wwPDB X-ray Structure Validation Summary Report (i)

Aug 27, 2023 – 10:19 PM EDT

PDB ID	:	3K1F
Title	:	Crystal structure of RNA Polymerase II in complex with TFIIB
Authors	:	Kostrewa, D.; Zeller, M.E.; Armache, K.-J.; Seizl, M.; Leike, K.; Thomm, M.; Cramer, P.
Deposited on	:	2009-09-27
Resolution	:	4.30 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:

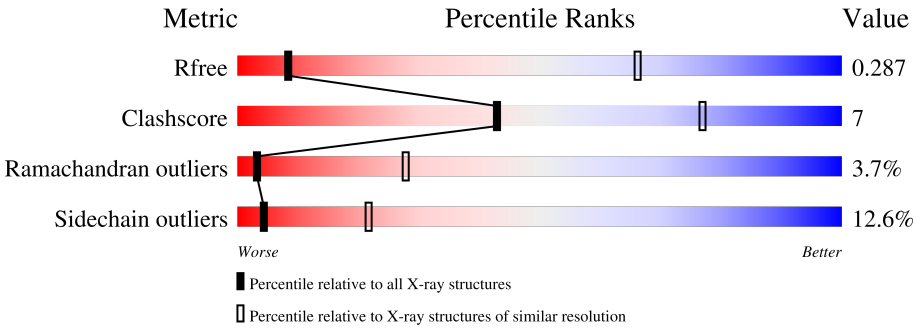
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 4.30 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.



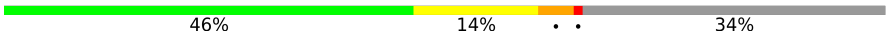
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1014 (4.80-3.80)
Clashscore	141614	1077 (4.80-3.80)
Ramachandran outliers	138981	1029 (4.80-3.80)
Sidechain outliers	138945	1012 (4.80-3.80)

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%

Mol	Chain	Length	Quality of chain
1	A	1733	59% 19% • 18%
2	B	1224	64% 24% • 8%
3	C	318	63% 19% • 16%
4	D	221	58% 21% • 19%
5	E	215	74% 23% •
6	F	155	38% 17% • 44%
7	G	171	71% 24% 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
8	H	146	
9	I	122	
10	J	70	
11	K	120	
12	L	70	
13	M	197	

2 Entry composition (i)

There are 14 unique types of molecules in this entry. The entry contains 32332 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	1416	11143	7021	1949	2111	62	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	1120	8910	5639	1560	1656	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 polymerase II subunit RPB4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	178	1434	887	257	288	2	0	0	0

- Molecule 5 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			
5	E	214	1752	1111	309	321	11	0	0	0

- Molecule 6 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			
6	F	87	705	451	119	132	3	0	0	0

- Molecule 7 is a protein called DNA-directed RNA polymerase II subunit RPB7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	G	171	Total	C	N	O	S	0	0	0
			1340	861	222	249	8			

- Molecule 8 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	H	134	Total	C	N	O	S	0	0	0
			1076	677	182	213	4			

- Molecule 9 is a protein called DNA-directed RNA polymerase II subunit RPB9.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
9	I	119	Total	C	N	O	S	0	0	0
			971	596	179	186	10			

- Molecule 10 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
10	J	65	Total	C	N	O	S	0	0	0
			532	339	93	94	6			

- Molecule 11 is a protein called DNA-directed RNA polymerase II subunit RPB11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
11	K	114	Total	C	N	O	S	0	0	0
			919	590	156	171	2			

- Molecule 12 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
12	L	46	Total	C	N	O	S	0	0	0
			363	224	72	63	4			

- Molecule 13 is a protein called Transcription initiation factor IIB.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
13	M	185	Total	C	N	O	S	0	0	0
			1083	662	207	210	4			

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

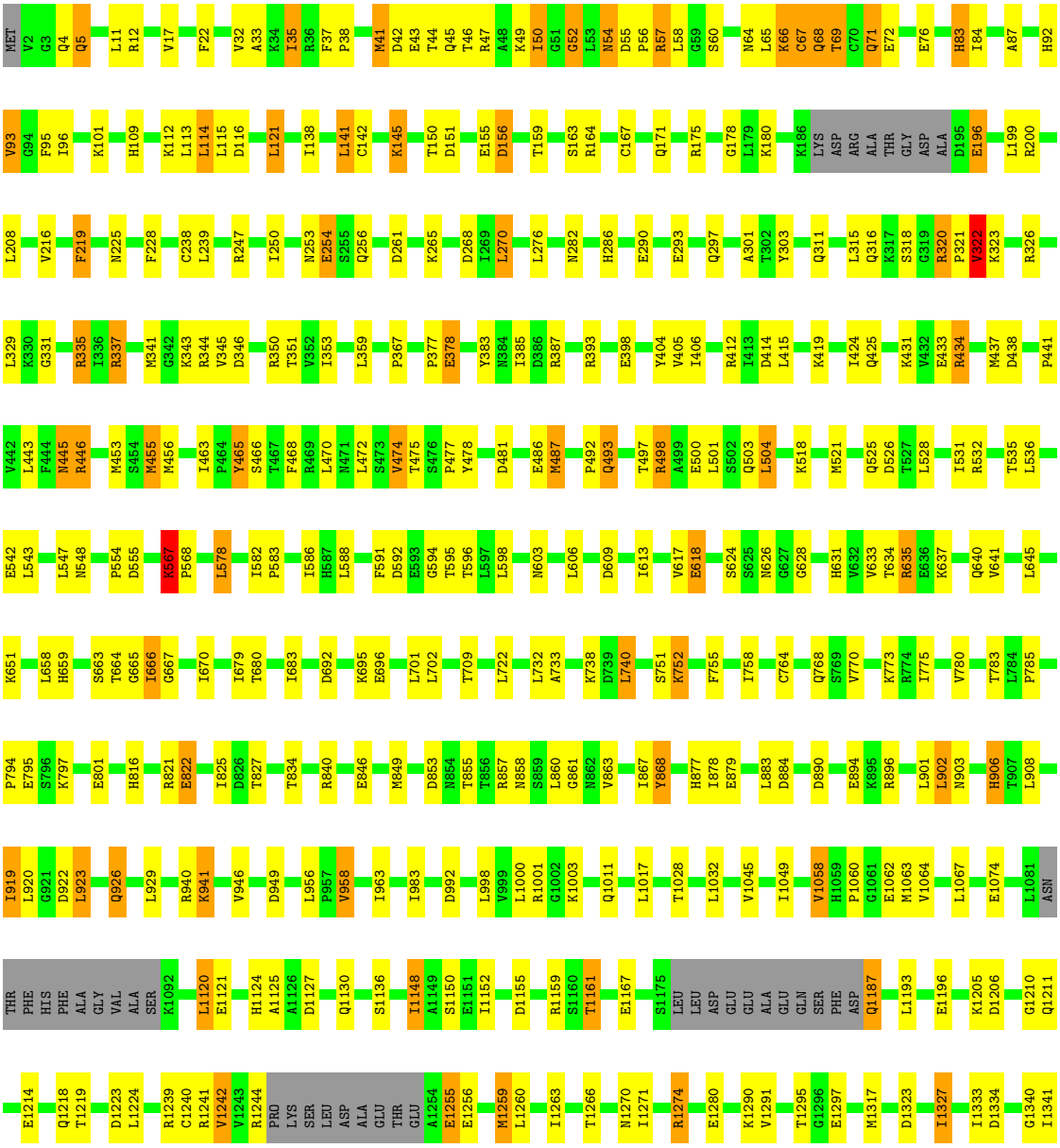
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
14	A	2	Total Zn 2 2	0	0
14	B	1	Total Zn 1 1	0	0
14	C	1	Total Zn 1 1	0	0
14	I	2	Total Zn 2 2	0	0
14	J	1	Total Zn 1 1	0	0
14	L	1	Total Zn 1 1	0	0
14	M	1	Total Zn 1 1	0	0

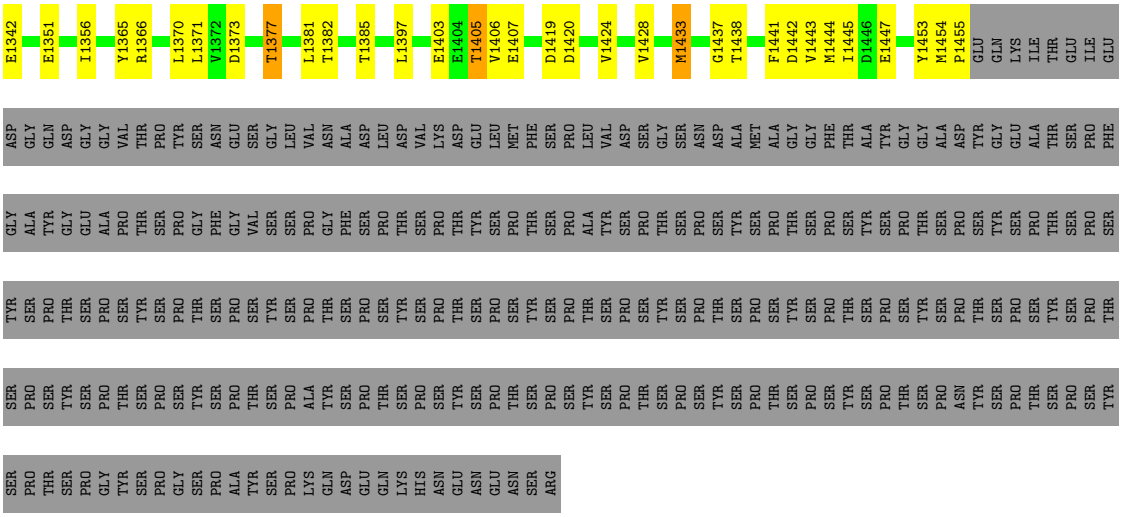
3 Residue-property plots (i)

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. 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. 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

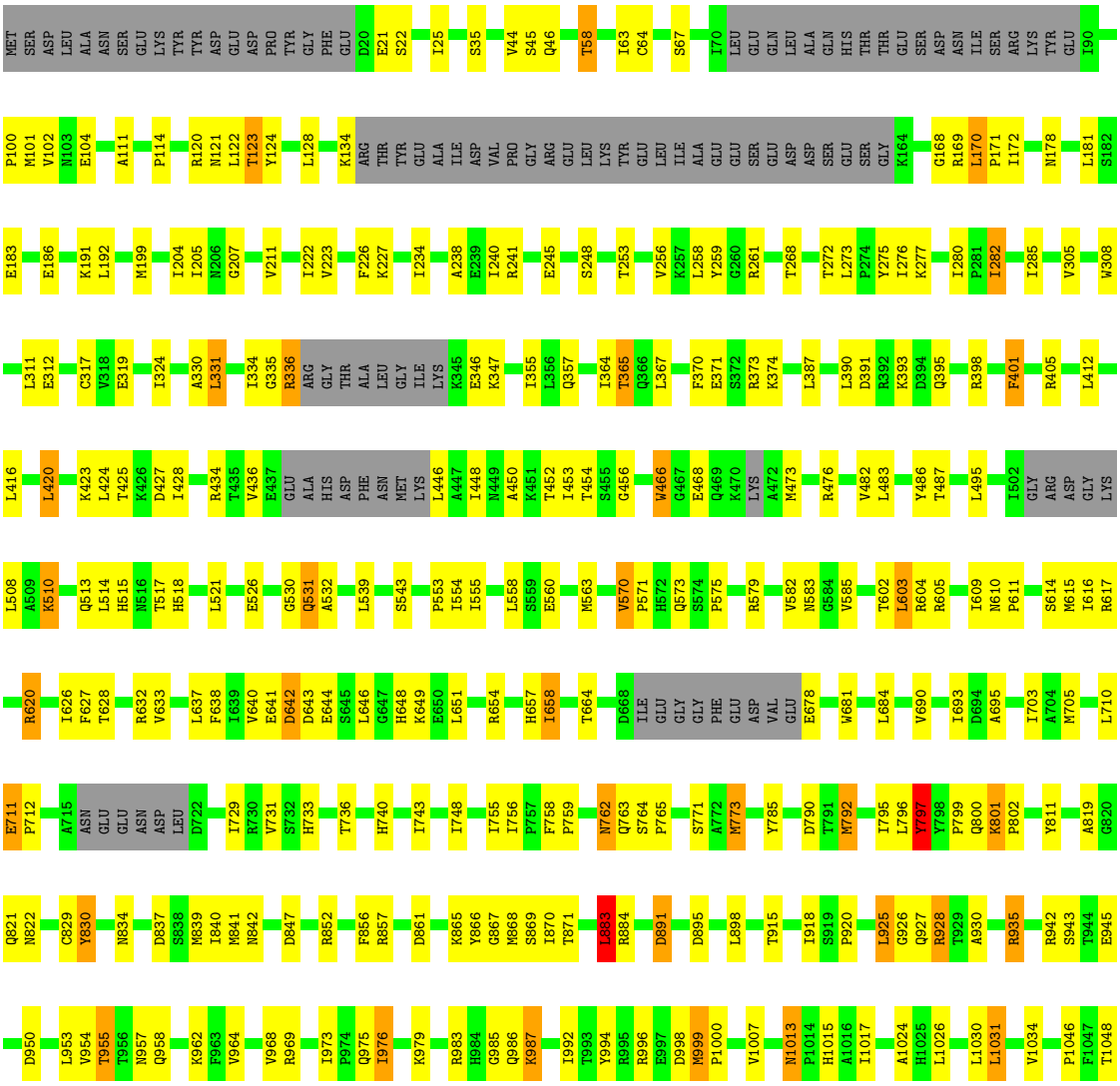
Chain A:

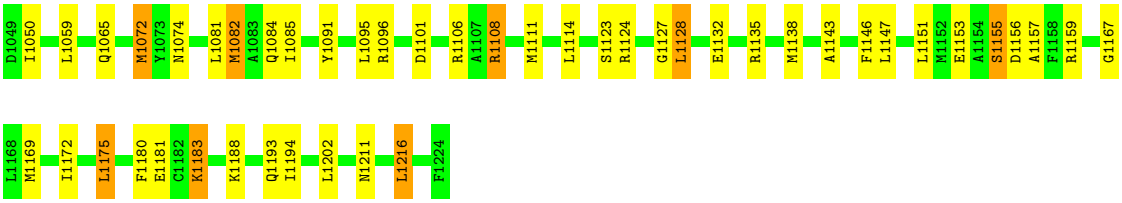




• Molecule 2: DNA-directed RNA polymerase II subunit RPB2

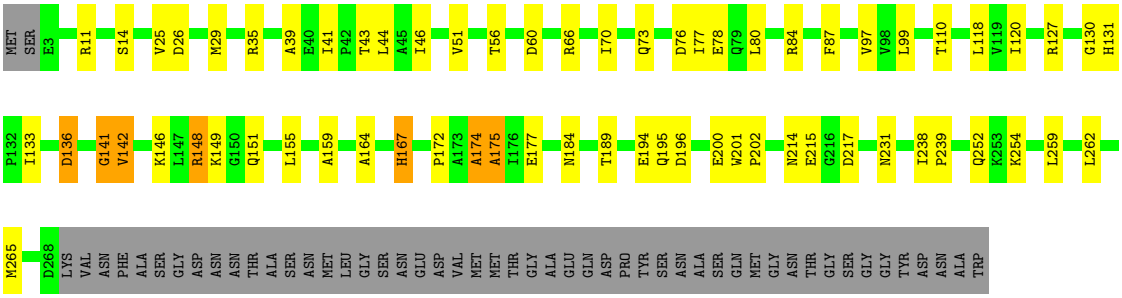
Chain B:





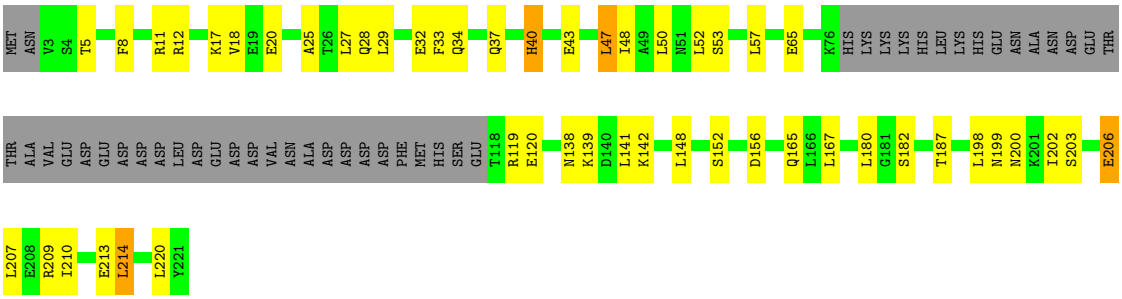
• Molecule 3: DNA-directed RNA polymerase II subunit RPB3

Chain C:



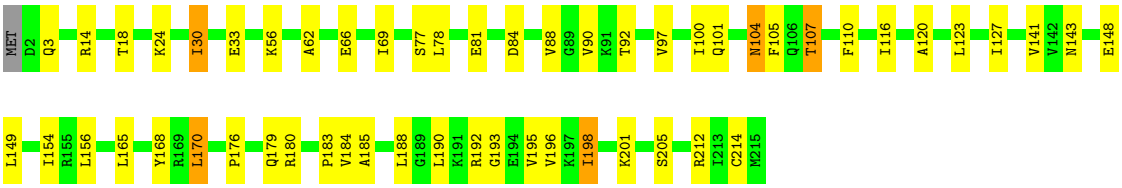
• Molecule 4: DNA-directed RNA polymerase II subunit RPB4

Chain D:



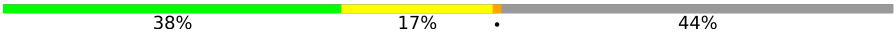
• Molecule 5: DNA-directed RNA polymerases I, II, and III subunit RPABC1

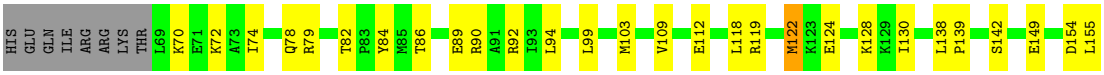
Chain E:



• Molecule 6: DNA-directed RNA polymerases I, II, and III subunit RPABC2

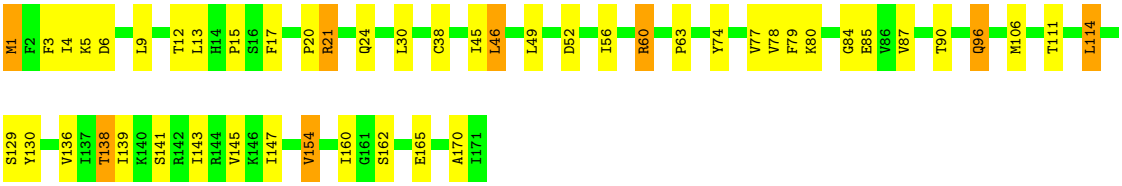
Chain F:





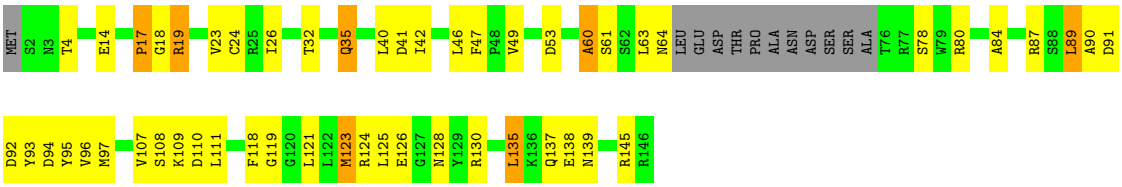
• Molecule 7: DNA-directed RNA polymerase II subunit RPB7

Chain G:



• Molecule 8: DNA-directed RNA polymerases I, II, and III subunit RPABC3

Chain H:

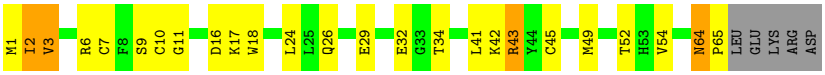


• Molecule 9: DNA-directed RNA polymerase II subunit RPB9

Chain I:

• Molecule 10: DNA-directed RNA polymerases I, II, and III subunit RPABC5

Chain J:

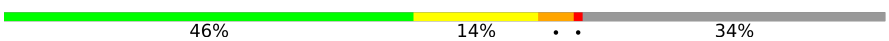


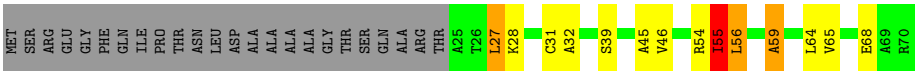
• Molecule 11: DNA-directed RNA polymerase II subunit RPB11

Chain K:

• Molecule 12: DNA-directed RNA polymerases I, II, and III subunit RPABC4

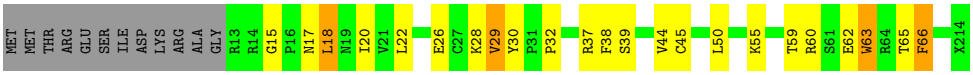
Chain L:





• Molecule 13: Transcription initiation factor IIB

Chain M:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	220.97Å 408.27Å 275.24Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.69 – 4.30 39.69 – 4.30	Depositor EDS
% Data completeness (in resolution range)	99.4 (39.69-4.30) 99.4 (39.69-4.30)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.13	Depositor
$< I/\sigma(I) >$ [1]	2.41 (at 4.28Å)	Xtriage
Refinement program	BUSTER 2.7.0	Depositor
R, R_{free}	0.220 , 0.255 0.251 , 0.287	Depositor DCC
R_{free} test set	2058 reflections (2.45%)	wwPDB-VP
Wilson B-factor (Å^2)	81.9	Xtriage
Anisotropy	0.780	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 111.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.83	EDS
Total number of atoms	32332	wwPDB-VP
Average B, all atoms (Å^2)	124.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.92% 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

5.1 Standard geometry

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	0.44	0/11342	0.72	0/15337
2	B	0.44	0/9084	0.69	0/12251
3	C	0.43	0/2133	0.71	0/2891
4	D	0.42	0/1444	0.68	0/1935
5	E	0.43	0/1788	0.66	0/2406
6	F	0.46	0/717	0.75	0/967
7	G	0.41	0/1368	0.74	0/1844
8	H	0.45	0/1094	0.74	0/1481
9	I	0.41	0/989	0.73	0/1331
10	J	0.45	0/541	0.74	0/727
11	K	0.44	0/937	0.64	0/1265
12	L	0.49	0/365	0.78	0/485
13	M	0.43	0/449	0.72	0/609
All	All	0.44	0/32251	0.71	0/43529

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	11143	0	11217	166	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	8910	0	8926	135	0
3	C	2095	0	2051	31	0
4	D	1434	0	1460	18	0
5	E	1752	0	1776	27	0
6	F	705	0	731	15	0
7	G	1340	0	1357	28	0
8	H	1076	0	1046	21	0
9	I	971	0	927	12	0
10	J	532	0	542	13	0
11	K	919	0	929	15	0
12	L	363	0	386	5	0
13	M	1083	0	578	12	0
14	A	2	0	0	0	0
14	B	1	0	0	0	0
14	C	1	0	0	0	0
14	I	2	0	0	0	0
14	J	1	0	0	0	0
14	L	1	0	0	0	0
14	M	1	0	0	0	0
All	All	32332	0	31926	434	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 7.

The worst 5 of 434 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:567:LYS:HB2	1:A:568:PRO:HD3	1.48	0.94
1:A:465:TYR:HB3	2:B:976:ILE:HG21	1.60	0.83
1:A:768:GLN:HG2	1:A:816:HIS:HA	1.62	0.82
1:A:225:ASN:HD22	1:A:228:PHE:H	1.28	0.81
3:C:142:VAL:H	10:J:16:ASP:HB3	1.46	0.79

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	1406/1733 (81%)	1215 (86%)	137 (10%)	54 (4%)	3 27
2	B	1102/1224 (90%)	938 (85%)	126 (11%)	38 (3%)	3 29
3	C	264/318 (83%)	219 (83%)	37 (14%)	8 (3%)	4 32
4	D	174/221 (79%)	147 (84%)	20 (12%)	7 (4%)	3 26
5	E	212/215 (99%)	192 (91%)	16 (8%)	4 (2%)	8 41
6	F	85/155 (55%)	74 (87%)	11 (13%)	0	100 100
7	G	169/171 (99%)	145 (86%)	21 (12%)	3 (2%)	8 42
8	H	130/146 (89%)	97 (75%)	18 (14%)	15 (12%)	0 6
9	I	117/122 (96%)	96 (82%)	19 (16%)	2 (2%)	9 43
10	J	63/70 (90%)	51 (81%)	7 (11%)	5 (8%)	1 14
11	K	112/120 (93%)	101 (90%)	10 (9%)	1 (1%)	17 56
12	L	44/70 (63%)	34 (77%)	5 (11%)	5 (11%)	0 7
13	M	54/197 (27%)	43 (80%)	8 (15%)	3 (6%)	2 21
All	All	3932/4762 (83%)	3352 (85%)	435 (11%)	145 (4%)	3 28

5 of 145 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	57	ARG
1	A	65	LEU
1	A	66	LYS
1	A	167	CYS
1	A	311	GLN

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	1239/1520 (82%)	1071 (86%)	168 (14%)	3 20
2	B	972/1061 (92%)	847 (87%)	125 (13%)	4 21
3	C	234/274 (85%)	210 (90%)	24 (10%)	7 27
4	D	160/200 (80%)	138 (86%)	22 (14%)	3 20
5	E	196/197 (100%)	181 (92%)	15 (8%)	13 39
6	F	77/137 (56%)	68 (88%)	9 (12%)	5 23
7	G	152/152 (100%)	135 (89%)	17 (11%)	6 25
8	H	118/128 (92%)	102 (86%)	16 (14%)	3 20
9	I	113/116 (97%)	103 (91%)	10 (9%)	10 34
10	J	60/65 (92%)	53 (88%)	7 (12%)	5 23
11	K	99/102 (97%)	87 (88%)	12 (12%)	5 23
12	L	40/57 (70%)	34 (85%)	6 (15%)	3 17
13	M	52/62 (84%)	40 (77%)	12 (23%)	1 5
All	All	3512/4071 (86%)	3069 (87%)	443 (13%)	4 22

5 of 443 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	615	MET
2	B	1132	GLU
13	M	45	CYS
9	I	35	VAL
2	B	729	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 70 such sidechains are listed below:

Mol	Chain	Res	Type
5	E	115	ASN
7	G	71	ASN
8	H	137	GLN
1	A	1128	GLN
1	A	1124	HIS

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 monosaccharides in this entry.

5.6 Ligand geometry

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

There are no such residues in this entry.

5.8 Polymer linkage issues

The following chains have linkage breaks:

Mol	Chain	Number of breaks
13	M	2

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	M	68:ASN	C	84:UNK	N	13.19
1	M	99:UNK	C	102:UNK	N	11.41

6 Fit of model and data

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.