wwPDB/EMDataBank EM Map/Model Validation Summary Report (i)

Jan 19, 2019 – 11:06 AM EST

PDB ID	:	6EEC
EMDB ID:	:	EMD-9041
Title	:	Mycobacterium tuberculosis RNAP promoter unwinding intermediate complex with RbpA/CarD and AP3 promoter captured by Corallopyronin
Authors	:	Darst, S.A.; Campbell, E.A.; Boyaci Selcuk, H.; Chen, J.
Deposited on	:	2018-08-13
Resolution	:	3.55 Å(reported)

This is a wwPDB/EMDataBank EM Map/Model Validation Summary Report for a publicly released PDB/EMDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the (i) symbol.

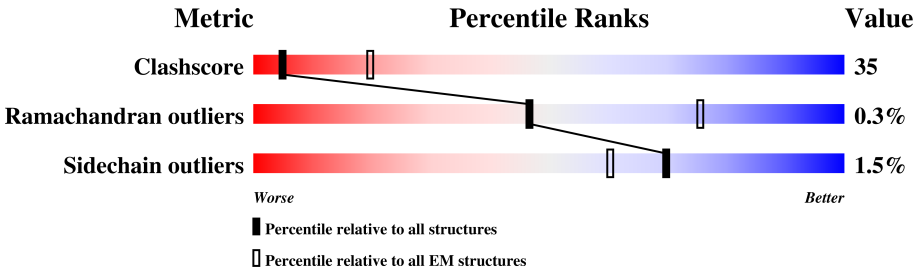
MolProbity	:	4.02b-467
Mogul	:	1.7.3 (157068), CSD as539be (2018)
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20031633

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 3.55 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)	EM structures (#Entries)
Clashscore	136327	1886
Ramachandran outliers	132723	1663
Sidechain outliers	132532	1531

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. 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	347	34% 30% 35%
1	B	347	31% 37% 32%
2	C	1179	44% 49% • 6%
3	D	1326	45% 49% • 5%
4	E	110	35% 41% 25%
5	F	531	31% 28% • 40%
6	J	111	55% 40% • •
7	O	90	10% 61% • 28%
8	P	90	12% 58% 30%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
9	M	162	43% 55%

2 Entry composition (i)

There are 12 unique types of molecules in this entry. The entry contains 29936 atoms, of which 40 are hydrogens and 0 are deuteriums.

In the tables below, 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 alpha.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	225	Total	C	N	O	S	0	0
			1716	1080	296	338	2		
1	B	237	Total	C	N	O	S	0	0
			1759	1112	298	346	3		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
2	C	1111	Total	C	N	O	S	0	0
			8593	5381	1507	1666	39		

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	1179	LEU	-	expression tag	UNP V9Z879
C	1180	ALA	-	expression tag	UNP V9Z879
C	1181	ARG	-	expression tag	UNP V9Z879
C	1182	HIS	-	expression tag	UNP V9Z879
C	1183	GLY	-	expression tag	UNP V9Z879
C	1184	GLY	-	expression tag	UNP V9Z879
C	1185	SER	-	expression tag	UNP V9Z879

- Molecule 3 is a protein called DNA-directed RNA polymerase subunit beta'.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	D	1266	Total	C	N	O	S	0	0
			9873	6184	1794	1853	42		

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	-1	GLY	-	expression tag	UNP A5U053
D	0	ALA	-	expression tag	UNP A5U053

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	1317	HIS	-	expression tag	UNP A5U053
D	1318	HIS	-	expression tag	UNP A5U053
D	1319	HIS	-	expression tag	UNP A5U053
D	1320	HIS	-	expression tag	UNP A5U053
D	1321	HIS	-	expression tag	UNP A5U053
D	1322	HIS	-	expression tag	UNP A5U053
D	1323	HIS	-	expression tag	UNP A5U053
D	1324	HIS	-	expression tag	UNP A5U053

- Molecule 4 is a protein called DNA-directed RNA polymerase subunit omega.

Mol	Chain	Residues	Atoms				AltConf	Trace
4	E	83	Total	C	N	O	0	0
			649	414	108	127		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	1	GLY	-	expression tag	UNP A0A0T9N9K3

- Molecule 5 is a protein called RNA polymerase sigma factor SigA.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	F	319	Total	C	N	O	S	0	0
			2518	1571	456	482	9		

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	-2	GLY	-	expression tag	UNP P9WGI0
F	-1	PRO	-	expression tag	UNP P9WGI0
F	0	HIS	-	expression tag	UNP P9WGI0

- Molecule 6 is a protein called RNA polymerase-binding protein RbpA.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	J	108	Total	C	N	O	S	0	0
			881	543	168	167	3		

- Molecule 7 is a DNA chain called DNA (65-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
7	O	65	Total	C	N	O	P	0	0
			1336	633	243	395	65		

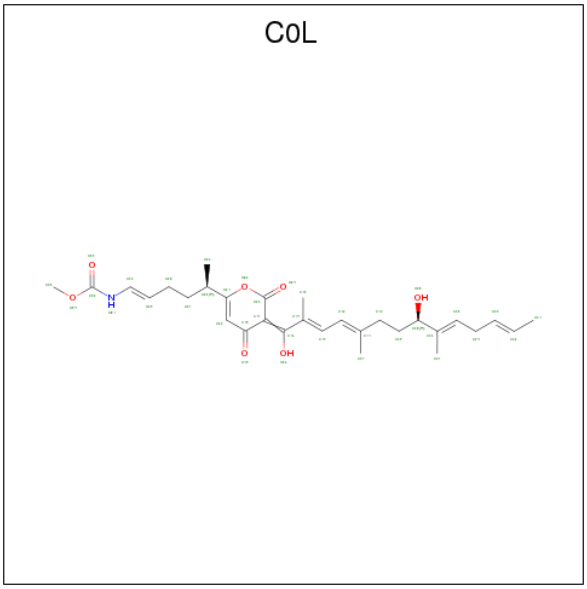
- Molecule 8 is a DNA chain called DNA (63-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
8	P	63	Total	C	N	O	P	0	0
			1289	610	242	374	63		

- Molecule 9 is a protein called RNA polymerase-binding transcription factor CarD.

Mol	Chain	Residues	Atoms					AltConf	Trace
9	M	159	Total	C	N	O	S	0	0
			1241	777	224	239	1		

- Molecule 10 is methyl [(1E,5R)-5-{(3E)-3-[(2E,4E,8R,9E,12E)-1,8-dihydroxy-2,5,9-trimethyltetradeca-2,4,9,12-tetraen-1-ylidene]-2,4-dioxo-3,4-dihydro-2H-pyran-6-yl}hex-1-en-1-yl]carbamate (three-letter code: C0L) (formula: $C_{30}H_{41}NO_7$).



Mol	Chain	Residues	Atoms					AltConf
10	C	1	Total	C	H	N	O	0
			78	30	40	1	7	

- Molecule 11 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		AltConf
11	D	2	Total	Zn	0
			2	2	

- Molecule 12 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

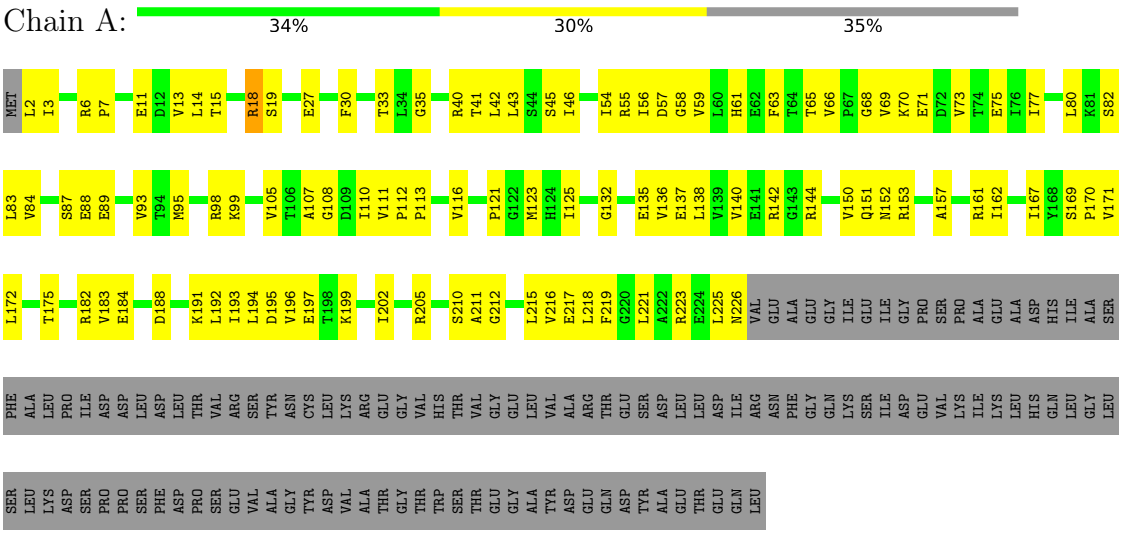
Mol	Chain	Residues	Atoms		AltConf
12	D	1	Total	Mg	0
			1	1	

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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.

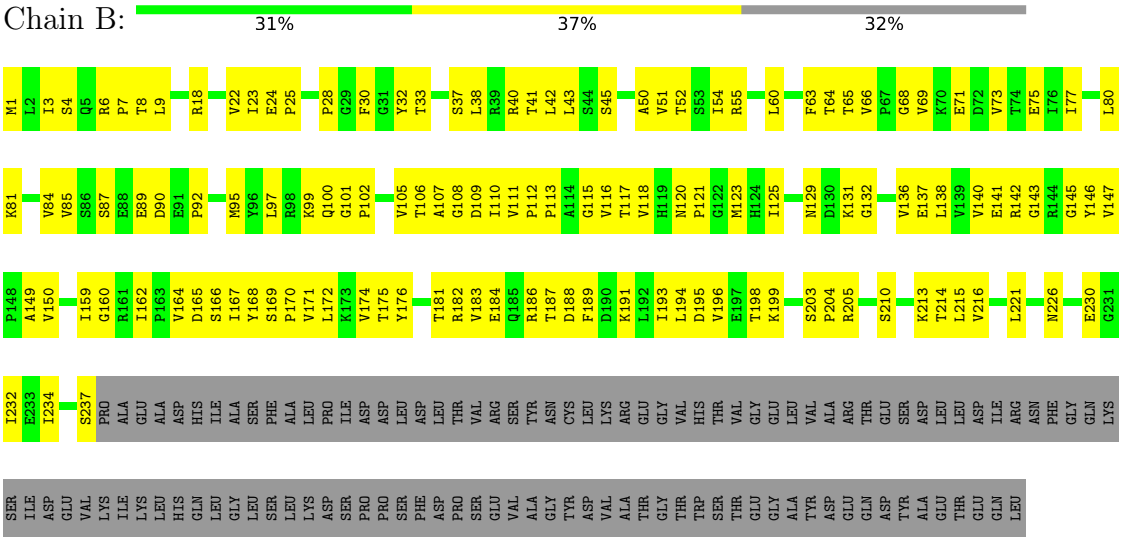
• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain A:



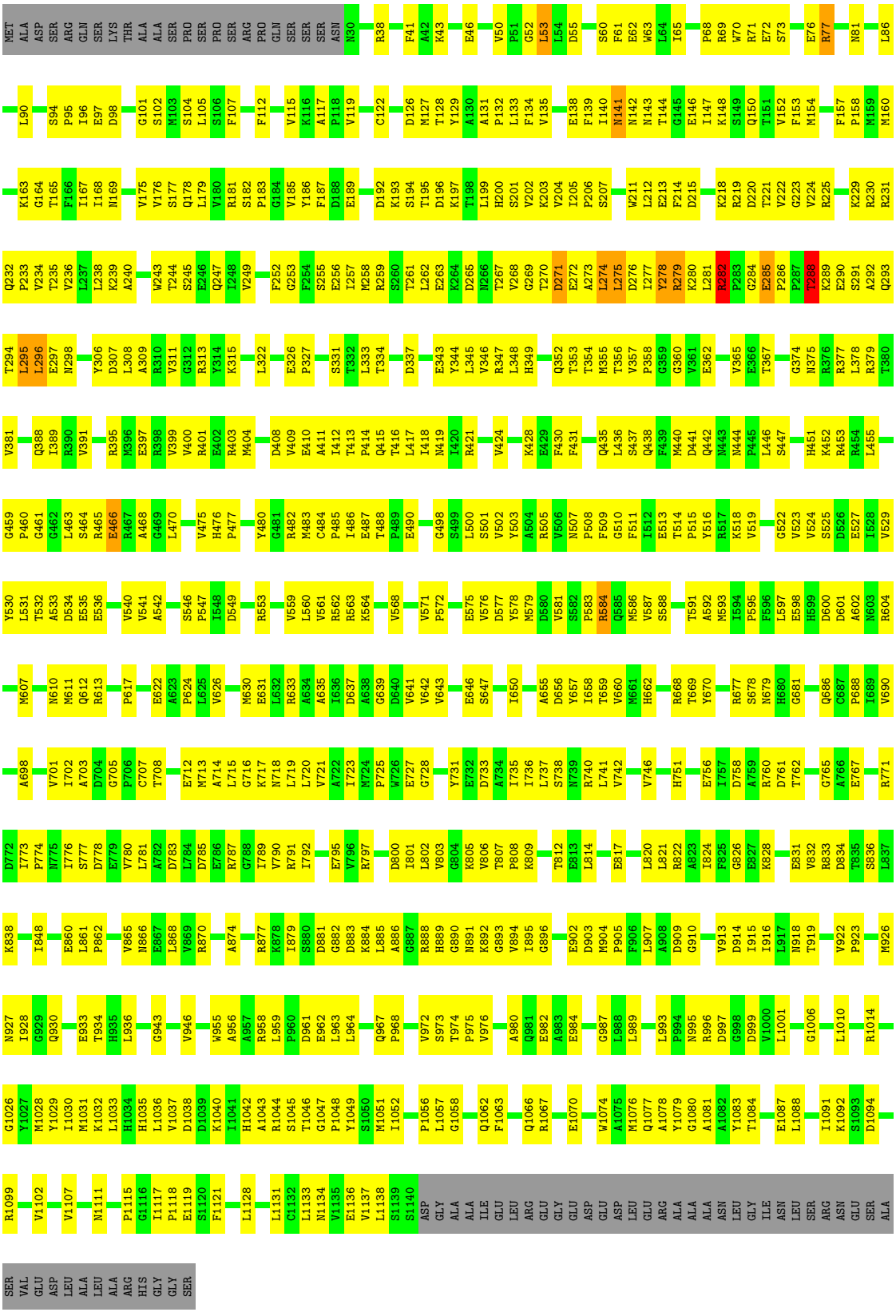
• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain B:



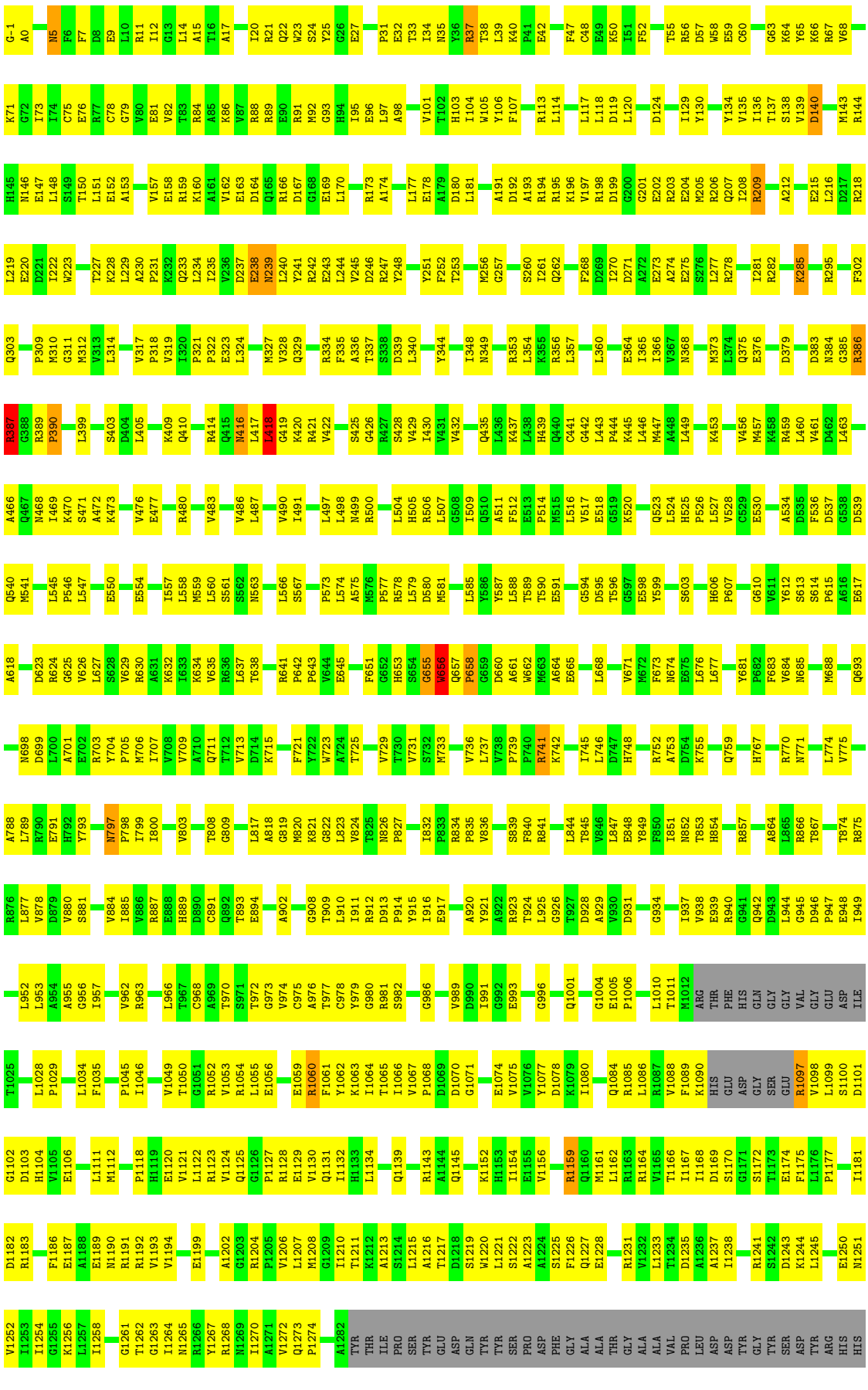
• Molecule 2: DNA-directed RNA polymerase subunit beta

Chain C:



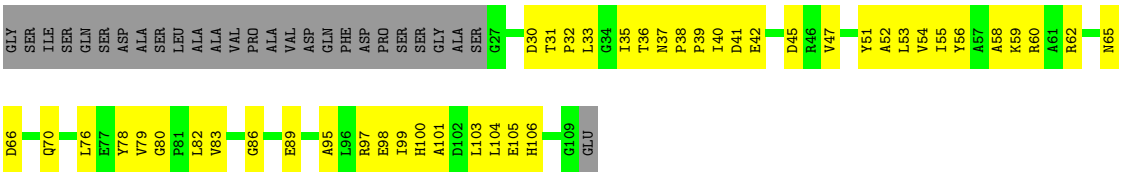
• Molecule 3: DNA-directed RNA polymerase subunit beta'

Chain D: 45% 49% • 5%



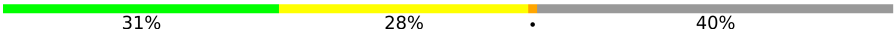
• Molecule 4: DNA-directed RNA polymerase subunit omega

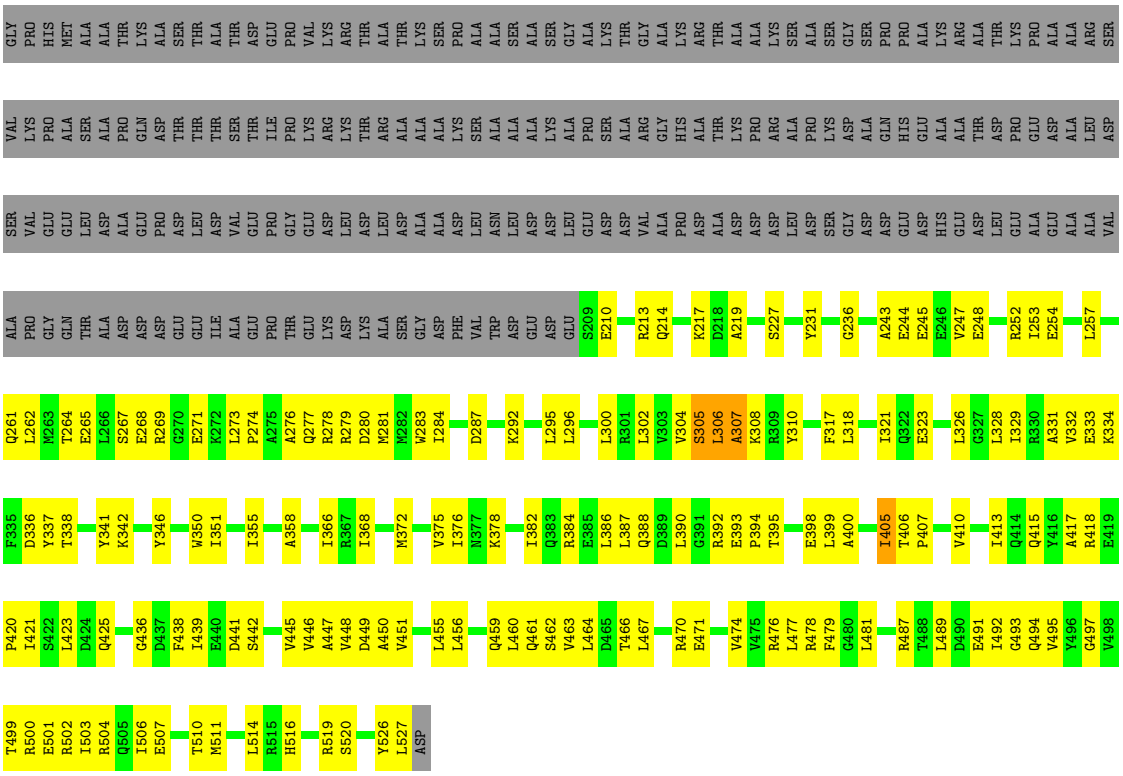
Chain E:



• Molecule 5: RNA polymerase sigma factor SigA

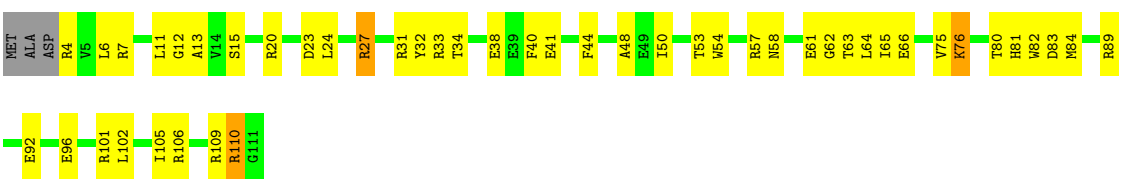
Chain F:





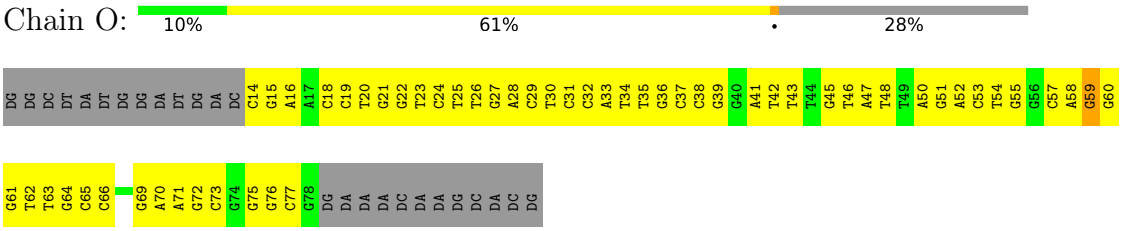
• Molecule 6: RNA polymerase-binding protein RbpA

Chain J:



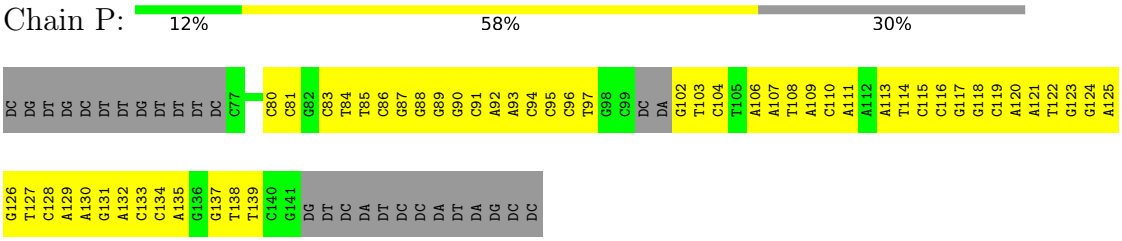
• Molecule 7: DNA (65-MER)

Chain O:



• Molecule 8: DNA (63-MER)

Chain P:



• Molecule 9: RNA polymerase-binding transcription factor CarD

Chain M:

4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of particles used	246409	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	69.9	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, C0L, 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	#$\|Z\|>2$	RMSZ	#$\|Z\|>2$
1	A	0.47	0/1742	0.55	0/2370
1	B	0.43	0/1786	0.53	0/2435
2	C	0.55	0/8751	0.61	5/11869 (0.0%)
3	D	0.56	1/10037 (0.0%)	0.59	6/13570 (0.0%)
4	E	0.44	0/662	0.52	0/901
5	F	0.46	0/2549	0.55	3/3438 (0.1%)
6	J	0.38	0/897	0.55	1/1210 (0.1%)
7	O	0.80	0/1497	0.97	2/2310 (0.1%)
8	P	0.76	0/1445	0.91	0/2224
9	M	0.36	0/1257	0.49	0/1700
All	All	0.55	1/30623 (0.0%)	0.63	17/42027 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	656	TRP	CB-CG	-7.29	1.37	1.50

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	F	306	LEU	C-N-CA	-7.20	103.71	121.70
3	D	419	GLY	N-CA-C	7.08	130.80	113.10
3	D	655	GLY	N-CA-C	-6.73	96.27	113.10
2	C	275	LEU	CA-CB-CG	-6.63	100.04	115.30
2	C	288	THR	N-CA-C	-6.59	93.21	111.00

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	1716	0	1756	121	0
1	B	1759	0	1783	145	0
2	C	8593	0	8517	695	0
3	D	9873	0	9938	738	0
4	E	649	0	645	54	0
5	F	2518	0	2540	182	0
6	J	881	0	861	62	0
7	O	1336	0	732	116	0
8	P	1289	0	706	88	0
9	M	1241	0	1259	111	0
10	C	38	40	0	1	0
11	D	2	0	0	0	0
12	D	1	0	0	0	0
All	All	29896	40	28737	2075	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 35.

The worst 5 of 2075 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:C:281:LEU:CD2	2:C:295:LEU:HD21	1.62	1.30
2:C:271:ASP:O	2:C:275:LEU:HD12	1.25	1.27
2:C:1067:ARG:CZ	3:D:418:LEU:CD2	2.13	1.25
2:C:278:TYR:CE1	2:C:292:ALA:HB2	1.73	1.23
2:C:1067:ARG:NH1	3:D:418:LEU:CD2	2.04	1.21

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 PDB entries followed by that with respect to all EM

entries.

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	223/347 (64%)	197 (88%)	26 (12%)	0	100	100
1	B	235/347 (68%)	194 (83%)	41 (17%)	0	100	100
2	C	1109/1179 (94%)	937 (84%)	168 (15%)	4 (0%)	36	74
3	D	1260/1326 (95%)	1141 (91%)	115 (9%)	4 (0%)	43	78
4	E	81/110 (74%)	76 (94%)	5 (6%)	0	100	100
5	F	317/531 (60%)	296 (93%)	20 (6%)	1 (0%)	43	78
6	J	106/111 (96%)	87 (82%)	19 (18%)	0	100	100
9	M	157/162 (97%)	144 (92%)	12 (8%)	1 (1%)	27	68
All	All	3488/4113 (85%)	3072 (88%)	406 (12%)	10 (0%)	47	78

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	D	418	LEU
3	D	658	PRO
2	C	274	LEU
3	D	653	HIS
2	C	53	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 PDB entries followed by that with respect to all EM entries.

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	194/297 (65%)	192 (99%)	2 (1%)	78	90
1	B	194/297 (65%)	193 (100%)	1 (0%)	90	96
2	C	935/997 (94%)	918 (98%)	17 (2%)	62	84
3	D	1042/1103 (94%)	1024 (98%)	18 (2%)	63	85
4	E	69/89 (78%)	68 (99%)	1 (1%)	69	87

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
5	F	264/429 (62%)	262 (99%)	2 (1%)	83 93
6	J	93/97 (96%)	89 (96%)	4 (4%)	32 67
9	M	129/131 (98%)	129 (100%)	0	100 100
All	All	2920/3440 (85%)	2875 (98%)	45 (2%)	70 87

5 of 45 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	5	ASN
3	D	285	LYS
6	J	4	ARG
3	D	140	ASP
3	D	386	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 39 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	239	ASN
3	D	505	HIS
5	F	516	HIS
3	D	307	ASN
3	D	416	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 4 ligands modelled in this entry, 3 are monoatomic - leaving 1 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
10	C0L	C	1201	-	36,38,38	2.60	13 (36%)	35,49,49	2.98	11 (31%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
10	C0L	C	1201	-	-	0/37/57/57	0/1/1/1

The worst 5 of 13 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
10	C	1201	C0L	O36-C16	-4.52	1.18	1.33
10	C	1201	C0L	C17-C18	-3.94	1.35	1.45
10	C	1201	C0L	C20-C18	-3.54	1.34	1.44
10	C	1201	C0L	O34-C35	-3.00	1.38	1.45
10	C	1201	C0L	C17-C23	-2.45	1.38	1.45

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
10	C	1201	C0L	C23-O22-C21	-6.86	118.03	122.45
10	C	1201	C0L	C35-O34-C32	-5.93	108.38	115.68
10	C	1201	C0L	O33-C32-N31	-4.32	119.36	125.46
10	C	1201	C0L	O34-C32-O33	-2.80	120.35	124.58
10	C	1201	C0L	C28-C27-C25	-2.78	109.25	114.47

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
10	C	1201	C0L	1	0

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.