wwPDB X-ray Structure Validation Summary Report (i)

Nov 13, 2023 – 12:37 PM JST

PDB ID	:	5X50
Title	:	RNA Polymerase II from Komagataella Pastoris (Type-2 crystal)
Authors	:	Ehara, H.; Umehara, T.; Sekine, S.; Yokoyama, S.
Deposited on	:	2017-02-14
Resolution	:	4.29 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:

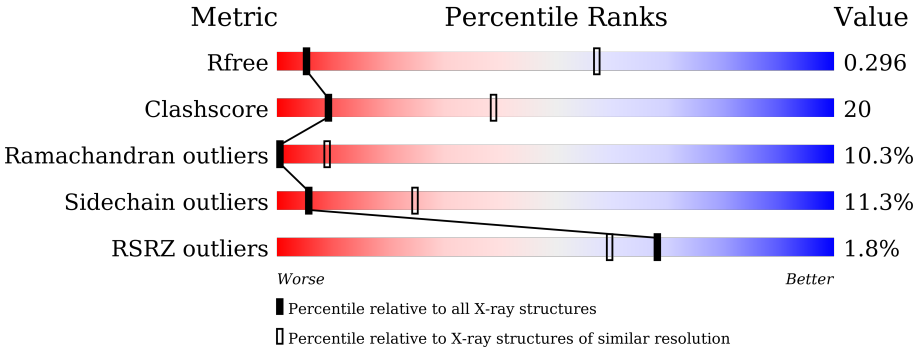
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 4.29 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)
RSRZ outliers	127900	1075 (4.90-3.70)

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	1743	2%, 44%, 30%, 6%, •, 19%
2	B	1227	2%, 46%, 34%, 8%, 12%
3	C	304	48%, 32%, 6%, 13%
4	D	186	%, 46%, 24%, 6%, •, 23%
5	E	214	60%, 34%, 5%, •
6	F	155	%, 26%, 25%, •, 46%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
7	G	171	55% 39% 6%
8	H	145	.% 55% 29% 6% • 10%
9	I	115	3% 56% 34% 6% • •
10	J	72	.% 29% 47% 10% • 11%
11	K	118	.% 49% 39% 7% 5%
12	L	73	4% 29% 27% 7% 37%

2 Entry composition (i)

There are 13 unique types of molecules in this entry. The entry contains 28437 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1409	10410	6594	1819	1943	54	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	1079	8241	5236	1442	1517	46	0	0	0

- Molecule 3 is a protein called RNA polymerase II third largest subunit B44, part of central core.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	263	1852	1181	308	355	8	0	0	0

- Molecule 4 is a protein called RNA polymerase II subunit B32.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	143	1004	647	170	185	2	0	0	0

- Molecule 5 is a protein called RNA polymerase subunit ABC27, common to RNA polymerases I, II, and III.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	E	214	1638	1045	288	297	8	0	0	0

- Molecule 6 is a protein called RNA polymerase subunit ABC23, common to RNA polymerases I, II, and III.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
6	F	84	637	406	109	119	3	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
7	G	171	1187	775	192	217	3	0	0	0

- Molecule 8 is a protein called RNA polymerase subunit ABC14.5, common to RNA polymerases I, II, and III.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
8	H	131	965	616	155	191	3	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
9	I	113	856	537	150	158	11	0	0	0

- Molecule 10 is a protein called RNA polymerase subunit ABC10-beta, common to RNA polymerases I, II, and III.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
10	J	64	500	327	87	80	6	0	0	0

- Molecule 11 is a protein called RNA polymerase II subunit B12.5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
11	K	112	820	535	137	146	2	0	0	0

- Molecule 12 is a protein called RNA polymerase subunit, found in RNA polymerase complexes I, II, and III.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
12	L	46	319	196	64	55	4	0	0	0

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

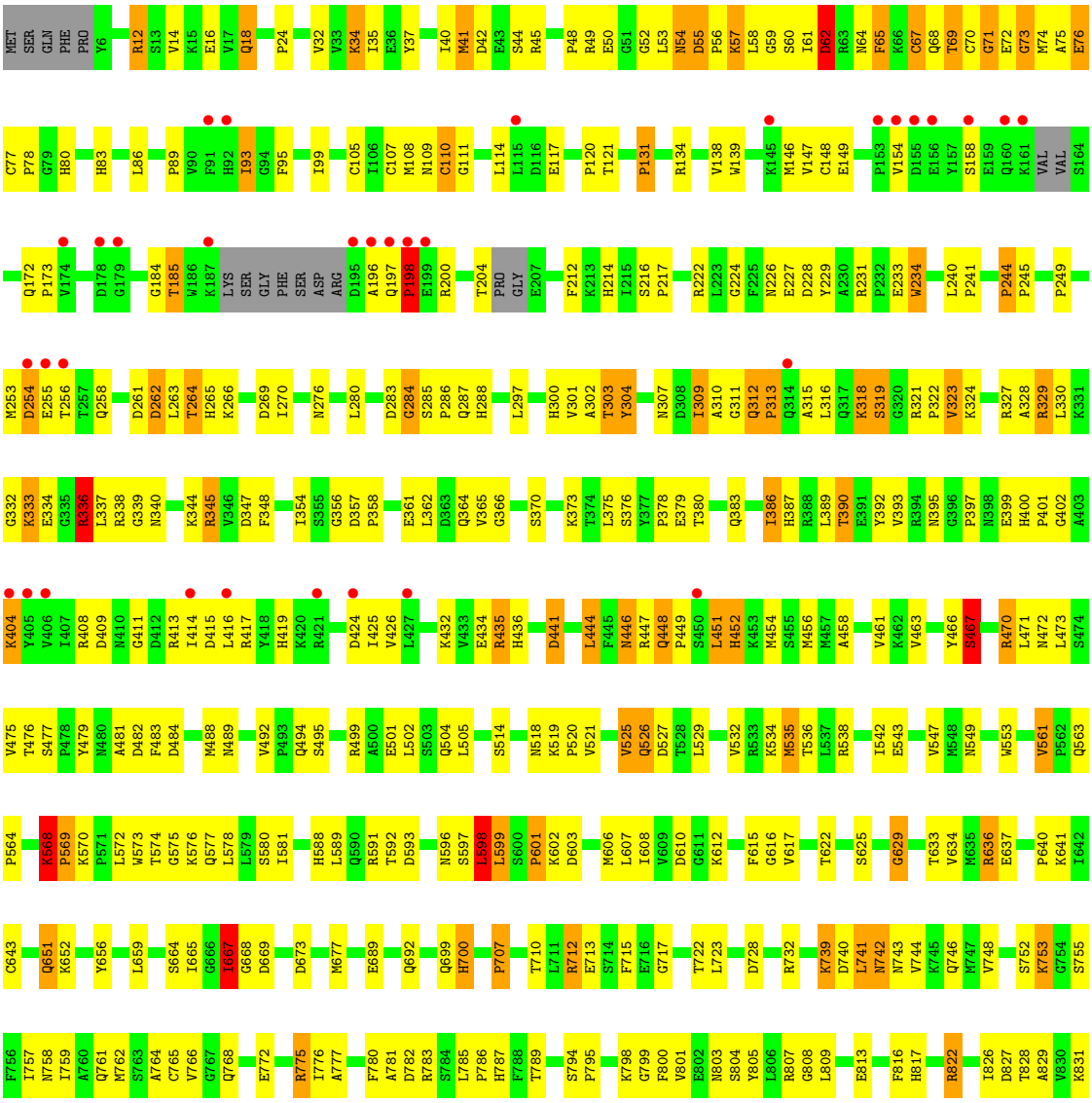
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
13	A	2	2	2	0	0
13	B	1	1	1	0	0
13	C	1	1	1	0	0
13	I	2	2	2	0	0
13	J	1	1	1	0	0
13	L	1	1	1	0	0

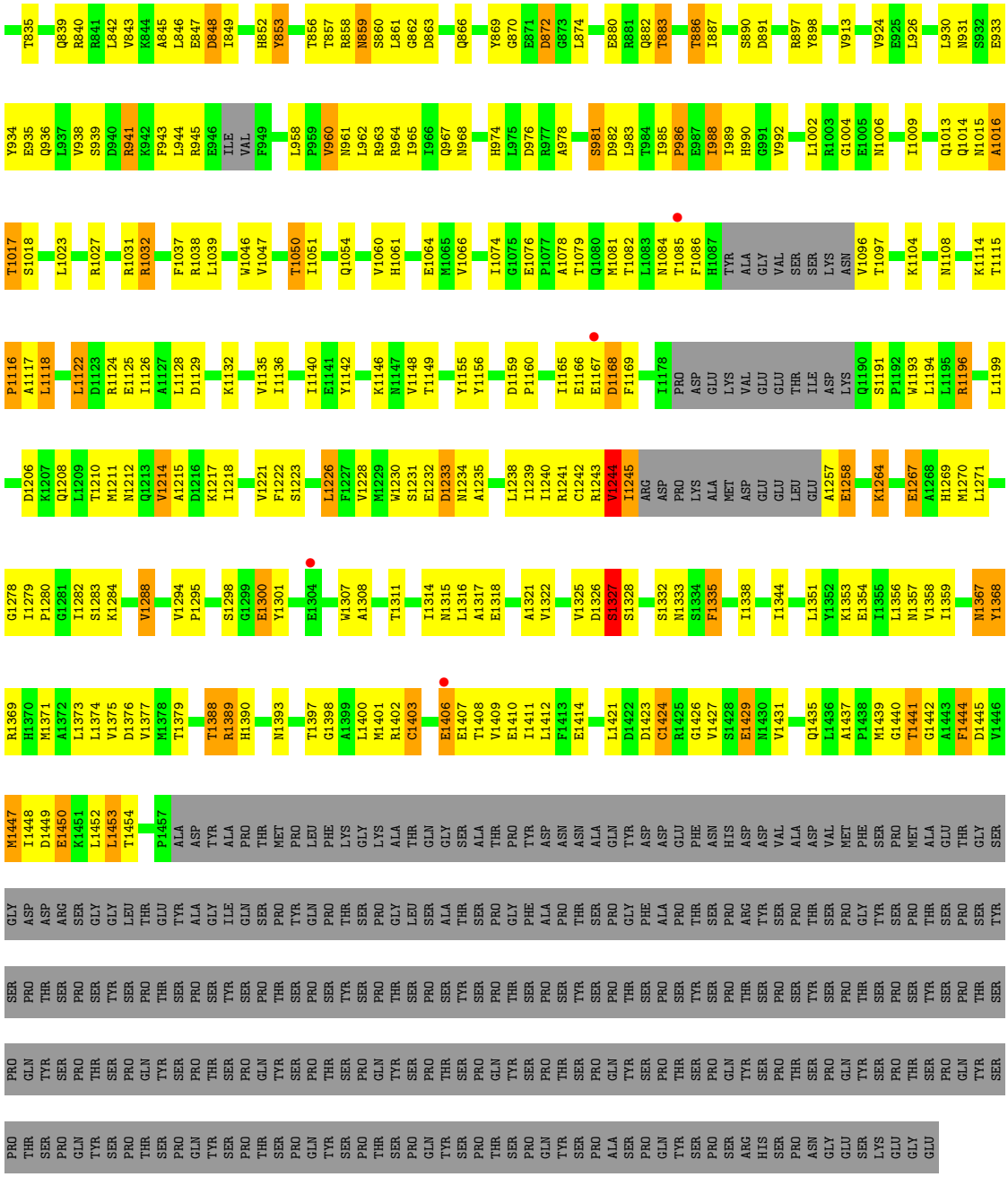
3 Residue-property plots (i)

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 subunit

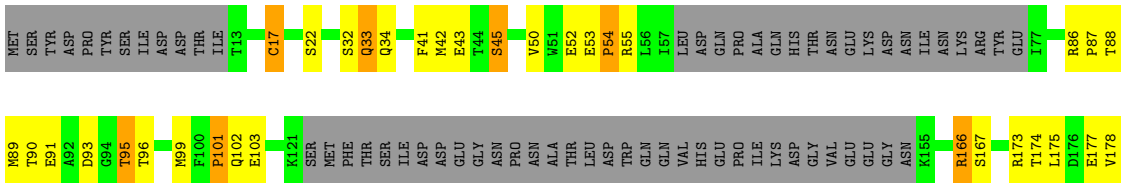
Chain A: 2% 44% 30% 6% 19%

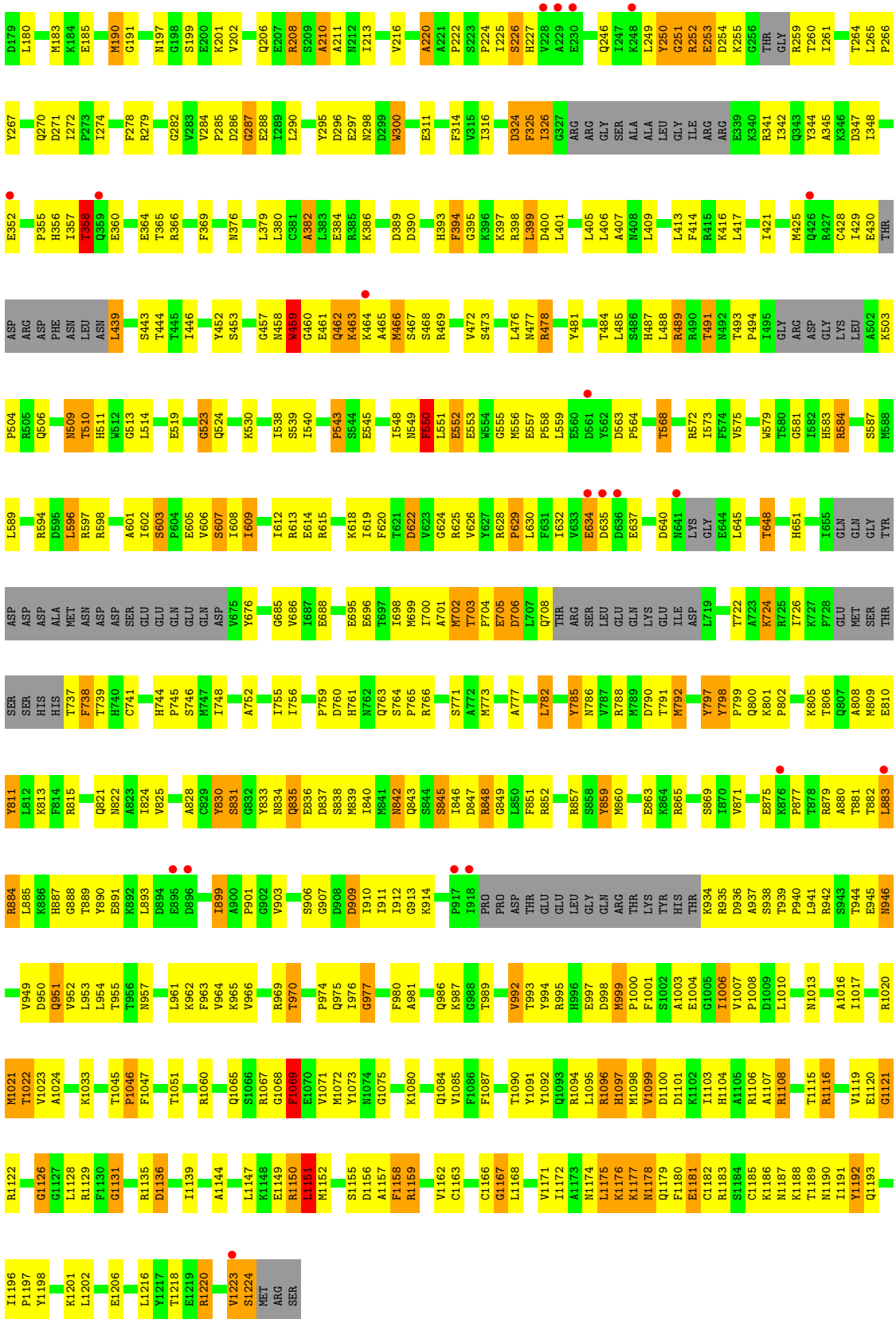




• Molecule 2: DNA-directed RNA polymerase subunit beta

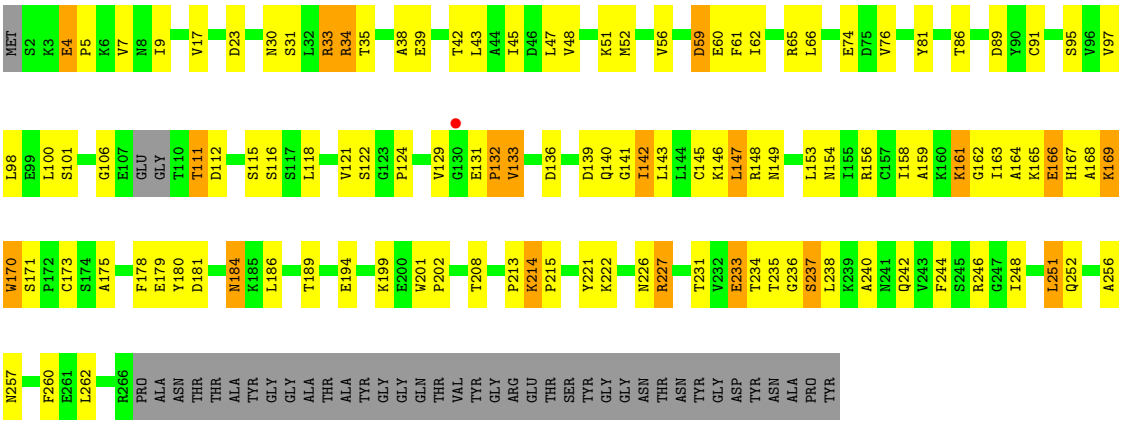
Chain B:





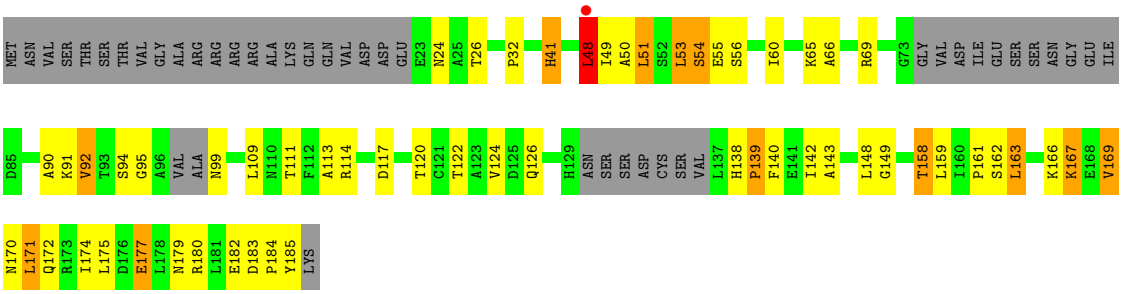
• Molecule 3: RNA polymerase II third largest subunit B44, part of central core

Chain C:



• Molecule 4: RNA polymerase II subunit B32

Chain D:

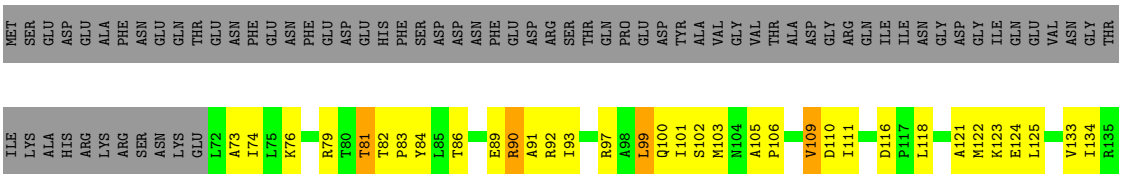


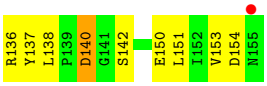
• Molecule 5: RNA polymerase subunit ABC27, common to RNA polymerases I, II, and III

Chain E:

• Molecule 6: RNA polymerase subunit ABC23, common to RNA polymerases I, II, and III

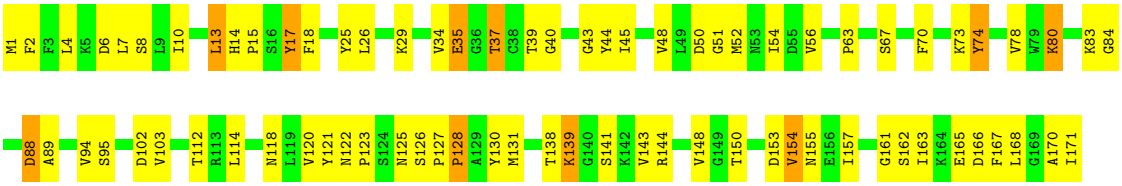
Chain F:





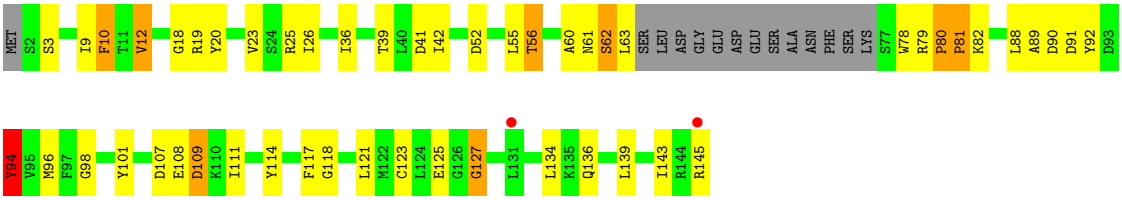
• Molecule 7: RNA polymerase II subunit

Chain G:



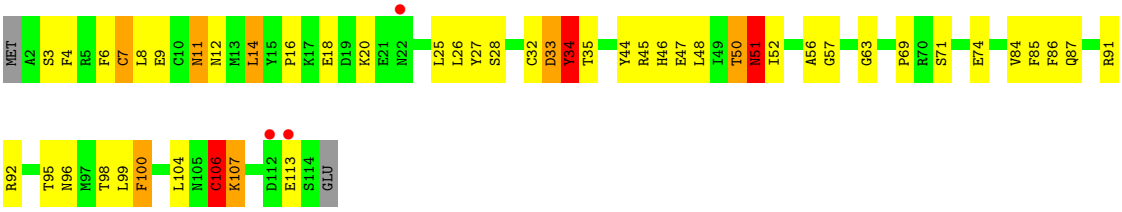
• Molecule 8: RNA polymerase subunit ABC14.5, common to RNA polymerases I, II, and III

Chain H:



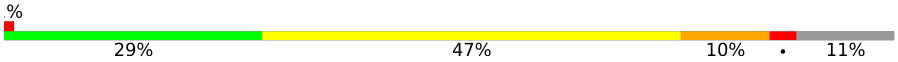
• Molecule 9: DNA-directed RNA polymerase subunit

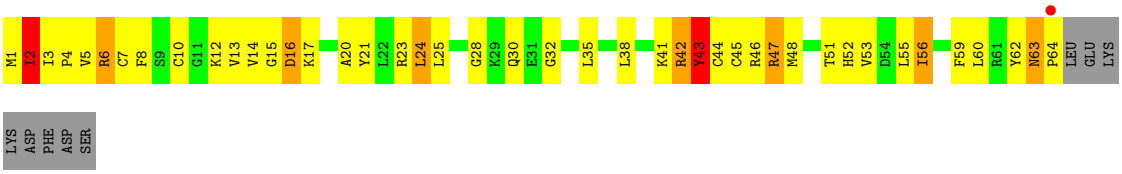
Chain I:



• Molecule 10: RNA polymerase subunit ABC10-beta, common to RNA polymerases I, II, and III

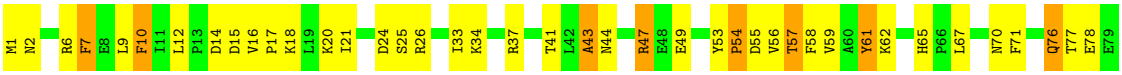
Chain J:

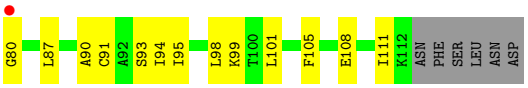




• Molecule 11: RNA polymerase II subunit B12.5

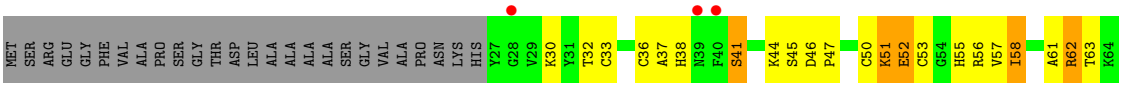
Chain K:

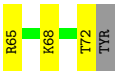




• Molecule 12: RNA polymerase subunit, found in RNA polymerase complexes I, II, and III

Chain L:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, α, β, γ	211.75Å 211.75Å 132.28Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	40.70 – 4.29 49.15 – 4.29	Depositor EDS
% Data completeness (in resolution range)	99.6 (40.70-4.29) 99.6 (49.15-4.29)	Depositor EDS
R_{merge}	0.19	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.51 (at 4.29Å)	Xtriage
Refinement program	PHENIX 1.10.1_2155	Depositor
R, R_{free}	0.235 , 0.296 0.235 , 0.296	Depositor DCC
R_{free} test set	1992 reflections (4.41%)	wwPDB-VP
Wilson B-factor (Å^2)	191.6	Xtriage
Anisotropy	0.254	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.24 , 174.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.37$, $< L^2 > = 0.20$	Xtriage
Estimated twinning fraction	0.124 for -h,-k,l 0.128 for h,-h-k,-l 0.125 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	28437	wwPDB-VP
Average B, all atoms (Å^2)	223.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.87% 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

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	0.27	0/10569	0.71	3/14276 (0.0%)
2	B	0.28	0/8386	0.69	3/11301 (0.0%)
3	C	0.28	0/1877	0.74	2/2543 (0.1%)
4	D	0.25	0/1010	0.69	0/1357
5	E	0.27	0/1668	0.66	0/2245
6	F	0.27	0/646	0.73	1/873 (0.1%)
7	G	0.26	0/1207	0.71	0/1629
8	H	0.28	0/980	0.67	0/1324
9	I	0.28	0/871	0.70	0/1175
10	J	0.33	0/509	0.77	0/684
11	K	0.25	0/836	0.63	0/1131
12	L	0.28	0/321	0.78	0/425
All	All	0.27	0/28880	0.70	9/38963 (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
10	J	0	1

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	1245	ILE	CG1-CB-CG2	-7.11	95.75	111.40
1	A	568	LYS	C-N-CD	6.23	141.49	128.40
2	B	399	LEU	CA-CB-CG	5.98	129.06	115.30
3	C	4	GLU	C-N-CD	5.74	140.45	128.40

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	251	LEU	CA-CB-CG	5.58	128.12	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
10	J	30	GLN	Peptide

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	10410	0	9856	448	0
2	B	8241	0	7943	372	0
3	C	1852	0	1639	79	0
4	D	1004	0	913	37	0
5	E	1638	0	1551	45	0
6	F	637	0	620	33	0
7	G	1187	0	1092	54	0
8	H	965	0	864	35	0
9	I	856	0	774	41	0
10	J	500	0	503	54	0
11	K	820	0	719	36	0
12	L	319	0	288	13	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
All	All	28437	0	26762	1096	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 20.

The worst 5 of 1096 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:285:PRO:HB2	9:I:11:ASN:HB3	1.30	1.13
1:A:256:THR:H	2:B:935:ARG:HH12	1.10	0.97
1:A:67:CYS:SG	1:A:80:HIS:NE2	2.41	0.93
2:B:777:ALA:HA	2:B:1095:LEU:HA	1.56	0.87
1:A:226:ASN:HD22	1:A:229:TYR:H	1.23	0.86

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	1393/1743 (80%)	972 (70%)	291 (21%)	130 (9%)	0 12
2	B	1055/1227 (86%)	723 (68%)	214 (20%)	118 (11%)	0 7
3	C	259/304 (85%)	186 (72%)	50 (19%)	23 (9%)	1 12
4	D	135/186 (73%)	98 (73%)	23 (17%)	14 (10%)	0 9
5	E	212/214 (99%)	150 (71%)	44 (21%)	18 (8%)	1 13
6	F	82/155 (53%)	58 (71%)	14 (17%)	10 (12%)	0 5
7	G	169/171 (99%)	129 (76%)	27 (16%)	13 (8%)	1 15
8	H	127/145 (88%)	95 (75%)	17 (13%)	15 (12%)	0 6
9	I	111/115 (96%)	71 (64%)	26 (23%)	14 (13%)	0 5
10	J	62/72 (86%)	37 (60%)	15 (24%)	10 (16%)	0 3
11	K	110/118 (93%)	83 (76%)	16 (14%)	11 (10%)	0 10
12	L	44/73 (60%)	20 (46%)	11 (25%)	13 (30%)	0 0
All	All	3759/4523 (83%)	2622 (70%)	748 (20%)	389 (10%)	0 9

5 of 389 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	42	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	48	PRO
1	A	54	ASN
1	A	57	LYS
1	A	67	CYS

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	988/1528 (65%)	885 (90%)	103 (10%)	7	27
2	B	822/1077 (76%)	731 (89%)	91 (11%)	6	25
3	C	155/264 (59%)	130 (84%)	25 (16%)	2	15
4	D	78/160 (49%)	65 (83%)	13 (17%)	2	14
5	E	155/197 (79%)	140 (90%)	15 (10%)	8	29
6	F	60/137 (44%)	54 (90%)	6 (10%)	7	28
7	G	102/148 (69%)	91 (89%)	11 (11%)	6	26
8	H	91/130 (70%)	84 (92%)	7 (8%)	13	39
9	I	82/109 (75%)	72 (88%)	10 (12%)	5	23
10	J	48/66 (73%)	40 (83%)	8 (17%)	2	14
11	K	67/109 (62%)	56 (84%)	11 (16%)	2	14
12	L	26/58 (45%)	25 (96%)	1 (4%)	33	58
All	All	2674/3983 (67%)	2373 (89%)	301 (11%)	6	25

5 of 301 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	185	TYR
10	J	43	TYR
5	E	72	SER
7	G	128	PRO
12	L	65	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 57 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	509	ASN
11	K	76	GLN
2	B	975	GLN
11	K	65	HIS
7	G	53	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	1409/1743 (80%)	-0.18	37 (2%)	56	46	110, 206, 318, 455	0
2	B	1079/1227 (87%)	-0.19	20 (1%)	66	58	115, 226, 316, 418	0
3	C	263/304 (86%)	-0.36	1 (0%)	92	87	148, 216, 311, 349	0
4	D	143/186 (76%)	-0.11	1 (0%)	87	82	167, 263, 321, 343	0
5	E	214/214 (100%)	-0.46	0	100	100	122, 197, 260, 300	0
6	F	84/155 (54%)	-0.46	1 (1%)	79	70	160, 222, 270, 312	0
7	G	171/171 (100%)	-0.35	0	100	100	156, 213, 267, 311	0
8	H	131/145 (90%)	-0.04	2 (1%)	73	64	161, 214, 298, 344	0
9	I	113/115 (98%)	0.02	3 (2%)	54	44	211, 276, 343, 386	0
10	J	64/72 (88%)	-0.43	1 (1%)	72	62	154, 214, 272, 288	0
11	K	112/118 (94%)	-0.43	1 (0%)	84	77	149, 202, 274, 293	0
12	L	46/73 (63%)	0.09	3 (6%)	18	15	223, 275, 329, 350	0
All	All	3829/4523 (84%)	-0.22	70 (1%)	68	60	110, 219, 314, 455	0

The worst 5 of 70 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	254	ASP	5.3
1	A	155	ASP	5.3
1	A	156	GLU	5.2
1	A	195	ASP	4.6
9	I	113	GLU	4.6

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	1801	1/1	0.82	0.14	299,299,299,299	0
13	ZN	I	201	1/1	0.88	0.08	239,239,239,239	0
13	ZN	C	401	1/1	0.89	0.05	283,283,283,283	0
13	ZN	L	101	1/1	0.89	0.07	318,318,318,318	0
13	ZN	I	202	1/1	0.91	0.07	242,242,242,242	0
13	ZN	J	101	1/1	0.94	0.24	158,158,158,158	0
13	ZN	A	1802	1/1	0.98	0.13	270,270,270,270	0
13	ZN	B	1301	1/1	0.99	0.12	248,248,248,248	0

6.5 Other polymers (i)

There are no such residues in this entry.