



wwPDB EM Validation Summary Report ⓘ

Oct 6, 2022 – 05:04 PM EDT

PDB ID : 6W5I
EMDB ID : EMD-21542
Title : Cryo-EM structure of MLL1 in complex with RbBP5, WDR5, SET1, and ASH2L bound to the nucleosome (Class01)
Authors : Park, S.H.; Lee, Y.T.; Ayoub, A.; Dou, Y.; Cho, U.
Deposited on : 2020-03-13
Resolution : 6.90 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 : 0.0.1.dev43
MolProbity : 4.02b-467
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ : 1.9.9
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.31.2

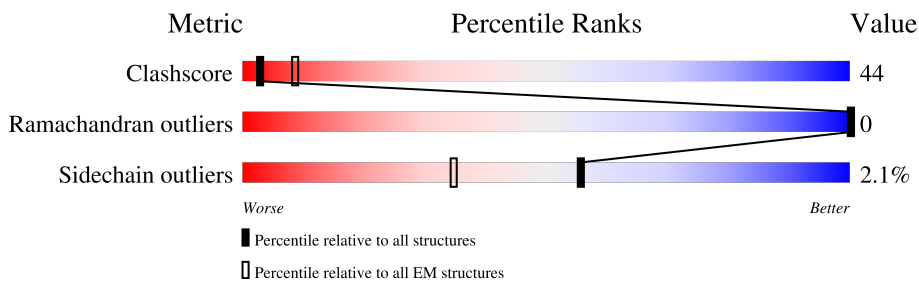
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 6.90 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.



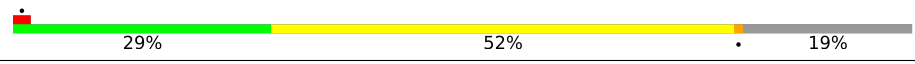
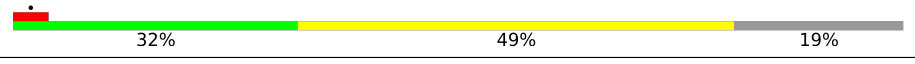
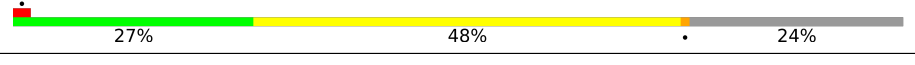
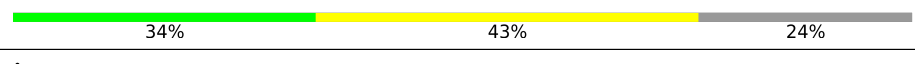
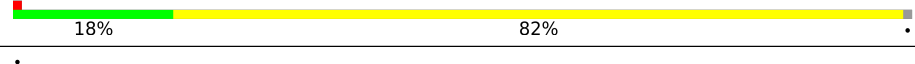
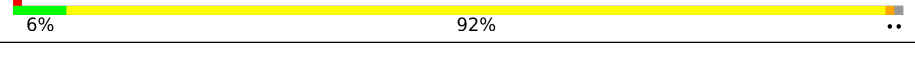
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

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 $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the EM map (all-atom inclusion $< 40\%$). The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	538	28% (Poor fit) 26% (0 outliers), 31% (1 outlier), 43% (2+ outliers or not modelled)
2	B	313	41% (0 outliers), 50% (1 outlier), 47% (2+ outliers or not modelled)
3	C	209	24% (0 outliers), 42% (1 outlier), 31% (2+ outliers or not modelled), 27% (not modelled)
4	D	534	8% (Poor fit), 19% (0 outliers), 21% (1 outlier), 59% (2+ outliers or not modelled)
5	G	136	15% (0 outliers), 53% (1 outlier), 28% (2+ outliers or not modelled)
5	K	136	21% (0 outliers), 48% (1 outlier), 28% (2+ outliers or not modelled)
6	H	103	24% (0 outliers), 51% (1 outlier), 20% (2+ outliers or not modelled)
6	L	103	26% (0 outliers), 49% (1 outlier), 5% (2+ outliers or not modelled), 20% (not modelled)

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
7	I	132	 29% 52% 19%
7	M	132	 32% 49% 19%
8	J	122	 27% 48% 24%
8	N	122	 34% 43% 24%
9	O	147	 18% 82%
10	P	147	 6% 92%

2 Entry composition

There are 10 unique types of molecules in this entry. The entry contains 19672 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
			Total	C	N	O	S		
1	A	308	2396	1512	424	445	15	0	0

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
			Total	C	N	O	S		
2	B	303	2342	1492	388	451	11	3	0

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

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	C	152	1220	774	217	216	13	0	0

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
			Total	C	N	O	S		
4	D	218	1750	1130	296	317	7	0	0

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 H3.2.

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

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	102	ALA	GLY	engineered mutation	UNP P84233
K	102	ALA	GLY	engineered mutation	UNP P84233

- Molecule 6 is a protein called Histone H4.

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

- Molecule 7 is a protein called Histone H2A.

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

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
I	-2	SER	-	expression tag	UNP Q6AZJ8
I	-1	ASN	-	expression tag	UNP Q6AZJ8
I	0	ALA	-	expression tag	UNP Q6AZJ8
M	-2	SER	-	expression tag	UNP Q6AZJ8
M	-1	ASN	-	expression tag	UNP Q6AZJ8
M	0	ALA	-	expression tag	UNP Q6AZJ8

- Molecule 8 is a protein called Histone H2B 1.1.

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

There are 2 discrepancies between the modelled and reference sequences:

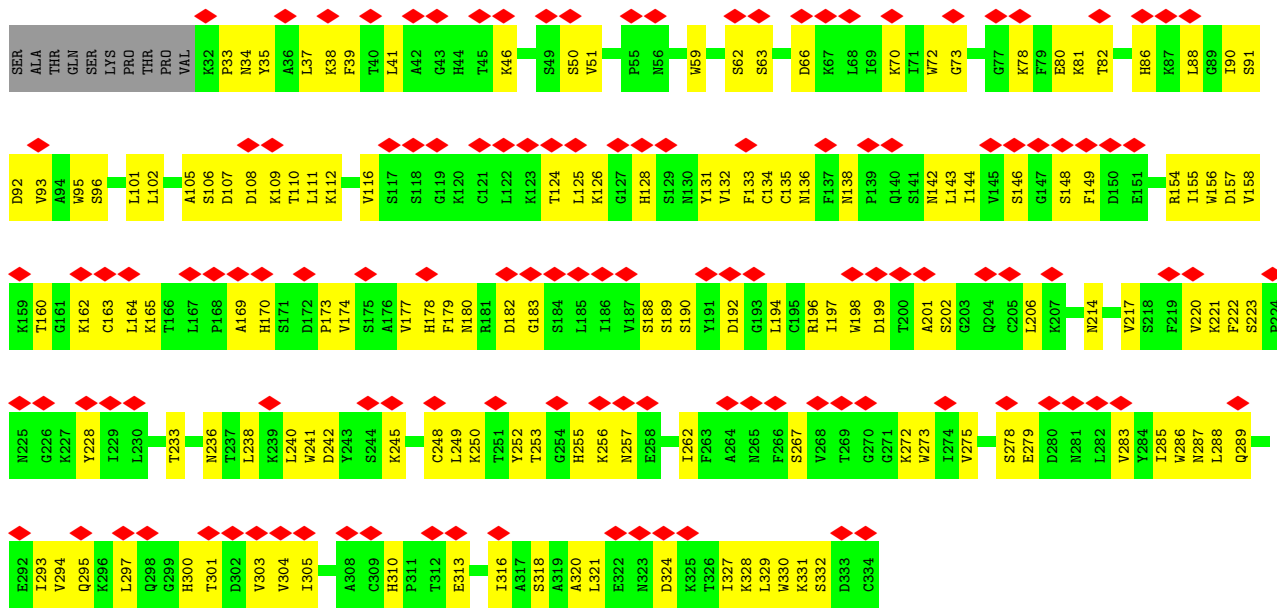
Chain	Residue	Modelled	Actual	Comment	Reference
J	29	THR	SER	conflict	UNP P02281
N	29	THR	SER	conflict	UNP P02281

- Molecule 9 is a DNA chain called DNA (147-MER).

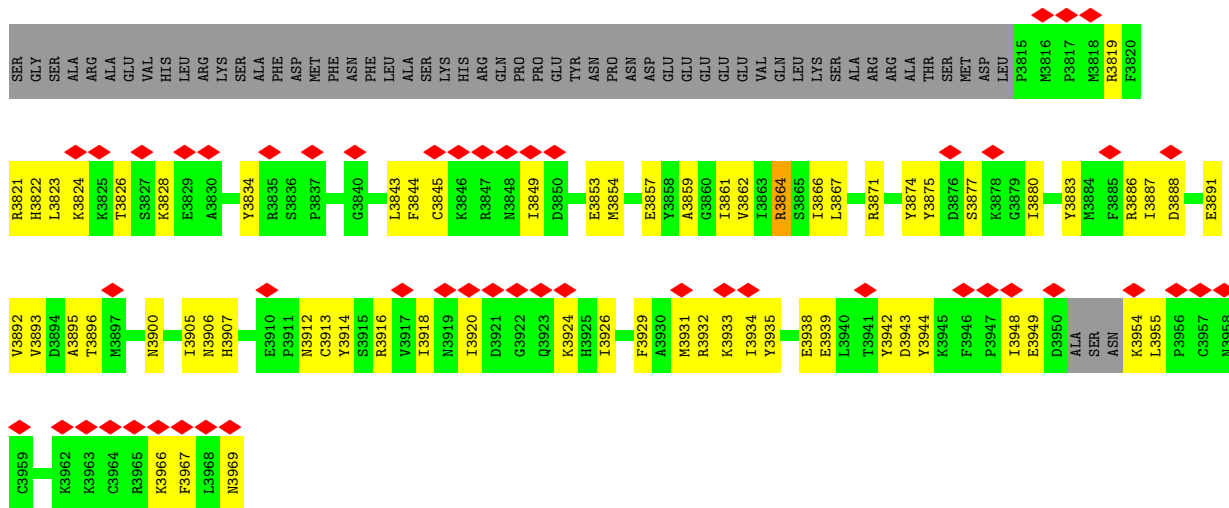
Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
9	O	146	2975	1413	540	876	146	0	0

- Molecule 10 is a DNA chain called DNA (147-MER).

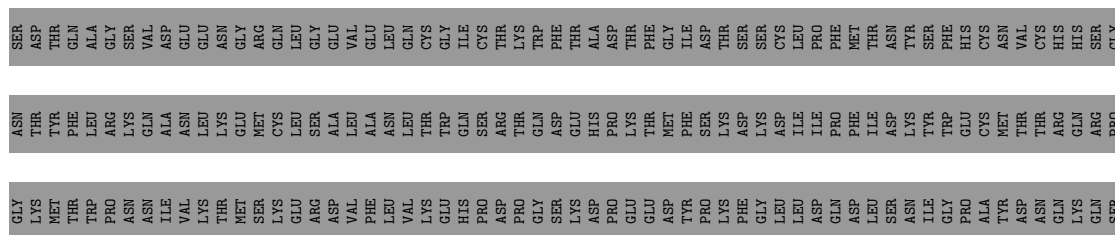
Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
10	P	146	3011	1425	564	876	146	0	0

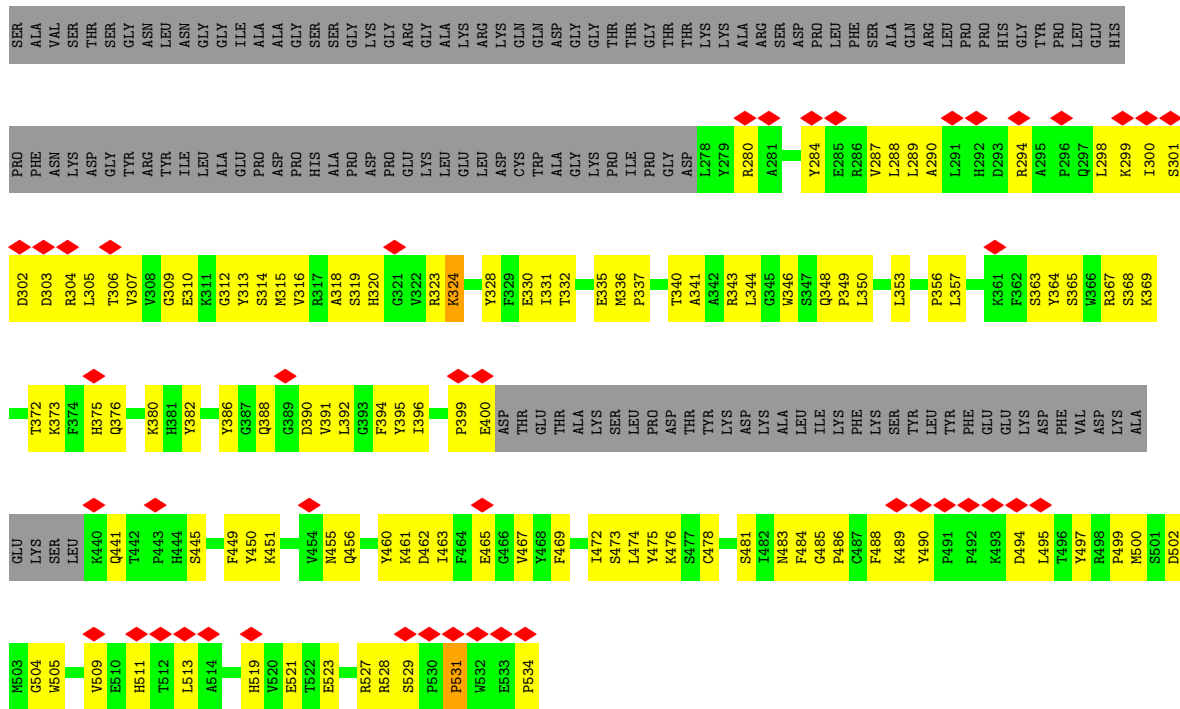


• Molecule 3: Histone-lysine N-methyltransferase 2A

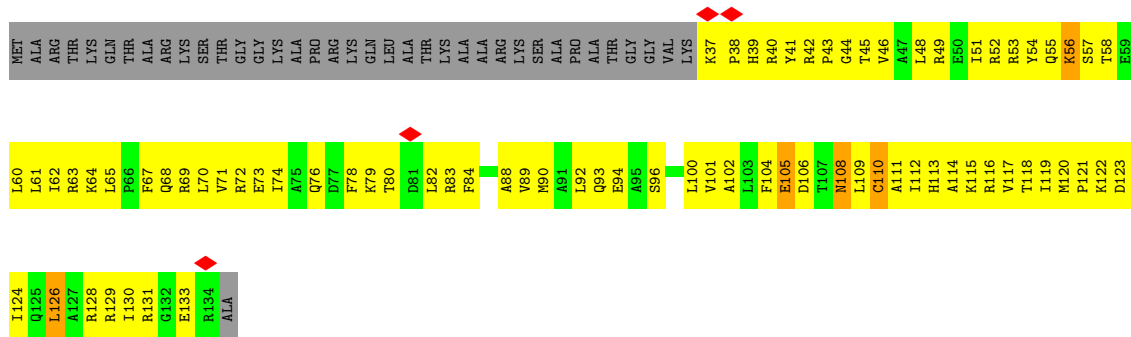
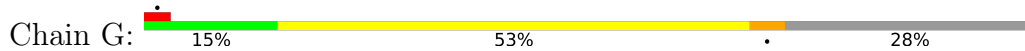


• Molecule 4: Set1/Ash2 histone methyltransferase complex subunit ASH2

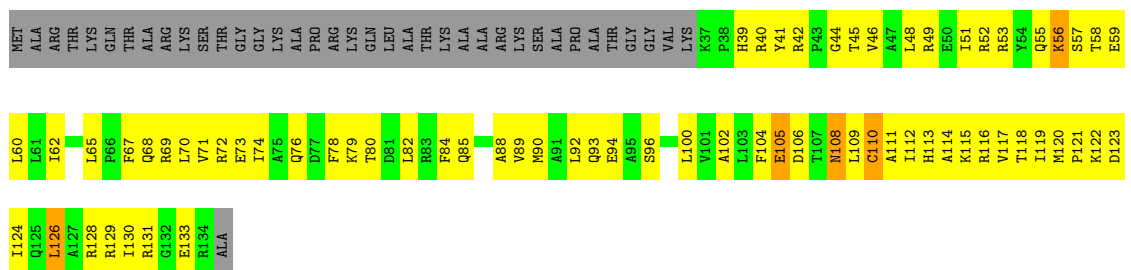
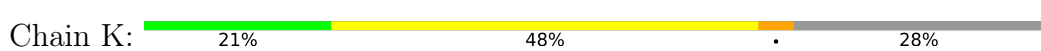




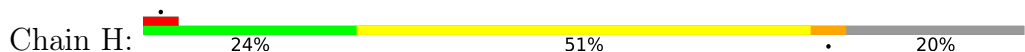
• Molecule 5: Histone H3.2

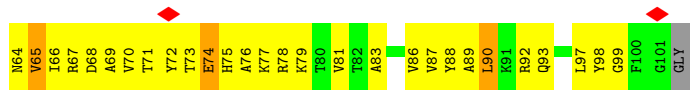
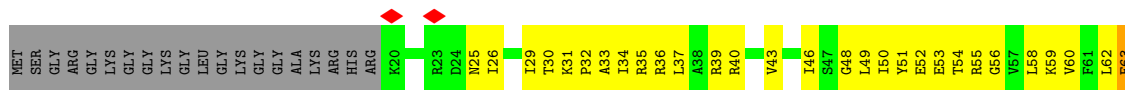


• Molecule 5: Histone H3.2

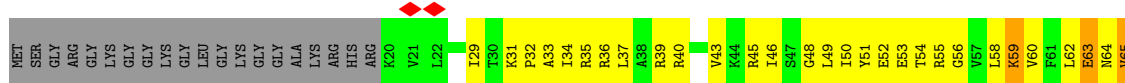


• Molecule 6: Histone H4

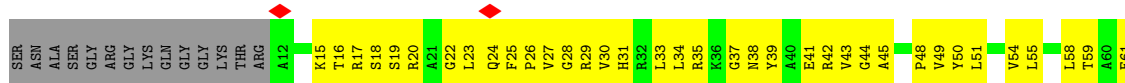




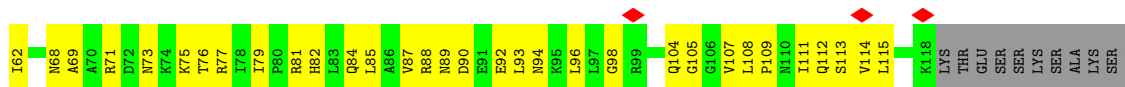
• Molecule 6: Histone H4



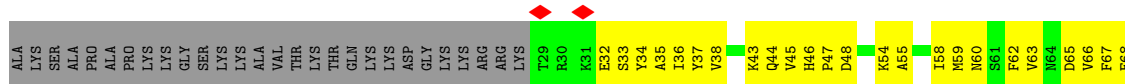
• Molecule 7: Histone H2A

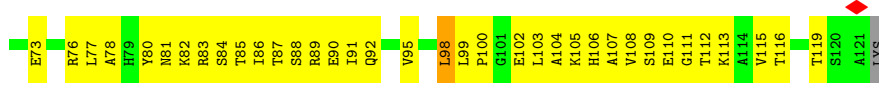


• Molecule 7: Histone H2A

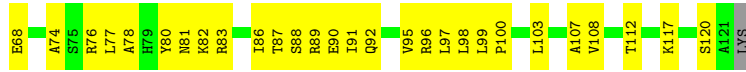
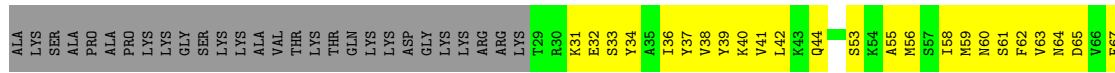
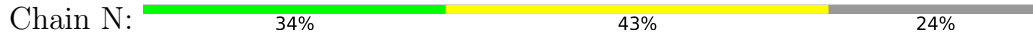


• Molecule 8: Histone H2B 1.1

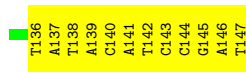




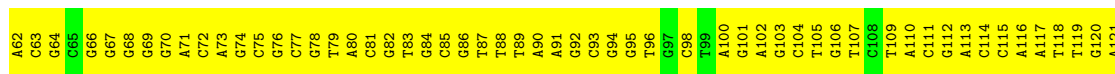
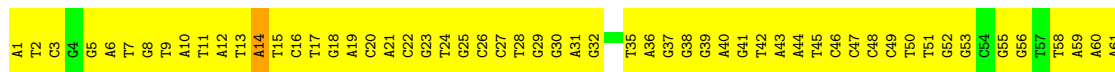
• Molecule 8: Histone H2B 1.1



• Molecule 9: DNA (147-MER)



• Molecule 10: DNA (147-MER)



4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	13086	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 ($e^-/\text{\AA}^2$)	64	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	0.068	Depositor
Minimum map value	-0.032	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.002	Depositor
Recommended contour level	0.0086	Depositor
Map size (\AA)	300.0, 300.0, 300.0	wwPDB
Map dimensions	300, 300, 300	wwPDB
Map angles ($^\circ$)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (\AA)	1.0, 1.0, 1.0	Depositor

5 Model quality [i](#)

5.1 Standard geometry [i](#)

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	0/2444	0.48	0/3315
2	B	0.28	0/2413	0.47	0/3275
3	C	0.28	0/1245	0.42	0/1666
4	D	0.29	0/1805	0.49	0/2449
5	G	0.31	0/813	0.43	0/1093
5	K	0.31	0/813	0.43	0/1093
6	H	0.33	0/660	0.50	0/885
6	L	0.34	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	0.50	0/20679	0.67	3/29225 (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
4	D	0	1

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
8	J	98	LEU	C-N-CA	-5.36	108.31	121.70
10	P	14	DA	O4'-C4'-C3'	-5.25	102.40	104.50
10	P	130	DG	O4'-C4'-C3'	-5.21	102.42	104.50

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
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	2396	0	2397	139	0
2	B	2342	0	2298	126	0
3	C	1220	0	1217	68	0
4	D	1750	0	1696	106	0
5	G	801	0	831	157	0
5	K	801	0	831	117	0
6	H	653	0	695	135	0
6	L	653	0	695	89	0
7	I	811	0	849	113	0
7	M	815	0	860	154	0
8	J	718	0	725	140	0
8	N	726	0	747	110	0
9	O	2975	0	1638	381	0
10	P	3011	0	1639	450	0
All	All	19672	0	17118	1599	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 44.

The worst 5 of 1599 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)
3:C:3866:ILE:HG21	4:D:313:TYR:CE2	1.55	1.40
7:M:77:ARG:HG2	10:P:132:DC:OP1	1.20	1.33
9:O:76:DG:O6	10:P:71:DA:N6	1.64	1.30
9:O:2:DT:C6	10:P:146:DA:C2	2.22	1.27
9:O:2:DT:H71	10:P:146:DA:C4	1.72	1.24

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	302/538 (56%)	285 (94%)	17 (6%)	0	100	100
2	B	304/313 (97%)	287 (94%)	17 (6%)	0	100	100
3	C	148/209 (71%)	143 (97%)	5 (3%)	0	100	100
4	D	214/534 (40%)	196 (92%)	18 (8%)	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	105/132 (80%)	102 (97%)	3 (3%)	0	100	100
7	M	105/132 (80%)	101 (96%)	4 (4%)	0	100	100
8	J	91/122 (75%)	86 (94%)	5 (6%)	0	100	100
8	N	91/122 (75%)	89 (98%)	2 (2%)	0	100	100
All	All	1712/2580 (66%)	1613 (94%)	99 (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	268/462 (58%)	265 (99%)	3 (1%)	73	84
2	B	263/274 (96%)	263 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	C	131/182 (72%)	130 (99%)	1 (1%)	81	89
4	D	186/460 (40%)	185 (100%)	1 (0%)	88	93
5	G	84/111 (76%)	78 (93%)	6 (7%)	14	39
5	K	84/111 (76%)	78 (93%)	6 (7%)	14	39
6	H	67/79 (85%)	61 (91%)	6 (9%)	9	30
6	L	67/79 (85%)	61 (91%)	6 (9%)	9	30
7	I	81/103 (79%)	80 (99%)	1 (1%)	71	83
7	M	82/103 (80%)	82 (100%)	0	100	100
8	J	77/102 (76%)	77 (100%)	0	100	100
8	N	79/102 (78%)	78 (99%)	1 (1%)	69	82
All	All	1469/2168 (68%)	1438 (98%)	31 (2%)	56	72

5 of 31 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
6	H	65	VAL
6	L	65	VAL
7	I	71	ARG
6	L	90	LEU
6	L	49	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 38 such sidechains are listed below:

Mol	Chain	Res	Type
6	L	25	ASN
8	N	46	HIS
7	M	31	HIS
7	M	104	GLN
8	N	64	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](#)

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.

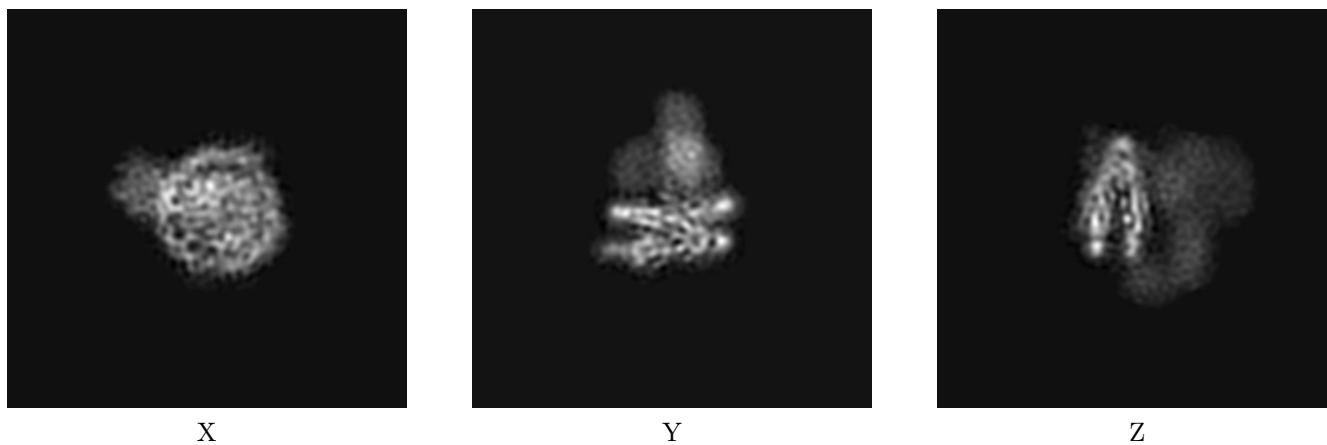
6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-21542. 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 [i](#)

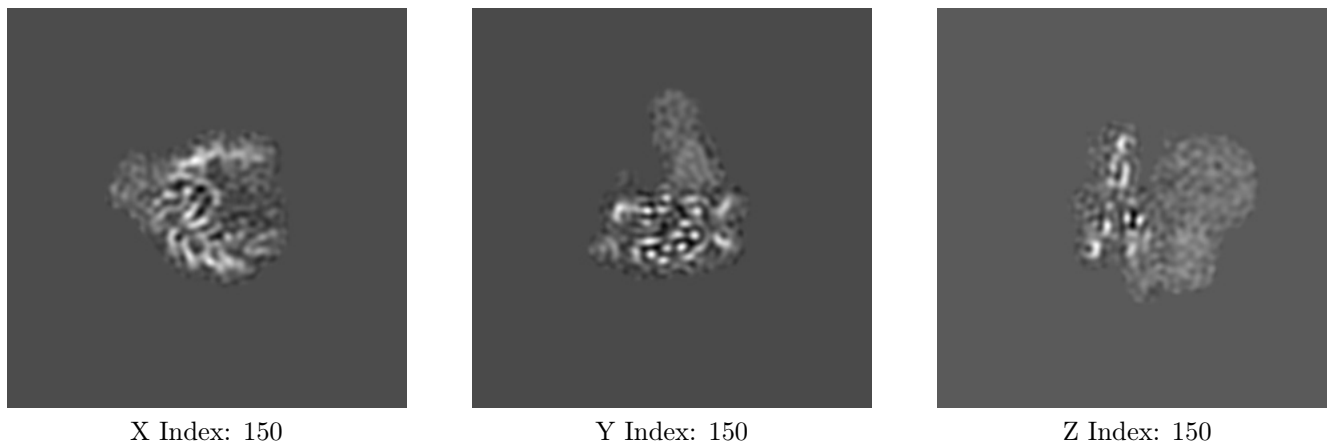
6.1.1 Primary map



The images above show the map projected in three orthogonal directions.

6.2 Central slices [i](#)

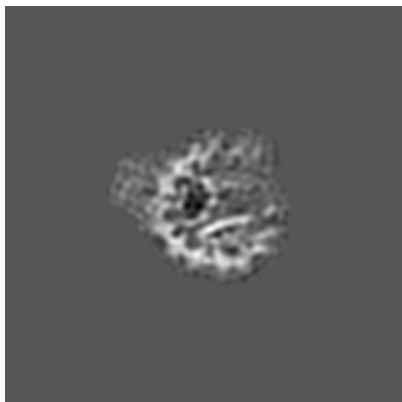
6.2.1 Primary map



The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices [i](#)

6.3.1 Primary map



X Index: 147



Y Index: 142

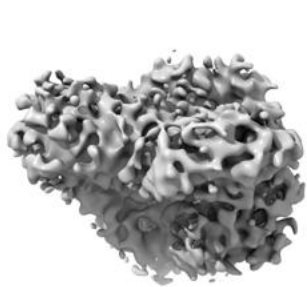


Z Index: 162

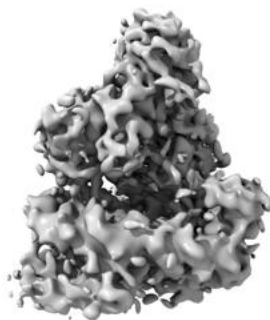
The images above show the largest variance slices of the map in three orthogonal directions.

6.4 Orthogonal surface views [i](#)

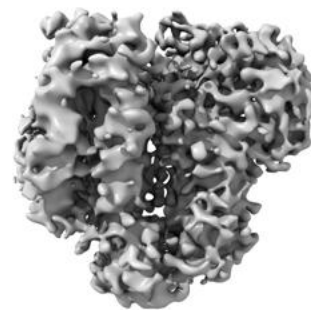
6.4.1 Primary map



X



Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.0086. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

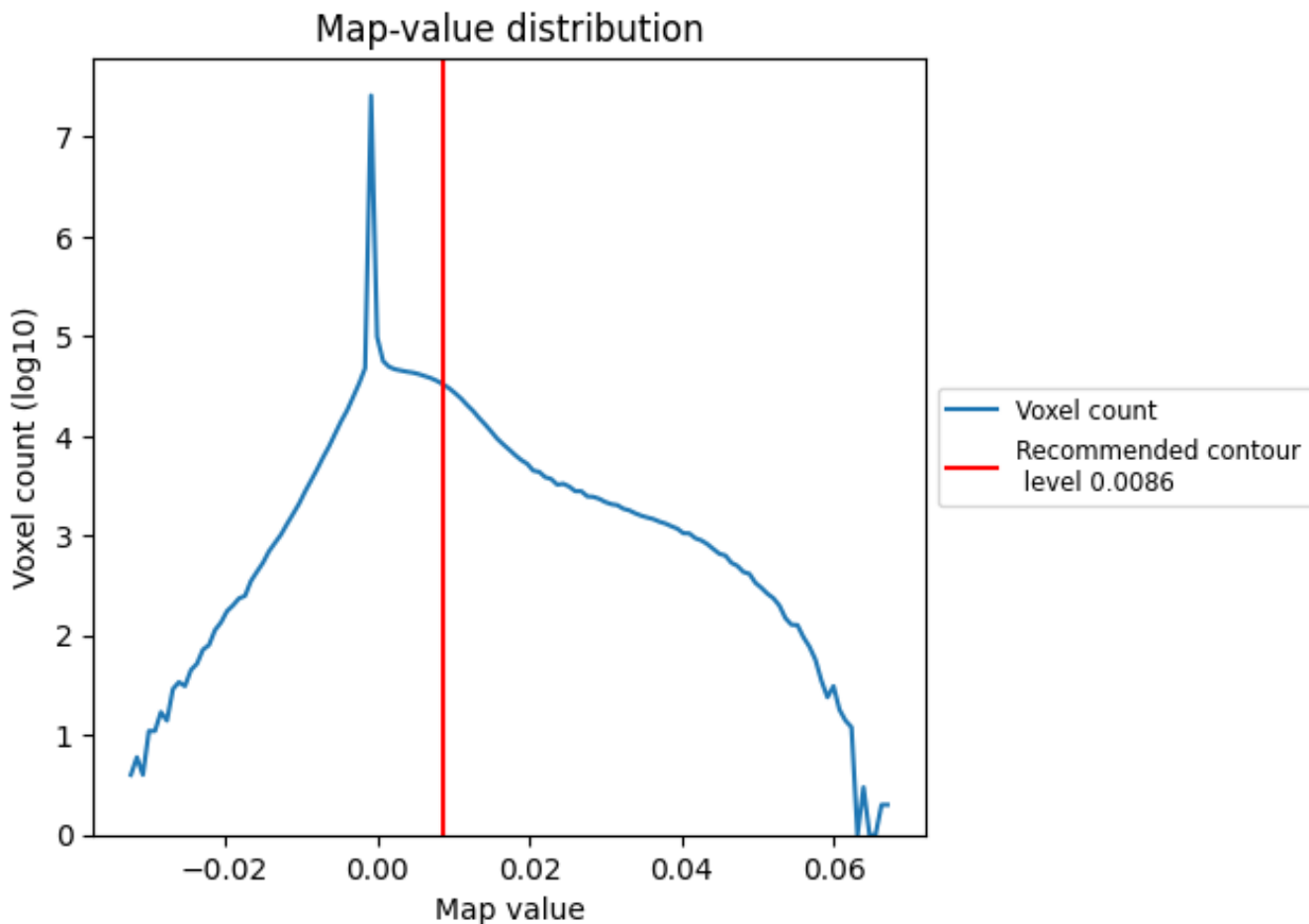
6.5 Mask visualisation

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

7 Map analysis [i](#)

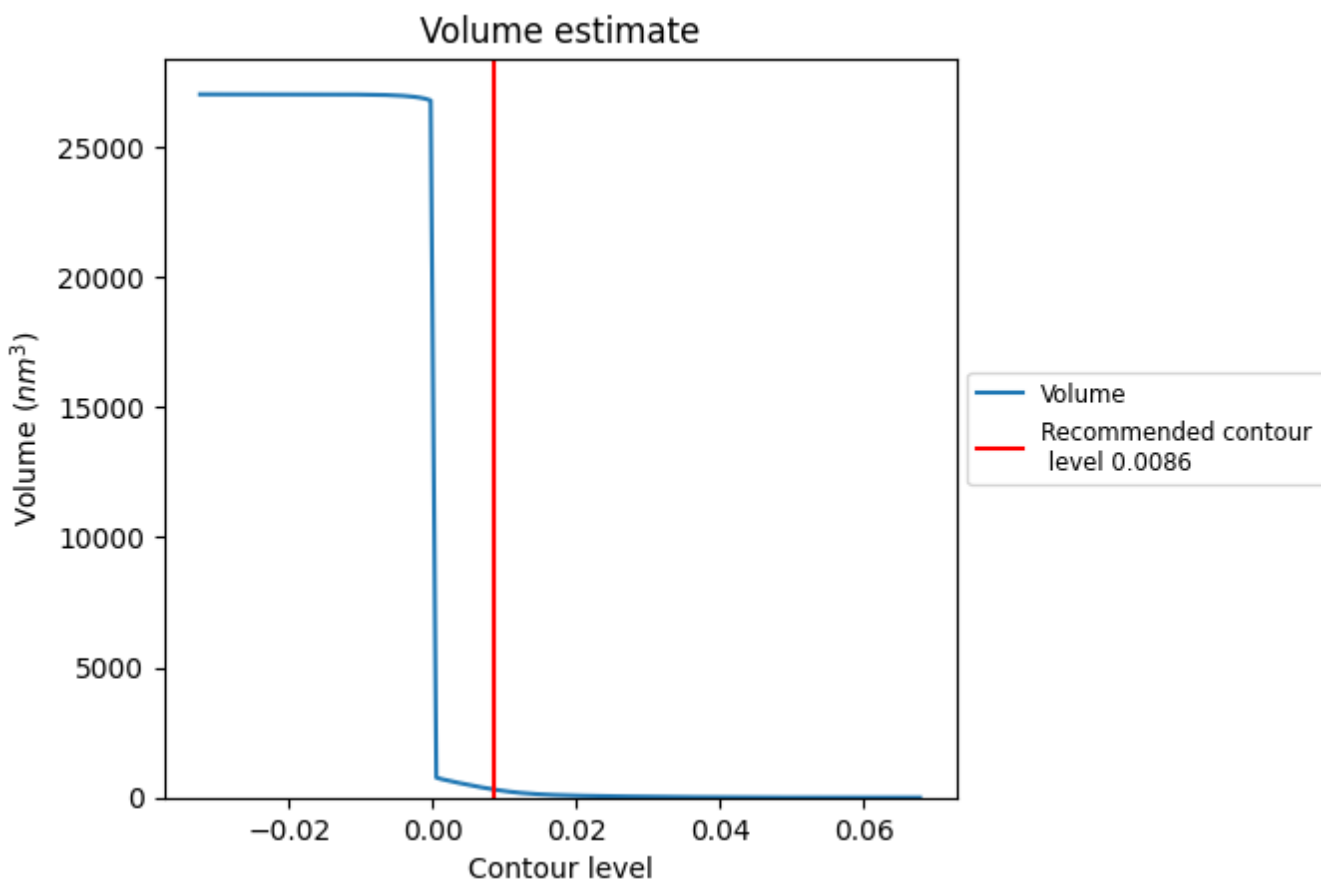
This section contains the results of statistical analysis of the map.

7.1 Map-value distribution [i](#)



The map-value distribution is plotted in 128 intervals along the x-axis. The y-axis is logarithmic. A spike in this graph at zero usually indicates that the volume has been masked.

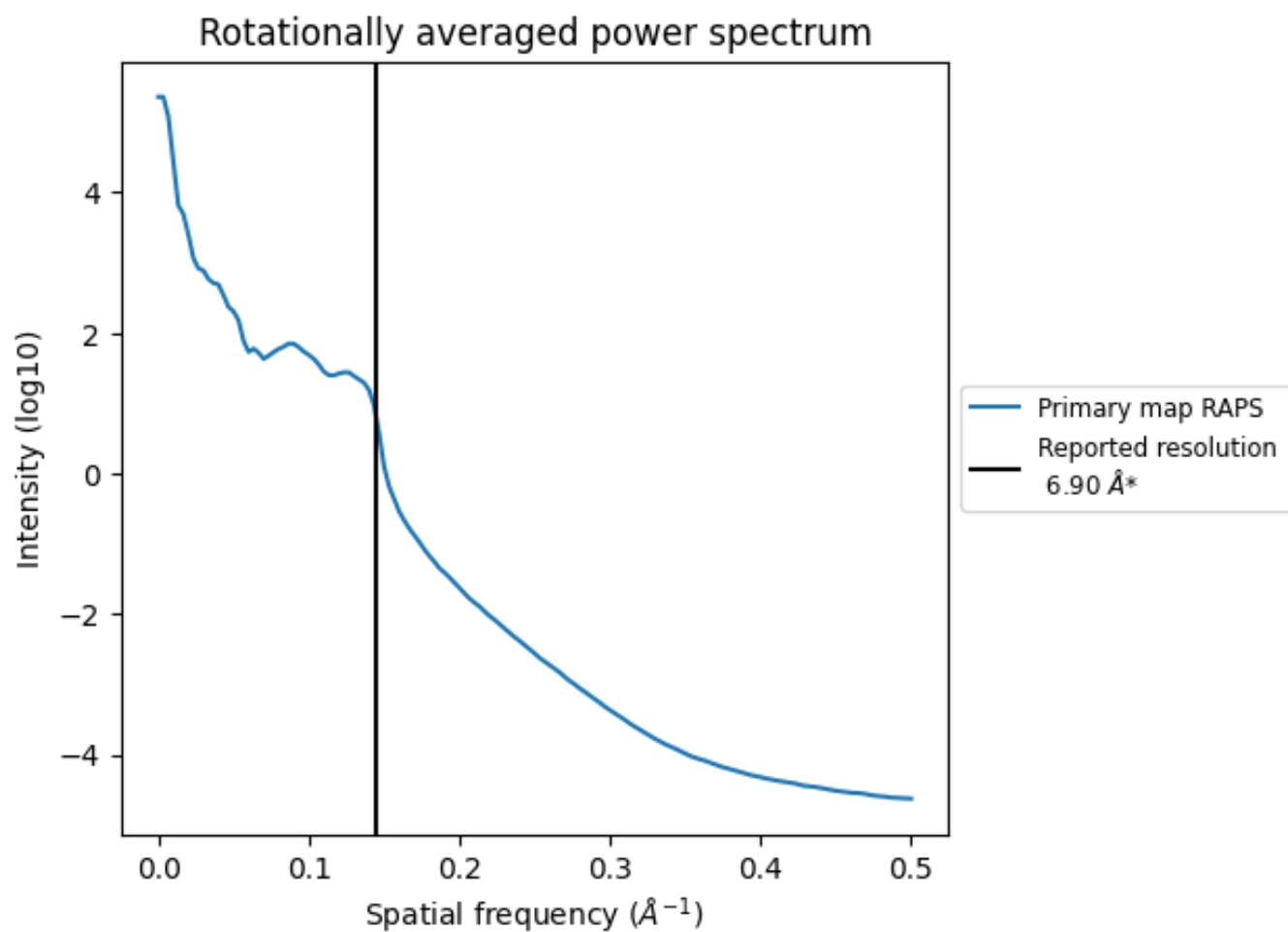
7.2 Volume estimate [\(i\)](#)



The volume at the recommended contour level is 309 nm³; this corresponds to an approximate mass of 279 kDa.

The volume estimate graph shows how the enclosed volume varies with the contour level. The recommended contour level is shown as a vertical line and the intersection between the line and the curve gives the volume of the enclosed surface at the given level.

7.3 Rotationally averaged power spectrum [i](#)

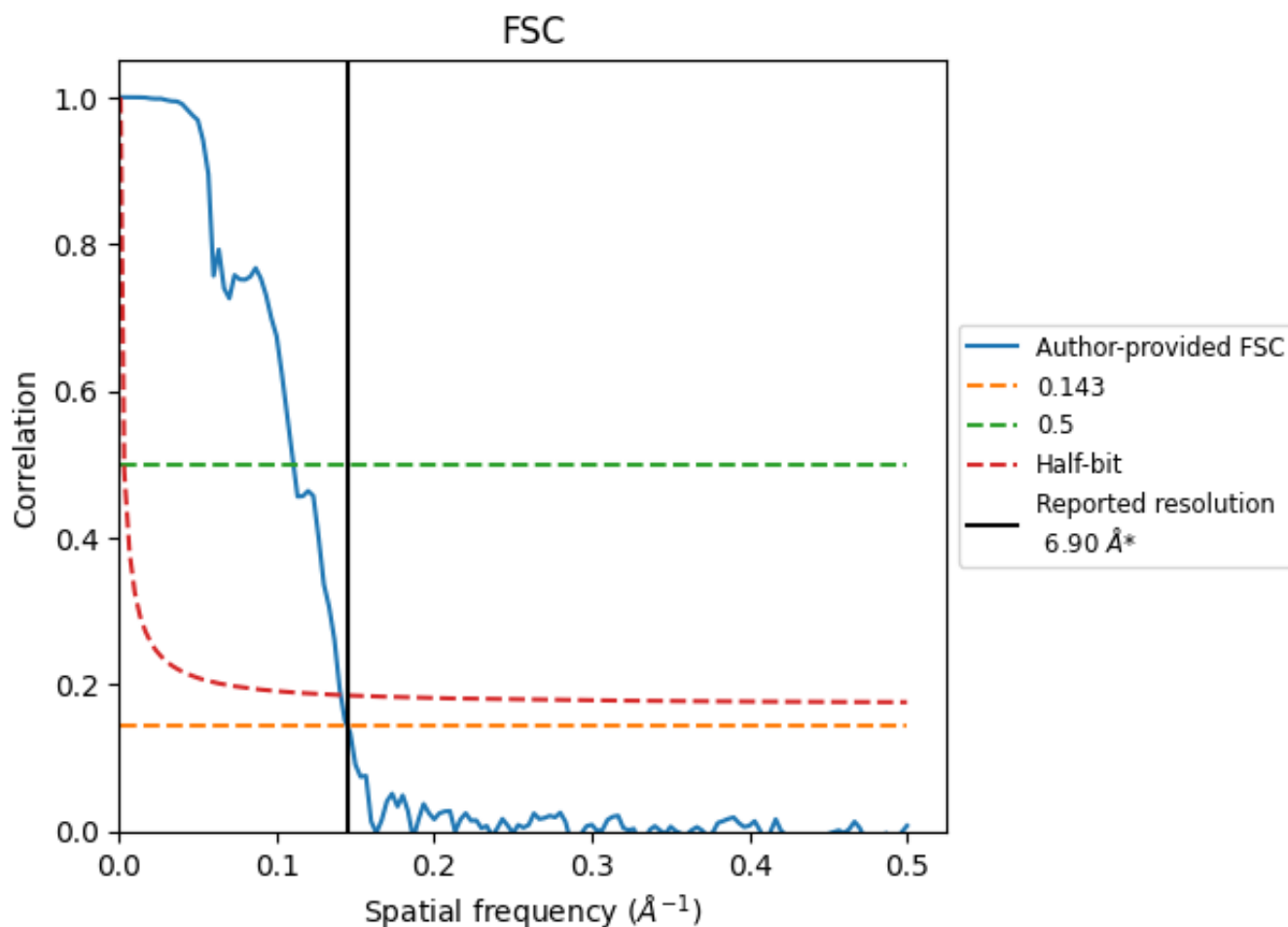


*Reported resolution corresponds to spatial frequency of 0.145\AA^{-1}

8 Fourier-Shell correlation [i](#)

Fourier-Shell Correlation (FSC) is the most commonly used method to estimate the resolution of single-particle and subtomogram-averaged maps. The shape of the curve depends on the imposed symmetry, mask and whether or not the two 3D reconstructions used were processed from a common reference. The reported resolution is shown as a black line. A curve is displayed for the half-bit criterion in addition to lines showing the 0.143 gold standard cut-off and 0.5 cut-off.

8.1 FSC [i](#)



*Reported resolution corresponds to spatial frequency of 0.145 Å⁻¹

8.2 Resolution estimates [i](#)

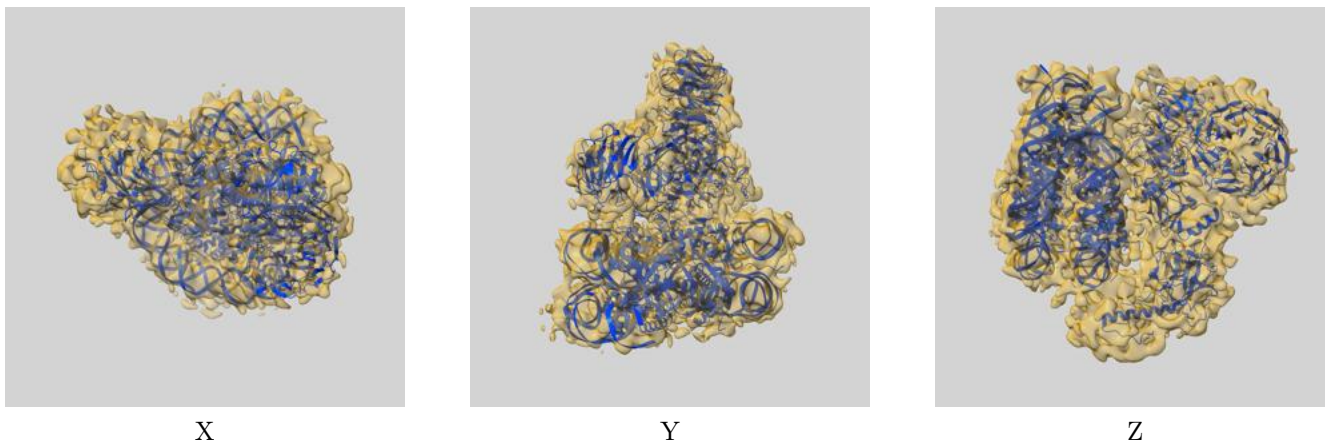
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	6.90	-	-
Author-provided FSC curve	6.90	9.03	7.11
Unmasked-calculated*	-	-	-

*Resolution estimate based on FSC curve calculated by comparison of deposited half-maps.

9 Map-model fit [i](#)

This section contains information regarding the fit between EMDB map EMD-21542 and PDB model 6W5I. Per-residue inclusion information can be found in section 3 on page 7.

9.1 Map-model overlay [i](#)



The images above show the 3D surface view of the map at the recommended contour level 0.0086 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

9.2 Q-score mapped to coordinate model [i](#)



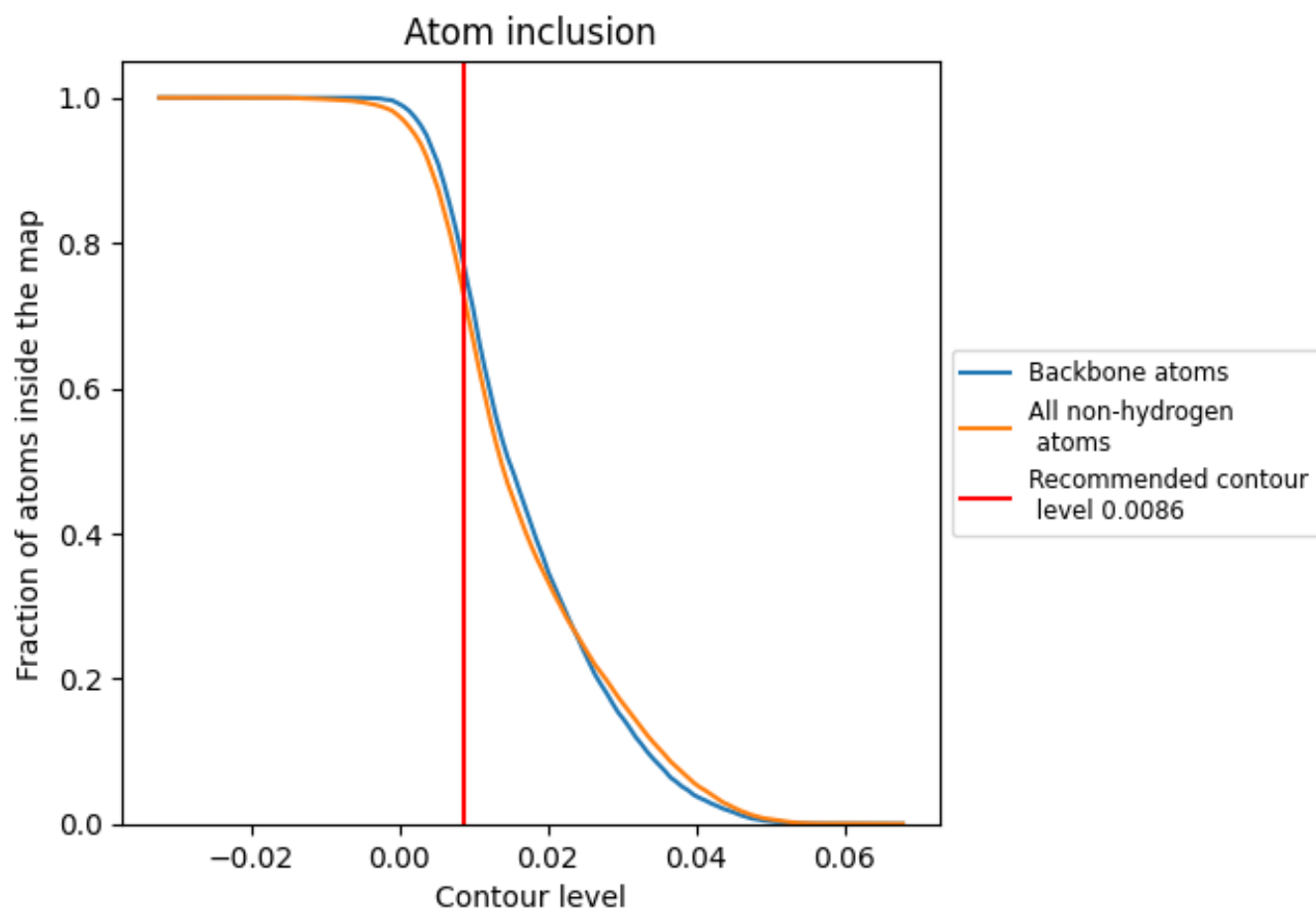
The images above show the model with each residue coloured according to its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions at which atoms can be resolved. Low Q-score values may therefore be expected for many entries.

9.3 Atom inclusion mapped to coordinate model [i](#)



The images above show the model with each residue coloured according to its atom inclusion. This shows to what extent they are inside the map at the recommended contour level (0.0086).































9.4 Atom inclusion [i](#)



At the recommended contour level, 77% of all backbone atoms, 73% of all non-hydrogen atoms, are inside the map.

9.5 Map-model fit summary

The table lists the average atom inclusion at the recommended contour level (0.0086) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	 0.7320	 0.1350
A	 0.4271	 0.0250
B	 0.5039	 0.0320
C	 0.5845	 0.0320
D	 0.6995	 0.0250
G	 0.7513	 0.1460
H	 0.7687	 0.1690
I	 0.7672	 0.1840
J	 0.7347	 0.1710
K	 0.7992	 0.1690
L	 0.7911	 0.1800
M	 0.7861	 0.1780
N	 0.7997	 0.1760
O	 0.9301	 0.2320
P	 0.9435	 0.2380

