, 241	0
2	B	1114/1224 (91%)	0.02	36 (3%) 47 26	60, 101, 178, 207	0
3	C	266/318 (83%)	-0.25	1 (0%) 92 86	72, 98, 138, 170	0
4	E	214/215 (99%)	-0.07	3 (1%) 75 59	79, 137, 185, 197	0
5	F	85/155 (54%)	-0.31	0 100 100	87, 119, 155, 169	0
6	H	133/146 (91%)	0.30	5 (3%) 40 21	100, 141, 173, 184	0
7	I	119/122 (97%)	-0.27	0 100 100	77, 107, 146, 158	0
8	J	65/70 (92%)	-0.17	0 100 100	64, 94, 131, 144	0
9	K	114/120 (95%)	-0.17	0 100 100	64, 101, 127, 143	0
10	L	46/70 (65%)	-0.03	2 (4%) 35 17	79, 135, 168, 176	0
11	R	5/5 (100%)	0.84	1 (20%) 1 0	211, 213, 220, 220	0
12	T	8/29 (27%)	0.42	0 100 100	193, 198, 203, 205	0
All	All	3574/4207 (84%)	0.07	186 (5%) 27 12	60, 107, 200, 241	0

The worst 5 of 186 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1087	ALA	9.2
1	A	1088	GLY	8.7
1	A	317	LYS	7.3
1	A	44	THR	7.2
1	A	1176	LEU	6.5

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(Å^2)	Q<0.9
13	ZN	A	1734	1/1	0.50	0.27	300,300,300,300	0
13	ZN	B	1307	1/1	0.81	0.06	202,202,202,202	0
13	ZN	A	1735	1/1	0.85	0.06	190,190,190,190	0
13	ZN	I	203	1/1	0.97	0.12	108,108,108,108	0
13	ZN	C	319	1/1	0.98	0.12	92,92,92,92	0
14	MG	A	2001	1/1	0.98	0.14	60,60,60,60	0
13	ZN	J	101	1/1	0.99	0.21	85,85,85,85	0
13	ZN	L	105	1/1	0.99	0.12	129,129,129,129	0
13	ZN	I	204	1/1	0.99	0.13	83,83,83,83	0

6.5 Other polymers (i)

There are no such residues in this entry.