wwPDB EM Validation Summary Report ⓘ

May 21, 2024 – 10:12 AM JST

PDB ID	:	8HYJ
EMDB ID	:	EMD-35086
Title	:	A cryo-EM structure of KTF1-bound polymerase V transcription elongation complex
Authors	:	Zhang, H.; Zhang, Y.
Deposited on	:	2023-01-06
Resolution	:	4.30 Å(reported)

This is a wwPDB EM 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/EMValidationReportHelp
with specific help available everywhere you see the ⓘ 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 ⓘ) were used in the production of this report:

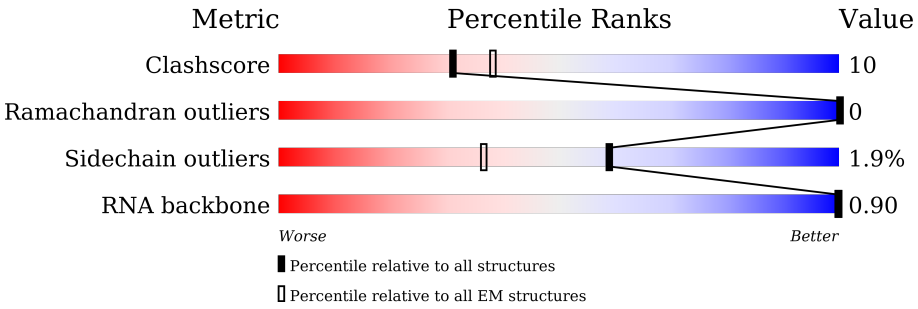
EMDB validation analysis	:	FAILED
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	FAILED
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 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.



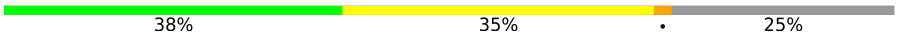
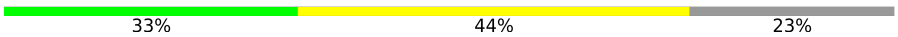
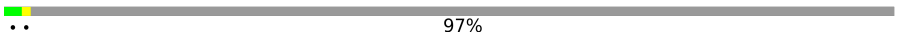
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826
RNA backbone	4643	859

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the 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	1976	45% 13% 42%
2	B	1172	73% 16% • 10%
3	C	319	70% 18% • 11%
4	D	205	39% 18% 43%
5	E	222	71% 23% 6%
6	F	144	43% 10% 47%
7	G	178	75% 25%
8	H	146	69% 15% 16%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
9	I	114	
10	J	71	
11	K	116	
12	L	51	
13	N	48	
14	P	30	
15	T	48	
16	W	1493	

2 Entry composition (i)

There are 18 unique types of molecules in this entry. The entry contains 28568 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 DNA-directed RNA polymerase V subunit 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	1141	8794	5540	1526	1672	56	0	0

- Molecule 2 is a protein called DNA-directed RNA polymerases IV and V subunit 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	B	1054	7912	5047	1426	1400	39	0	0

- Molecule 3 is a protein called DNA-directed RNA polymerases IV and V subunit 3B.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	C	285	2123	1351	360	398	14	0	0

- Molecule 4 is a protein called DNA-directed RNA polymerases IV and V subunit 4.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	D	117	896	566	142	181	7	0	0

- Molecule 5 is a protein called DNA-directed RNA polymerase V subunit 5A.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
5	E	209	1652	1050	283	316	3	0	0

- Molecule 6 is a protein called DNA-directed RNA polymerases II, IV and V subunit 6A.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
6	F	76	554	355	97	99	3	0	0

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

Mol	Chain	Residues	Atoms					AltConf	Trace
7	G	178	Total	C	N	O	S	0	0
			1417	926	227	253	11		

- Molecule 8 is a protein called DNA-directed RNA polymerases II and V subunit 8A.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	H	123	Total	C	N	O	S	0	0
			904	597	142	160	5		

- Molecule 9 is a protein called DNA-directed RNA polymerases II, IV and V subunit 9A.

Mol	Chain	Residues	Atoms					AltConf	Trace
9	I	98	Total	C	N	O	S	0	0
			702	436	131	124	11		

- Molecule 10 is a protein called DNA-directed RNA polymerases II, IV and V subunit 10.

Mol	Chain	Residues	Atoms					AltConf	Trace
10	J	62	Total	C	N	O	S	0	0
			484	316	83	79	6		

- Molecule 11 is a protein called DNA-directed RNA polymerases II, IV and V subunit 11.

Mol	Chain	Residues	Atoms					AltConf	Trace
11	K	100	Total	C	N	O	S	0	0
			756	482	138	135	1		

- Molecule 12 is a protein called DNA-directed RNA polymerases II, IV and V subunit 12.

Mol	Chain	Residues	Atoms					AltConf	Trace
12	L	42	Total	C	N	O	S	0	0
			309	192	54	59	4		

- Molecule 13 is a DNA chain called DNA (48-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
13	N	36	Total	C	N	O	P	0	0
			749	356	139	218	36		

- Molecule 14 is a RNA chain called RNA (30-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
14	P	10	Total	C	N	O	P	0	0
			218	97	42	69	10		

- Molecule 15 is a DNA chain called DNA (48-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
15	T	37	Total	C	N	O	P	0	0
			746	356	133	220	37		

- Molecule 16 is a protein called Protein RNA-directed DNA methylation 3.

Mol	Chain	Residues	Atoms					AltConf	Trace
16	W	44	Total	C	N	O	S	0	0
			347	223	63	60	1		

- Molecule 17 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		AltConf
17	A	1	Total	Mg	0
			1	1	

- Molecule 18 is ZINC ION (three-letter code: ZN) (formula: Zn).

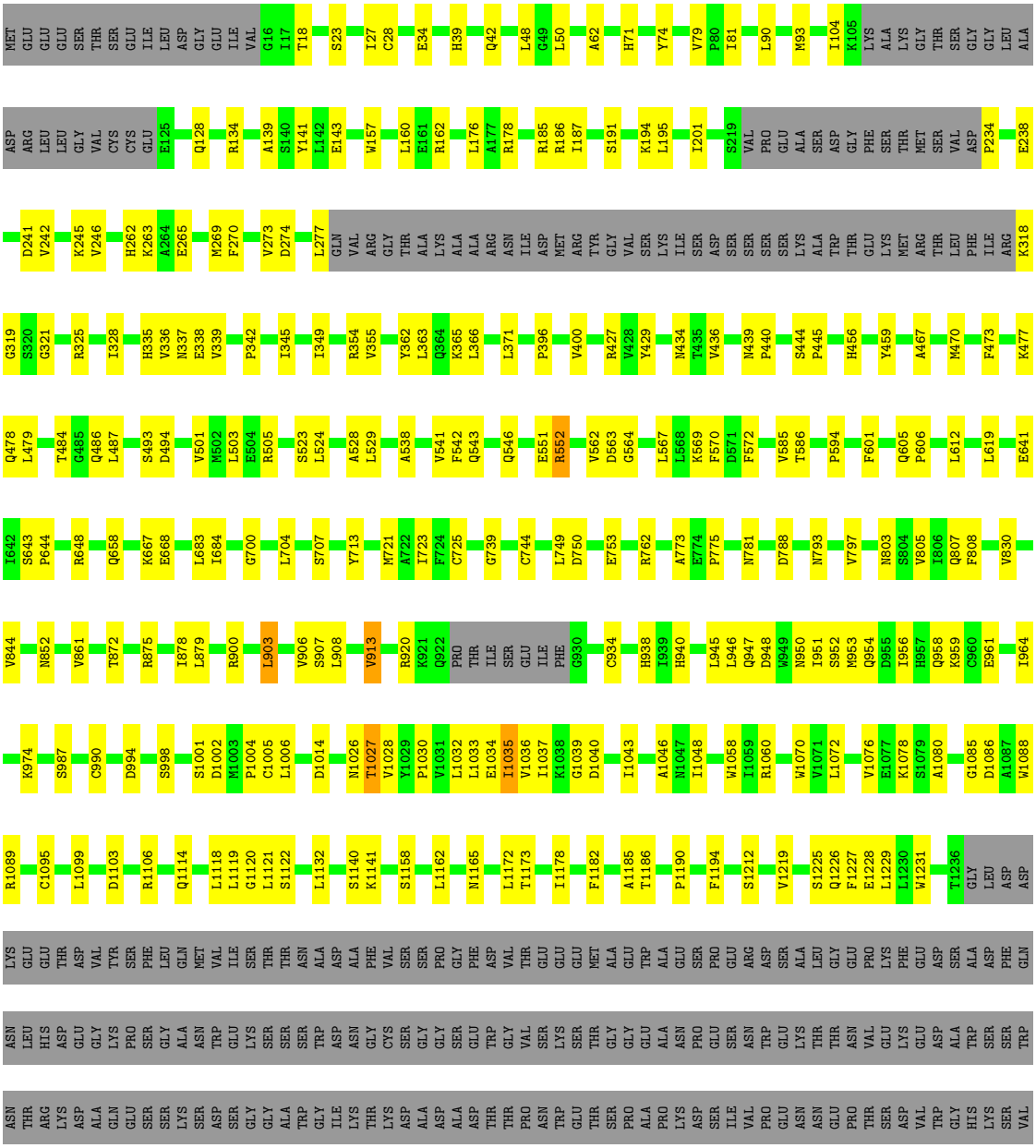
Mol	Chain	Residues	Atoms		AltConf
18	C	1	Total	Zn	0
			1	1	
18	I	1	Total	Zn	0
			1	1	
18	J	1	Total	Zn	0
			1	1	
18	L	1	Total	Zn	0
			1	1	

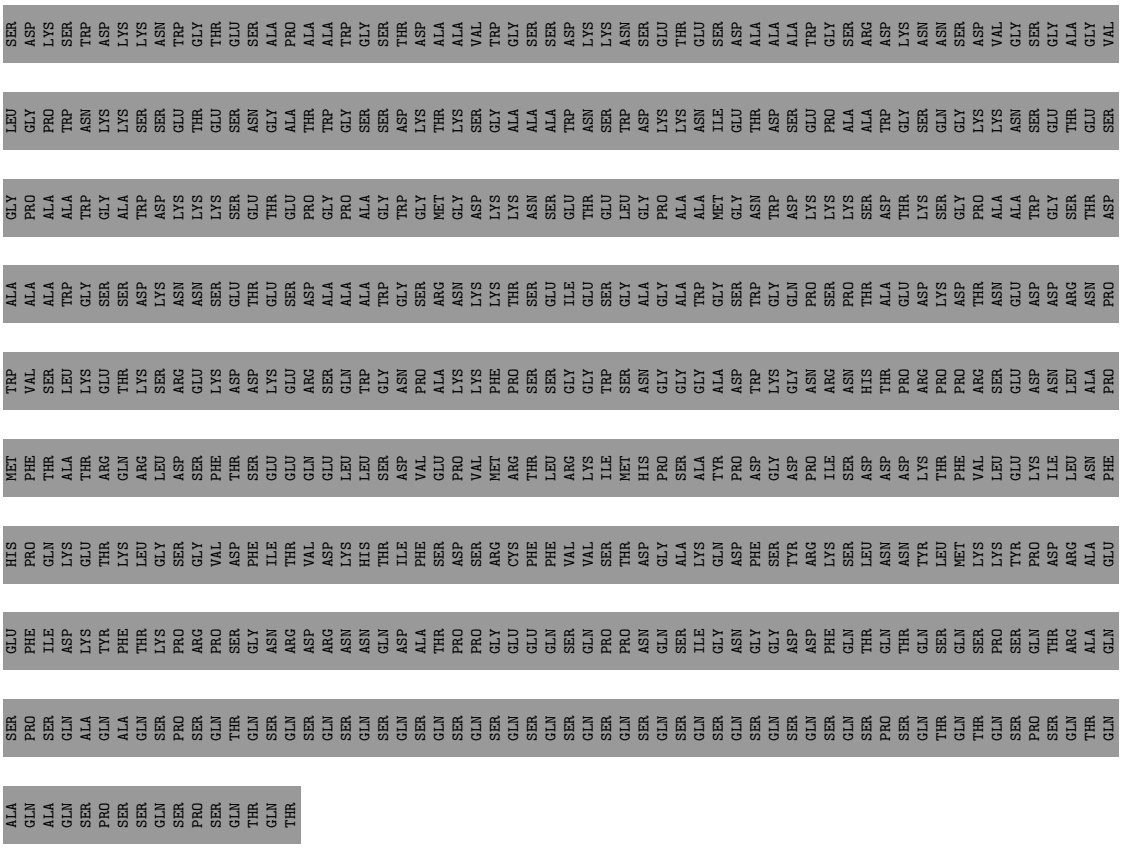
3 Residue-property plots (i)

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 V subunit 1

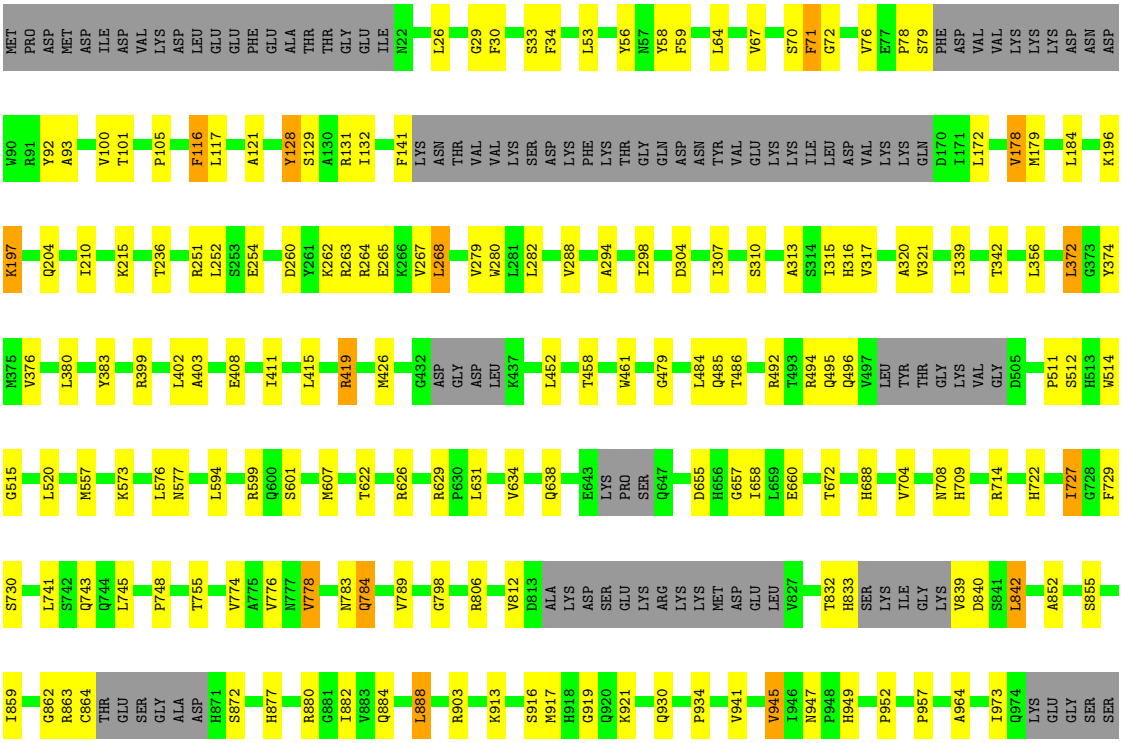
Chain A:

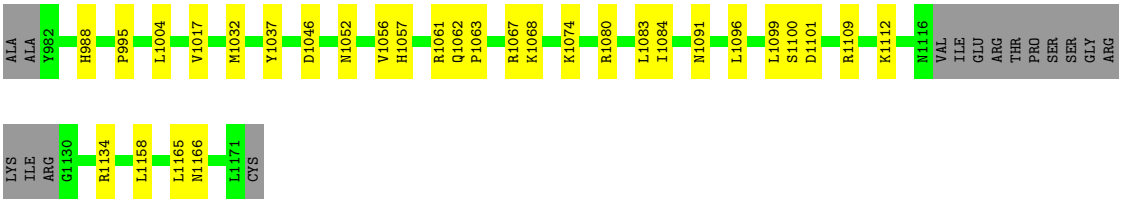




• Molecule 2: DNA-directed RNA polymerases IV and V subunit 2

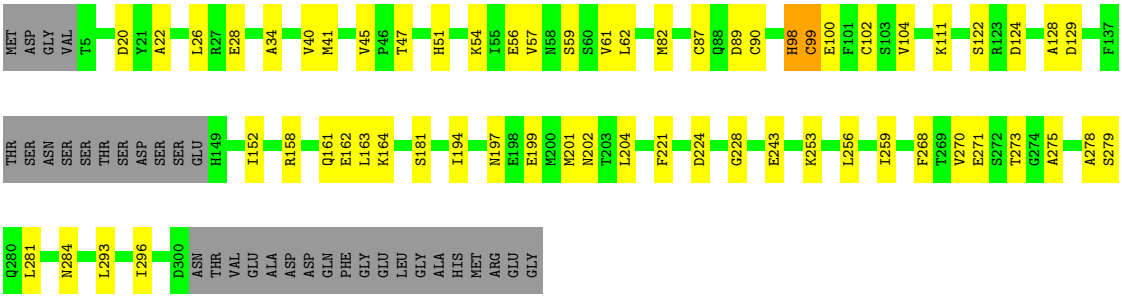
Chain B:





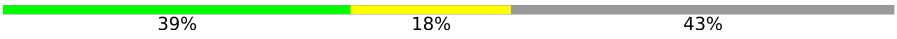
• Molecule 3: DNA-directed RNA polymerases IV and V subunit 3B

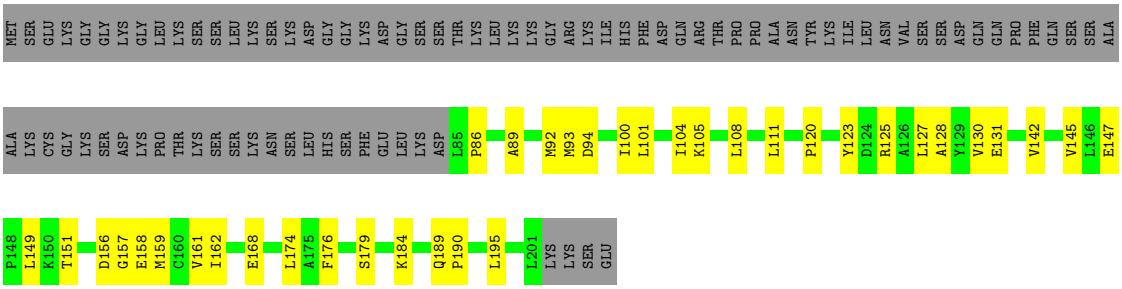
Chain C:



• Molecule 4: DNA-directed RNA polymerases IV and V subunit 4

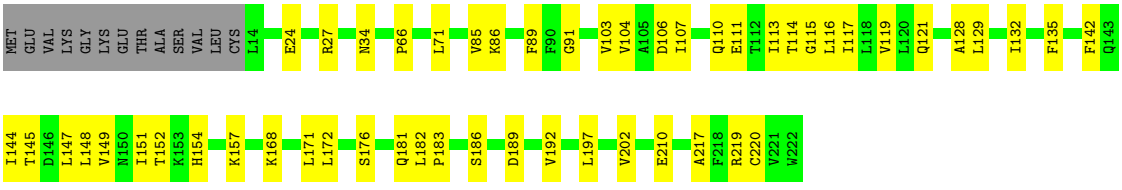
Chain D:





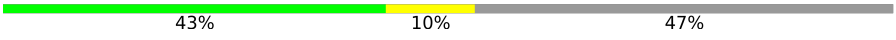
• Molecule 5: DNA-directed RNA polymerase V subunit 5A

Chain E:



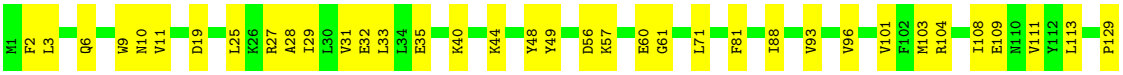
• Molecule 6: DNA-directed RNA polymerases II, IV and V subunit 6A

Chain F:



• Molecule 7: DNA-directed RNA polymerase V subunit 7

Chain G:

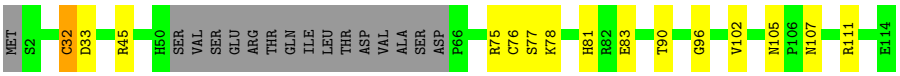


• Molecule 8: DNA-directed RNA polymerases II and V subunit 8A

Chain H:

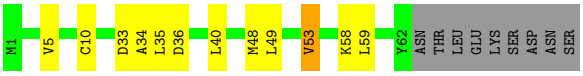
• Molecule 9: DNA-directed RNA polymerases II, IV and V subunit 9A

Chain I:



• Molecule 10: DNA-directed RNA polymerases II, IV and V subunit 10

Chain J:

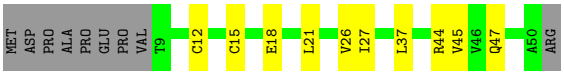


• Molecule 11: DNA-directed RNA polymerases II, IV and V subunit 11

Chain K:

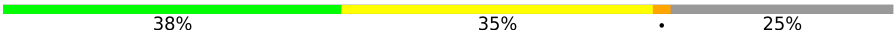
• Molecule 12: DNA-directed RNA polymerases II, IV and V subunit 12

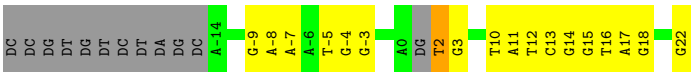
Chain L:



• Molecule 13: DNA (48-MER)

Chain N:



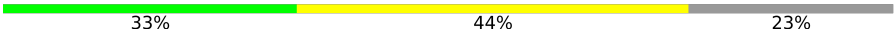


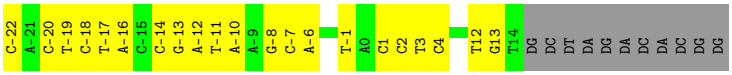
• Molecule 14: RNA (30-MER)

Chain P:

• Molecule 15: DNA (48-MER)

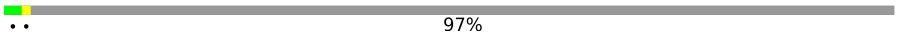
Chain T:

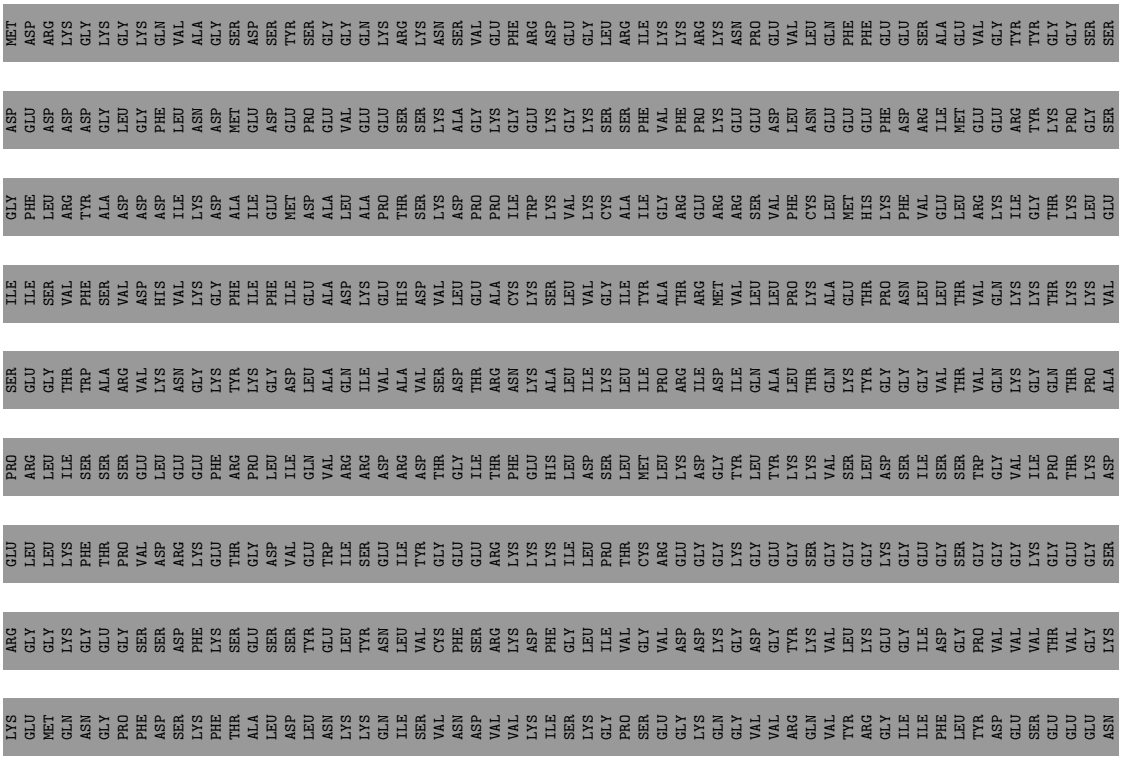


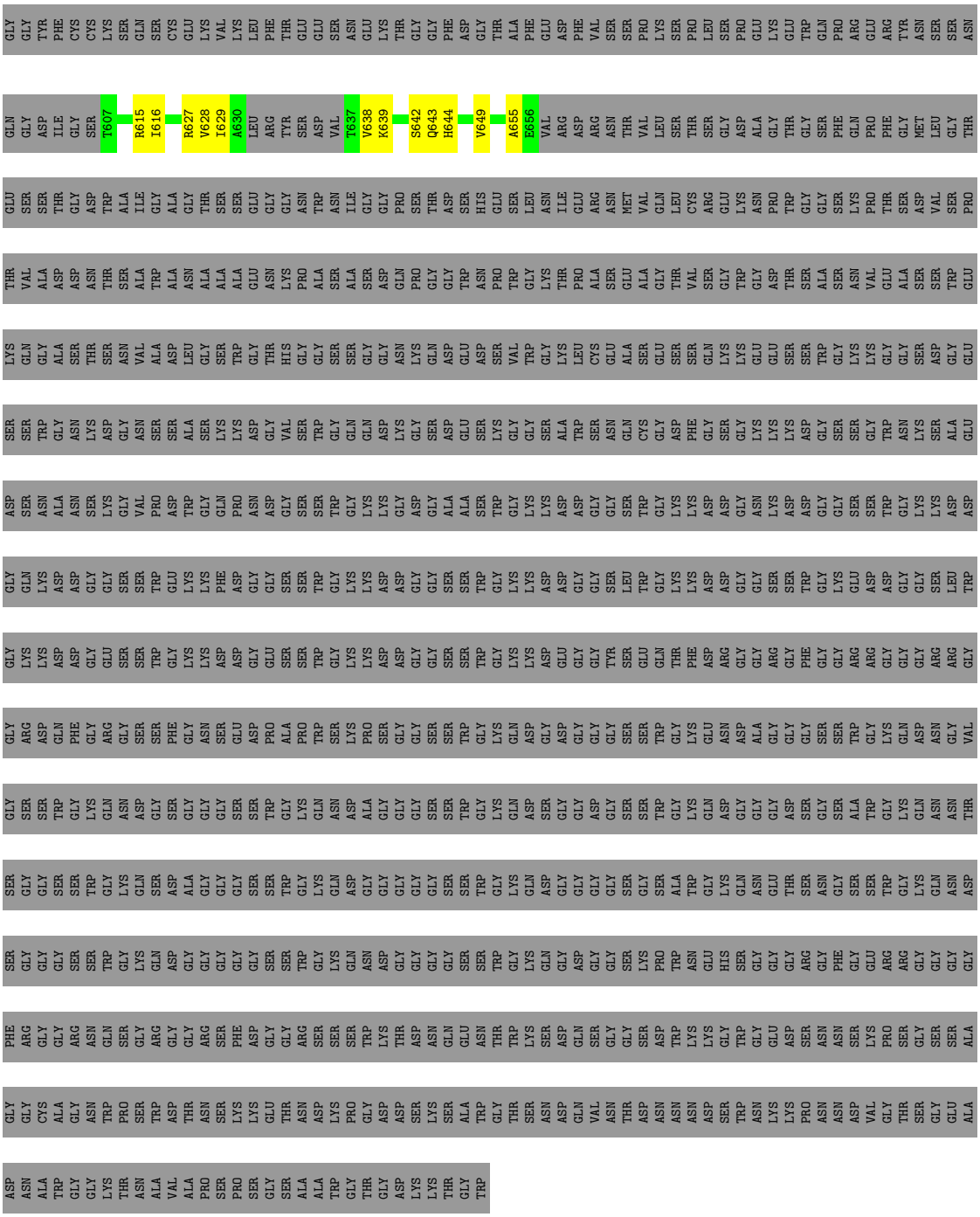


• Molecule 16: Protein RNA-directed DNA methylation 3

Chain W:






T607
R615
I616
R627
V628
I629
A630
T637
V638
K639
S642
Q643
H644
V649
A655
E656

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	30359	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING ONLY	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	50	Depositor
Minimum defocus (nm)	1200	Depositor
Maximum defocus (nm)	2200	Depositor
Magnification	Not provided	
Image detector	GATAN K3 (6k x 4k)	Depositor

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	0.39	0/8949	0.52	4/12093 (0.0%)
2	B	1.02	0/8074	0.90	15/10929 (0.1%)
3	C	1.11	0/2156	0.93	3/2929 (0.1%)
4	D	0.27	0/909	0.45	0/1232
5	E	0.24	0/1678	0.42	0/2271
6	F	0.25	0/564	0.41	0/767
7	G	0.27	0/1448	0.46	0/1951
8	H	0.26	0/919	0.42	0/1244
9	I	0.28	0/716	0.52	0/973
10	J	1.20	0/492	1.01	2/666 (0.3%)
11	K	1.06	0/770	0.82	0/1045
12	L	1.07	0/312	1.05	2/420 (0.5%)
13	N	0.54	1/840 (0.1%)	0.83	1/1295 (0.1%)
14	P	0.25	0/244	0.66	0/379
15	T	0.54	0/834	0.85	0/1281
16	W	0.88	1/351 (0.3%)	0.98	1/470 (0.2%)
All	All	0.73	2/29256 (0.0%)	0.72	28/39945 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
16	W	642	SER	CA-CB	-5.84	1.44	1.52
13	N	2	DT	O3'-P	5.41	1.67	1.61

The worst 5 of 28 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	304	ASP	N-CA-C	-7.36	91.14	111.00
2	B	72	GLY	N-CA-C	7.34	131.45	113.10
2	B	798	GLY	N-CA-C	6.89	130.33	113.10

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
12	L	44	ARG	N-CA-C	-6.70	92.92	111.00
1	A	903	LEU	CA-CB-CG	6.40	130.02	115.30

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	8794	0	8748	214	0
2	B	7912	0	7591	148	0
3	C	2123	0	2089	42	0
4	D	896	0	899	25	0
5	E	1652	0	1634	58	0
6	F	554	0	516	12	0
7	G	1417	0	1459	31	0
8	H	904	0	851	13	0
9	I	702	0	609	9	0
10	J	484	0	486	7	0
11	K	756	0	716	9	0
12	L	309	0	282	5	0
13	N	749	0	409	20	0
14	P	218	0	109	3	0
15	T	746	0	416	31	0
16	W	347	0	366	6	0
17	A	1	0	0	0	0
18	C	1	0	0	0	0
18	I	1	0	0	0	0
18	J	1	0	0	0	0
18	L	1	0	0	0	0
All	All	28568	0	27180	541	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 10.

The worst 5 of 541 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:116:PHE:CZ	2:B:121:ALA:HB2	1.36	1.56
1:A:1060:ARG:NH2	5:E:210:GLU:HG2	1.47	1.29
1:A:1060:ARG:HH22	5:E:210:GLU:CG	1.47	1.27
2:B:116:PHE:CZ	2:B:121:ALA:CB	2.20	1.25
2:B:116:PHE:HZ	2:B:121:ALA:CB	1.51	1.23

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	1131/1976 (57%)	1077 (95%)	54 (5%)	0	100	100
2	B	1032/1172 (88%)	966 (94%)	66 (6%)	0	100	100
3	C	281/319 (88%)	233 (83%)	48 (17%)	0	100	100
4	D	115/205 (56%)	109 (95%)	6 (5%)	0	100	100
5	E	207/222 (93%)	189 (91%)	18 (9%)	0	100	100
6	F	74/144 (51%)	74 (100%)	0	0	100	100
7	G	176/178 (99%)	166 (94%)	10 (6%)	0	100	100
8	H	117/146 (80%)	114 (97%)	3 (3%)	0	100	100
9	I	94/114 (82%)	86 (92%)	8 (8%)	0	100	100
10	J	60/71 (84%)	52 (87%)	8 (13%)	0	100	100
11	K	98/116 (84%)	96 (98%)	2 (2%)	0	100	100
12	L	40/51 (78%)	40 (100%)	0	0	100	100
16	W	40/1493 (3%)	39 (98%)	1 (2%)	0	100	100
All	All	3465/6207 (56%)	3241 (94%)	224 (6%)	0	100	100

There are no Ramachandran outliers to report.

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	974/1710 (57%)	967 (99%)	7 (1%)	84 90
2	B	786/1029 (76%)	764 (97%)	22 (3%)	43 65
3	C	221/282 (78%)	214 (97%)	7 (3%)	39 62
4	D	103/181 (57%)	103 (100%)	0	100 100
5	E	181/204 (89%)	181 (100%)	0	100 100
6	F	50/128 (39%)	50 (100%)	0	100 100
7	G	154/155 (99%)	154 (100%)	0	100 100
8	H	84/127 (66%)	84 (100%)	0	100 100
9	I	66/104 (64%)	65 (98%)	1 (2%)	65 80
10	J	49/66 (74%)	45 (92%)	4 (8%)	11 37
11	K	72/105 (69%)	66 (92%)	6 (8%)	11 37
12	L	30/45 (67%)	28 (93%)	2 (7%)	16 43
16	W	38/1156 (3%)	35 (92%)	3 (8%)	12 38
All	All	2808/5292 (53%)	2756 (98%)	52 (2%)	59 75

5 of 52 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	1046	ASP
3	C	243	GLU
16	W	616	ILE
3	C	28	GLU
3	C	98	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 59 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	678	GLN
11	K	55	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	955	GLN
10	J	61	ASN
8	H	46	ASN

5.3.3 RNA

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
14	P	9/30 (30%)	3 (33%)	0

All (3) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
14	P	8	G
14	P	9	G
14	P	10	U

There are no RNA pucker outliers to report.

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 5 ligands modelled in this entry, 5 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.