



wwPDB/EMDatabank EM Map/Model Validation Summary Report ⓘ

May 1, 2017 – 03:44 PM EDT

PDB ID : 5MS0
EMDB ID: : EMD-3561
Title : pseudo-atomic model of the RNA polymerase lambda-based antitermination complex solved by cryo-EM
Authors : Said, N.; Krupp, F.
Deposited on : 2016-12-29
Resolution : 9.80 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

<http://wwpdb.org/validation/2016/EMValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

MolProbity : 4.02b-467
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20029077

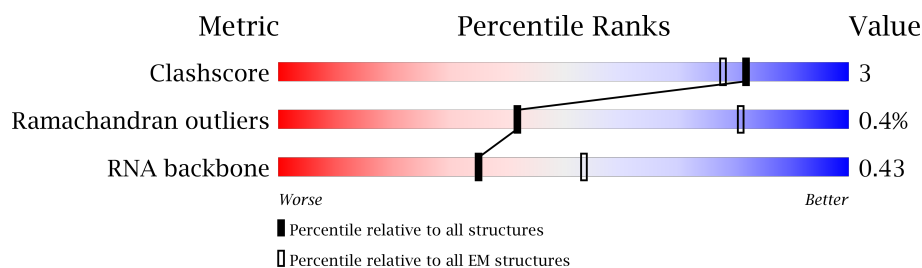
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 9.80 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)	EM structures (#Entries)
Clashscore	125131	1336
Ramachandran outliers	121729	1120
RNA backbone	3398	335

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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	N	85	85% 14% .
2	R	29	34% 45% 21%
3	A	329	67% . 29%
3	B	329	67% . 29%
4	C	1342	94% . .
5	D	1416	93% . .
6	E	100	92% 6% .
7	F	183	91% 7% ..
8	H	14	36% 50% 14%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
9	I	27	<div><div></div><div>93%</div><div>7%</div></div>
10	J	39	<div><div></div><div>67%</div><div>28%</div><div>5%</div></div>
11	L	139	<div><div></div><div>93%</div><div>6%</div><div></div></div>
12	M	497	<div><div></div><div>80%</div><div>5%</div><div>14%</div></div>
13	O	91	<div><div></div><div>95%</div><div></div><div>...</div></div>

2 Entry composition

There are 15 unique types of molecules in this entry. The entry contains 22698 atoms, of which 0 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 Antitermination protein.

Mol	Chain	Residues	Atoms				AltConf	Trace
1	N	85	Total	C	N	O	0	0
			424	254	85	85		

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
N	2	ALA	-	expression tag	UNP A0A0L6XR33
N	56	ALA	ASP	conflict	UNP A0A0L6XR33
N	57	ASP	LEU	conflict	UNP A0A0L6XR33
N	58	LEU	THR	conflict	UNP A0A0L6XR33
N	59	THR	VAL	conflict	UNP A0A0L6XR33
N	60	VAL	LEU	conflict	UNP A0A0L6XR33

- Molecule 2 is a RNA chain called nascent RNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	R	29	Total	C	N	O	P	0	0
			623	279	117	198	29		

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

Mol	Chain	Residues	Atoms				AltConf	Trace
3	A	235	Total	C	N	O	0	0
			1159	689	235	235		
3	B	235	Total	C	N	O	0	0
			1159	689	235	235		

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

Mol	Chain	Residues	Atoms				AltConf	Trace
4	C	1291	Total	C	N	O	0	0
			6355	3773	1291	1291		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	516	VAL	ASP	conflict	UNP P0A8V2

- Molecule 5 is a protein called DNA-directed RNA polymerase subunit beta'.

Mol	Chain	Residues	Atoms				AltConf	Trace
5	D	1364	Total	C	N	O	0	0
			6709	3981	1364	1364		

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	1408	LEU	-	expression tag	UNP P0A8T8
D	1409	GLU	-	expression tag	UNP P0A8T8
D	1410	VAL	-	expression tag	UNP P0A8T8
D	1411	HIS	-	expression tag	UNP P0A8T8
D	1412	HIS	-	expression tag	UNP P0A8T8
D	1413	HIS	-	expression tag	UNP P0A8T8
D	1414	HIS	-	expression tag	UNP P0A8T8
D	1415	HIS	-	expression tag	UNP P0A8T8
D	1416	HIS	-	expression tag	UNP P0A8T8

- Molecule 6 is a protein called 30S ribosomal protein S10.

Mol	Chain	Residues	Atoms				AltConf	Trace
6	E	100	Total	C	N	O	0	0
			496	296	100	100		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	46	SER	LYS	conflict	UNP P0A7R5

- Molecule 7 is a protein called Transcription termination/antitermination protein NusG.

Mol	Chain	Residues	Atoms				AltConf	Trace
7	F	181	Total	C	N	O	0	0
			891	529	181	181		

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	-1	GLY	-	expression tag	UNP P0AFG0
F	0	ALA	-	expression tag	UNP P0AFG0

- Molecule 8 is a RNA chain called RNA transcription bubble.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	H	14	Total	C	N	O	P	0	0
			301	135	59	93	14		

- Molecule 9 is a DNA chain called DNAI.

Mol	Chain	Residues	Atoms					AltConf	Trace
9	I	27	Total	C	N	O	P	0	0
			552	266	103	158	25		

- Molecule 10 is a DNA chain called DNAIL.

Mol	Chain	Residues	Atoms					AltConf	Trace
10	J	39	Total	C	N	O	P	0	0
			792	380	142	232	38		

- Molecule 11 is a protein called N utilization substance protein B homolog.

Mol	Chain	Residues	Atoms				AltConf	Trace
11	L	139	Total	C	N	O	0	0
			690	412	139	139		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
L	2	GLU	LYS	conflict	UNP B7MD74

- Molecule 12 is a protein called Transcription termination/antitermination protein NusA.

Mol	Chain	Residues	Atoms				AltConf	Trace
12	M	425	Total	C	N	O	0	0
			2104	1254	425	425		

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
M	-1	GLY	-	expression tag	UNP P0AFF6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
M	0	ALA	-	expression tag	UNP P0AFF6
M	462	GLN	GLU	conflict	UNP P0AFF6

- Molecule 13 is a protein called DNA-directed RNA polymerase subunit omega.

Mol	Chain	Residues	Atoms				AltConf	Trace
13	O	89	Total	C	N	O	0	0
			440	262	89	89		

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

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


- Molecule 15 is ZINC ION (three-letter code: ZN) (formula: Zn).

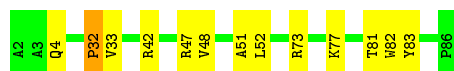
Mol	Chain	Residues	Atoms		AltConf
15	D	2	Total	Zn	0
			2	2	

3 Residue-property plots

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: Antitermination protein

Chain N: 



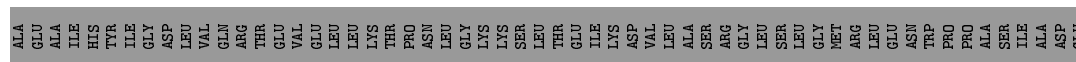
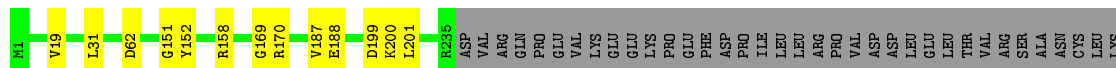
- Molecule 2: nascent RNA

Chain R: 



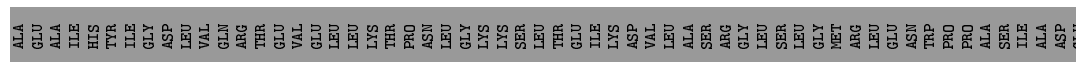
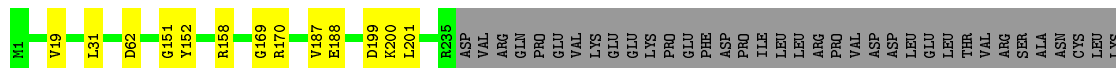
- Molecule 3: DNA-directed RNA polymerase subunit alpha

Chain A: 



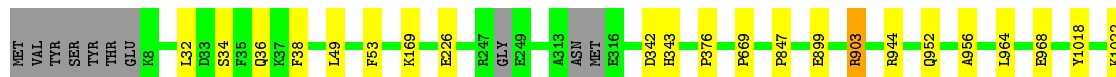
- Molecule 3: DNA-directed RNA polymerase subunit alpha

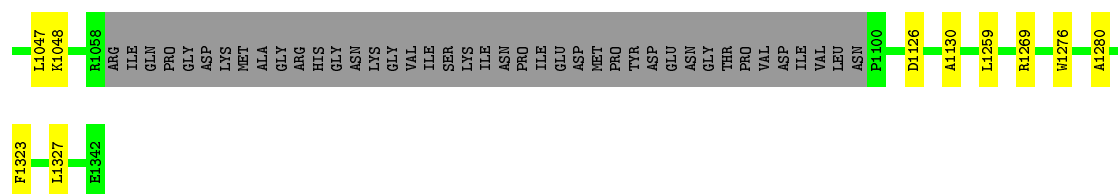
Chain B: 



- Molecule 4: DNA-directed RNA polymerase subunit beta

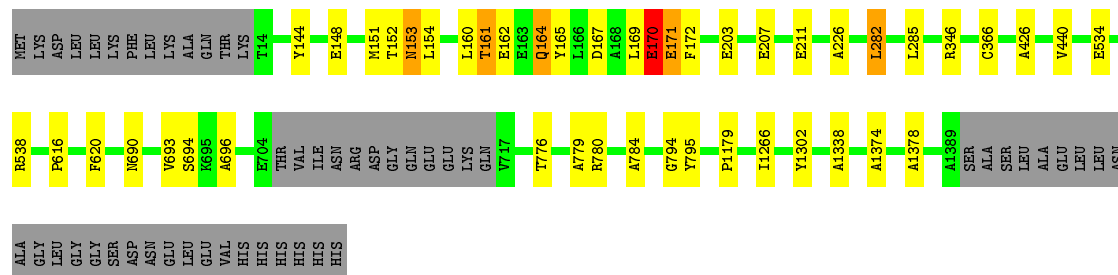
Chain C: 





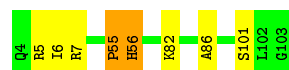
- Molecule 5: DNA-directed RNA polymerase subunit beta'

Chain D: 93%



- Molecule 6: 30S ribosomal protein S10

Chain E: 92% 6%



- Molecule 7: Transcription termination/antitermination protein NusG

Chain F: 91% 7%



- Molecule 8: RNA transcription bubble

Chain H: 36% 50% 14%



- Molecule 9: DNAI

Chain I: 93% 7%



- Molecule 10: DNAII

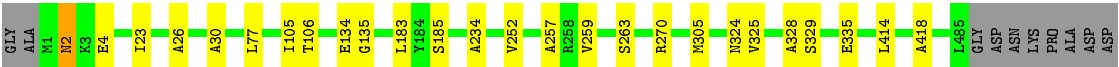
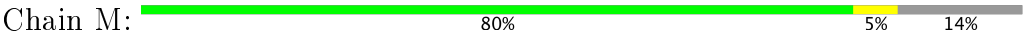
Chain J: 67% 28% 5%



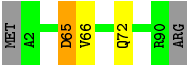
- Molecule 11: N utilization substance protein B homolog



- Molecule 12: Transcription termination/antitermination protein NusA



- Molecule 13: DNA-directed RNA polymerase subunit omega



4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of particles used	23983	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	Not provided	Depositor
Microscope	FEI POLARA 300	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	Not provided	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	Not provided	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, 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	N	0.31	0/422	0.75	2/587 (0.3%)
10	J	0.73	1/887 (0.1%)	1.13	4/1366 (0.3%)
11	L	0.29	0/689	0.60	2/960 (0.2%)
12	M	0.29	0/2103	0.65	2/2930 (0.1%)
13	O	0.32	0/438	0.74	2/607 (0.3%)
2	R	0.55	1/698 (0.1%)	1.69	14/1084 (1.3%)
3	A	0.27	0/1153	0.55	0/1595
3	B	0.26	0/1153	0.55	0/1595
4	C	0.27	0/6348	0.54	2/8816 (0.0%)
5	D	0.28	0/6701	0.57	7/9302 (0.1%)
6	E	0.27	0/495	0.60	1/689 (0.1%)
7	F	0.90	3/890 (0.3%)	0.96	4/1236 (0.3%)
8	H	0.27	0/337	0.96	1/523 (0.2%)
9	I	0.59	0/619	0.99	0/953
All	All	0.37	5/22933 (0.0%)	0.71	41/32243 (0.1%)

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
1	N	0	2
11	L	0	1
12	M	0	3
13	O	0	1
3	A	0	2
3	B	0	2
4	C	0	6
5	D	0	6
6	E	0	3

Continued on next page...

Continued from previous page...

Mol	Chain	#Chirality outliers	#Planarity outliers
7	F	0	2
All	All	0	28

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	F	92	PHE	C-N	17.02	1.73	1.34
7	F	93	ILE	N-CA	16.56	1.79	1.46
2	R	7	A	OP3-P	-10.50	1.48	1.61
7	F	123	ARG	C-N	7.73	1.49	1.34
10	J	16	DA	C3'-O3'	6.14	1.51	1.44

The worst 5 of 41 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	R	12	U	OP1-P-OP2	18.55	147.43	119.60
2	R	12	U	O5'-P-OP1	-17.79	89.35	110.70
7	F	92	PHE	C-N-CA	17.15	164.58	121.70
2	R	12	U	O5'-P-OP2	-15.76	91.51	105.70
2	R	11	U	OP1-P-O3'	-14.96	72.28	105.20

There are no chirality outliers.

5 of 28 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	158	ARG	Peptide
3	A	19	VAL	Peptide
3	B	19	VAL	Peptide
1	N	42	ARG	Peptide
1	N	82	TRP	Peptide

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	N	424	0	194	7	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	R	623	0	314	1	0
3	A	1159	0	525	6	0
3	B	1159	0	525	6	0
4	C	6355	0	2858	14	0
5	D	6709	0	3134	24	0
6	E	496	0	225	3	0
7	F	891	0	393	6	0
8	H	301	0	154	7	0
9	I	552	0	309	1	0
10	J	792	0	442	11	0
11	L	690	0	334	5	0
12	M	2104	0	1001	13	0
13	O	440	0	212	0	0
14	D	1	0	0	0	0
15	D	2	0	0	0	0
All	All	22698	0	10620	96	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 3.

The worst 5 of 96 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)
7:F:93:ILE:N	7:F:93:ILE:CA	1.79	1.44
7:F:92:PHE:C	7:F:93:ILE:N	1.73	1.40
3:B:169:GLY:C	3:B:170:ARG:N	2.08	1.07
3:A:169:GLY:C	3:A:170:ARG:N	2.08	1.07
3:B:187:VAL:C	3:B:188:GLU:N	2.09	1.05

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	N	82/85 (96%)	68 (83%)	11 (13%)	3 (4%)	4	33
3	A	224/329 (68%)	192 (86%)	31 (14%)	1 (0%)	38	77
3	B	224/329 (68%)	192 (86%)	31 (14%)	1 (0%)	38	77
4	C	1277/1342 (95%)	1132 (89%)	142 (11%)	3 (0%)	51	84
5	D	1348/1416 (95%)	1218 (90%)	125 (9%)	5 (0%)	38	77
6	E	98/100 (98%)	89 (91%)	7 (7%)	2 (2%)	9	46
7	F	179/183 (98%)	160 (89%)	18 (10%)	1 (1%)	28	71
11	L	137/139 (99%)	124 (90%)	13 (10%)	0	100	100
12	M	423/497 (85%)	370 (88%)	52 (12%)	1 (0%)	51	84
13	O	85/91 (93%)	70 (82%)	14 (16%)	1 (1%)	15	57
All	All	4077/4511 (90%)	3615 (89%)	444 (11%)	18 (0%)	42	77

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	N	33	VAL
4	C	1048	LYS
5	D	148	GLU
5	D	170	GLU
5	D	171	GLU

5.3.2 Protein sidechains ⓘ

There are no protein residues with a non-rotameric sidechain to report in this entry.

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	R	29/29 (100%)	15 (51%)	4 (13%)
8	H	13/14 (92%)	2 (15%)	1 (7%)
All	All	42/43 (97%)	17 (40%)	5 (11%)

5 of 17 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	R	8	U
2	R	11	U
2	R	12	U

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	R	13	A
2	R	14	A

All (5) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	R	7	A
2	R	10	U
2	R	11	U
2	R	34	G
8	H	12	A

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 carbohydrates in this entry.

5.6 Ligand geometry ⓘ

Of 3 ligands modelled in this entry, 3 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
5	D	6
3	B	5
3	A	5
4	C	3
1	N	1
13	O	1
7	F	1

The worst 5 of 22 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	D	146:VAL	C	147:ILE	N	9.08
1	O	76:GLU	C	77:ALA	N	9.08
1	D	175:GLU	C	176:PHE	N	8.86
1	C	1170:MET	C	1171:ARG	N	7.38
1	D	191:SER	C	192:MET	N	5.59