



wwPDB EM Validation Summary Report ⓘ

Nov 7, 2022 – 02:20 AM EST

PDB ID : 6NN6
EMDB ID : EMD-0458
Title : Structure of Dot1L-H2BK120ub nucleosome complex
Authors : Anderson, C.J.; Baird, M.R.; Hsu, A.; Barbour, E.H.; Koyama, Y.; Borgnia, M.J.; McGinty, R.K.
Deposited on : 2019-01-14
Resolution : 3.90 Å (reported)
Based on initial models : 1NW3, 1UBQ, 3LZ0

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

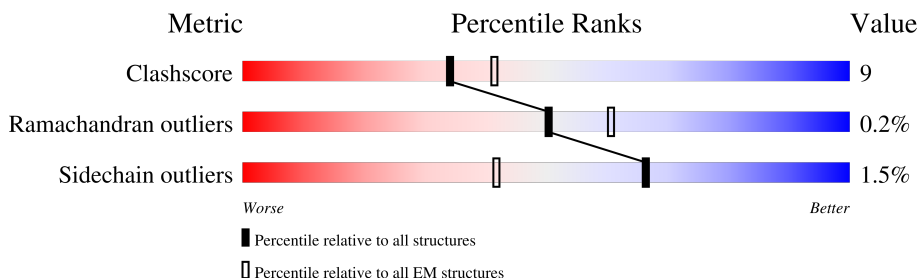
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 3.90 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	135	
1	E	135	
2	B	102	
2	F	102	
3	C	129	
3	G	129	
4	D	124	
4	H	124	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	I	145	<div><div></div><div>81%19%</div></div>
6	J	145	<div><div></div><div>80%20%</div></div>
7	K	416	<div><div></div><div>54%62%17%21%</div></div>
8	L	80	<div><div></div><div>40%72%18%8%</div></div>

2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 15125 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 Histone H3.2.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	97	Total	C	N	O	S	0	0
			802	506	155	138	3		
1	E	97	Total	C	N	O	S	0	0
			801	504	155	139	3		

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	102	ALA	GLY	engineered mutation	UNP P84233
E	102	ALA	GLY	engineered mutation	UNP P84233

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	80	Total	C	N	O	S	0	0
			641	405	125	110	1		
2	F	78	Total	C	N	O	S	0	0
			619	391	120	107	1		

- Molecule 3 is a protein called Histone H2A type 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	104	Total	C	N	O		0	0
			804	507	157	140			
3	G	105	Total	C	N	O		0	0
			809	510	158	141			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	99	ARG	GLY	engineered mutation	UNP P06897
C	123	SER	ALA	engineered mutation	UNP P06897
G	99	ARG	GLY	engineered mutation	UNP P06897

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	123	SER	ALA	engineered mutation	UNP P06897

- Molecule 4 is a protein called Histone H2B 1.1.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	D	93	Total	C	N	O	S	0	0
			723	454	129	137	3		
4	H	93	Total	C	N	O	S	0	0
			723	454	129	137	3		

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	2	GLY	-	expression tag	UNP P02281
D	3	SER	-	expression tag	UNP P02281
D	32	THR	SER	engineered mutation	UNP P02281
D	120	CYS	LYS	engineered mutation	UNP P02281
H	2	GLY	-	expression tag	UNP P02281
H	3	SER	-	expression tag	UNP P02281
H	32	THR	SER	engineered mutation	UNP P02281
H	120	CYS	LYS	engineered mutation	UNP P02281

- Molecule 5 is a DNA chain called DNA (145-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
5	I	145	Total	C	N	O	P	0	0
			2952	1404	537	867	144		

- Molecule 6 is a DNA chain called DNA (145-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
6	J	145	Total	C	N	O	P	0	0
			2987	1416	558	869	144		

- Molecule 7 is a protein called Histone-lysine N-methyltransferase, H3 lysine-79 specific.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	K	328	Total	C	N	O	S	0	0
			2671	1706	455	498	12		

There are 2 discrepancies between the modelled and reference sequences:

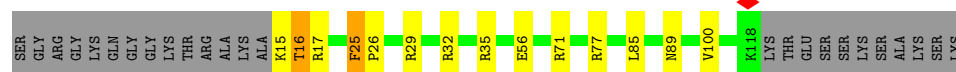
Chain	Residue	Modelled	Actual	Comment	Reference
K	1	GLY	-	expression tag	UNP Q8TEK3
K	2	SER	-	expression tag	UNP Q8TEK3

- Molecule 8 is a protein called Ubiquitin.

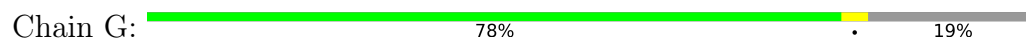
Mol	Chain	Residues	Atoms					AltConf	Trace
8	L	74	Total	C	N	O	S	0	0
			593	374	103	115	1		

There are 4 discrepancies between the modelled and reference sequences:

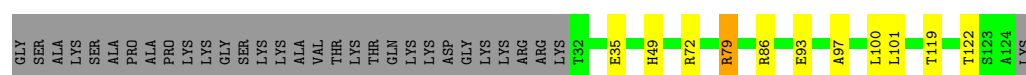
Chain	Residue	Modelled	Actual	Comment	Reference
L	-3	GLY	-	expression tag	UNP F5H388
L	-2	SER	-	expression tag	UNP F5H388
L	0	SER	GLY	engineered mutation	UNP F5H388
L	76	CYS	-	expression tag	UNP F5H388



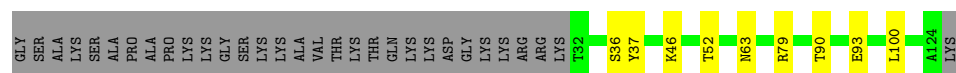
- Molecule 3: Histone H2A type 1



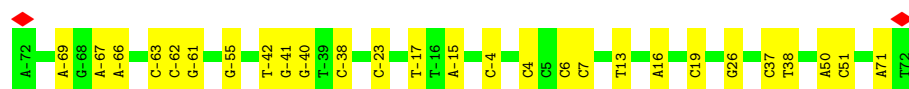
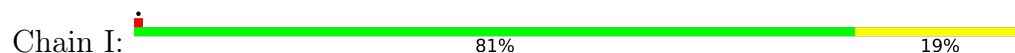
- Molecule 4: Histone H2B 1.1



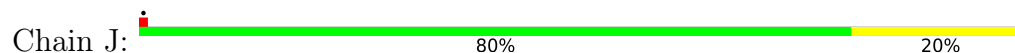
- Molecule 4: Histone H2B 1.1



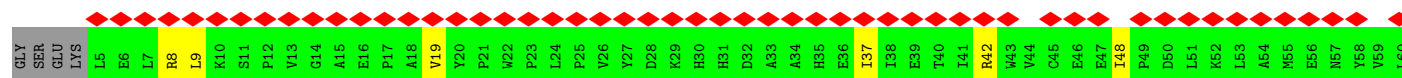
- Molecule 5: DNA (145-MER)

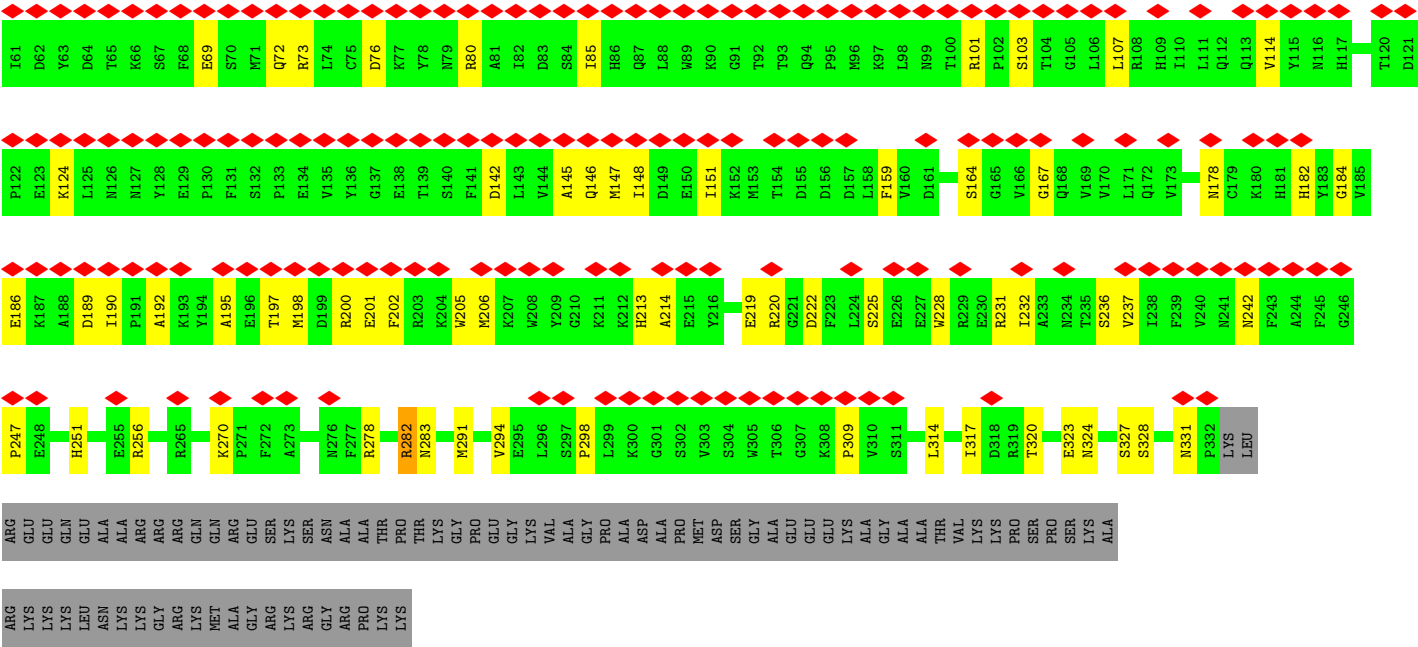


- Molecule 6: DNA (145-MER)

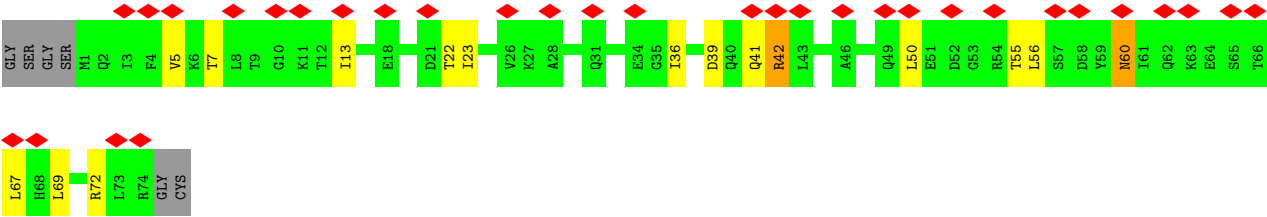


- Molecule 7: Histone-lysine N-methyltransferase, H3 lysine-79 specific





● Molecule 8: Ubiquitin



4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	39558	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	48	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	FEI FALCON III (4k x 4k)	Depositor
Maximum map value	0.194	Depositor
Minimum map value	-0.115	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.004	Depositor
Recommended contour level	0.03	Depositor
Map size (Å)	356.40002, 356.40002, 356.40002	wwPDB
Map dimensions	330, 330, 330	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.08, 1.08, 1.08	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	0.52	0/814	0.59	0/1092
1	E	0.53	0/812	0.58	0/1088
2	B	0.61	0/648	0.62	0/868
2	F	0.61	0/626	0.63	0/837
3	C	0.52	0/814	0.61	0/1099
3	G	0.49	0/819	0.59	0/1106
4	D	0.55	0/734	0.61	0/990
4	H	0.57	0/734	0.58	0/990
5	I	0.96	0/3308	1.01	1/5099 (0.0%)
6	J	0.92	0/3354	1.02	1/5180 (0.0%)
7	K	0.29	0/2741	0.50	0/3718
8	L	0.28	0/599	0.56	0/806
All	All	0.71	0/16003	0.80	2/22873 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	J	-59	DT	O4'-C4'-C3'	-5.17	102.43	104.50
5	I	-63	DC	O4'-C1'-N1	5.01	111.51	108.00

There are no chirality outliers.

There are no planarity outliers.

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	802	0	841	9	0
1	E	801	0	838	11	0
2	B	641	0	684	10	0
2	F	619	0	659	7	0
3	C	804	0	859	14	0
3	G	809	0	864	3	0
4	D	723	0	739	10	0
4	H	723	0	739	4	0
5	I	2952	0	1629	23	0
6	J	2987	0	1630	27	0
7	K	2671	0	2617	42	0
8	L	593	0	623	9	0
All	All	15125	0	12722	132	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 132 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)
4:D:79:ARG:HG2	4:D:79:ARG:HH11	1.45	0.81
4:D:79:ARG:HG2	4:D:79:ARG:NH1	2.02	0.74
5:I:6:DC:H42	6:J:-6:DG:H1	1.40	0.69
8:L:22:THR:HA	8:L:55:THR:HA	1.75	0.69
1:E:68:GLN:HE21	1:E:72:ARG:HH21	1.42	0.67

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	95/135 (70%)	92 (97%)	3 (3%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	E	95/135 (70%)	90 (95%)	4 (4%)	1 (1%)	14	51
2	B	78/102 (76%)	75 (96%)	3 (4%)	0	100	100
2	F	76/102 (74%)	73 (96%)	3 (4%)	0	100	100
3	C	102/129 (79%)	98 (96%)	3 (3%)	1 (1%)	15	52
3	G	103/129 (80%)	99 (96%)	4 (4%)	0	100	100
4	D	91/124 (73%)	89 (98%)	2 (2%)	0	100	100
4	H	91/124 (73%)	88 (97%)	3 (3%)	0	100	100
7	K	326/416 (78%)	312 (96%)	14 (4%)	0	100	100
8	L	72/80 (90%)	68 (94%)	4 (6%)	0	100	100
All	All	1129/1476 (76%)	1084 (96%)	43 (4%)	2 (0%)	50	79

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	16	THR
1	E	63	ARG

5.3.2 Protein sidechains ⓘ

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	85/110 (77%)	83 (98%)	2 (2%)	49	69
1	E	84/110 (76%)	81 (96%)	3 (4%)	35	61
2	B	66/78 (85%)	65 (98%)	1 (2%)	65	80
2	F	63/78 (81%)	63 (100%)	0	100	100
3	C	83/101 (82%)	82 (99%)	1 (1%)	71	83
3	G	83/101 (82%)	83 (100%)	0	100	100
4	D	79/103 (77%)	77 (98%)	2 (2%)	47	69
4	H	79/103 (77%)	77 (98%)	2 (2%)	47	69
7	K	294/360 (82%)	292 (99%)	2 (1%)	84	90

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
8	L	68/71 (96%)	66 (97%)	2 (3%)	42	65
All	All	984/1215 (81%)	969 (98%)	15 (2%)	66	80

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	53	ARG
8	L	42	ARG
1	E	63	ARG
8	L	60	ASN
7	K	42	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
7	K	324	ASN
8	L	41	GLN
8	L	60	ASN
1	E	68	GLN
1	A	108	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.

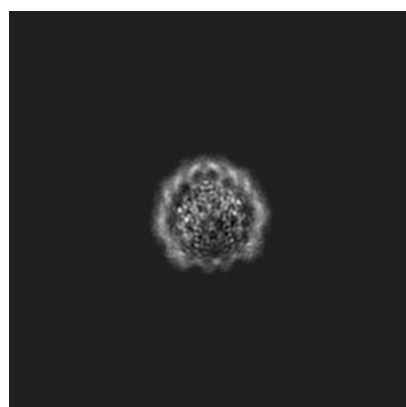
6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-0458. These allow visual inspection of the internal detail of the map and identification of artifacts.

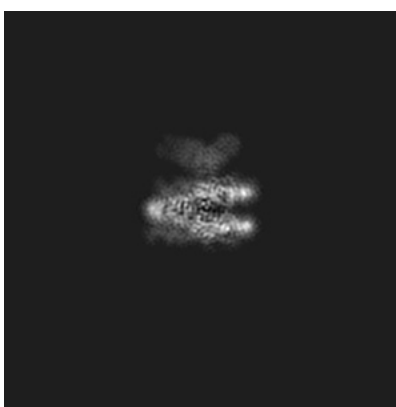
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections [i](#)

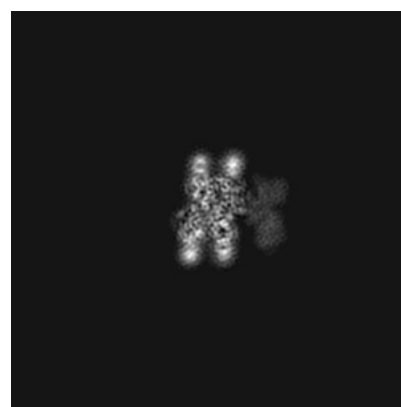
6.1.1 Primary map



X



Y

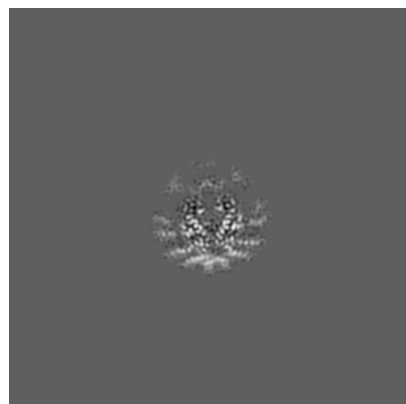


Z

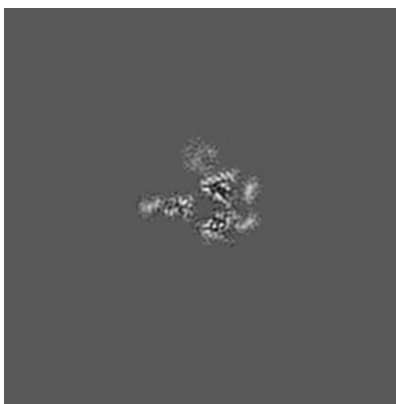
The images above show the map projected in three orthogonal directions.

6.2 Central slices [i](#)

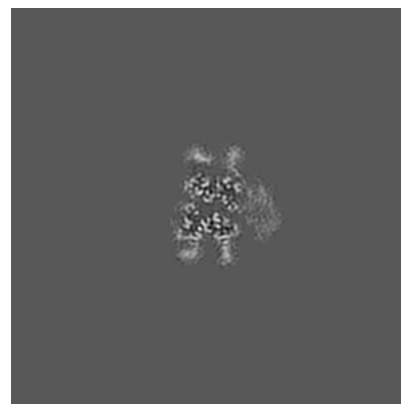
6.2.1 Primary map



X Index: 165



Y Index: 165

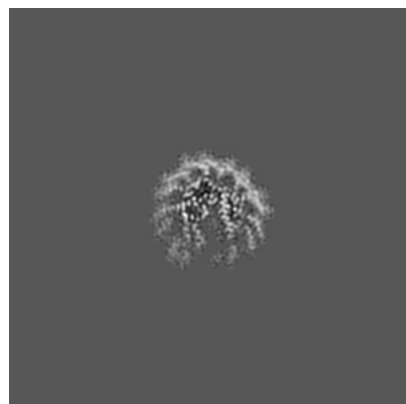


Z Index: 165

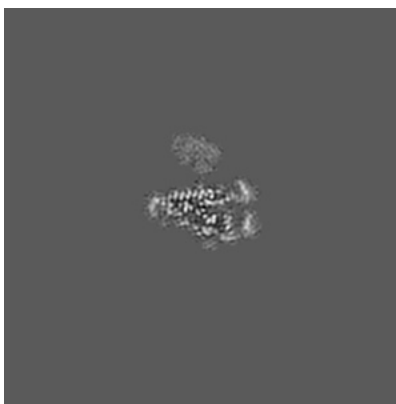
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices [i](#)

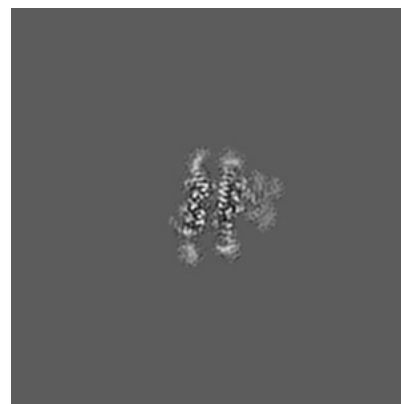
6.3.1 Primary map



X Index: 153



Y Index: 153

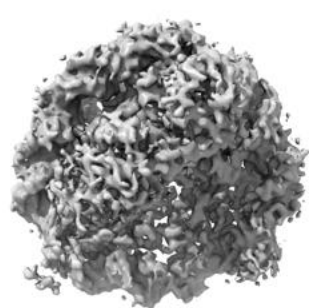


Z Index: 173

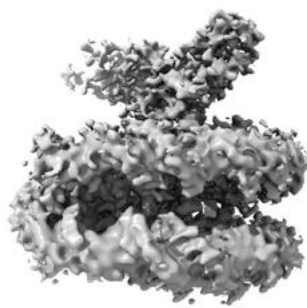
The images above show the largest variance slices of the map in three orthogonal directions.

6.4 Orthogonal surface views [i](#)

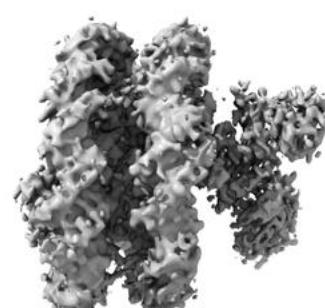
6.4.1 Primary map



X



Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.03. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

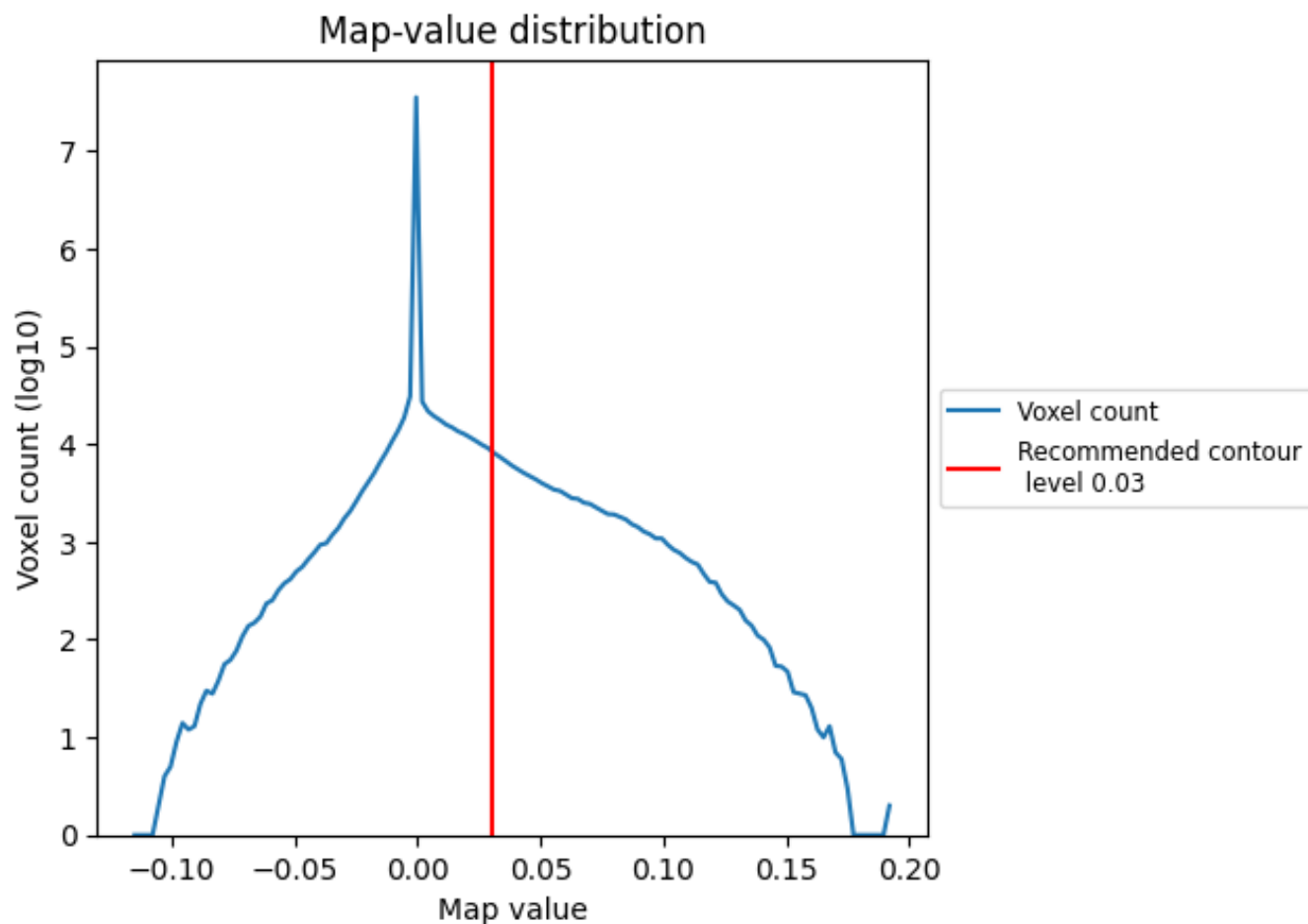
6.5 Mask visualisation

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

7 Map analysis [i](#)

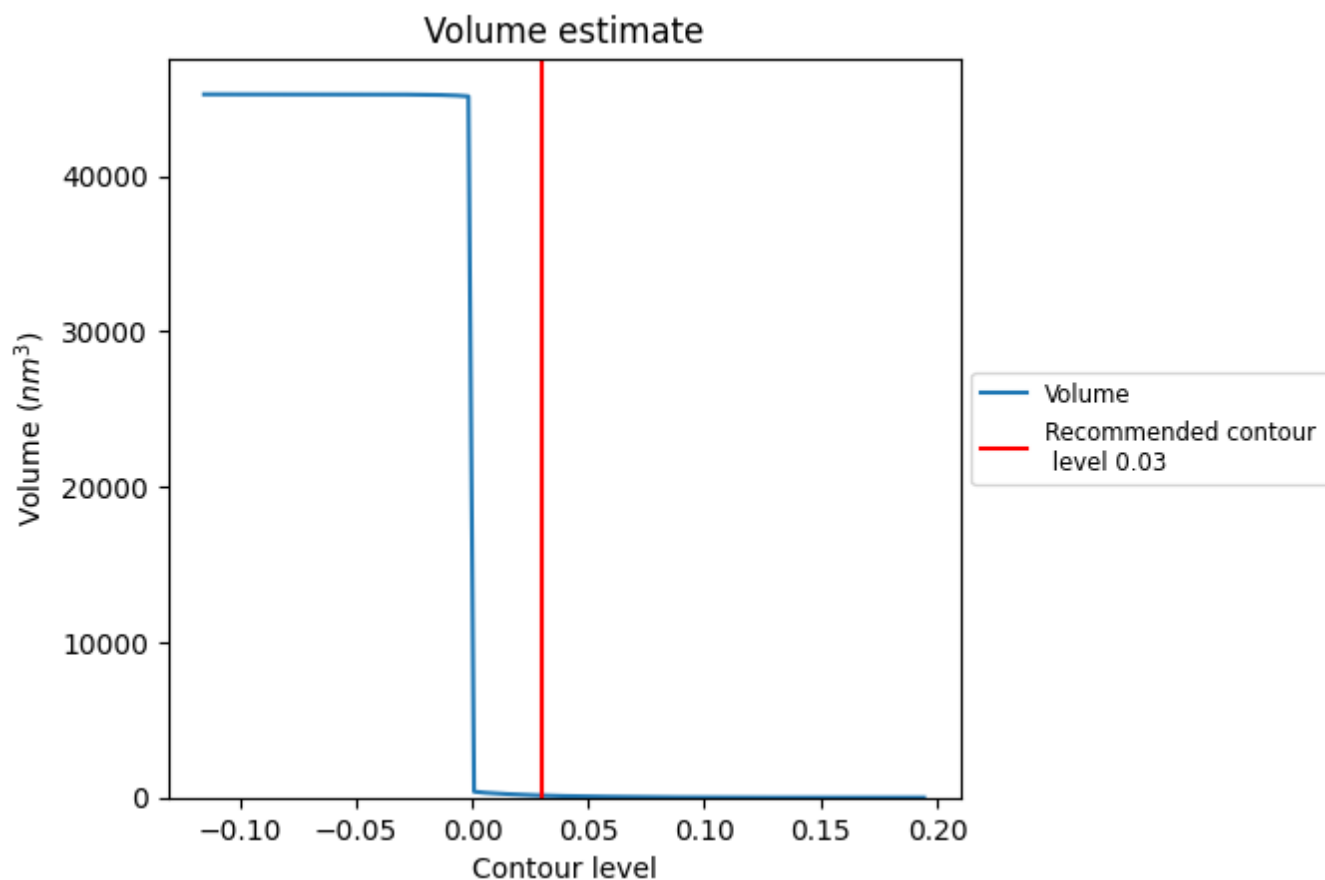
This section contains the results of statistical analysis of the map.

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.

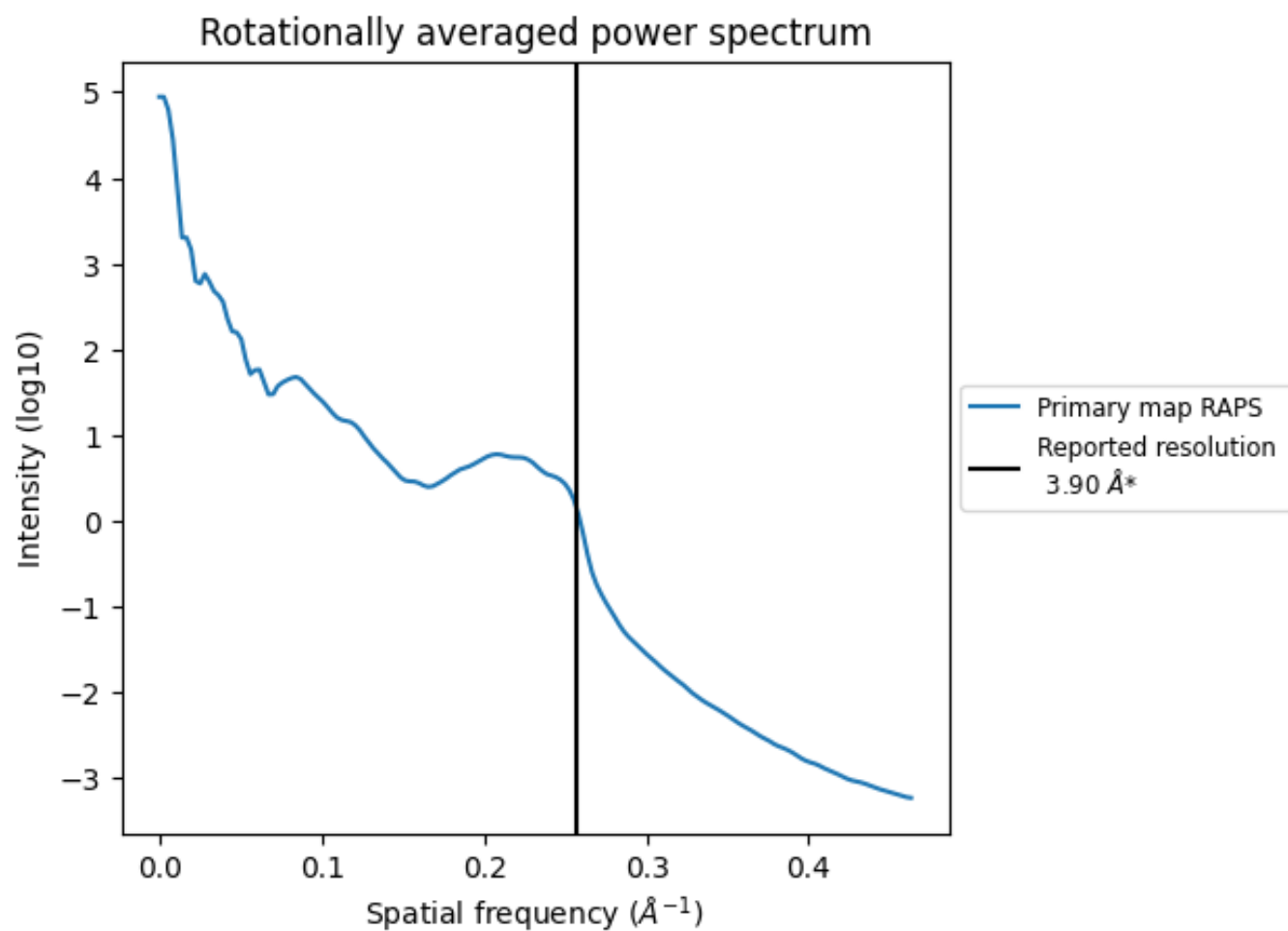
7.2 Volume estimate [i](#)



The volume at the recommended contour level is 134 nm³; this corresponds to an approximate mass of 121 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 ⓘ

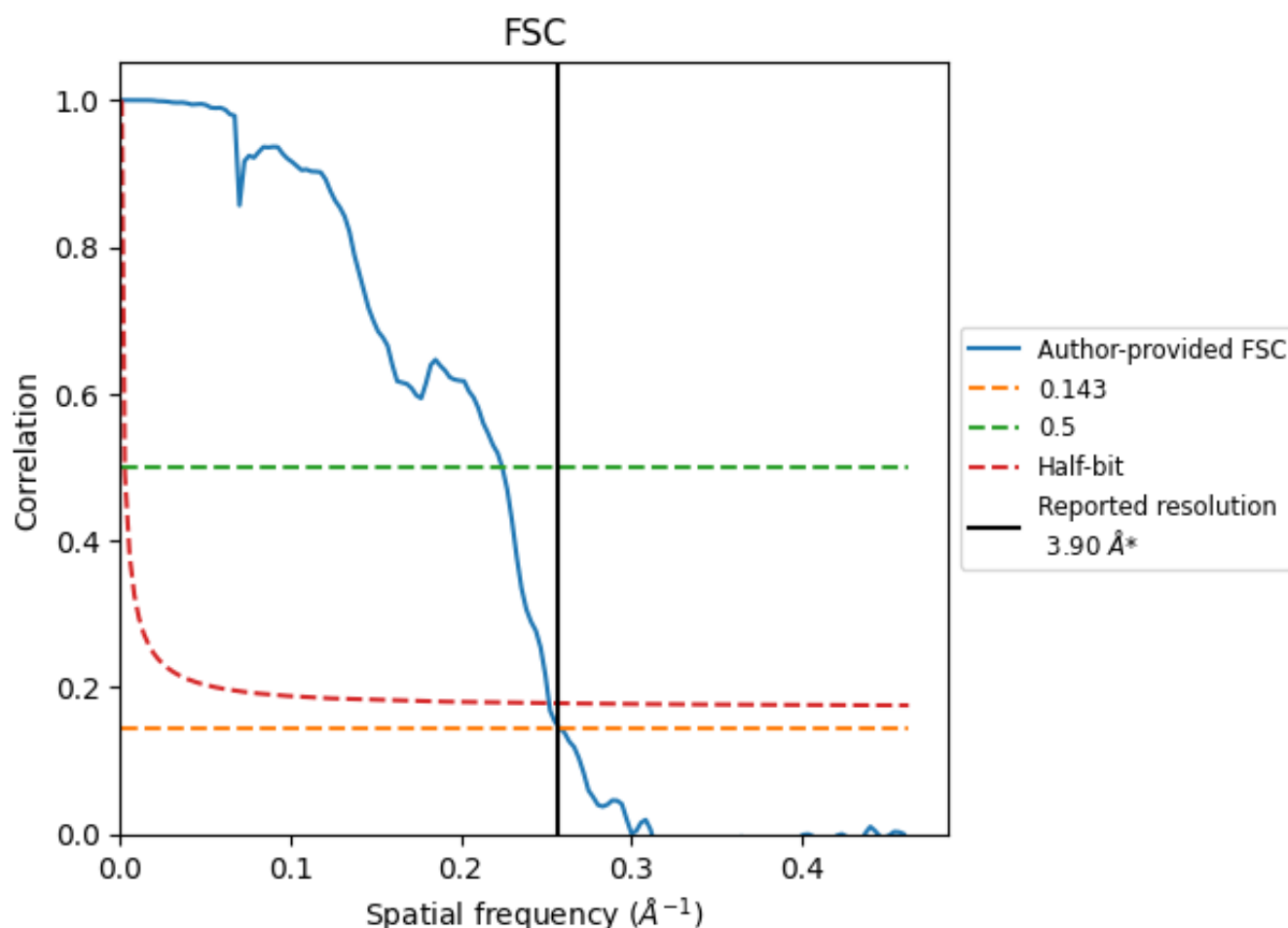


*Reported resolution corresponds to spatial frequency of 0.256 Å⁻¹

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 0.256 Å⁻¹

8.2 Resolution estimates [i](#)

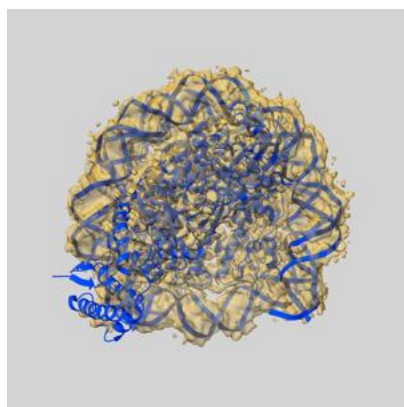
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.90	-	-
Author-provided FSC curve	3.87	4.46	3.97
Unmasked-calculated*	-	-	-

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

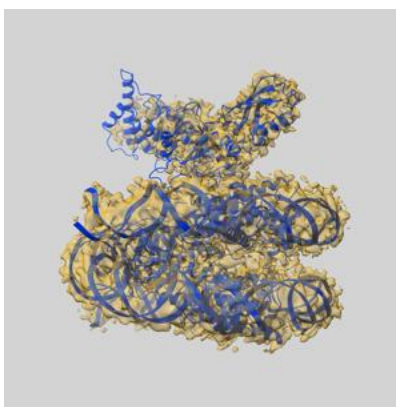
9 Map-model fit [i](#)

This section contains information regarding the fit between EMDB map EMD-0458 and PDB model 6NN6. Per-residue inclusion information can be found in section [3](#) on page [7](#).

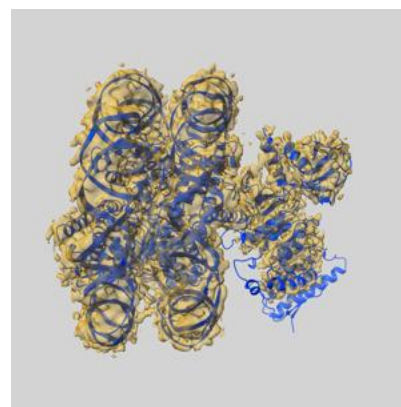
9.1 Map-model overlay [i](#)



X



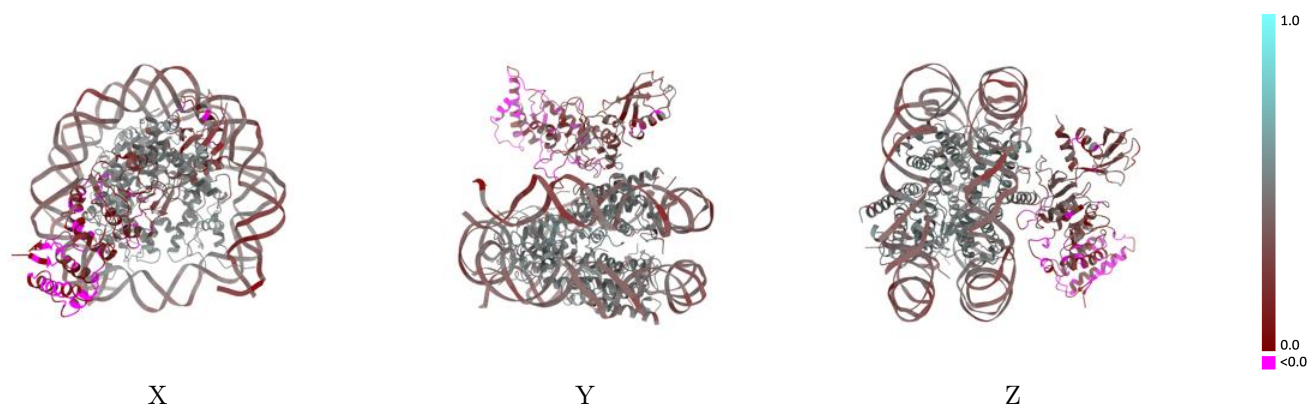
Y



Z

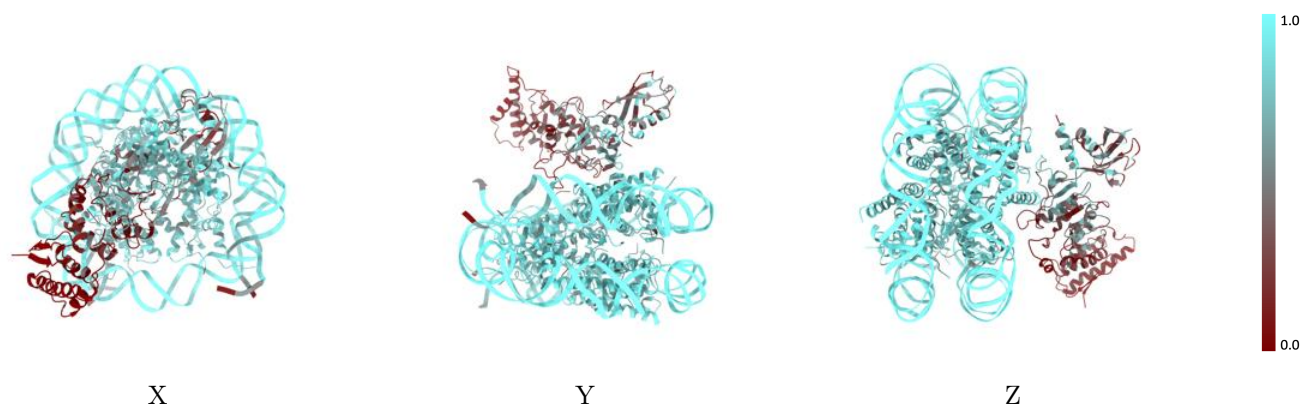
The images above show the 3D surface view of the map at the recommended contour level 0.03 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

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



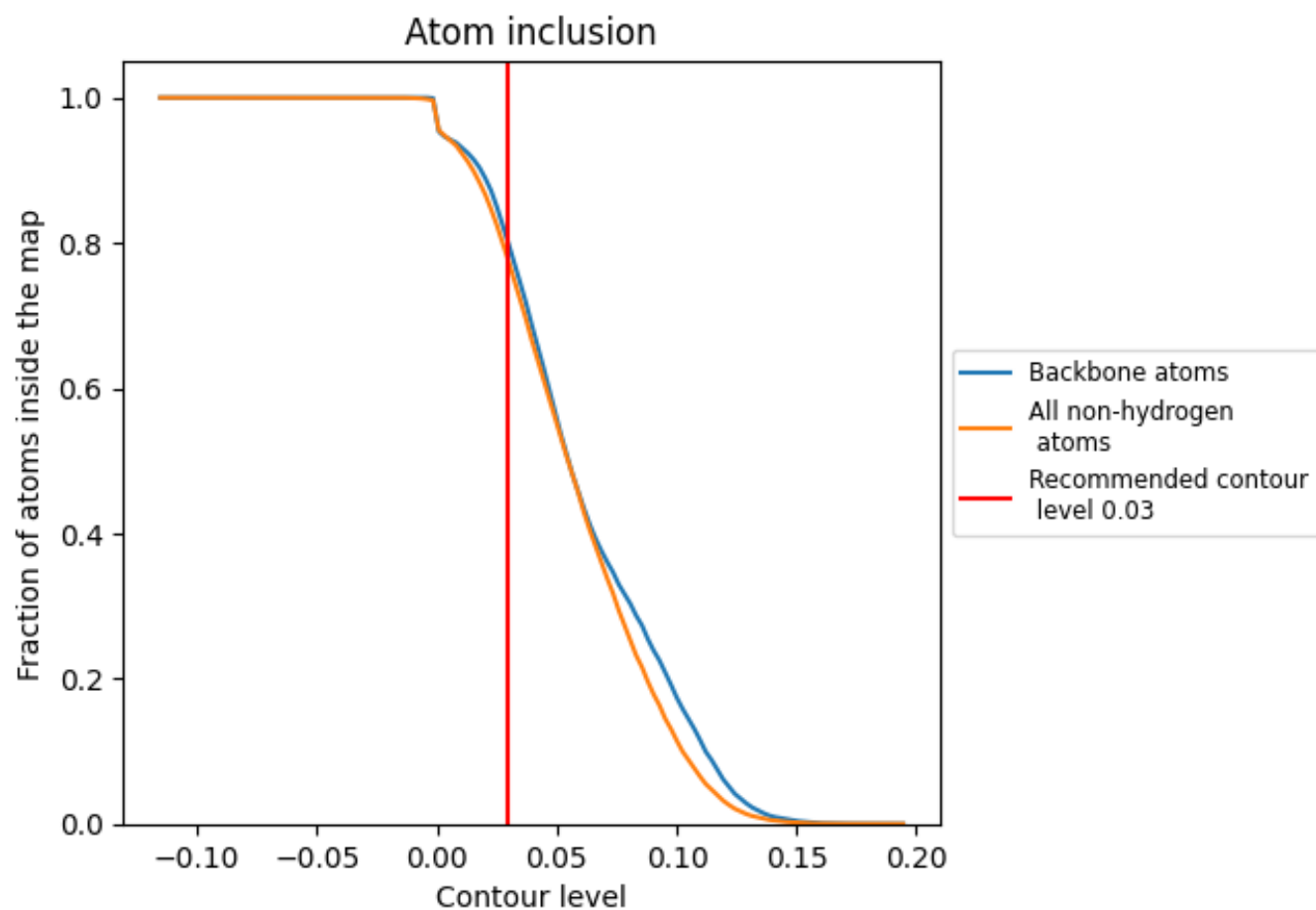
The images above show the model with each residue coloured according to its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions 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 (0.03).



























9.4 Atom inclusion [i](#)



At the recommended contour level, 80% of all backbone atoms, 77% 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 (0.03) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	 0.7741	 0.3710
A	 0.8431	 0.4860
B	 0.8813	 0.5070
C	 0.8383	 0.4800
D	 0.8626	 0.4810
E	 0.8623	 0.4990
F	 0.8908	 0.5000
G	 0.8622	 0.4910
H	 0.8612	 0.4770
I	 0.9448	 0.3620
J	 0.9394	 0.3510
K	 0.2746	 0.1630
L	 0.4469	 0.2750

