wwPDB EM Validation Summary Report

Feb 26, 2025 – 04:08 PM EST

PDB ID	:	7MBN
EMDB ID	:	EMD-23739
Title	:	Cryo-EM structure of MLL1-NCP (H3K4M) complex, mode02
Authors	:	Park, S.H.; Ayoub, A.; Lee, Y.T.; Dou, Y.; Cho, U.
Deposited on	:	2021-04-01
Resolution	:	4.02 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 (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 (i)) were used in the production of this report:

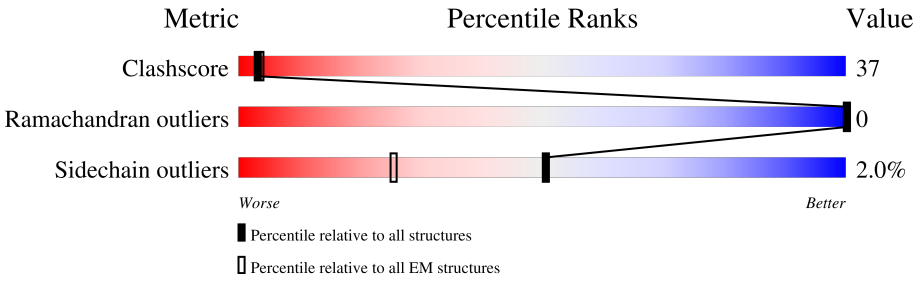
EMDB validation analysis	:	0.0.1.dev117
MolProbity	:	4.02b-467
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
MapQ	:	FAILED
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.41.4

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 4.02 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.



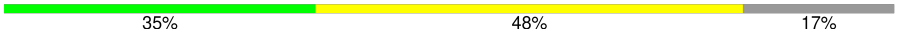
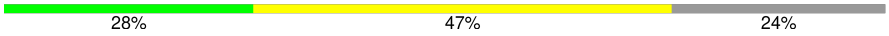
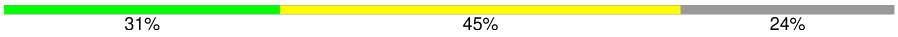
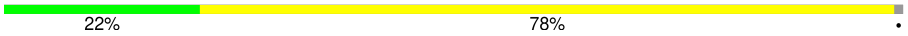
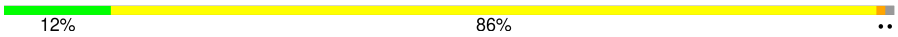
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	210492	15764
Ramachandran outliers	207382	16835
Sidechain outliers	206894	16415

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	538	28% 33% 38%
2	B	313	50% 47% .
3	C	209	42% 31% 27%
4	D	534	22% 26% 52%
5	G	136	18% 50% . 28%
5	K	136	20% 49% . 28%
6	H	103	23% 50% 6% 20%
6	L	103	20% 54% 5% 20%
7	I	129	32% 50% . 17%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
7	M	129	
8	J	123	
8	N	123	
9	O	147	
10	P	147	

2 Entry composition (i)

There are 10 unique types of molecules in this entry. The entry contains 20222 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 Retinoblastoma-binding protein 5.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	333	Total	C	N	O	S	0	0
			2613	1648	457	493	15		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	SER	-	expression tag	UNP Q15291

- Molecule 2 is a protein called WD repeat-containing protein 5.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	303	Total	C	N	O	S	3	0
			2342	1492	388	451	11		

- Molecule 3 is a protein called Histone-lysine N-methyltransferase 2A.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	153	Total	C	N	O	S	0	0
			1228	780	218	217	13		

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	3761	SER	-	expression tag	UNP Q03164
C	3861	ILE	ASN	conflict	UNP Q03164
C	3867	LEU	GLN	conflict	UNP Q03164

- Molecule 4 is a protein called Set1/Ash2 histone methyltransferase complex subunit ASH2.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	D	257	Total	C	N	O	S	0	0
			2075	1342	343	383	7		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	1	SER	-	expression tag	UNP Q9UBL3

- Molecule 5 is a protein called Histone H3.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	G	98	Total	C	N	O	S	0	0
			801	506	153	139	3		
5	K	98	Total	C	N	O	S	0	0
			801	506	153	139	3		

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	4	MET	LYS	conflict	UNP A0A310TTQ1
K	4	MET	LYS	conflict	UNP A0A310TTQ1

- Molecule 6 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	H	82	Total	C	N	O	S	0	0
			653	413	127	112	1		
6	L	82	Total	C	N	O	S	0	0
			653	413	127	112	1		

- Molecule 7 is a protein called Histone H2A.

Mol	Chain	Residues	Atoms				AltConf	Trace
7	I	107	Total	C	N	O	0	0
			811	510	158	143		
7	M	107	Total	C	N	O	0	0
			815	513	159	143		

- Molecule 8 is a protein called Histone H2B 1.1.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	J	93	Total	C	N	O	S	0	0
			718	451	128	137	2		
8	N	93	Total	C	N	O	S	0	0
			726	457	130	137	2		

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
J	0	MET	-	initiating methionine	UNP P02281
J	29	THR	SER	engineered mutation	UNP P02281
N	0	MET	-	initiating methionine	UNP P02281
N	29	THR	SER	engineered mutation	UNP P02281

- Molecule 9 is a DNA chain called DNA (146-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
9	O	146	Total	C	N	O	P	0	0
			2975	1413	540	876	146		

- Molecule 10 is a DNA chain called DNA (146-MER).

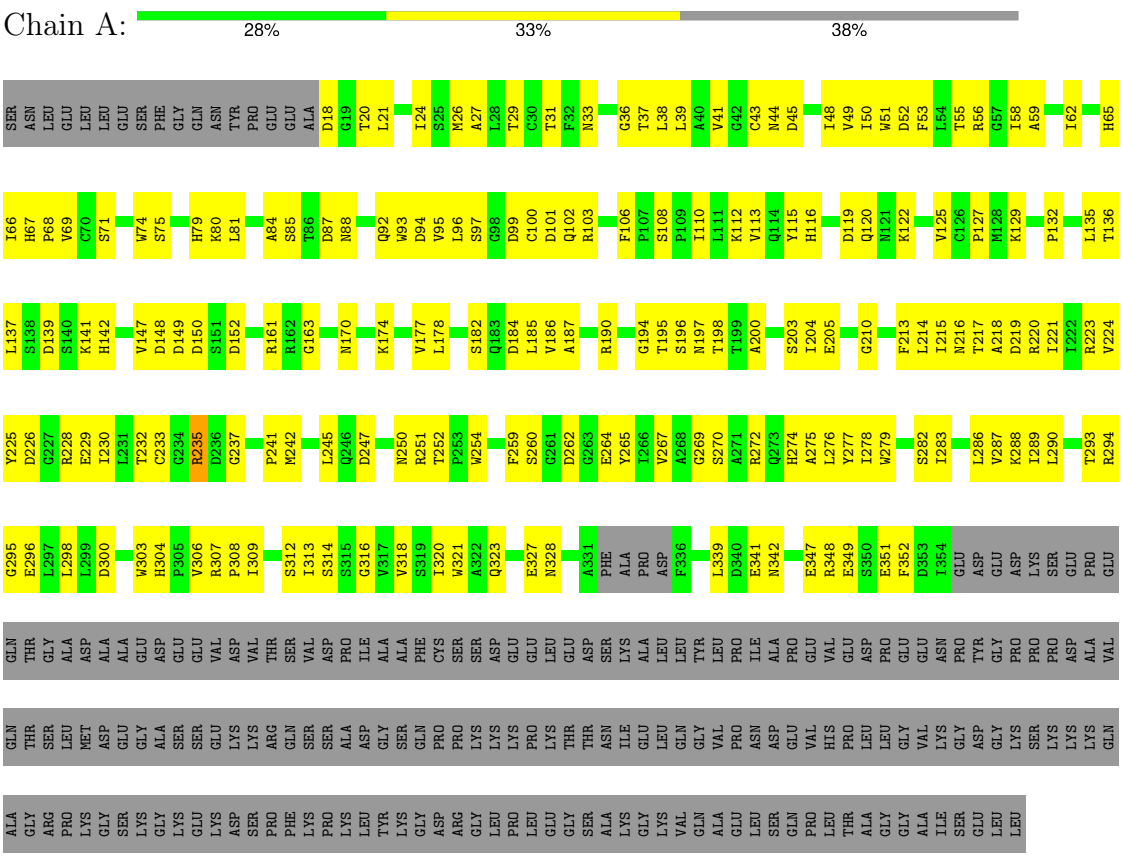
Mol	Chain	Residues	Atoms					AltConf	Trace
10	P	146	Total	C	N	O	P	0	0
			3011	1425	564	876	146		

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.

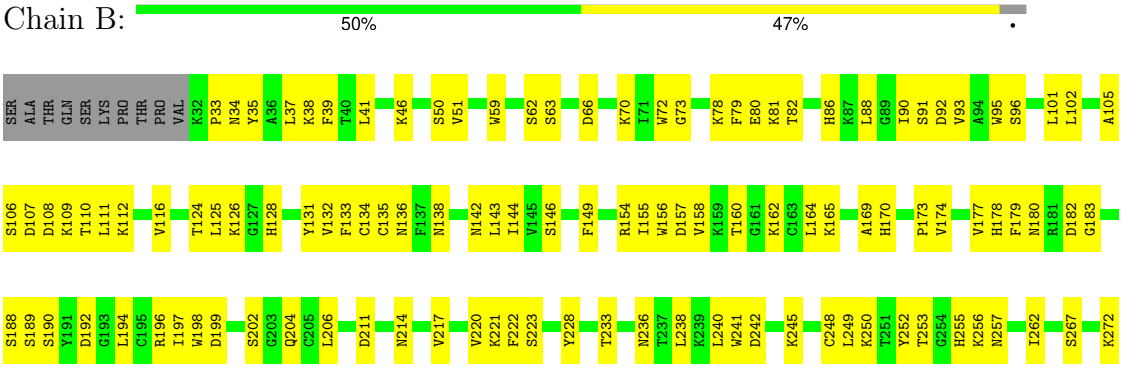
• Molecule 1: Retinoblastoma-binding protein 5

Chain A:



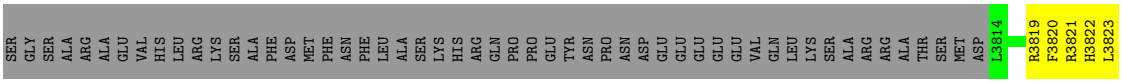
• Molecule 2: WD repeat-containing protein 5

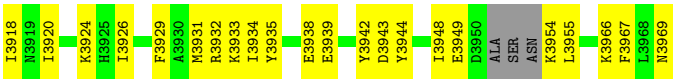
Chain B:



• Molecule 3: Histone-lysine N-methyltransferase 2A

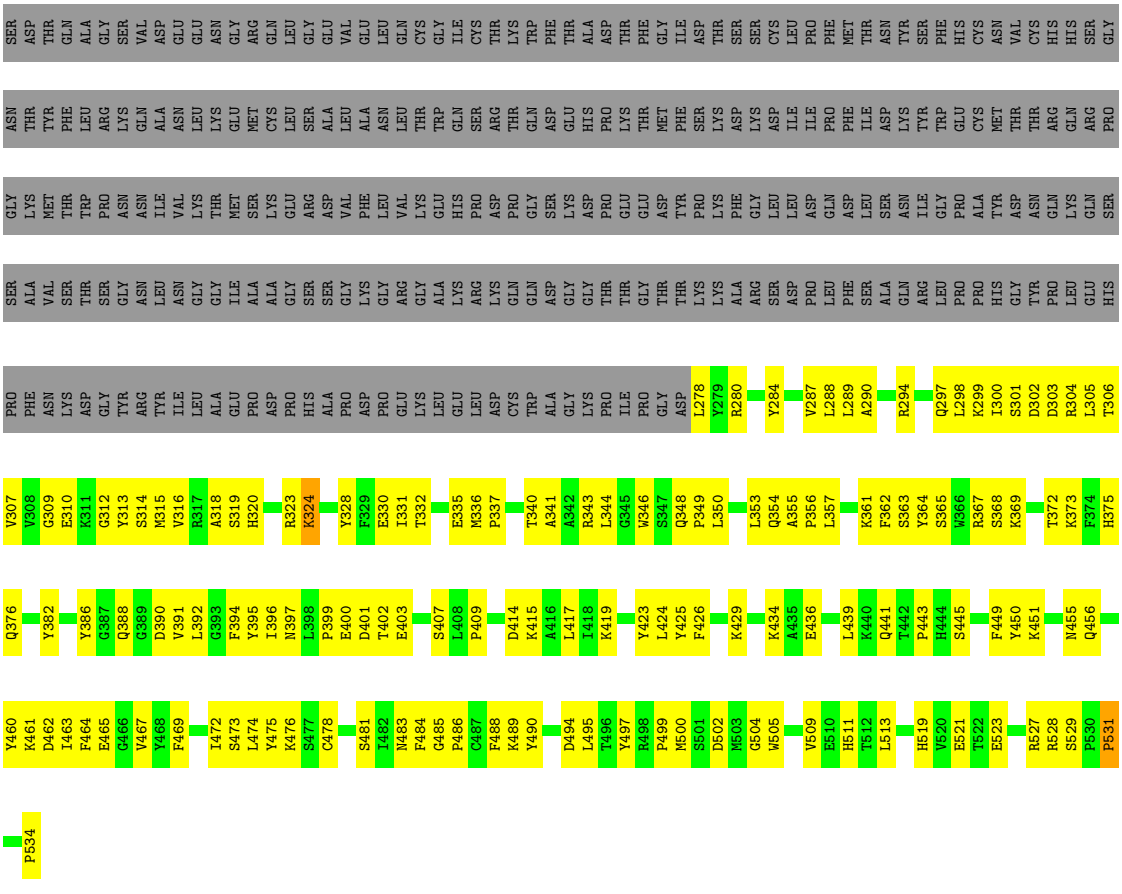
Chain C:





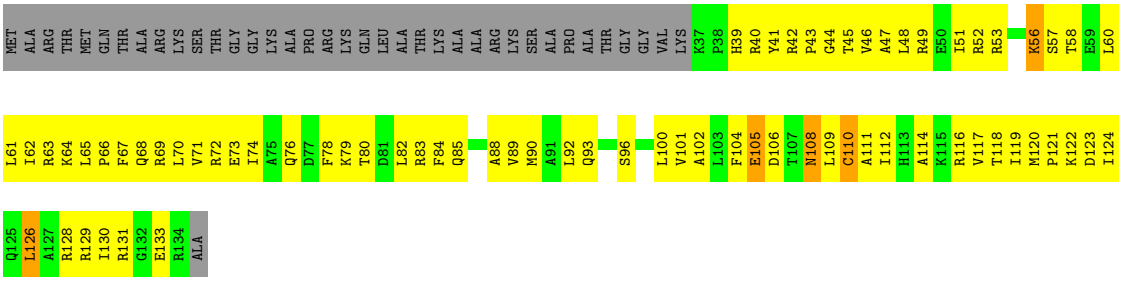
• Molecule 4: Set1/Ash2 histone methyltransferase complex subunit ASH2

Chain D:



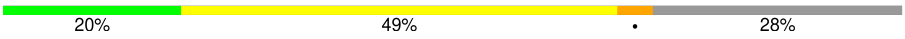
• Molecule 5: Histone H3

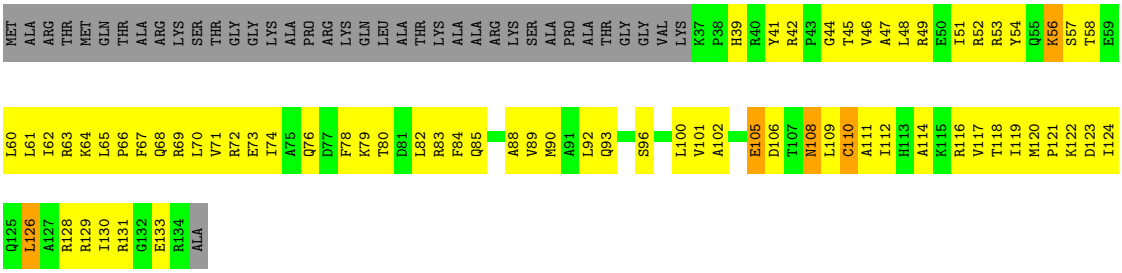
Chain G:



• Molecule 5: Histone H3

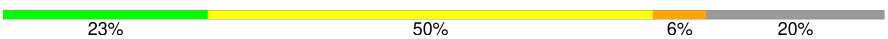
Chain K:

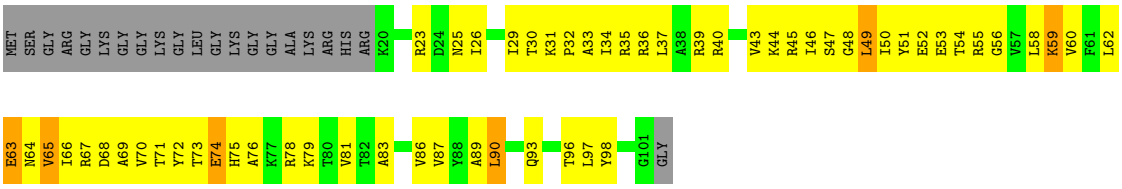




• Molecule 6: Histone H4

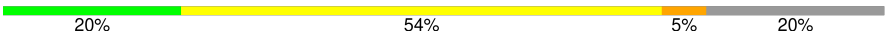
Chain H:

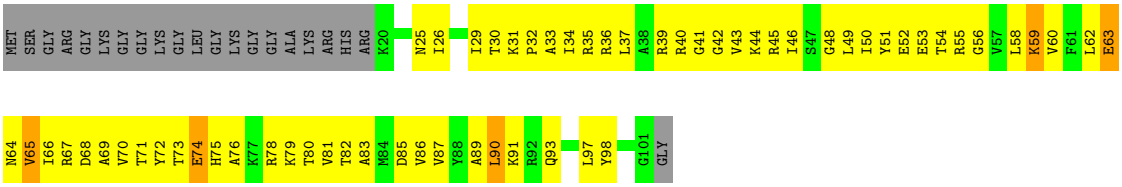




• Molecule 6: Histone H4

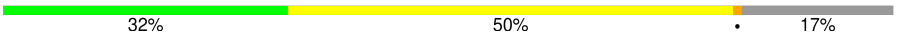
Chain L:

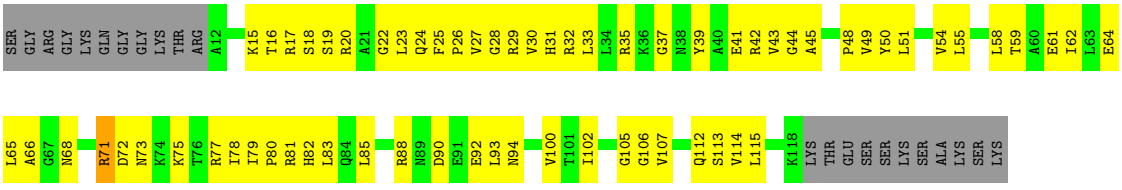




• Molecule 7: Histone H2A

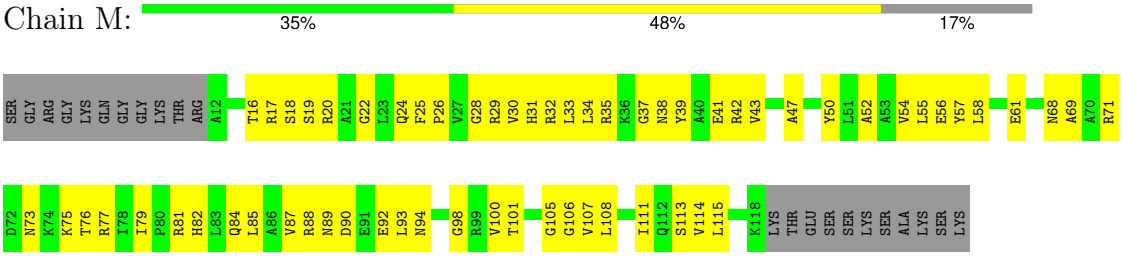
Chain I:





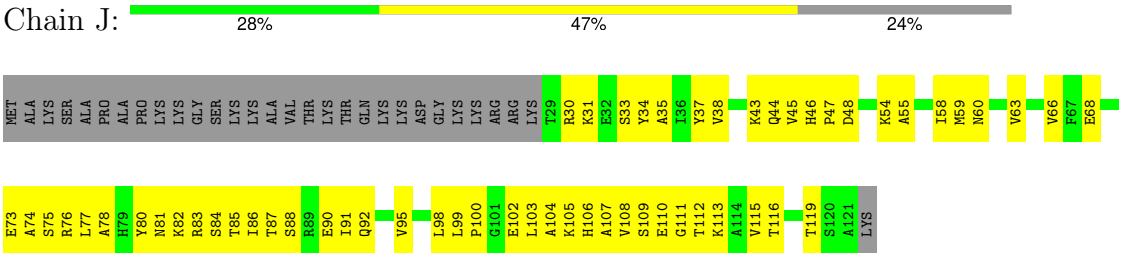
• Molecule 7: Histone H2A

Chain M:



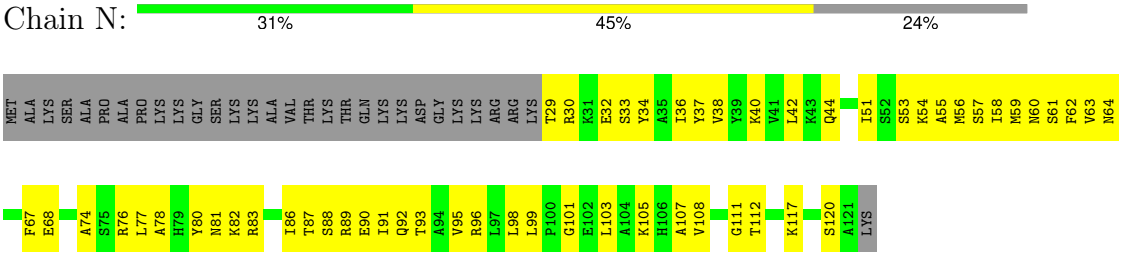
• Molecule 8: Histone H2B 1.1

Chain J:



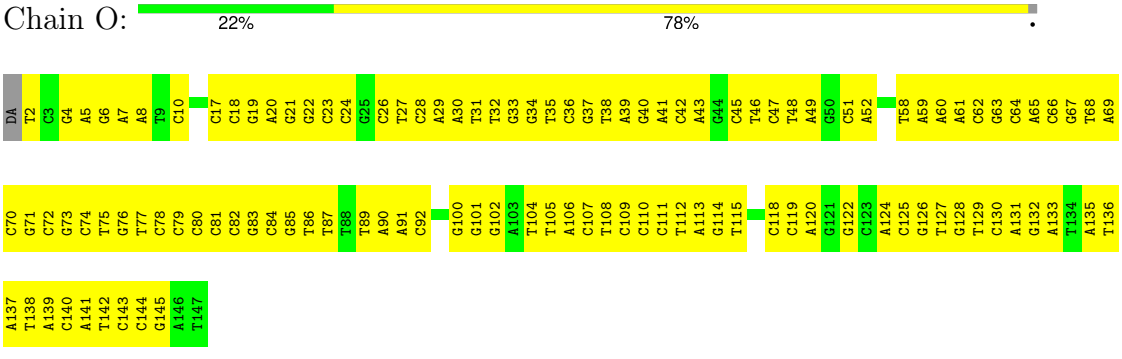
• Molecule 8: Histone H2B 1.1

Chain N:



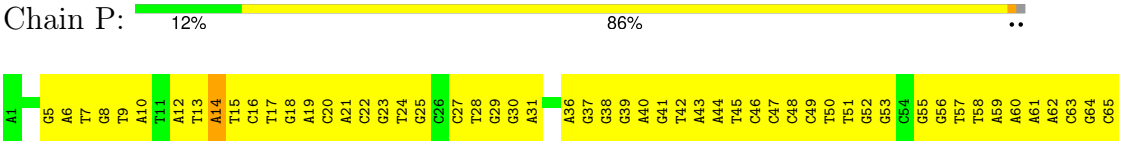
• Molecule 9: DNA (146-MER)

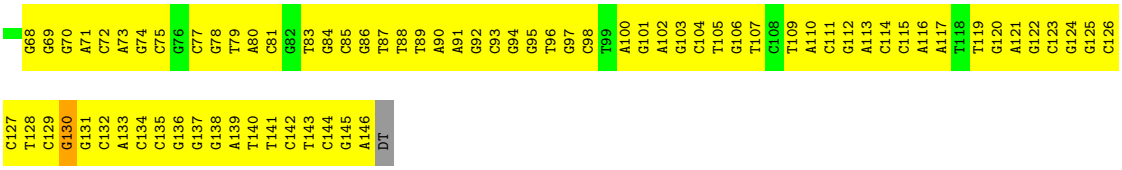
Chain O:



• Molecule 10: DNA (146-MER)

Chain P:




G68 G69 G70 A71 C72 A73 G74 C75 G76 C77 G78 T79 A80 C81 G82 T83 G84 C85 G86 T87 T88 T89 A90 A91 G92 C93 G94 G95 T96 G97 C98 T99 A100 G101 A102 G103 C104 T105 G106 T107 C108 T109 A110 C111 G112 A113 C114 C115 A116 A117 T118 T119 G120 A121 G122 C123 G124 G125 C126
C127 T128 C129 G130 G131 C132 A133 C134 C135 G136 G137 G138 A139 T140 T141 C142 T143 C144 G145 A146 DT

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	30847	Depositor
Resolution determination method	Not provided	
CTF correction method	PHASE FLIPPING ONLY	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	53.4	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

5 Model quality

5.1 Standard geometry

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	0.29	0/2667	0.48	0/3620
2	B	0.28	0/2413	0.47	0/3275
3	C	0.28	0/1253	0.41	0/1678
4	D	0.29	0/2138	0.49	0/2897
5	G	0.30	0/813	0.43	0/1093
5	K	0.30	0/813	0.43	0/1093
6	H	0.33	0/660	0.50	0/885
6	L	0.33	0/660	0.50	0/885
7	I	0.30	0/821	0.46	0/1112
7	M	0.29	0/825	0.45	0/1116
8	J	0.32	0/729	0.47	1/985 (0.1%)
8	N	0.31	0/737	0.45	0/993
9	O	0.75	0/3333	0.95	0/5137
10	P	0.77	0/3381	0.93	2/5221 (0.0%)
All	All	0.49	0/21243	0.67	3/29990 (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
1	A	0	1
4	D	0	1
All	All	0	2

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
8	J	98	LEU	C-N-CA	-5.37	108.28	121.70
10	P	14	DA	O4'-C4'-C3'	-5.20	102.42	104.50
10	P	130	DG	O4'-C4'-C3'	-5.15	102.44	104.50

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	328	ASN	Peptide
4	D	531	PRO	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	2613	0	2582	164	0
2	B	2342	0	2298	115	0
3	C	1228	0	1227	62	0
4	D	2075	0	2021	139	0
5	G	801	0	830	147	0
5	K	801	0	831	127	0
6	H	653	0	695	108	0
6	L	653	0	695	100	0
7	I	811	0	849	78	0
7	M	815	0	860	108	0
8	J	718	0	725	68	0
8	N	726	0	747	84	0
9	O	2975	0	1639	235	0
10	P	3011	0	1639	251	0
All	All	20222	0	17638	1364	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 37.

The worst 5 of 1364 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)
5:G:69:ARG:NH1	9:O:91:DA:OP2	1.60	1.34
6:H:78:ARG:HB3	9:O:102:DG:OP1	1.27	1.31
1:A:348:ARG:CA	4:D:475:TYR:OH	1.88	1.21
6:H:36:ARG:HH22	10:P:61:DA:P	1.65	1.19
1:A:348:ARG:HA	4:D:475:TYR:OH	1.36	1.18

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	329/538 (61%)	307 (93%)	22 (7%)	0	100	100
2	B	304/313 (97%)	287 (94%)	17 (6%)	0	100	100
3	C	149/209 (71%)	144 (97%)	5 (3%)	0	100	100
4	D	255/534 (48%)	231 (91%)	24 (9%)	0	100	100
5	G	96/136 (71%)	92 (96%)	4 (4%)	0	100	100
5	K	96/136 (71%)	92 (96%)	4 (4%)	0	100	100
6	H	80/103 (78%)	70 (88%)	10 (12%)	0	100	100
6	L	80/103 (78%)	70 (88%)	10 (12%)	0	100	100
7	I	105/129 (81%)	102 (97%)	3 (3%)	0	100	100
7	M	105/129 (81%)	101 (96%)	4 (4%)	0	100	100
8	J	91/123 (74%)	86 (94%)	5 (6%)	0	100	100
8	N	91/123 (74%)	89 (98%)	2 (2%)	0	100	100
All	All	1781/2576 (69%)	1671 (94%)	110 (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	291/462 (63%)	288 (99%)	3 (1%)	73 81
2	B	263/274 (96%)	263 (100%)	0	100 100
3	C	132/182 (72%)	132 (100%)	0	100 100
4	D	222/460 (48%)	221 (100%)	1 (0%)	86 90
5	G	84/111 (76%)	78 (93%)	6 (7%)	12 35
5	K	84/111 (76%)	78 (93%)	6 (7%)	12 35
6	H	67/79 (85%)	61 (91%)	6 (9%)	8 27
6	L	67/79 (85%)	61 (91%)	6 (9%)	8 27
7	I	81/101 (80%)	80 (99%)	1 (1%)	67 79
7	M	82/101 (81%)	82 (100%)	0	100 100
8	J	77/103 (75%)	77 (100%)	0	100 100
8	N	79/103 (77%)	78 (99%)	1 (1%)	65 77
All	All	1529/2166 (71%)	1499 (98%)	30 (2%)	50 68

5 of 30 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
6	H	74	GLU
6	L	74	GLU
5	K	56	LYS
8	N	42	LEU
6	L	59	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 37 such sidechains are listed below:

Mol	Chain	Res	Type
5	K	68	GLN
8	N	60	ASN
6	L	25	ASN
7	M	112	GLN
4	D	483	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 oligosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

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 Map visualisation

This section contains visualisations of the EMDB entry EMD-23739. These allow visual inspection of the internal detail of the map and identification of artifacts.

No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections

This section was not generated.

6.2 Central slices

This section was not generated.

6.3 Largest variance slices

This section was not generated.

6.4 Orthogonal standard-deviation projections (False-color)

This section was not generated.

6.5 Orthogonal surface views

This section was not generated.

6.6 Mask visualisation

This section was not generated. No masks/segmentation were deposited.

7 Map analysis

This section contains the results of statistical analysis of the map.

7.1 Map-value distribution

This section was not generated.

7.2 Volume estimate versus contour level

This section was not generated.

7.3 Rotationally averaged power spectrum

This section was not generated. The rotationally averaged power spectrum had issues being displayed.

8 Fourier-Shell correlation (i)

This section was not generated. No FSC curve or half-maps provided.

9 Map-model fit

This section was not generated.