



wwPDB EM Validation Summary Report ⓘ

May 18, 2025 – 05:49 PM EDT

PDB ID : 7MBM / pdb_00007mbm
EMDB ID : EMD-23738
Title : Cryo-EM structure of MLL1-NCP (H3K4M) complex, mode01
Authors : Park, S.H.; Ayoub, A.; Lee, Y.T.; Dou, Y.; Cho, U.
Deposited on : 2021-04-01
Resolution : 4.76 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.dev118
MolProbity : 4-5-2 with Phenix2.0rc1
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)
MapQ : 1.9.13
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.43.1

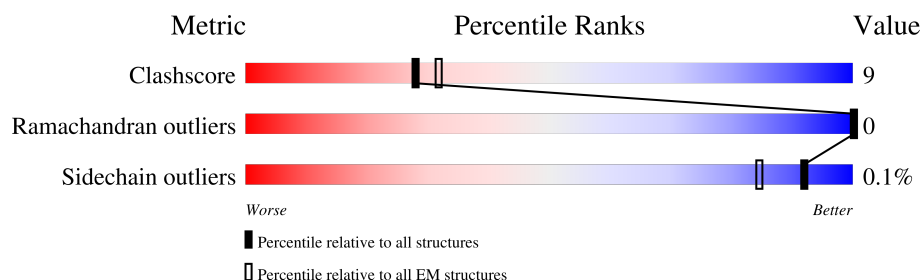
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 4.76 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 $\leq 5\%$. 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	
2	B	313	
3	C	209	
4	D	534	
5	G	136	
5	K	136	
6	H	103	
6	L	103	

Continued on next page...

Continued from previous page...

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

2 Entry composition

There are 10 unique types of molecules in this entry. The entry contains 19757 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	339	Total	C	N	O	S	0	0
			2665	1681	466	503	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	300	Total	C	N	O	S	0	0
			2326	1485	388	444	9		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	179	Total	C	N	O	S	0	0
			1435	904	266	250	15		

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	176	Total	C	N	O	S	0	0
			1399	906	234	253	6		

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	97	Total	C	N	O	S	0	0
			802	506	155	138	3		
5	K	97	Total	C	N	O	S	0	0
			802	506	155	138	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	80	Total	C	N	O	S	0	0
			641	405	125	110	1		
6	L	78	Total	C	N	O	S	0	0
			622	393	120	108	1		

- Molecule 7 is a protein called Histone H2A.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	I	107	Total	C	N	O	S	0	0
			825	519	163	143			
7	M	105	Total	C	N	O	S	0	0
			809	510	158	141			

- Molecule 8 is a protein called Histone H2B 1.1.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	J	94	Total	C	N	O	S	0	0
			741	466	135	138	2		
8	N	95	Total	C	N	O	S	0	0
			745	469	134	140	2		

There are 4 discrepancies between the modelled and reference sequences:

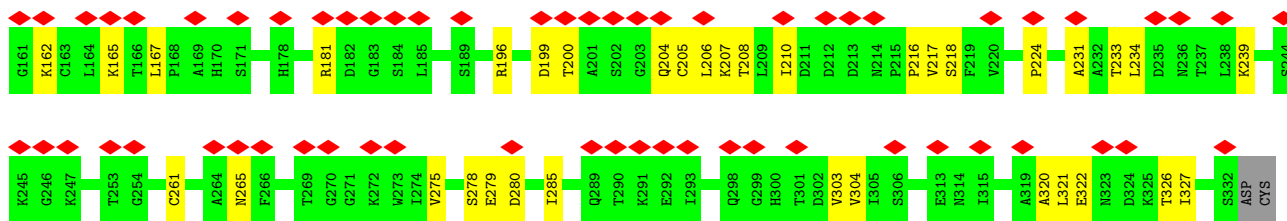
Chain	Residue	Modelled	Actual	Comment	Reference
J	0	MET	-	initiating methionine	UNP P02281
J	29	THR	SER	engineered mutation	UNP P02281
N	0	MET	-	initiating methionine	UNP P02281
N	29	THR	SER	engineered mutation	UNP P02281

- Molecule 9 is a DNA chain called DNA (145-MER).

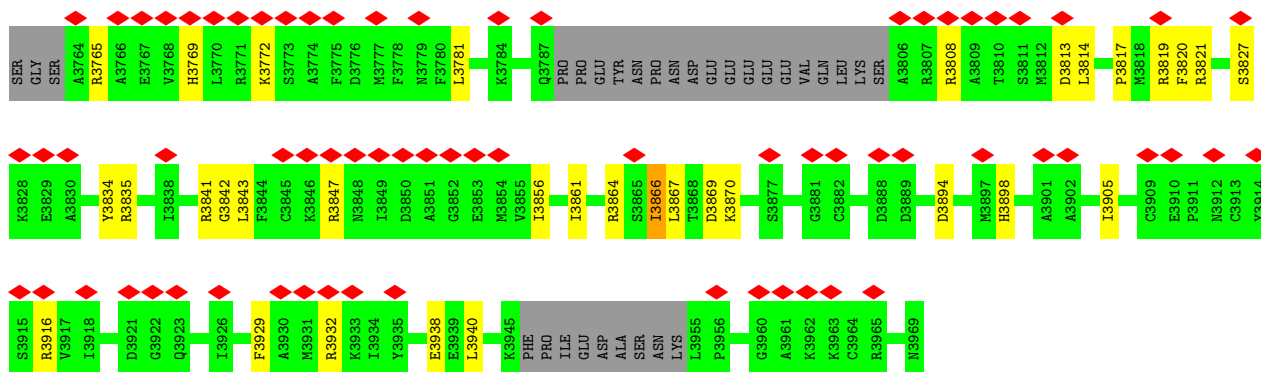
Mol	Chain	Residues	Atoms					AltConf	Trace
9	O	145	Total 2990	C 1415	N 559	O 871	P 145	0	0

- Molecule 10 is a DNA chain called DNA (145-MER).

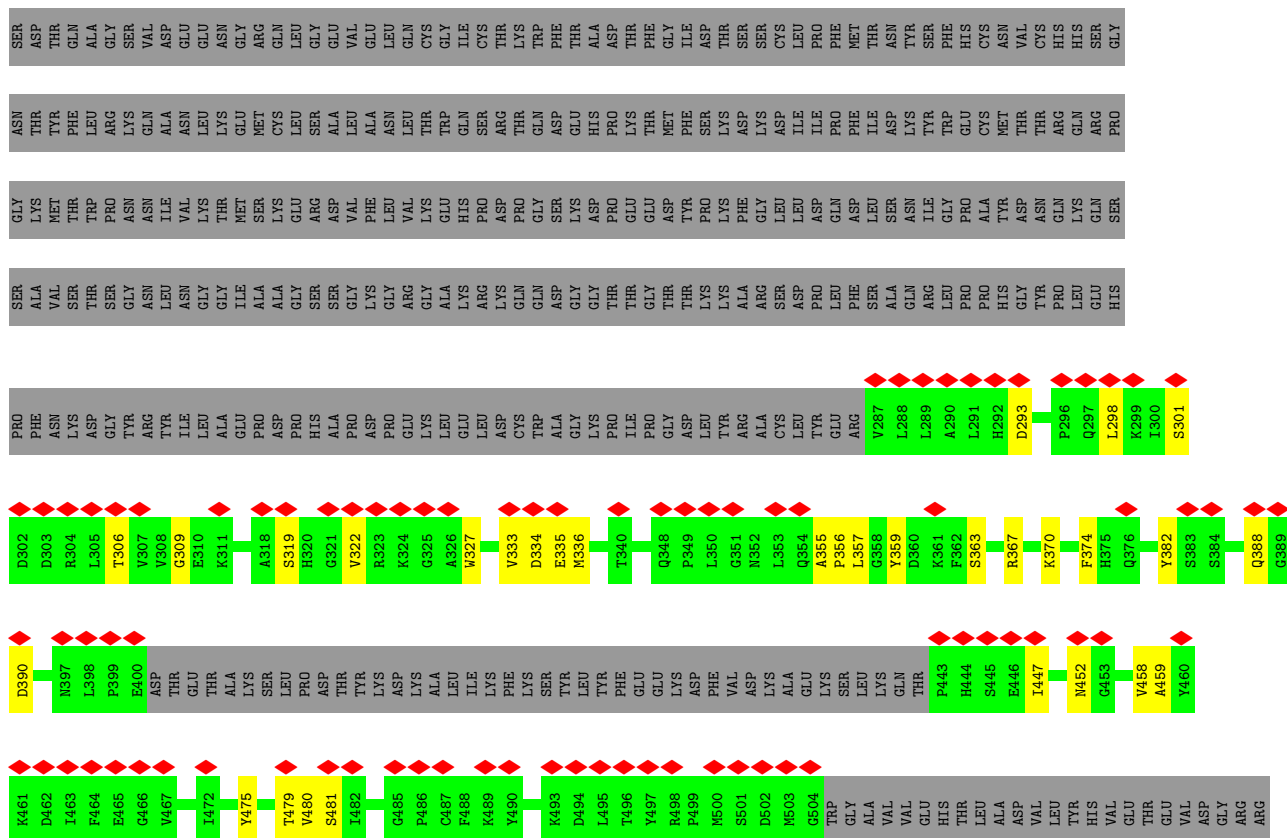
Mol	Chain	Residues	Atoms					AltConf	Trace
10	P	145	Total 2955	C 1403	N 538	O 869	P 145	0	0



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



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



SER
PRO
PRO
TRP
GLU
PRO

• Molecule 5: Histone H3

Chain G:  54% 17% 29%

MET ALA ARG THR MET GLN THR ALA ARG LYS SER THR GLY GLY LYS PRO ARG LYS GLN LEU THR LYS ALA ALA ARG LYS SER SER ALA PRO ALA THR THR GLY VAL LYS LYS P38 H39 R40 V45 Q55 K56 S57 T58 L61 R72 D81 L82 R83 E94 A98

F104 E105 D106 L109 C110 H113 T118 K122 D123 L126 R129 I130 R131 R134 ALA

• Molecule 5: Histone H3

Chain K:  58% 13% 29%

MET ALA ARG THR MET GLN THR ALA ARG LYS SER THR GLY GLY LYS PRO ARG LYS GLN LEU THR LYS ALA ALA ARG LYS SER SER ALA PRO ALA THR THR GLY VAL LYS LYS P38 H39 R40 K56 L61 I74 D81 A98 D106 T107 N108 L109 C110 A111 I112 H113

K122 D123 L126 R129 I130 R131 R134 ALA

• Molecule 6: Histone H4

Chain H:  65% 13% 22%

MET SER GLY ARG GLY LYS GLY GLY LYS LYS LEU GLY LYS GLY GLY ALA LYS ARG HIS ARG LYS VAL L22 R23 K31 R40 R45 T71 Y72 H75 A76 K77 T80 Y88 R92 Q93 G94 R95 T96 L97 G101 GLY

• Molecule 6: Histone H4

Chain L:  64% 12% 24%

MET SER GLY ARG GLY LYS GLY GLY LYS LYS LEU GLY LYS GLY GLY ALA LYS ARG HIS ARG LYS VAL L24 D24 L37 G42 S47 G48 I66 R67 D68 T71 Y72 H75 A76 K77 Y88 R92 G101 GLY

• Molecule 7: Histone H2A

Chain I:  73% 10% 17%

SER GLY ARG GLY LYS GLY GLY LYS LYS THR R11 A12 P26 A45 L63 E64 N68 T76 R81 R88 E92 R99 V100 T101 L108 P109 Q112 P117 LYS THR GLU SER SER LYS SER ALA LYS LYS

• Molecule 7: Histone H2A

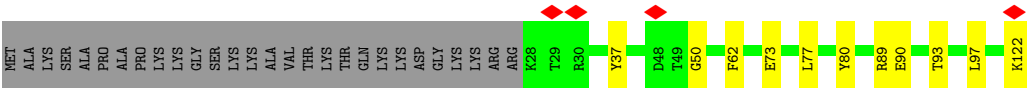
Chain M:  74% 8% 19%



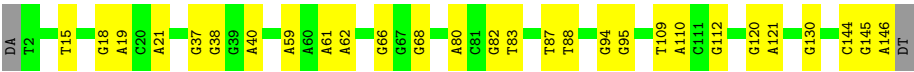
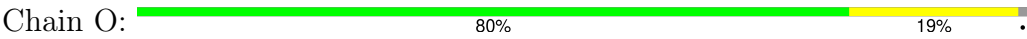
• Molecule 8: Histone H2B 1.1



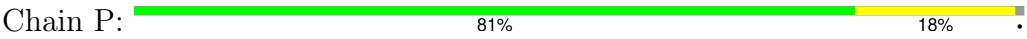
• Molecule 8: Histone H2B 1.1



• Molecule 9: DNA (145-MER)



• Molecule 10: DNA (145-MER)



4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	30322	Depositor
Resolution determination method	Not provided	
CTF correction method	PHASE FLIPPING ONLY	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	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
Maximum map value	0.051	Depositor
Minimum map value	-0.024	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.001	Depositor
Recommended contour level	0.0082	Depositor
Map size (\AA)	370.99997, 370.99997, 370.99997	wwPDB
Map dimensions	350, 350, 350	wwPDB
Map angles ($^\circ$)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (\AA)	1.06, 1.06, 1.06	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	0.21	0/2723	0.53	1/3699 (0.0%)
2	B	0.18	0/2382	0.42	0/3231
3	C	0.28	0/1461	0.55	1/1951 (0.1%)
4	D	0.18	0/1442	0.43	0/1950
5	G	0.32	0/814	0.54	0/1092
5	K	0.32	0/814	0.54	0/1092
6	H	0.33	0/648	0.65	0/868
6	L	0.35	0/629	0.58	0/843
7	I	0.30	0/835	0.52	0/1127
7	M	0.29	0/819	0.52	0/1106
8	J	0.31	0/752	0.51	0/1011
8	N	0.32	0/756	0.53	0/1015
9	O	0.29	0/3357	0.48	0/5184
10	P	0.28	0/3311	0.51	0/5103
All	All	0.27	0/20743	0.51	2/29272 (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
3	C	0	1
6	H	0	1
All	All	0	3

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	195	THR	N-CA-CB	5.99	120.61	110.49
3	C	3866	ILE	N-CA-C	-5.05	108.37	113.47

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	251	ARG	Peptide
3	C	3864	ARG	Peptide
6	H	31	LYS	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	A	2665	0	2631	84	0
2	B	2326	0	2308	95	0
3	C	1435	0	1437	26	0
4	D	1399	0	1363	24	0
5	G	802	0	841	45	0
5	K	802	0	841	36	0
6	H	641	0	684	35	0
6	L	622	0	660	44	0
7	I	825	0	882	17	0
7	M	809	0	864	14	0
8	J	741	0	768	36	0
8	N	745	0	773	44	0
9	O	2990	0	1628	40	0
10	P	2955	0	1627	40	0
All	All	19757	0	17307	338	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 9.

The worst 5 of 338 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)
1:A:334:PRO:HD3	2:B:208:THR:CB	1.39	1.47
1:A:334:PRO:CG	2:B:208:THR:O	1.67	1.41
1:A:334:PRO:CD	2:B:208:THR:HB	1.54	1.38
1:A:334:PRO:HG3	2:B:208:THR:N	1.37	1.35

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:334:PRO:CD	2:B:208:THR:O	1.78	1.32

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	337/538 (63%)	291 (86%)	46 (14%)	0	100	100
2	B	298/313 (95%)	270 (91%)	28 (9%)	0	100	100
3	C	173/209 (83%)	153 (88%)	20 (12%)	0	100	100
4	D	172/534 (32%)	158 (92%)	14 (8%)	0	100	100
5	G	95/136 (70%)	91 (96%)	4 (4%)	0	100	100
5	K	95/136 (70%)	90 (95%)	5 (5%)	0	100	100
6	H	78/103 (76%)	69 (88%)	9 (12%)	0	100	100
6	L	76/103 (74%)	72 (95%)	4 (5%)	0	100	100
7	I	105/129 (81%)	98 (93%)	7 (7%)	0	100	100
7	M	103/129 (80%)	98 (95%)	5 (5%)	0	100	100
8	J	92/123 (75%)	88 (96%)	4 (4%)	0	100	100
8	N	93/123 (76%)	89 (96%)	4 (4%)	0	100	100
All	All	1717/2576 (67%)	1567 (91%)	150 (9%)	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	296/462 (64%)	295 (100%)	1 (0%)	91	92
2	B	262/274 (96%)	262 (100%)	0	100	100
3	C	151/182 (83%)	151 (100%)	0	100	100
4	D	149/460 (32%)	149 (100%)	0	100	100
5	G	85/111 (77%)	85 (100%)	0	100	100
5	K	85/111 (77%)	85 (100%)	0	100	100
6	H	66/79 (84%)	66 (100%)	0	100	100
6	L	64/79 (81%)	64 (100%)	0	100	100
7	I	84/101 (83%)	84 (100%)	0	100	100
7	M	83/101 (82%)	83 (100%)	0	100	100
8	J	81/103 (79%)	81 (100%)	0	100	100
8	N	81/103 (79%)	81 (100%)	0	100	100
All	All	1487/2166 (69%)	1486 (100%)	1 (0%)	92	95

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	195	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
4	D	388	GLN
7	M	24	GLN
8	N	81	ASN
7	M	73	ASN
8	J	46	HIS

5.3.3 RNA ⓘ

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.

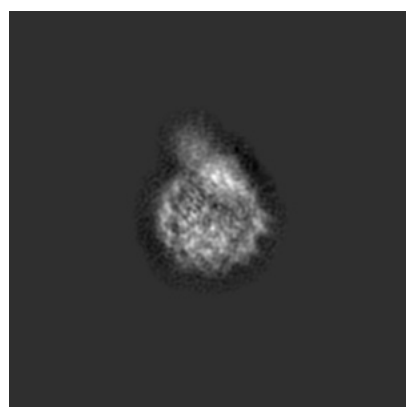
6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-23738. These allow visual inspection of the internal detail of the map and identification of artifacts.

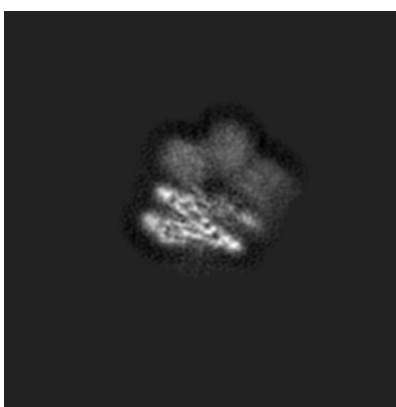
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections [i](#)

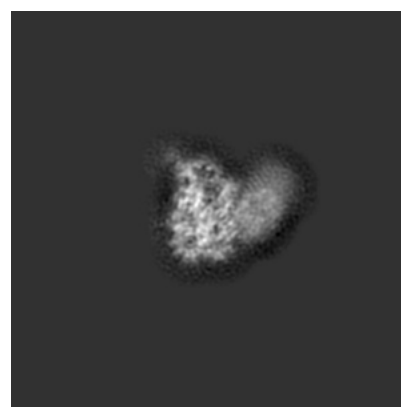
6.1.1 Primary map



X



Y

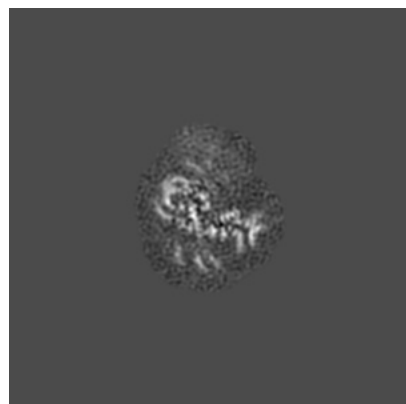


Z

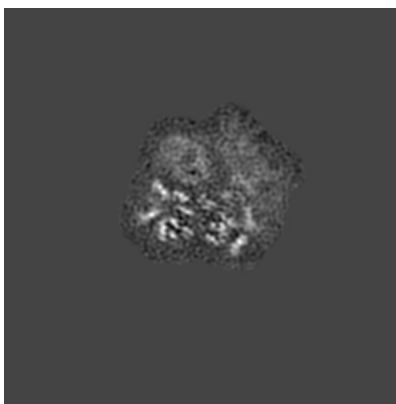
The images above show the map projected in three orthogonal directions.

6.2 Central slices [i](#)

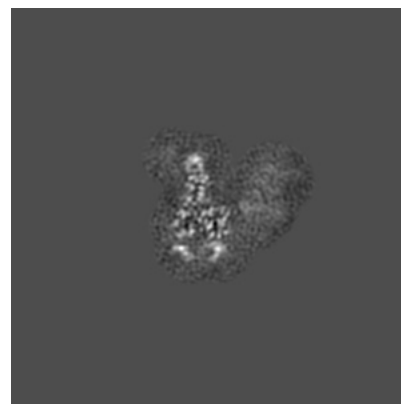
6.2.1 Primary map



X Index: 175



Y Index: 175

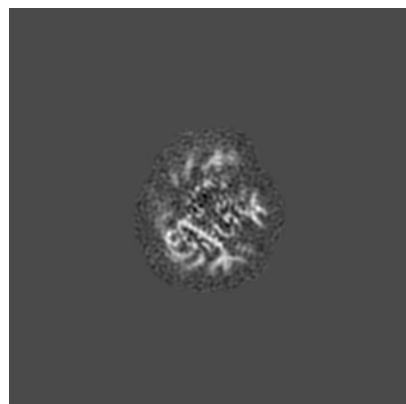


Z Index: 175

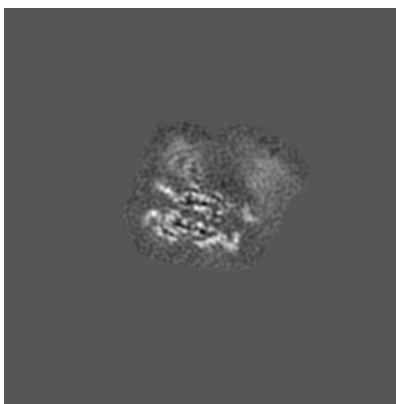
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices [i](#)

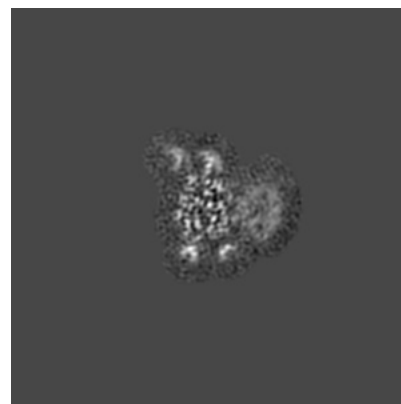
6.3.1 Primary map



X Index: 163



Y Index: 160

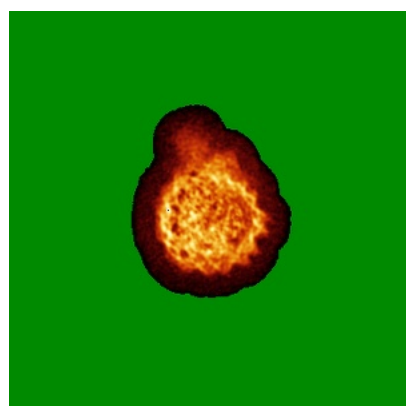


Z Index: 157

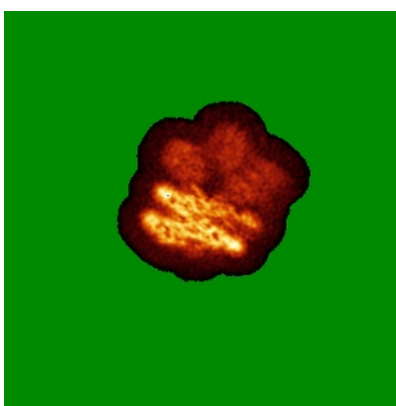
The images above show the largest variance slices of the map in three orthogonal directions.

6.4 Orthogonal standard-deviation projections (False-color) [i](#)

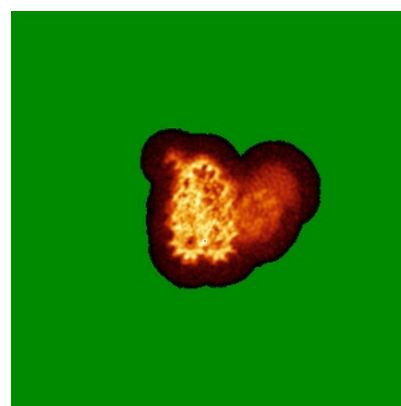
6.4.1 Primary map



X



Y

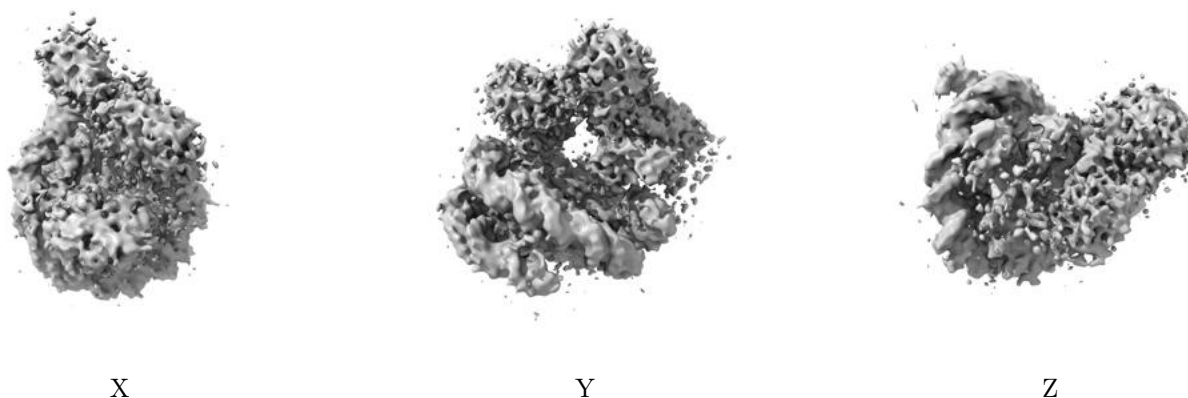


Z

The images above show the map standard deviation projections with false color in three orthogonal directions. Minimum values are shown in green, max in blue, and dark to light orange shades represent small to large values respectively.

6.5 Orthogonal surface views [i](#)

6.5.1 Primary map



The images above show the 3D surface view of the map at the recommended contour level 0.0082. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

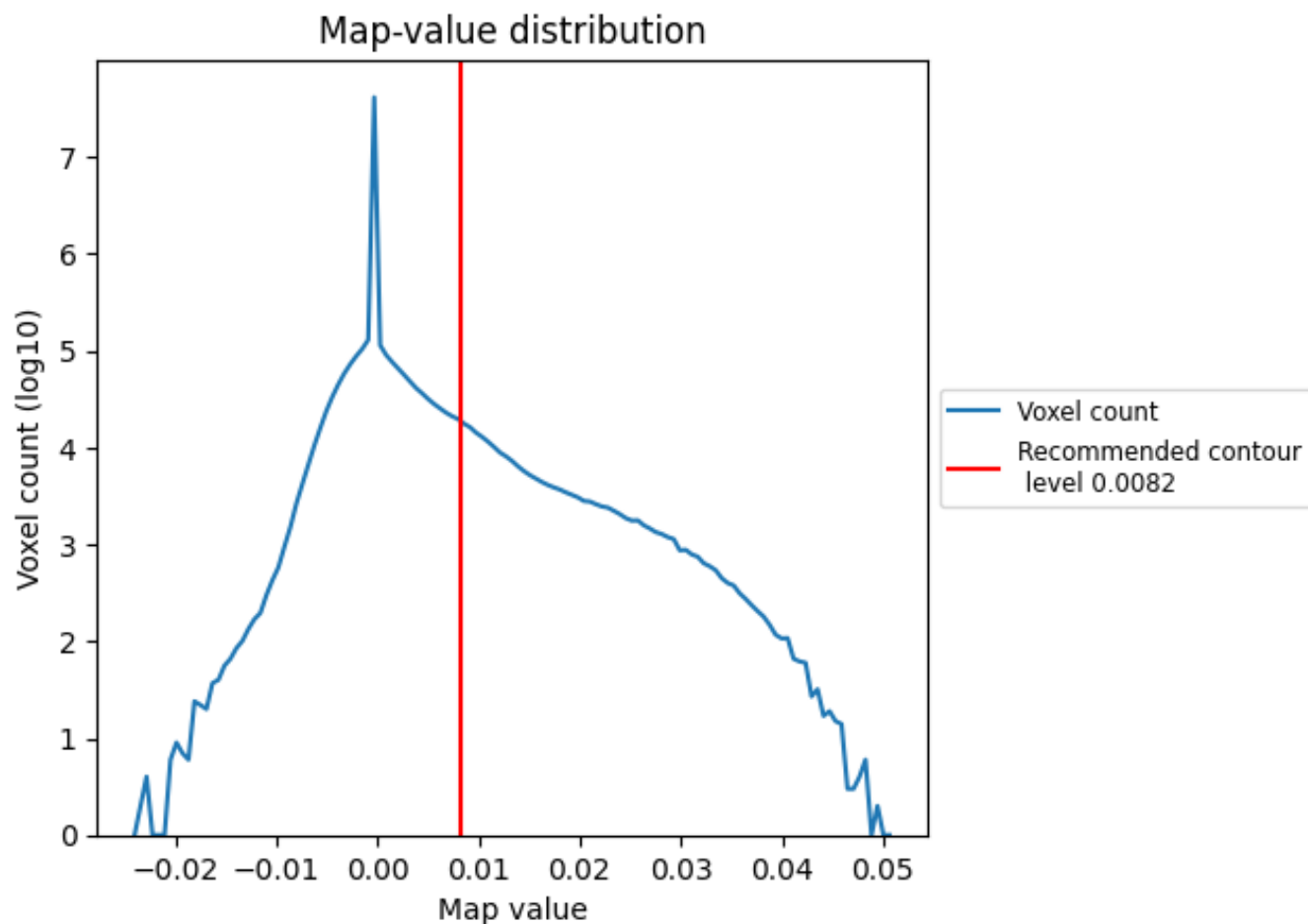
6.6 Mask visualisation [i](#)

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

7 Map analysis [i](#)

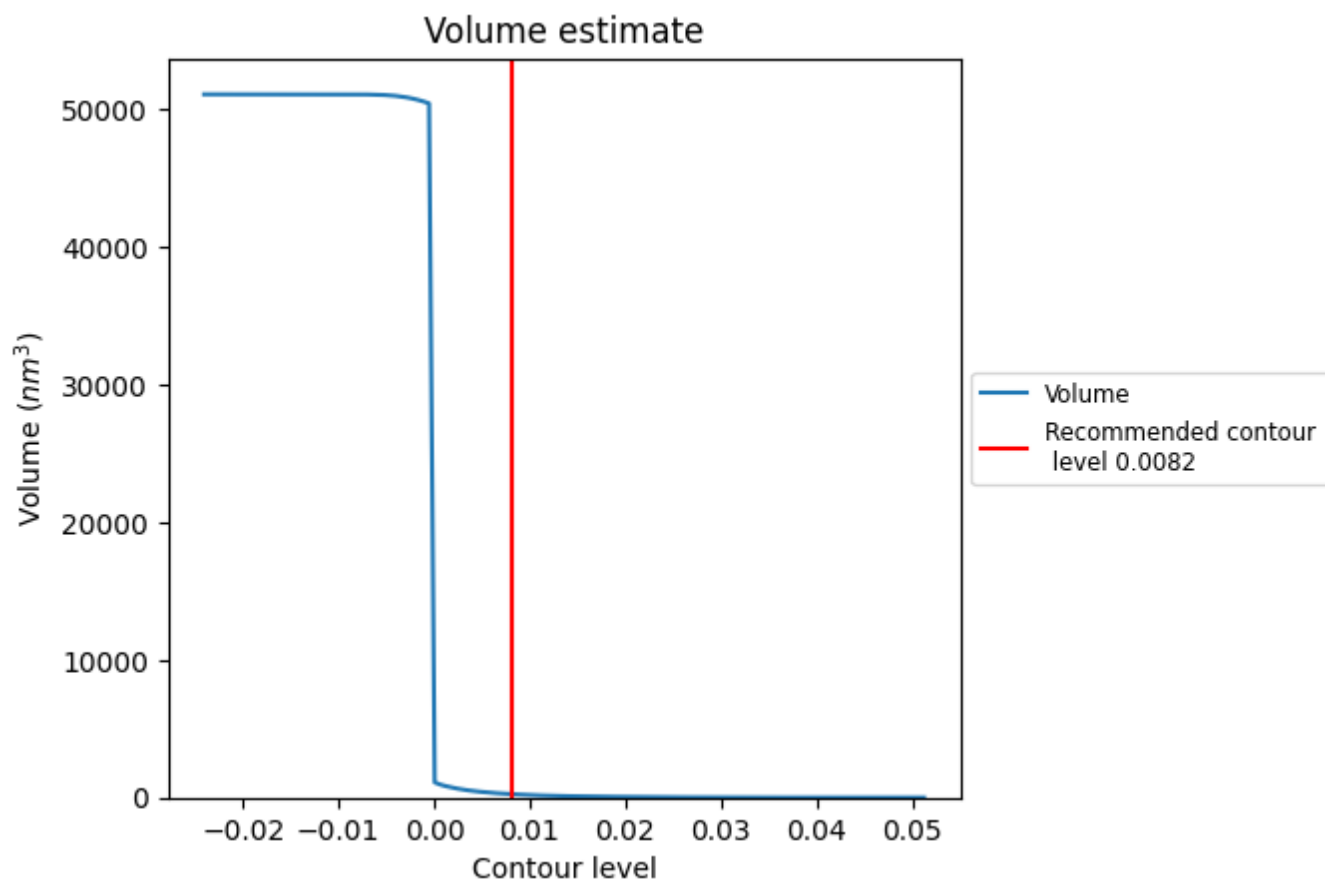
This section contains the results of statistical analysis of the map.

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.

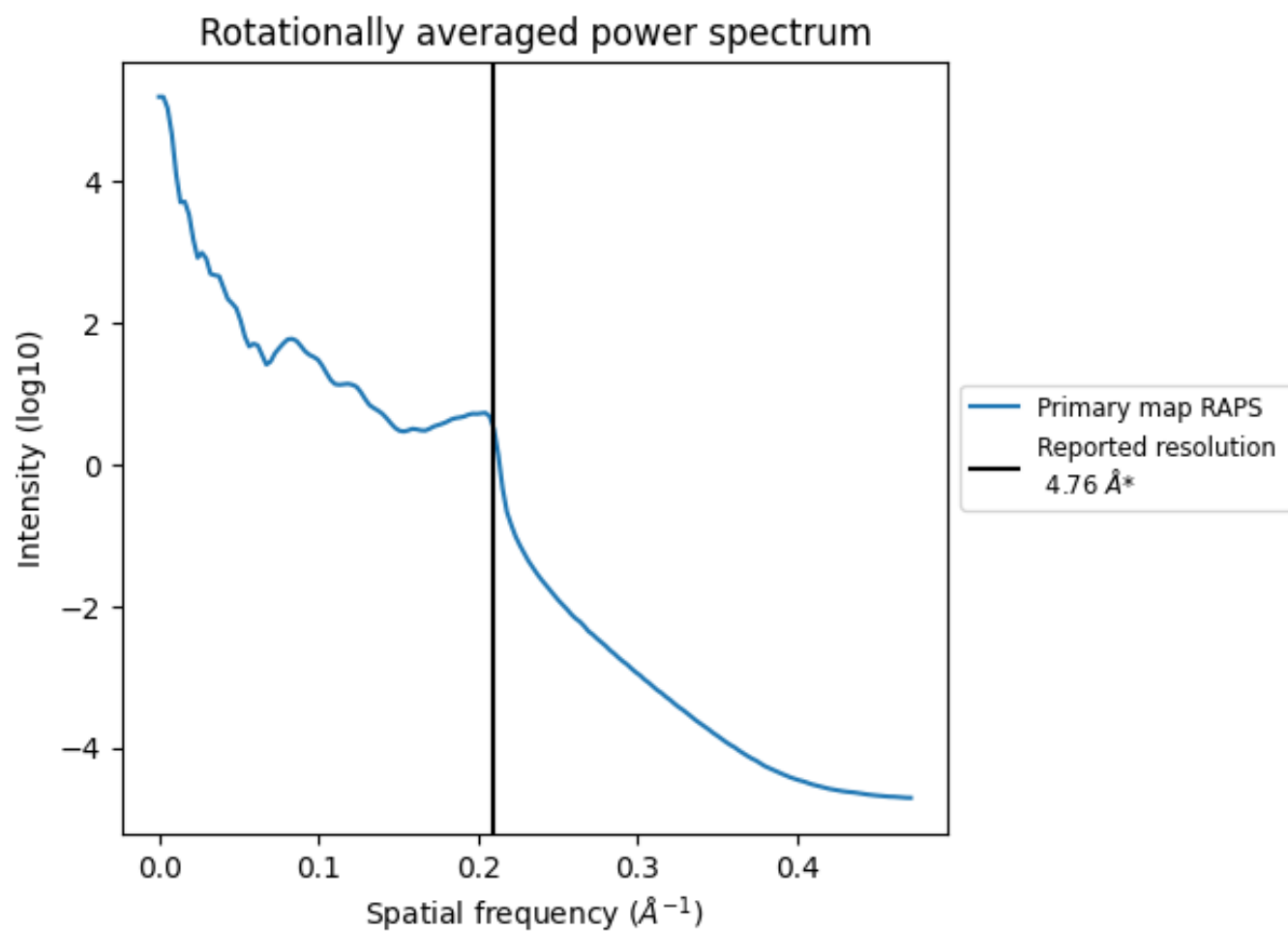
7.2 Volume estimate [i](#)



The volume at the recommended contour level is 245 nm^3 ; this corresponds to an approximate mass of 221 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 ⓘ

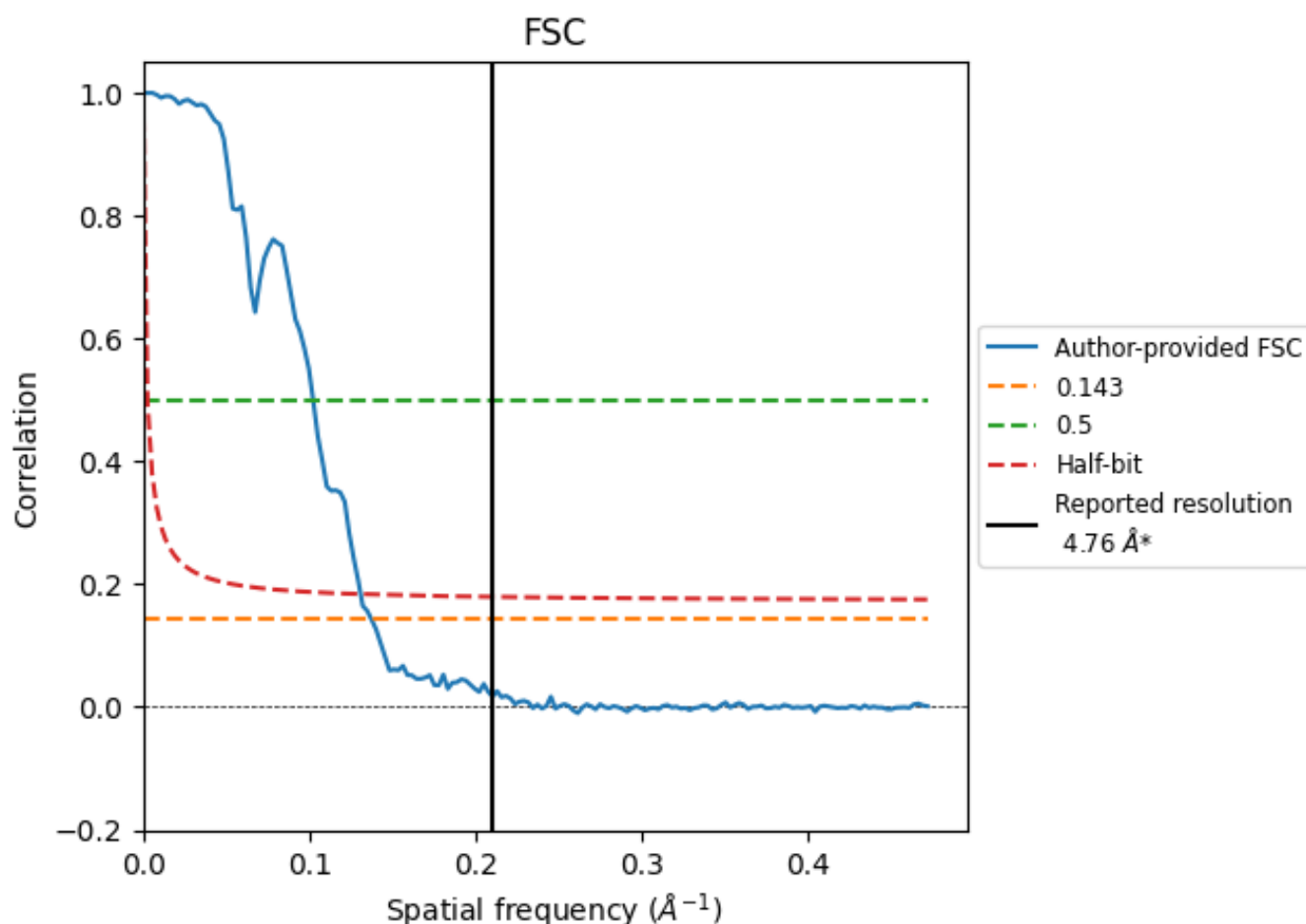


*Reported resolution corresponds to spatial frequency of 0.210 Å⁻¹

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 0.210 Å⁻¹

8.2 Resolution estimates [i](#)

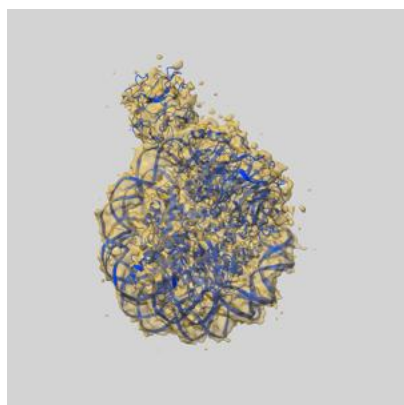
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	-	-	-
Author-provided FSC curve	7.29	9.78	7.65
Unmasked-calculated*	-	-	-

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

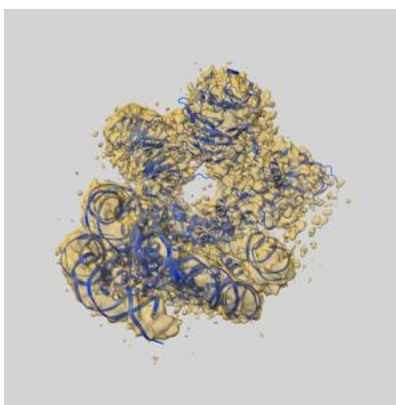
9 Map-model fit [i](#)

This section contains information regarding the fit between EMDB map EMD-23738 and PDB model 7MBM. Per-residue inclusion information can be found in section [3](#) on page [7](#).

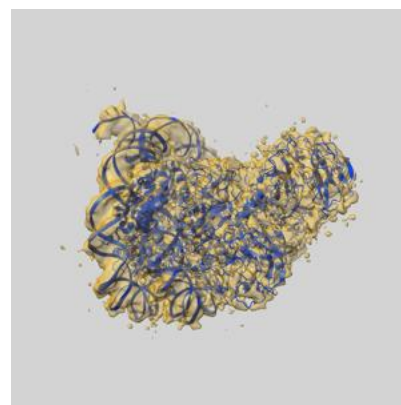
9.1 Map-model overlay [i](#)



X



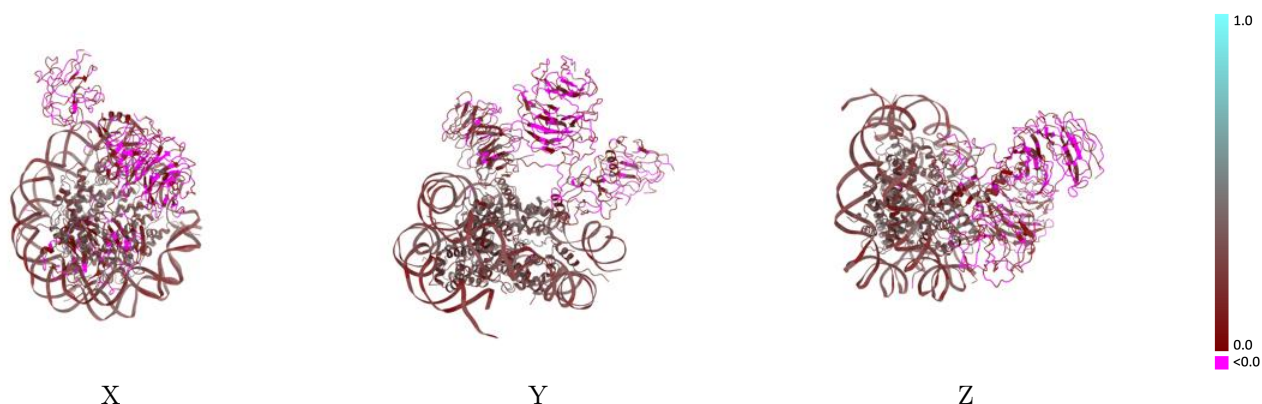
Y



Z

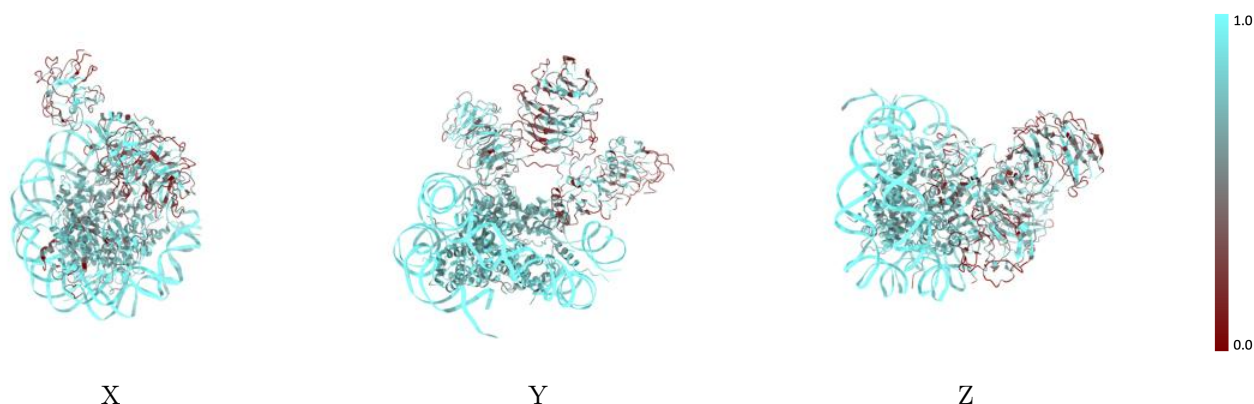
The images above show the 3D surface view of the map at the recommended contour level 0.0082 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

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



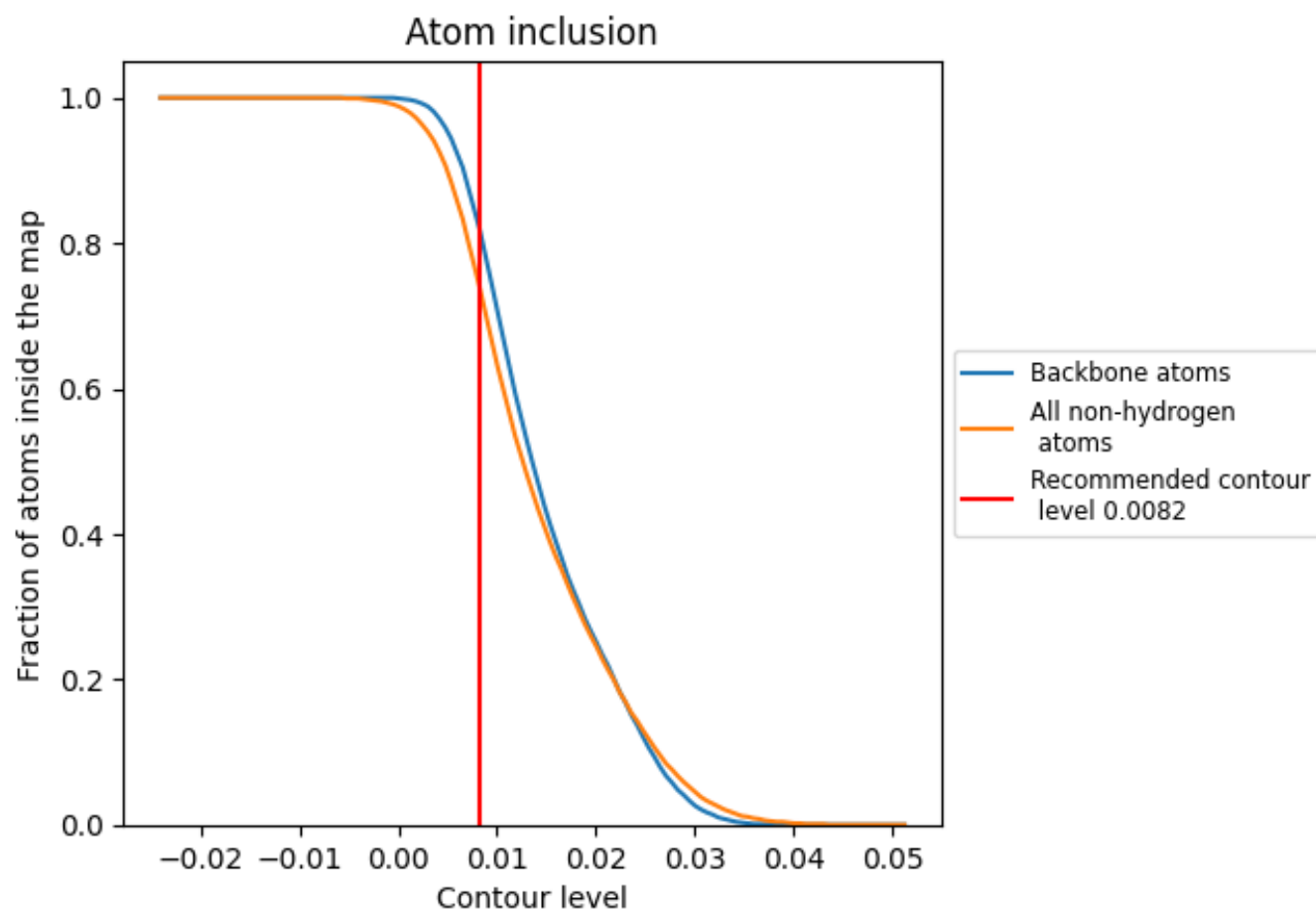
The images above show the model with each residue coloured according to its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions 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 (0.0082).































9.4 Atom inclusion [i](#)



At the recommended contour level, 82% of all backbone atoms, 74% 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 (0.0082) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	 0.7420	 0.1980
A	 0.6290	 0.1450
B	 0.5060	 0.0240
C	 0.5130	 0.1070
D	 0.4430	 0.0400
G	 0.7960	 0.2880
H	 0.8110	 0.3150
I	 0.7880	 0.3000
J	 0.7760	 0.2800
K	 0.7980	 0.2820
L	 0.8330	 0.3160
M	 0.8190	 0.3040
N	 0.8120	 0.3020
O	 0.9400	 0.2510
P	 0.9500	 0.2470

