wwPDB EM Validation Summary Report (i)

May 12, 2024 – 09:08 am BST

PDB ID	:	6T7A
EMDB ID	:	EMD-10391
Title	:	Structure of human Sox11 transcription factor in complex with a nucleosome
Authors	:	Dodonova, S.O.; Zhu, F.; Dienemann, C.; Taipale, J.; Cramer, P.
Deposited on	:	2019-10-21
Resolution	:	3.70 Å(reported)
Based on initial model	:	6T79

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:

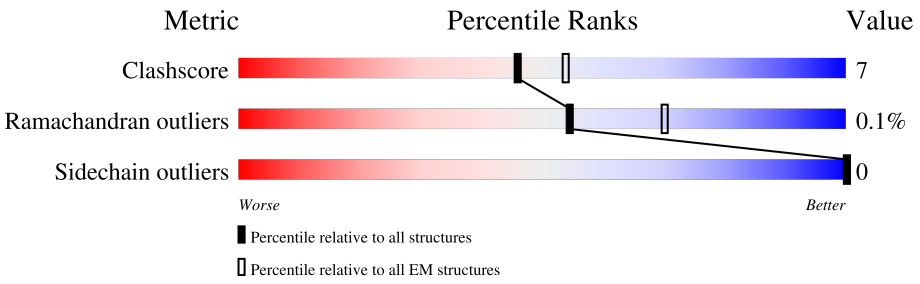
EMDB validation analysis	:	0.0.1.dev92
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.13
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 3.70 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 <=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	136	61% 11% 28%
1	E	136	60% 11% 29%
2	B	103	5% 64% 19% 17%
2	F	103	7% 65% 19% 16%
3	C	151	62% 8% 30%
3	G	151	60% 10% 30%
4	D	126	66% 10% 25%
4	H	126	66% 10% 25%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	I	147	54% 10% • 35%
6	J	147	53% 12% 35%
7	K	109	14% 71% 29%

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 10338 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
			Total	C	N	O	S		
1	A	98	809	511	157	139	2	0	0
1	E	97	800	505	155	138	2	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	110	ALA	CYS	engineered mutation	UNP Q71DI3
E	110	ALA	CYS	engineered mutation	UNP Q71DI3

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	B	86	698	440	141	116	1	0	0
2	F	87	703	442	142	118	1	0	0

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

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
3	C	106	819	517	160	142	0	0
3	G	106	819	517	160	142	0	0

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-21	MET	-	initiating methionine	UNP P04908
C	-20	GLY	-	expression tag	UNP P04908
C	-19	SER	-	expression tag	UNP P04908

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	-18	SER	-	expression tag	UNP P04908
C	-17	HIS	-	expression tag	UNP P04908
C	-16	HIS	-	expression tag	UNP P04908
C	-15	HIS	-	expression tag	UNP P04908
C	-14	HIS	-	expression tag	UNP P04908
C	-13	HIS	-	expression tag	UNP P04908
C	-12	HIS	-	expression tag	UNP P04908
C	-11	GLU	-	expression tag	UNP P04908
C	-10	ASN	-	expression tag	UNP P04908
C	-9	LEU	-	expression tag	UNP P04908
C	-8	TYR	-	expression tag	UNP P04908
C	-7	PHE	-	expression tag	UNP P04908
C	-6	GLN	-	expression tag	UNP P04908
C	-5	SER	-	expression tag	UNP P04908
C	-4	ASN	-	expression tag	UNP P04908
C	-3	ALA	-	expression tag	UNP P04908
C	-2	PRO	-	expression tag	UNP P04908
C	-1	TRP	-	expression tag	UNP P04908
G	-21	MET	-	initiating methionine	UNP P04908
G	-20	GLY	-	expression tag	UNP P04908
G	-19	SER	-	expression tag	UNP P04908
G	-18	SER	-	expression tag	UNP P04908
G	-17	HIS	-	expression tag	UNP P04908
G	-16	HIS	-	expression tag	UNP P04908
G	-15	HIS	-	expression tag	UNP P04908
G	-14	HIS	-	expression tag	UNP P04908
G	-13	HIS	-	expression tag	UNP P04908
G	-12	HIS	-	expression tag	UNP P04908
G	-11	GLU	-	expression tag	UNP P04908
G	-10	ASN	-	expression tag	UNP P04908
G	-9	LEU	-	expression tag	UNP P04908
G	-8	TYR	-	expression tag	UNP P04908
G	-7	PHE	-	expression tag	UNP P04908
G	-6	GLN	-	expression tag	UNP P04908
G	-5	SER	-	expression tag	UNP P04908
G	-4	ASN	-	expression tag	UNP P04908
G	-3	ALA	-	expression tag	UNP P04908
G	-2	PRO	-	expression tag	UNP P04908
G	-1	TRP	-	expression tag	UNP P04908

- Molecule 4 is a protein called Histone H2B type 1-K.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	D	95	744	467	136	139	2	0	0
4	H	95	744	467	136	139	2	0	0

- Molecule 5 is a DNA chain called DNA (147-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
5	I	95	1936	936	303	602	95	0	0

- Molecule 6 is a DNA chain called DNA (147-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
6	J	95	1959	933	393	538	95	0	0

- Molecule 7 is a protein called Transcription factor SOX-11.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
7	K	77	307	154	77	76	0	0

There are 3 discrepancies between the modelled and reference sequences:

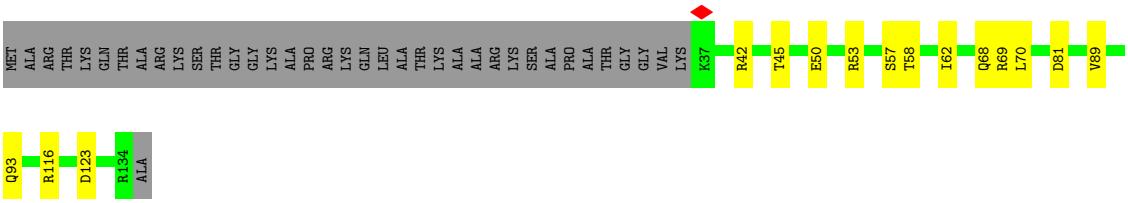
Chain	Residue	Modelled	Actual	Comment	Reference
K	30	SER	-	expression tag	UNP P35716
K	31	ASN	-	expression tag	UNP P35716
K	32	ALA	-	expression tag	UNP P35716

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 and atom inclusion in map density. 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. A red diamond above a residue indicates a poor fit to the EM map for this residue (all-atom inclusion < 40%). 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.

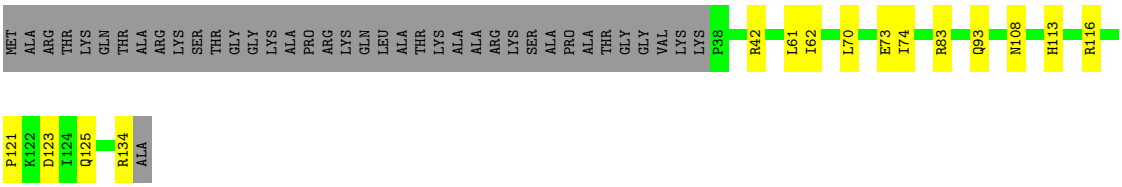
- Molecule 1: Histone H3.2

Chain A:



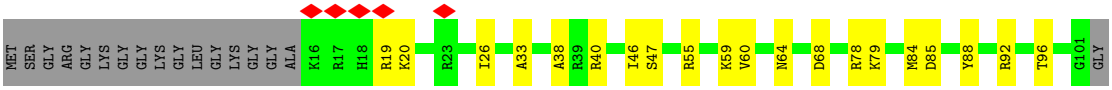
- Molecule 1: Histone H3.2

Chain E:



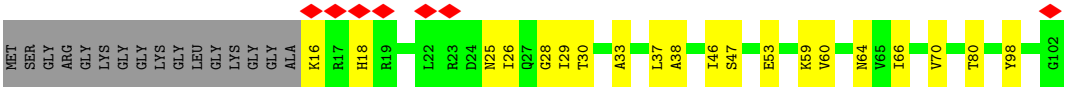
- Molecule 2: Histone H4

Chain B:



- Molecule 2: Histone H4

Chain F:



- Molecule 3: Histone H2A type 1-B/E

Chain C: 62% 8% 30%

MET GLY SER SER HIS HIS HIS HIS HIS HIS GLU ASN LEU TYR PHE GLN SER ASN ALA PRO TRP MET SER GLY ARG GLY LYS GLN GLY GLY LYS ALA ARG ALA K13 T16 S19 R20 Q24 L63 E64 L65 D72 I78 I79 P80 A86 V100 L115 K118

LYS THR GLU SER HIS HIS LYS ALA LYS GLY LYS

• Molecule 3: Histone H2A type 1-B/E

Chain G: 60% 10% 30%

MET GLY SER SER HIS HIS HIS HIS HIS HIS GLU ASN LEU TYR PHE GLN SER ASN ALA PRO TRP MET SER GLY ARG GLY LYS GLN GLY GLY LYS ALA ARG ALA K13 T16 S19 R20 Q24 R29 N38 Y39 S40 E41 R42 V54 T59 L65 D72 R81

A86 V100 K118 LYS THR GLU SER HIS HIS LYS ALA LYS GLY LYS

• Molecule 4: Histone H2B type 1-K

Chain D: 66% 10% 25%

MET PRO GLU PRO ALA LYS SER ALA PRO ALA PRO LYS LYS GLY SER LYS LYS ALA VAL THR LYS ALA GLN LYS LYS ASP GLY LYS LYS ARG K27 R28 S29 R30 K31 K40 V41 L42 K43 I51 I58 H79 R83 V95 R96 L97 L98 L99 K117 A121 LYS

• Molecule 4: Histone H2B type 1-K

Chain H: 66% 10% 25%

MET PRO GLU PRO ALA LYS SER ALA PRO ALA PRO LYS LYS GLY SER LYS LYS ALA VAL THR LYS ALA GLN LYS LYS ASP GLY LYS LYS ARG K27 R30 K31 E32 S33 K40 K43 M59 R69 E73 V95 L98 L99 A107 K117 A121 LYS

• Molecule 5: DNA (147-MER)

Chain I: 54% 10% • 35%

DA DT DC DT DA DC DA DC DG DA DC DG DC DT DC DT DT DC DC DG DA DT C23 T27 T28 T29 A39 C53 T54 G69 T79 T80 A81 T87 T88 T92 T99 T100 C101 G117 DA DC DA DC DT DC DG DG DA DG DA

DT DC DG DG DA DA DG DA DG DC DA DC DA DC DG DT DG DA DT

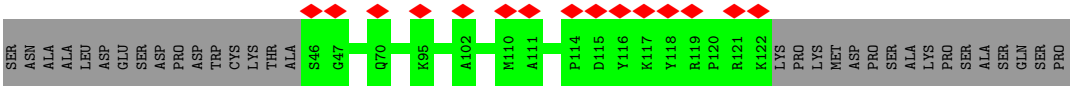
• Molecule 6: DNA (147-MER)

Chain J: 53% 12% 35%

DA DT DC DA DC DG DT DG DT DG DC DT DC DT DT DC DC DG DA DT DC DT DC DC DG DA DG DT DG DT C29 G30 T31 T32 T38 T44 A63 C69 A81 A85 A86 G87 A88 T91 T101 A102 A103 C104 G105 G123 DA

• Molecule 7: Transcription factor SOX-11

Chain K:



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	202142	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 ($e^-/Å^2$)	1.125	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	3500	Depositor
Magnification	130000	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	0.258	Depositor
Minimum map value	-0.130	Depositor
Average map value	0.003	Depositor
Map value standard deviation	0.021	Depositor
Recommended contour level	0.04	Depositor
Map size (Å)	147.0, 147.0, 147.0	wwPDB
Map dimensions	140, 140, 140	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.05, 1.05, 1.05	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	0.34	0/821	0.46	0/1101
1	E	0.33	0/812	0.48	0/1089
2	B	0.35	0/706	0.52	0/943
2	F	0.38	0/711	0.55	0/948
3	C	0.33	0/829	0.51	0/1118
3	G	0.33	0/829	0.51	0/1118
4	D	0.31	0/755	0.50	0/1014
4	H	0.31	0/755	0.50	0/1014
5	I	0.70	0/2158	1.16	3/3331 (0.1%)
6	J	0.74	0/2210	0.92	0/3406
7	K	0.27	0/306	0.50	0/380
All	All	0.53	0/10892	0.79	3/15462 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	I	87	DT	P-O3'-C3'	5.30	126.06	119.70
5	I	92	DT	O4'-C4'-C3'	-5.23	102.41	104.50
5	I	88	DT	C3'-C2'-C1'	-5.17	96.29	102.50

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	809	0	851	19	0
1	E	800	0	839	21	0
2	B	698	0	752	23	0
2	F	703	0	753	35	0
3	C	819	0	879	12	0
3	G	819	0	879	20	0
4	D	744	0	769	15	0
4	H	744	0	769	15	0
5	I	1936	0	1094	20	0
6	J	1959	0	1062	14	0
7	K	307	0	75	0	0
All	All	10338	0	8722	123	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 7.

The worst 5 of 123 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:70:LEU:HD11	2:B:26:ILE:CD1	1.69	1.21
1:A:70:LEU:CD1	2:B:26:ILE:HD13	1.75	1.16
1:E:73:GLU:OE1	2:F:25:ASN:HB2	1.36	1.13
1:A:70:LEU:HD11	2:B:26:ILE:HD13	1.30	1.11
1:E:70:LEU:HD11	2:F:26:ILE:HD13	1.36	1.03

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	96/136 (71%)	95 (99%)	1 (1%)	0	100 100
1	E	95/136 (70%)	94 (99%)	1 (1%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	84/103 (82%)	84 (100%)	0	0	100	100
2	F	85/103 (82%)	83 (98%)	1 (1%)	1 (1%)	13	48
3	C	104/151 (69%)	101 (97%)	3 (3%)	0	100	100
3	G	104/151 (69%)	101 (97%)	3 (3%)	0	100	100
4	D	93/126 (74%)	91 (98%)	2 (2%)	0	100	100
4	H	93/126 (74%)	91 (98%)	2 (2%)	0	100	100
7	K	75/109 (69%)	73 (97%)	2 (3%)	0	100	100
All	All	829/1141 (73%)	813 (98%)	15 (2%)	1 (0%)	54	83

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	F	29	ILE

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	85/110 (77%)	85 (100%)	0	100	100
1	E	84/110 (76%)	84 (100%)	0	100	100
2	B	72/79 (91%)	72 (100%)	0	100	100
2	F	72/79 (91%)	72 (100%)	0	100	100
3	C	84/119 (71%)	84 (100%)	0	100	100
3	G	84/119 (71%)	84 (100%)	0	100	100
4	D	81/105 (77%)	81 (100%)	0	100	100
4	H	81/105 (77%)	81 (100%)	0	100	100
All	All	643/826 (78%)	643 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes.

such sidechains are listed below:

Mol	Chain	Res	Type
1	E	93	GLN
1	E	108	ASN
4	H	79	HIS
1	E	113	HIS
1	A	113	HIS

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.

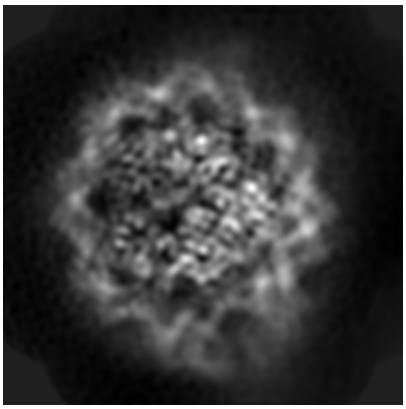
6 Map visualisation (i)

This section contains visualisations of the EMDB entry EMD-10391. These allow visual inspection of the internal detail of the map and identification of artifacts.

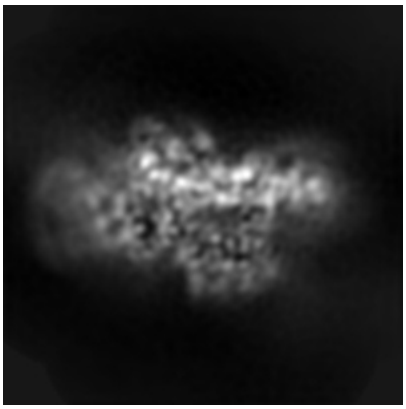
Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

6.1 Orthogonal projections (i)

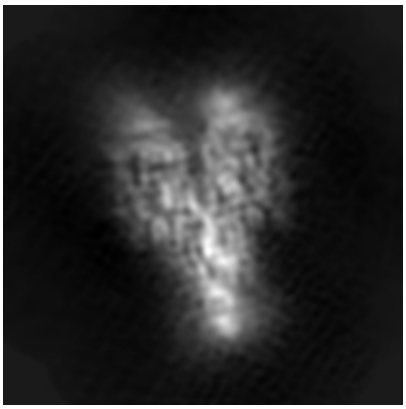
6.1.1 Primary map



X

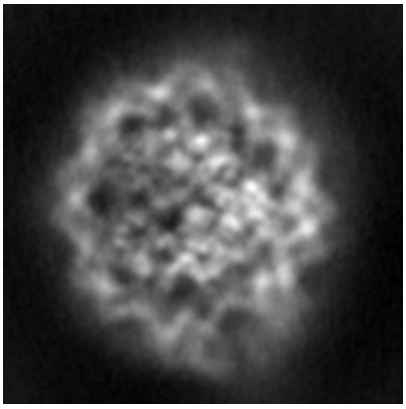


Y

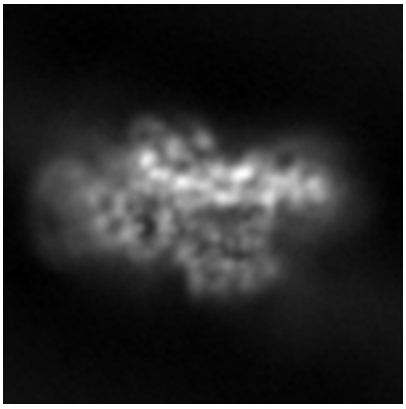


Z

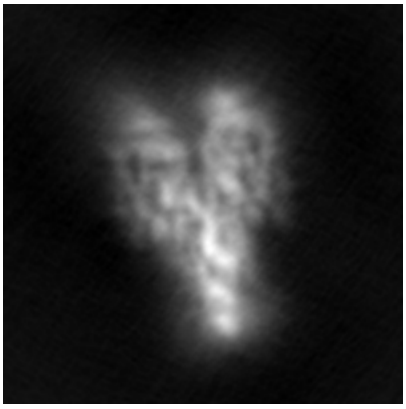
6.1.2 Raw map



X



Y

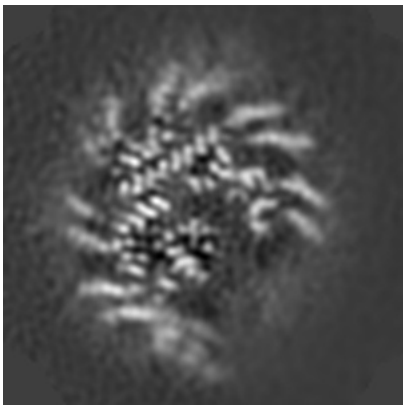


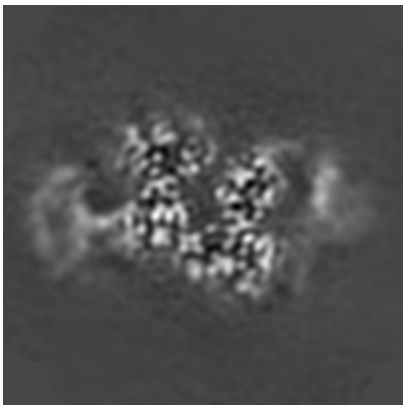
Z

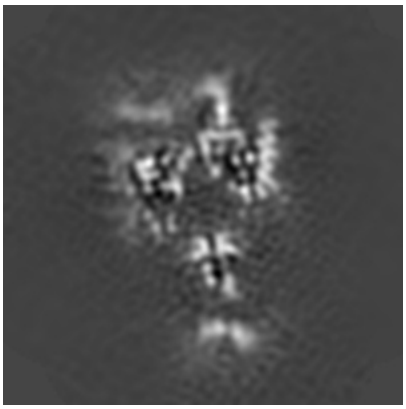
The images above show the map projected in three orthogonal directions.

6.2 Central slices (i)

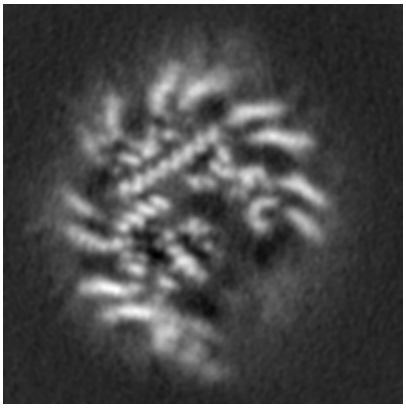
6.2.1 Primary map

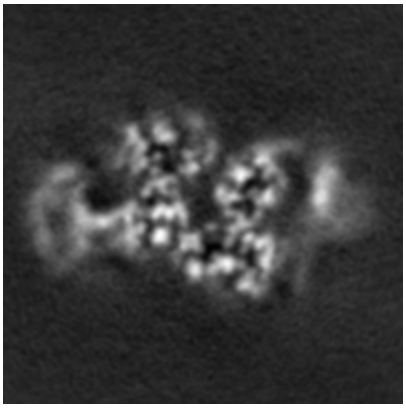

X Index: 70

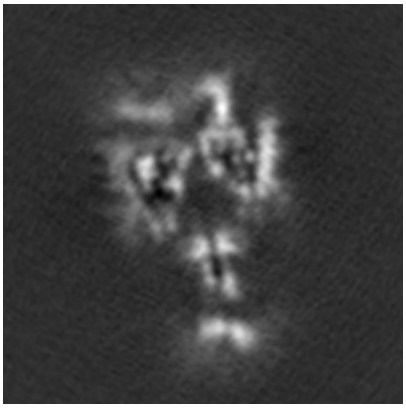

Y Index: 70


Z Index: 70

6.2.2 Raw map


X Index: 70

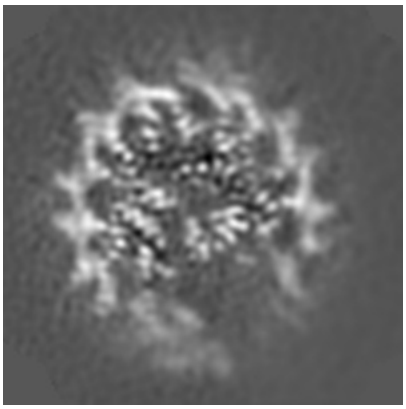

Y Index: 70

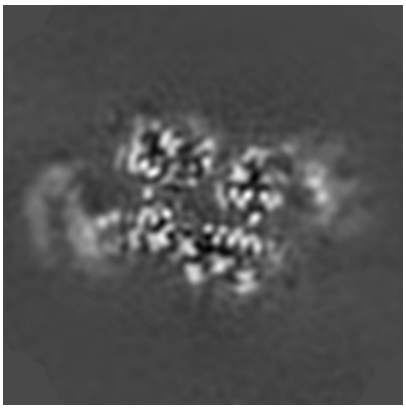

Z Index: 70

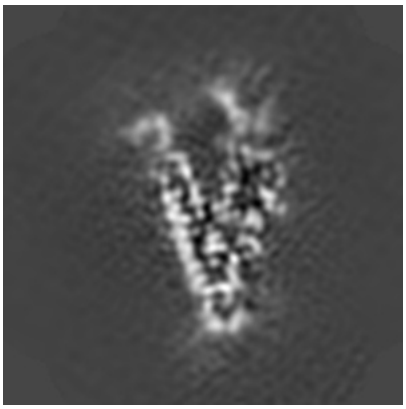
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

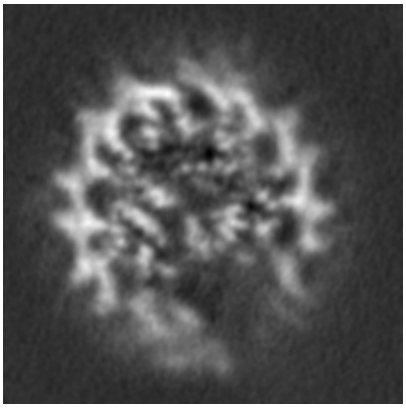
6.3.1 Primary map

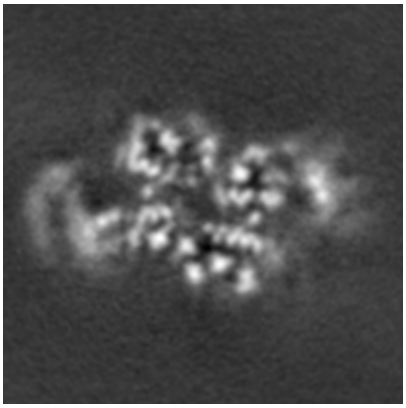

X Index: 77

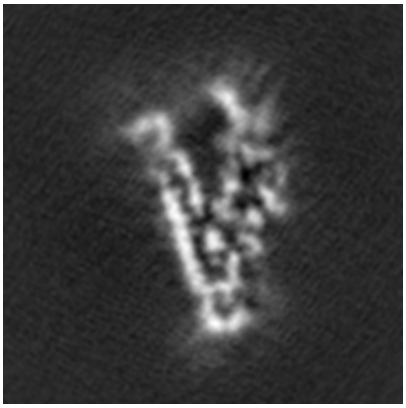

Y Index: 73


Z Index: 56

6.3.2 Raw map


X Index: 77

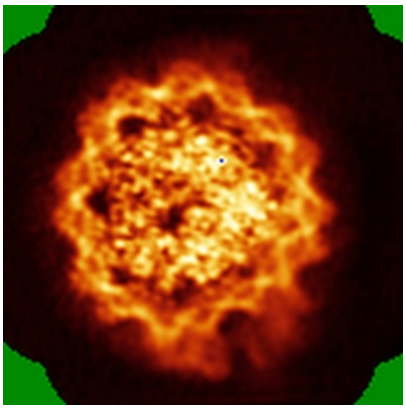

Y Index: 73

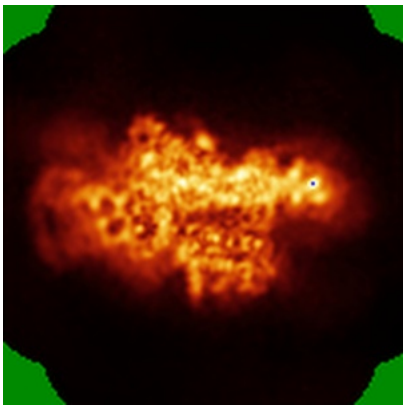

Z Index: 56

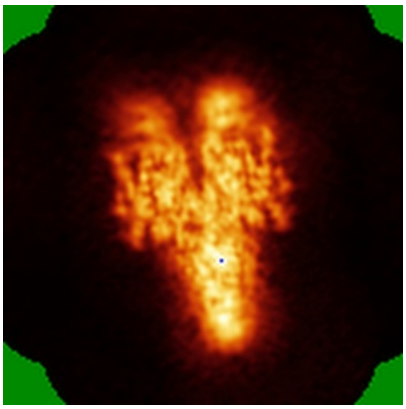
The images above show the largest variance slices of the map in three orthogonal directions.

6.4 Orthogonal standard-deviation projections (False-color) (i)

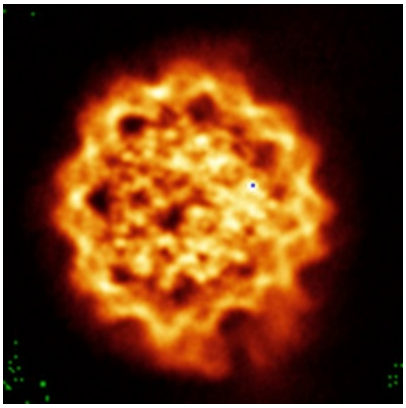
6.4.1 Primary map

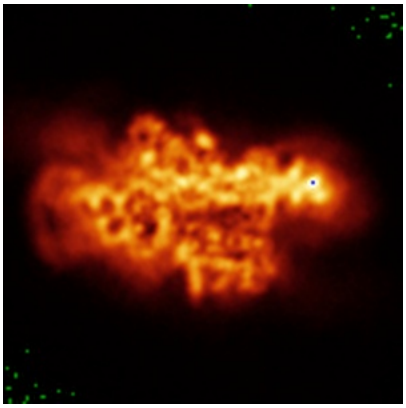

X

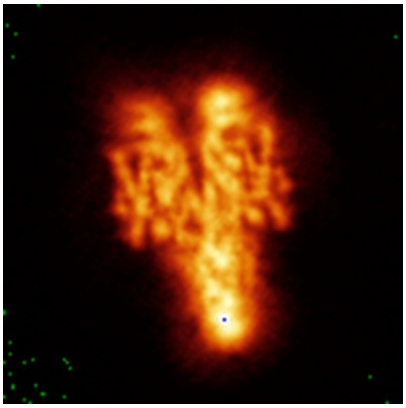

Y


Z

6.4.2 Raw map


X

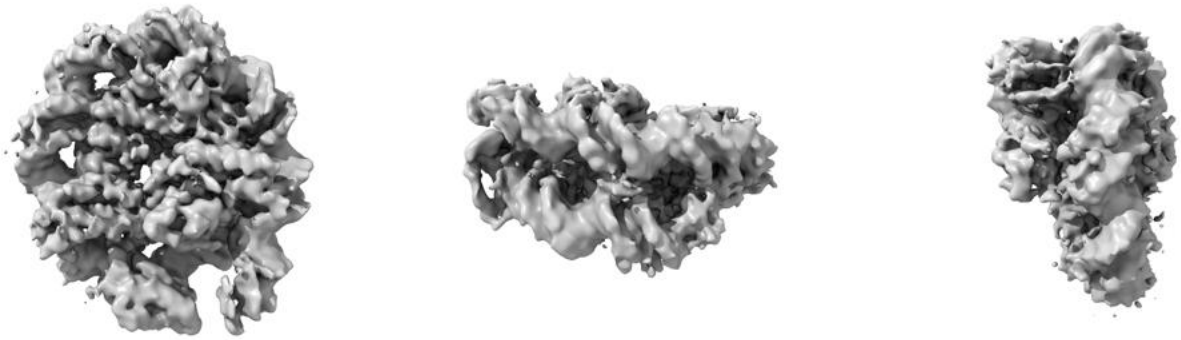

Y


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)

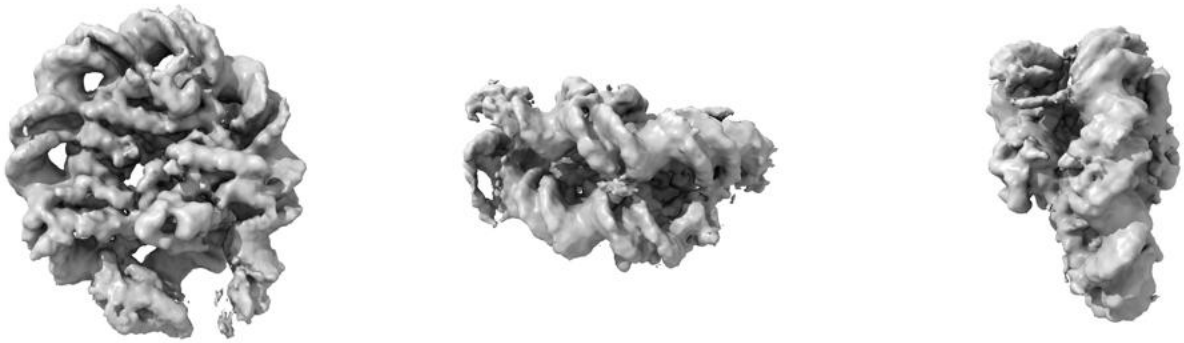
6.5.1 Primary map



X Y Z

The images above show the 3D surface view of the map at the recommended contour level 0.04. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

6.5.2 Raw map



X Y Z

These images show the 3D surface of the raw map. The raw map's contour level was selected so that its surface encloses the same volume as the primary map does at its recommended contour level.

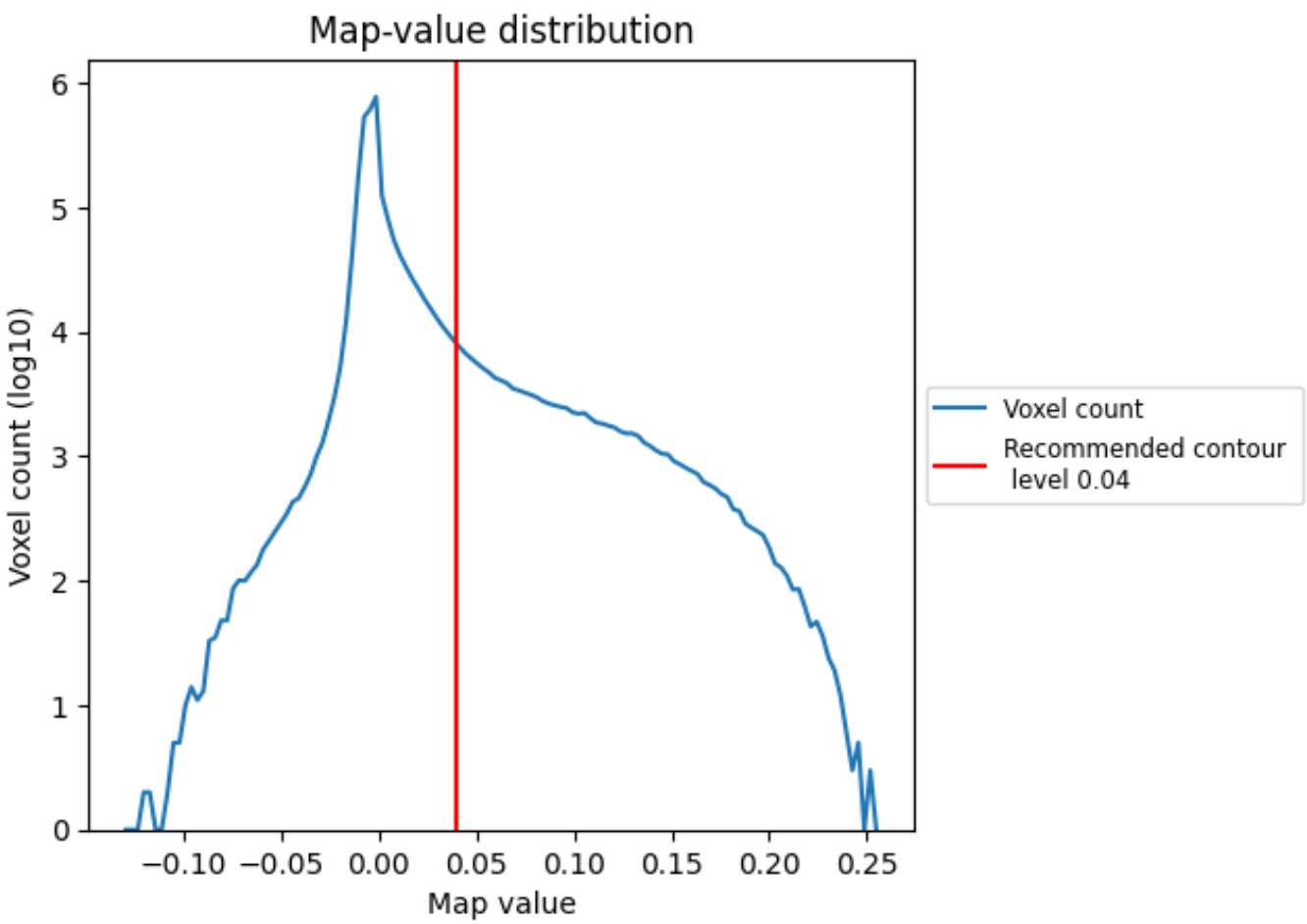
6.6 Mask visualisation (i)

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

7 Map analysis (i)

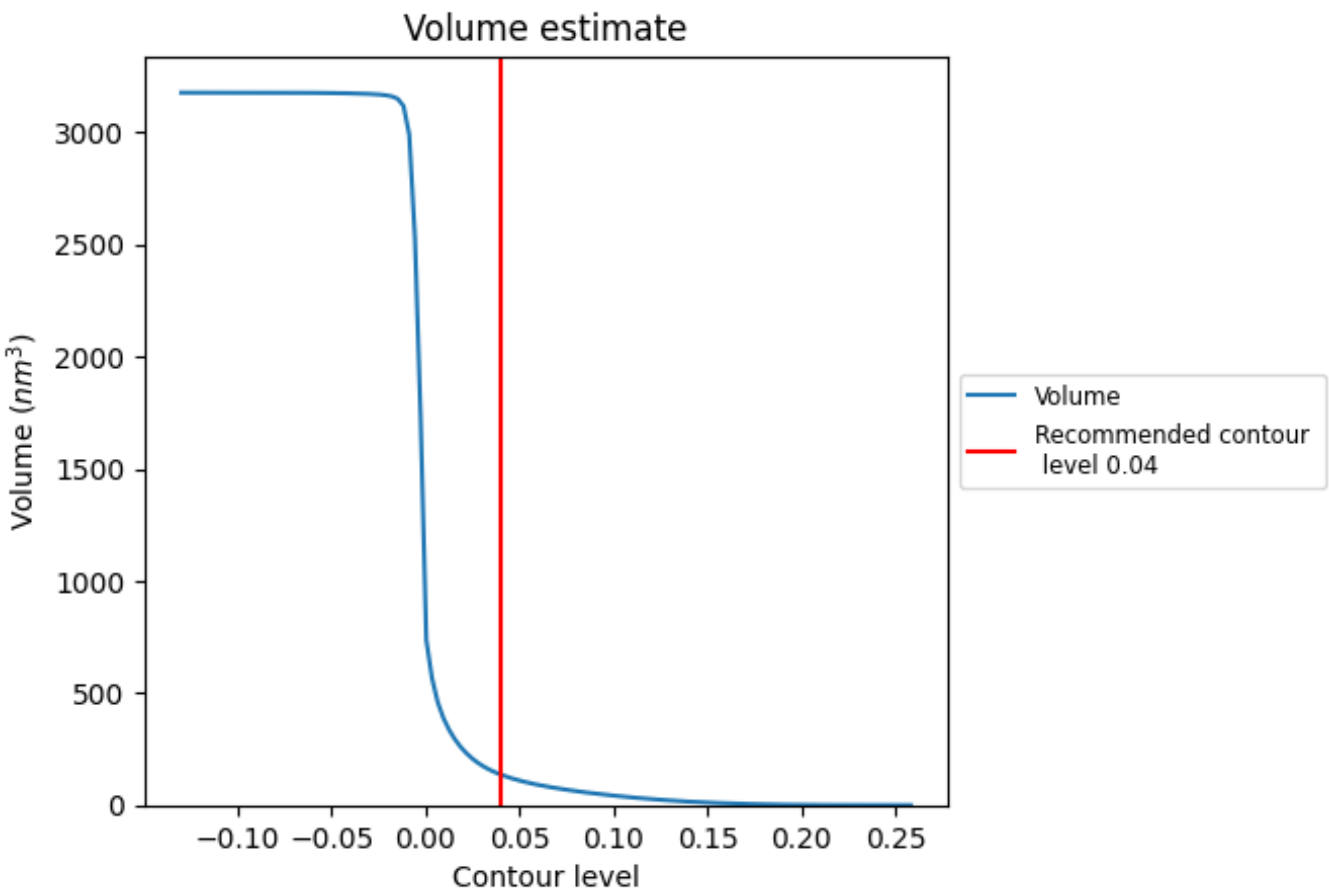
This section contains the results of statistical analysis of the map.

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.

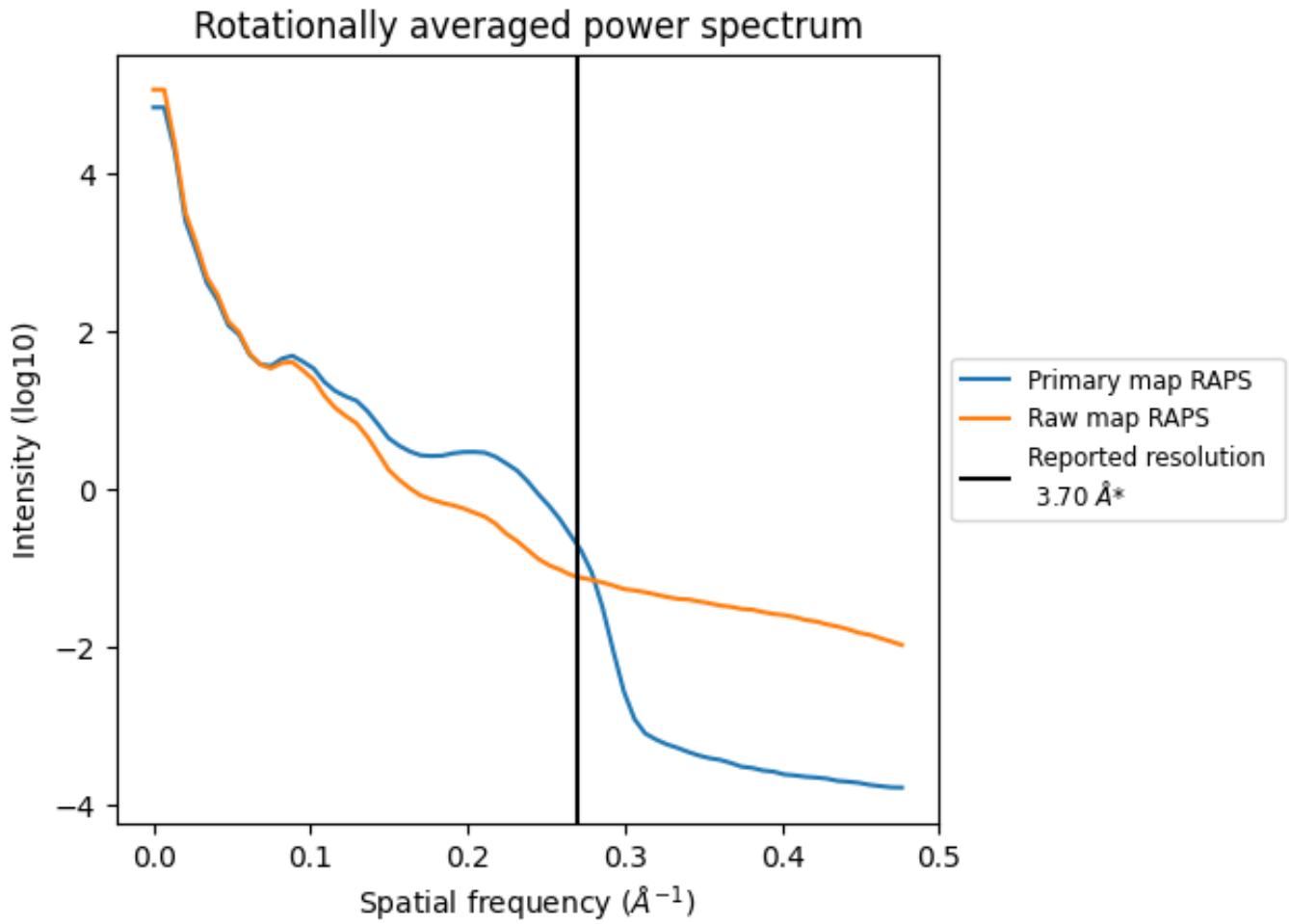
7.2 Volume estimate (i)



The volume at the recommended contour level is 136 nm^3; this corresponds to an approximate mass of 123 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)

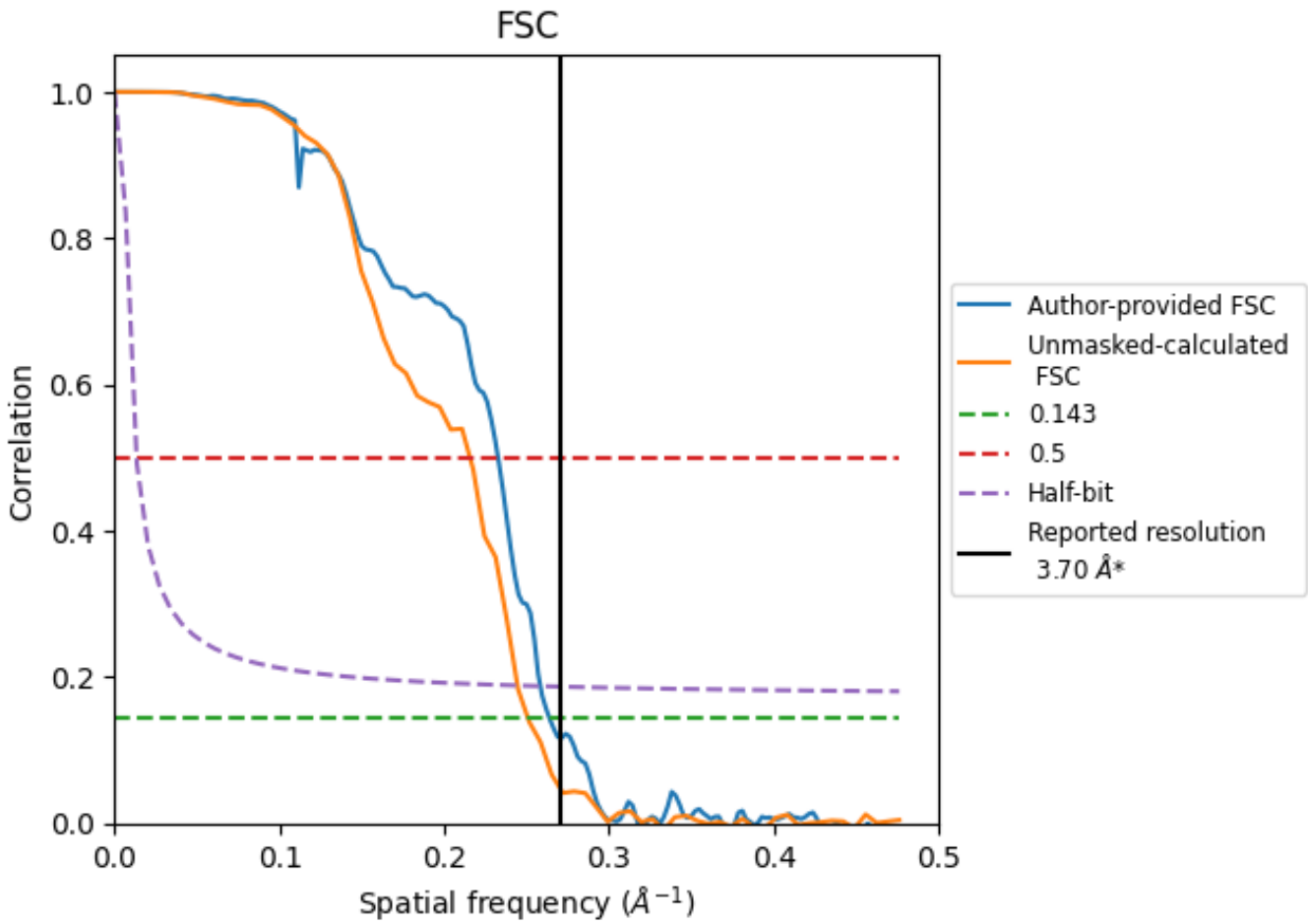


*Reported resolution corresponds to spatial frequency of 0.270 Å^{-1}

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 0.270 Å^{-1}

8.2 Resolution estimates (i)

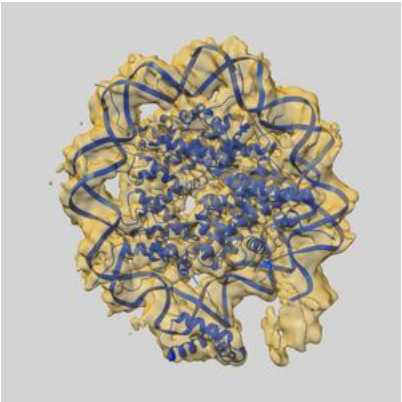
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.70	-	-
Author-provided FSC curve	3.79	4.30	3.86
Unmasked-calculated*	3.99	4.64	4.09

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

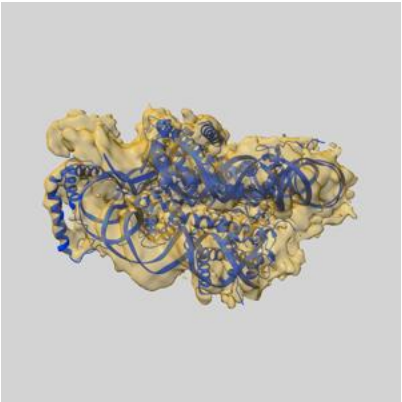
9 Map-model fit (i)

This section contains information regarding the fit between EMDB map EMD-10391 and PDB model 6T7A. Per-residue inclusion information can be found in section 3 on page 7.

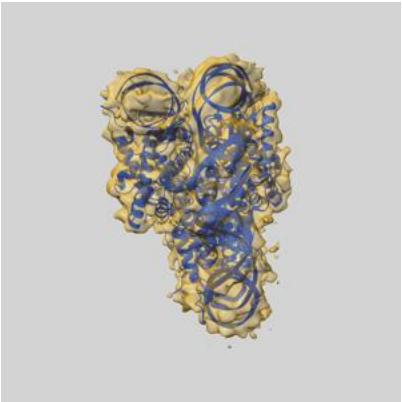
9.1 Map-model overlay (i)



X



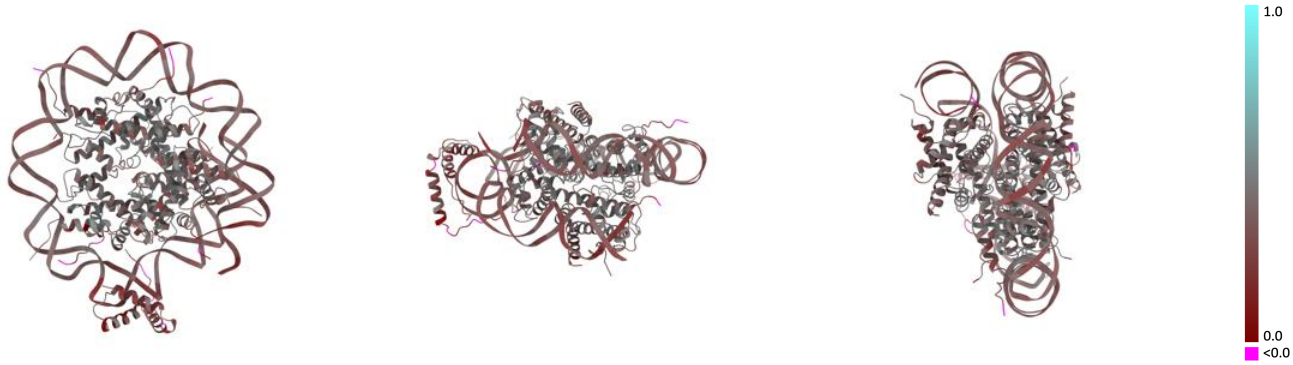
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.04 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.

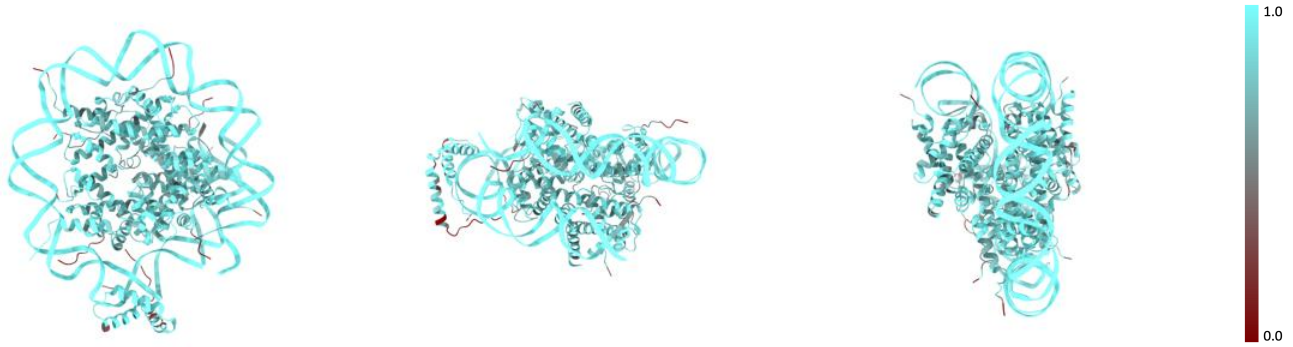
9.2 Q-score mapped to coordinate model (i)



X Y Z

The images above show the model with each residue coloured according 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.

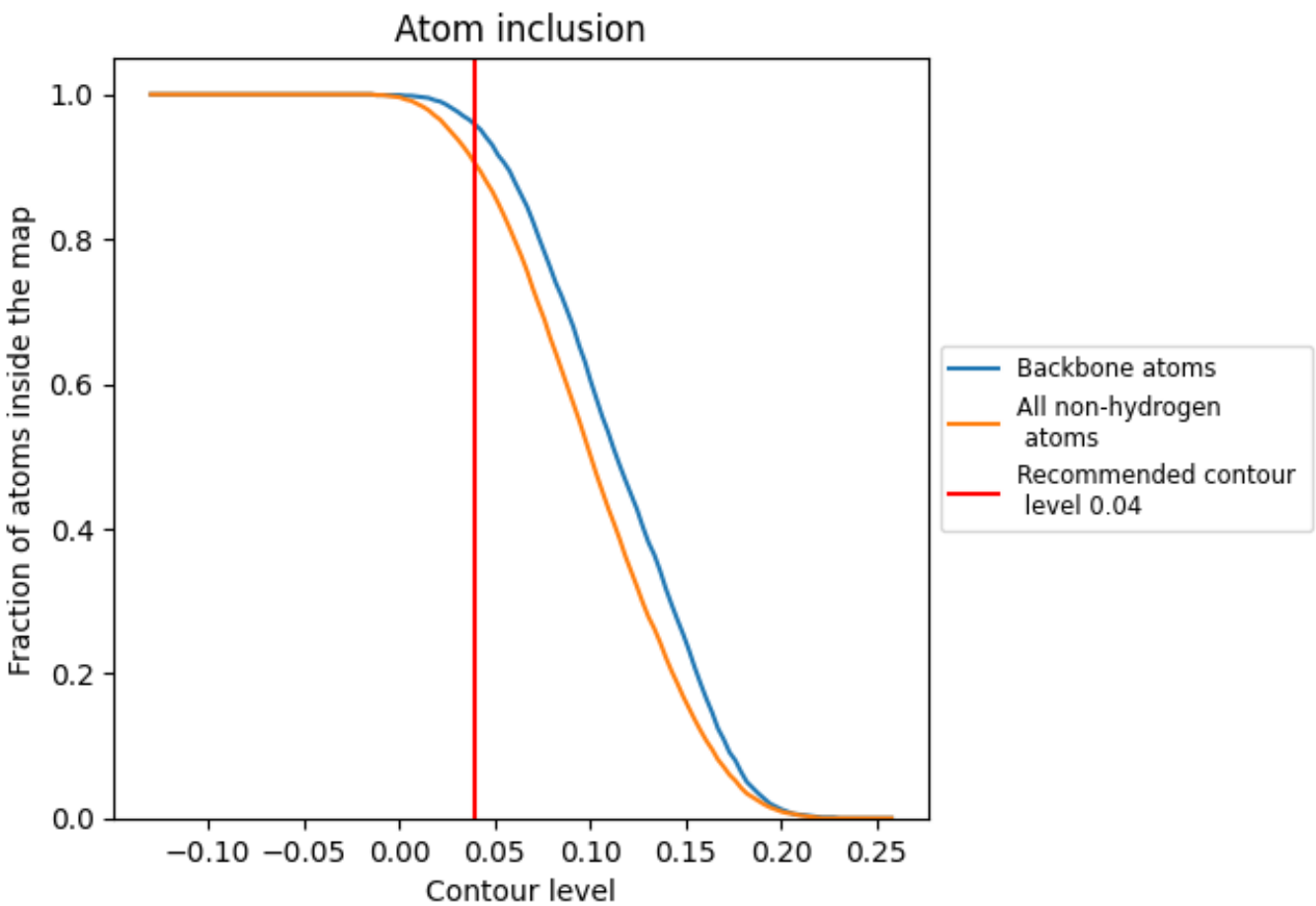
9.3 Atom inclusion mapped to coordinate model (i)



X Y Z

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 (0.04).

9.4 Atom inclusion (i)



At the recommended contour level, 96% of all backbone atoms, 90% of all non-hydrogen atoms, are inside the map.

9.5 Map-model fit summary (i)

The table lists the average atom inclusion at the recommended contour level (0.04) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.9040	0.3700
A	0.8830	0.4160
B	0.8680	0.4070
C	0.8730	0.4130
D	0.8520	0.3890
E	0.9000	0.4200
F	0.8230	0.3770
G	0.8460	0.4040
H	0.8700	0.3840
I	0.9770	0.3350
J	0.9720	0.3290
K	0.7650	0.2270