wwPDB EM Validation Summary Report (i)

Sep 13, 2022 – 05:47 pm BST

PDB ID	:	7PP4
EMDB ID	:	EMD-13579
Title	:	Cryo-EM structure of Mycobacterium tuberculosis RNA polymerase holoenzyme comprising sigma factor SigB
Authors	:	Brodolin, K.
Deposited on	:	2021-09-13
Resolution	:	3.84 Å(reported)
Based on initial model	:	6FBV

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:

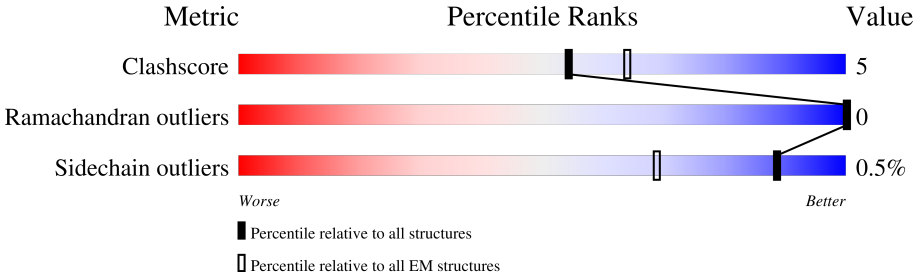
EMDB validation analysis	:	0.0.1.dev8
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.30

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 3.84 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	347	64% 36%
1	b	347	66% • 34%
2	c	1174	• 95% 5%
3	d	1322	• 96% •
4	e	110	• 75% 25%
5	f	343	6% 41% • 59%

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 23816 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 DNA-directed RNA polymerase subunit alpha.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	a	223	1700	1068	294	336	2	0	0
1	b	230	1749	1099	301	347	2	0	0

- Molecule 2 is a protein called DNA-directed RNA polymerase subunit beta.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	c	1112	8640	5408	1515	1678	39	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
c	5	MET	-	initiating methionine	UNP P9WGY9
c	6	VAL	ILE	conflict	UNP P9WGY9

- Molecule 3 is a protein called DNA-directed RNA polymerase subunit beta'.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	d	1269	9951	6231	1805	1874	41	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
d	1317	HIS	-	expression tag	UNP P9WGY7
d	1318	HIS	-	expression tag	UNP P9WGY7
d	1319	HIS	-	expression tag	UNP P9WGY7
d	1320	HIS	-	expression tag	UNP P9WGY7
d	1321	HIS	-	expression tag	UNP P9WGY7
d	1322	HIS	-	expression tag	UNP P9WGY7

- Molecule 4 is a protein called DNA-directed RNA polymerase subunit omega.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
4	e	83	654	417	108	129	0	0

- Molecule 5 is a protein called RNA polymerase sigma factor SigB.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
5	f	142	1119	705	210	201	3	0	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
f	-19	MET	-	initiating methionine	UNP P9WGI5
f	-18	GLY	-	expression tag	UNP P9WGI5
f	-17	SER	-	expression tag	UNP P9WGI5
f	-16	SER	-	expression tag	UNP P9WGI5
f	-15	HIS	-	expression tag	UNP P9WGI5
f	-14	HIS	-	expression tag	UNP P9WGI5
f	-13	HIS	-	expression tag	UNP P9WGI5
f	-12	HIS	-	expression tag	UNP P9WGI5
f	-11	HIS	-	expression tag	UNP P9WGI5
f	-10	HIS	-	expression tag	UNP P9WGI5
f	-9	SER	-	expression tag	UNP P9WGI5
f	-8	SER	-	expression tag	UNP P9WGI5
f	-7	GLY	-	expression tag	UNP P9WGI5
f	-6	LEU	-	expression tag	UNP P9WGI5
f	-5	VAL	-	expression tag	UNP P9WGI5
f	-4	PRO	-	expression tag	UNP P9WGI5
f	-3	ARG	-	expression tag	UNP P9WGI5
f	-2	GLY	-	expression tag	UNP P9WGI5
f	-1	SER	-	expression tag	UNP P9WGI5
f	0	HIS	-	expression tag	UNP P9WGI5

- Molecule 6 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		AltConf
			Total	Zn	
6	d	2	2	2	0

- Molecule 7 is MAGNESIUM ION (three-letter code: MG) (formula: Mg) (labeled as "Ligand

of Interest" by depositor).

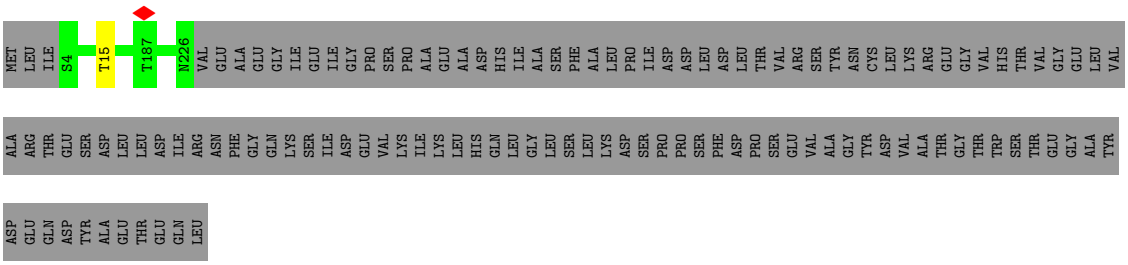
Mol	Chain	Residues	Atoms		AltConf
			Total	Mg	
7	d	1	1	1	0

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.

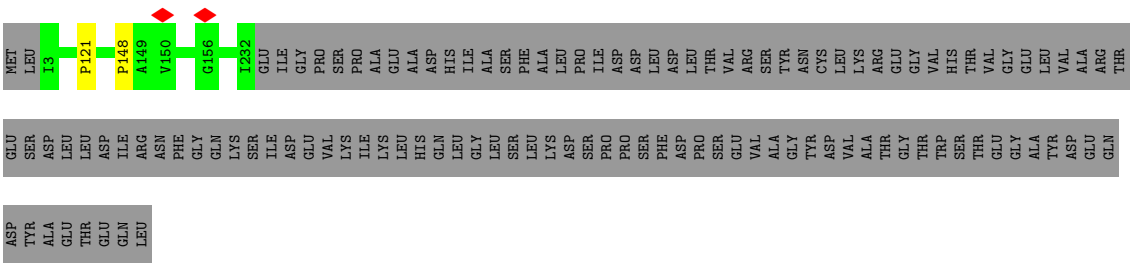
• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain a: 64% 36%



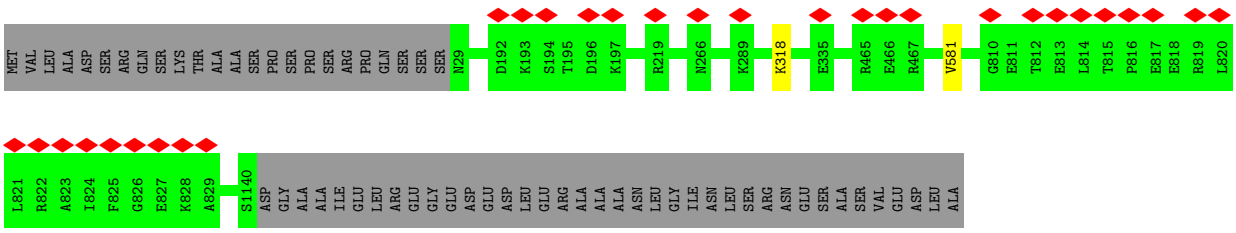
• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain b: 66% 34%



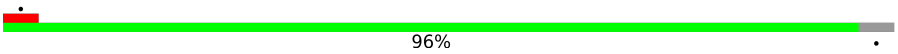
• Molecule 2: DNA-directed RNA polymerase subunit beta

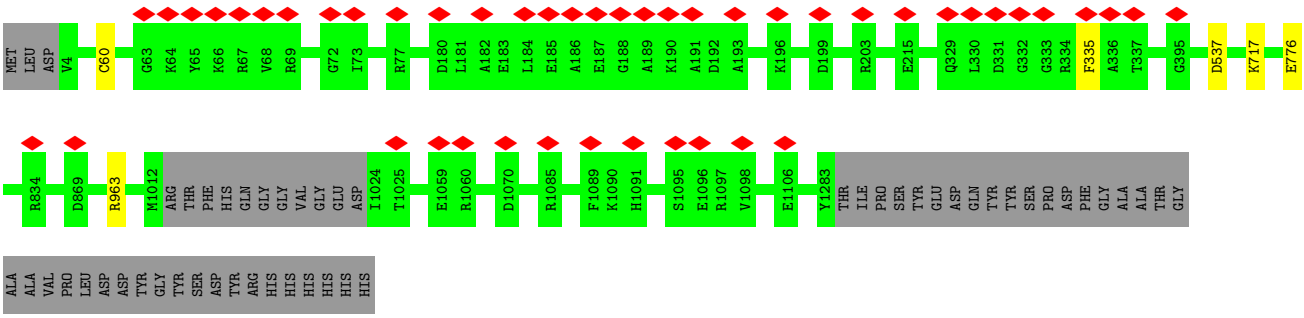
Chain c: 95% 5%



• Molecule 3: DNA-directed RNA polymerase subunit beta'

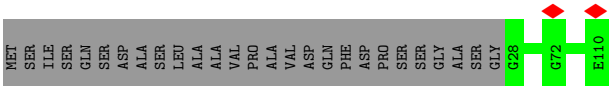
Chain d:





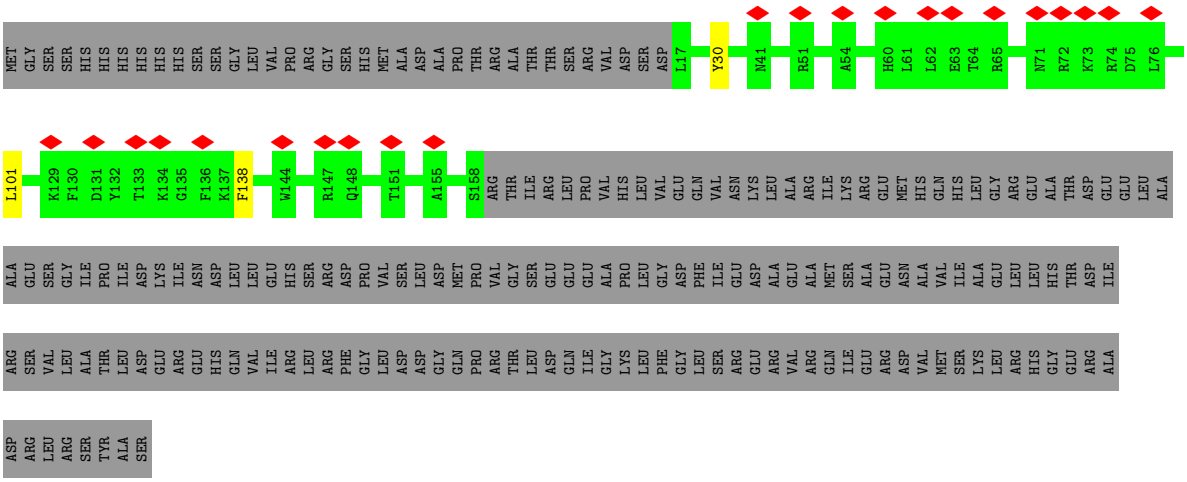
- Molecule 4: DNA-directed RNA polymerase subunit omega

Chain e:



- Molecule 5: RNA polymerase sigma factor SigB

Chain f:



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	115112	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 ($e^-/Å^2$)	49.6	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	2.139	Depositor
Minimum map value	-0.850	Depositor
Average map value	-0.004	Depositor
Map value standard deviation	0.046	Depositor
Recommended contour level	0.3	Depositor
Map size (Å)	495.0, 495.0, 495.0	wwPDB
Map dimensions	450, 450, 450	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.1, 1.1, 1.1	Depositor

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, MG

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	0.24	0/1726	0.51	0/2348
1	b	0.25	0/1775	0.51	0/2414
2	c	0.24	0/8799	0.49	0/11929
3	d	0.26	0/10118	0.52	0/13675
4	e	0.24	0/667	0.44	0/908
5	f	0.27	0/1134	0.55	0/1525
All	All	0.25	0/24219	0.51	0/32799

There are no bond length outliers.

There are no bond angle outliers.

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	1700	0	1734	0	0
1	b	1749	0	1781	0	0
2	c	8640	0	8584	0	0
3	d	9951	0	10035	0	0
4	e	654	0	648	0	0
5	f	1119	0	1155	0	0
6	d	2	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	d	1	0	0	0	0
All	All	23816	0	23937	0	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 5.

There are no clashes within the asymmetric unit.

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	221/347 (64%)	212 (96%)	9 (4%)	0	100	100
1	b	228/347 (66%)	222 (97%)	6 (3%)	0	100	100
2	c	1110/1174 (94%)	1076 (97%)	34 (3%)	0	100	100
3	d	1265/1322 (96%)	1220 (96%)	45 (4%)	0	100	100
4	e	81/110 (74%)	76 (94%)	5 (6%)	0	100	100
5	f	140/343 (41%)	138 (99%)	2 (1%)	0	100	100
All	All	3045/3643 (84%)	2944 (97%)	101 (3%)	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	192/297 (65%)	191 (100%)	1 (0%)	88	94
1	b	197/297 (66%)	195 (99%)	2 (1%)	76	86
2	c	946/995 (95%)	944 (100%)	2 (0%)	93	97
3	d	1058/1101 (96%)	1052 (99%)	6 (1%)	86	92
4	e	70/90 (78%)	70 (100%)	0	100	100
5	f	111/284 (39%)	108 (97%)	3 (3%)	44	68
All	All	2574/3064 (84%)	2560 (100%)	14 (0%)	89	94

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	d	537	ASP
3	d	717	LYS
5	f	138	PHE
5	f	30	TYR
5	f	101	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
3	d	369	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)

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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.

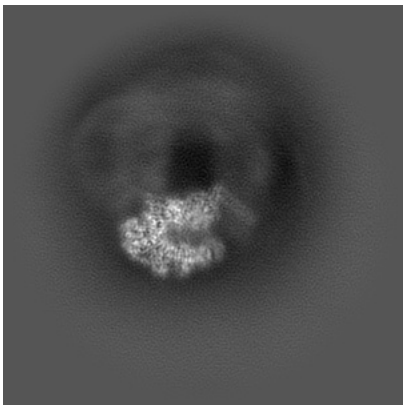
6 Map visualisation (i)

This section contains visualisations of the EMDB entry EMD-13579. These allow visual inspection of the internal detail of the map and identification of artifacts.

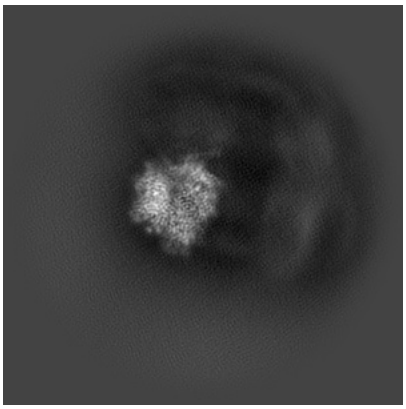
Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

6.1 Orthogonal projections (i)

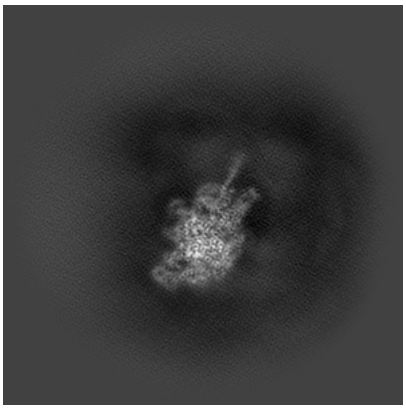
6.1.1 Primary map



X

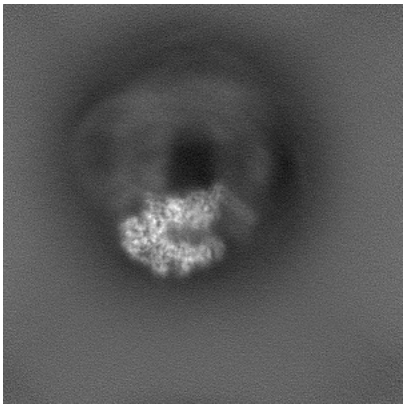


Y

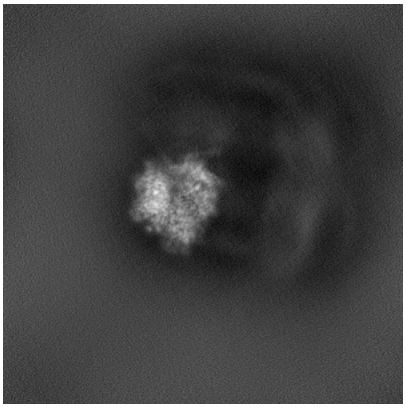


Z

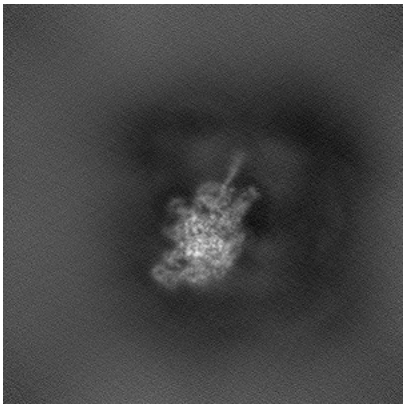
6.1.2 Raw map



X



Y

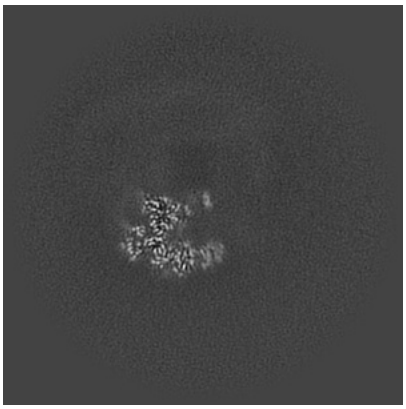


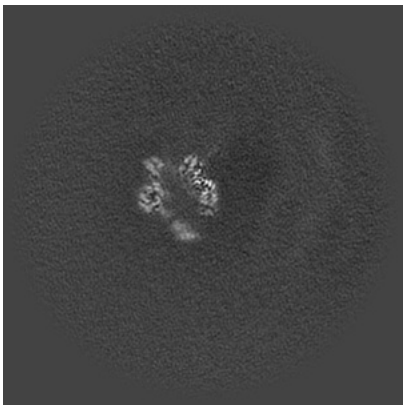
Z

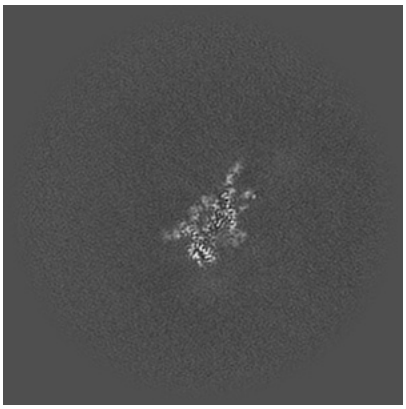
The images above show the map projected in three orthogonal directions.

6.2 Central slices (i)

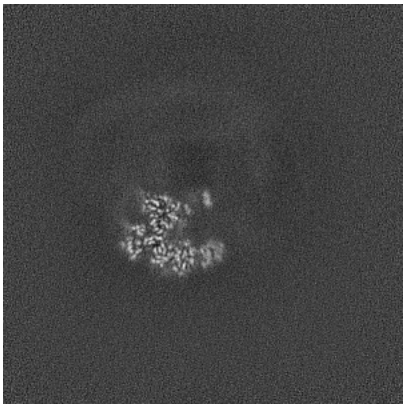
6.2.1 Primary map

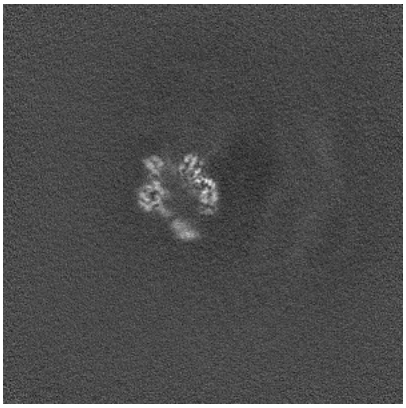

X Index: 225

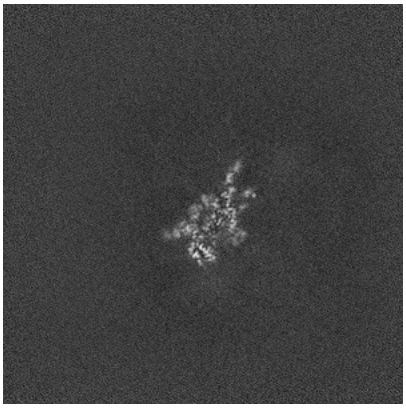

Y Index: 225


Z Index: 225

6.2.2 Raw map


X Index: 225

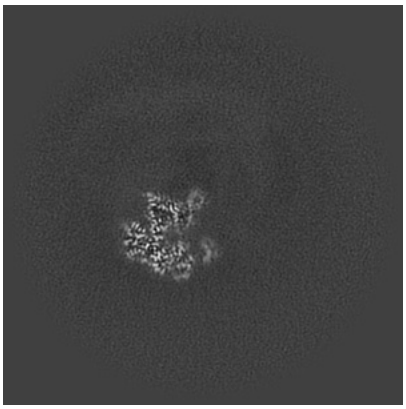

Y Index: 225


Z Index: 225

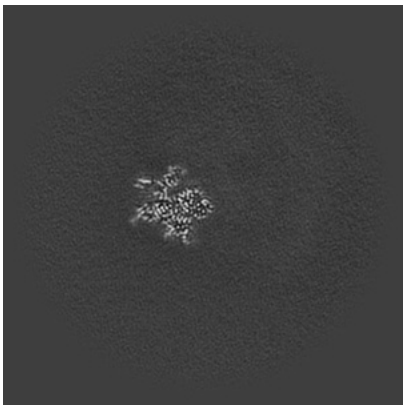
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

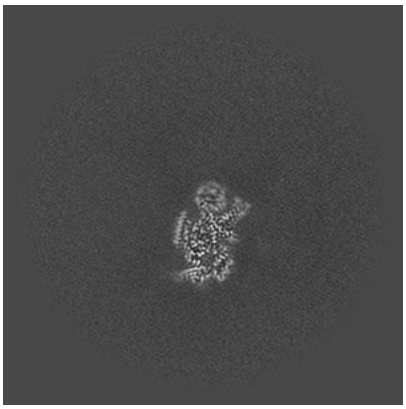
6.3.1 Primary map



X Index: 217

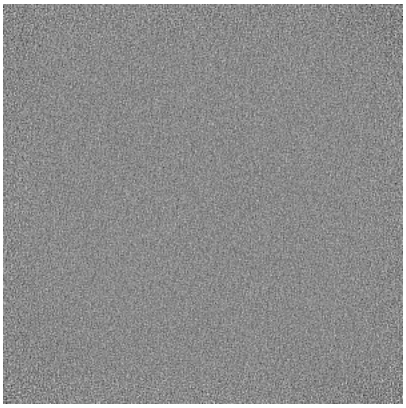


Y Index: 170

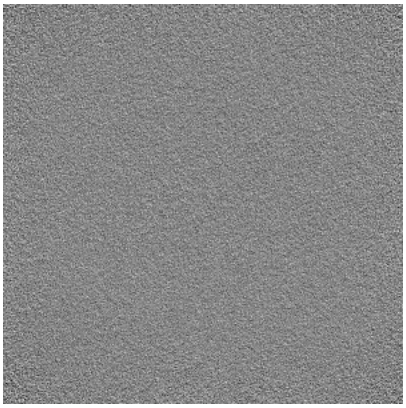


Z Index: 173

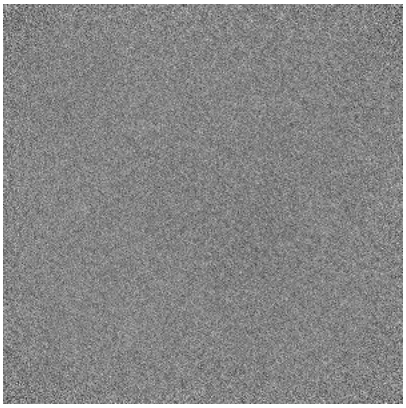
6.3.2 Raw map



X Index: 0



Y Index: 0

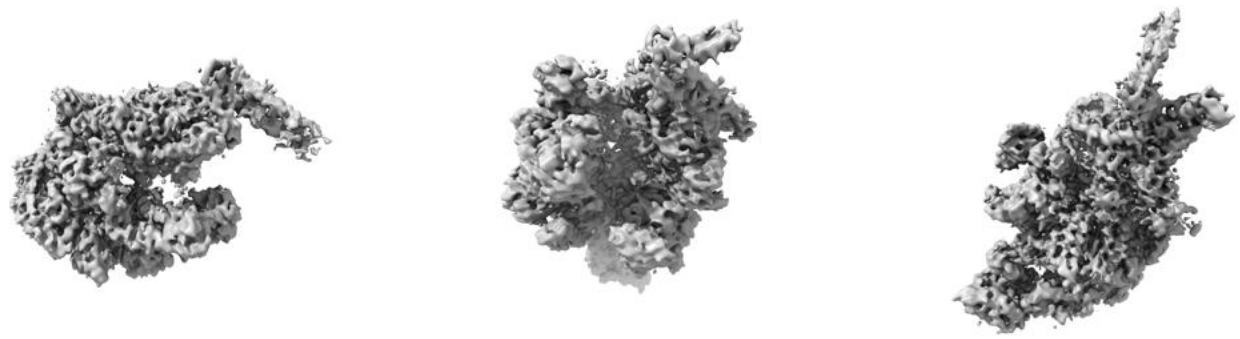


Z Index: 0

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 0.3. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

6.4.2 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.

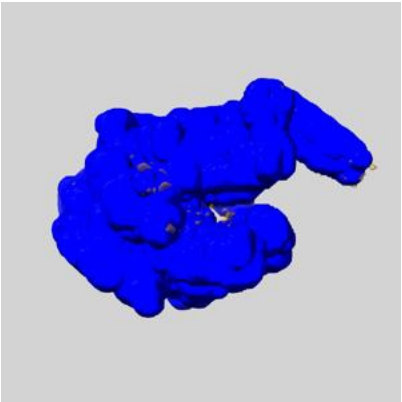
6.5 Mask visualisation (i)

This section shows the 3D surface view of the primary map at 50% transparency overlaid with the specified mask at 0% transparency

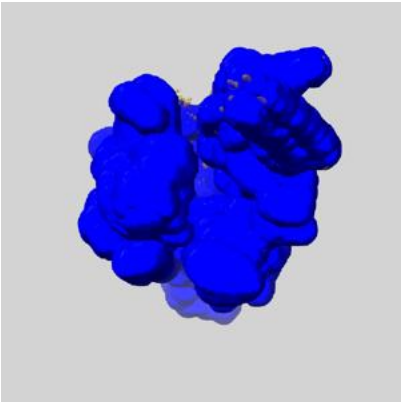
A mask typically either:

- Encompasses the whole structure
- Separates out a domain, a functional unit, a monomer or an area of interest from a larger structure

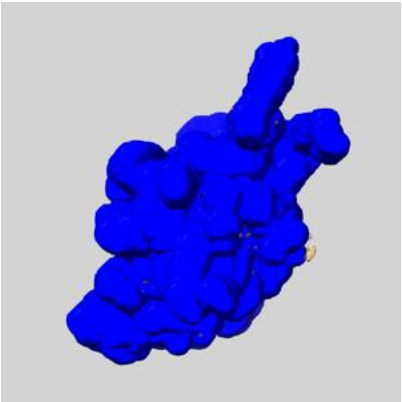
6.5.1 emd_13579_msk_1.map (i)



X



Y

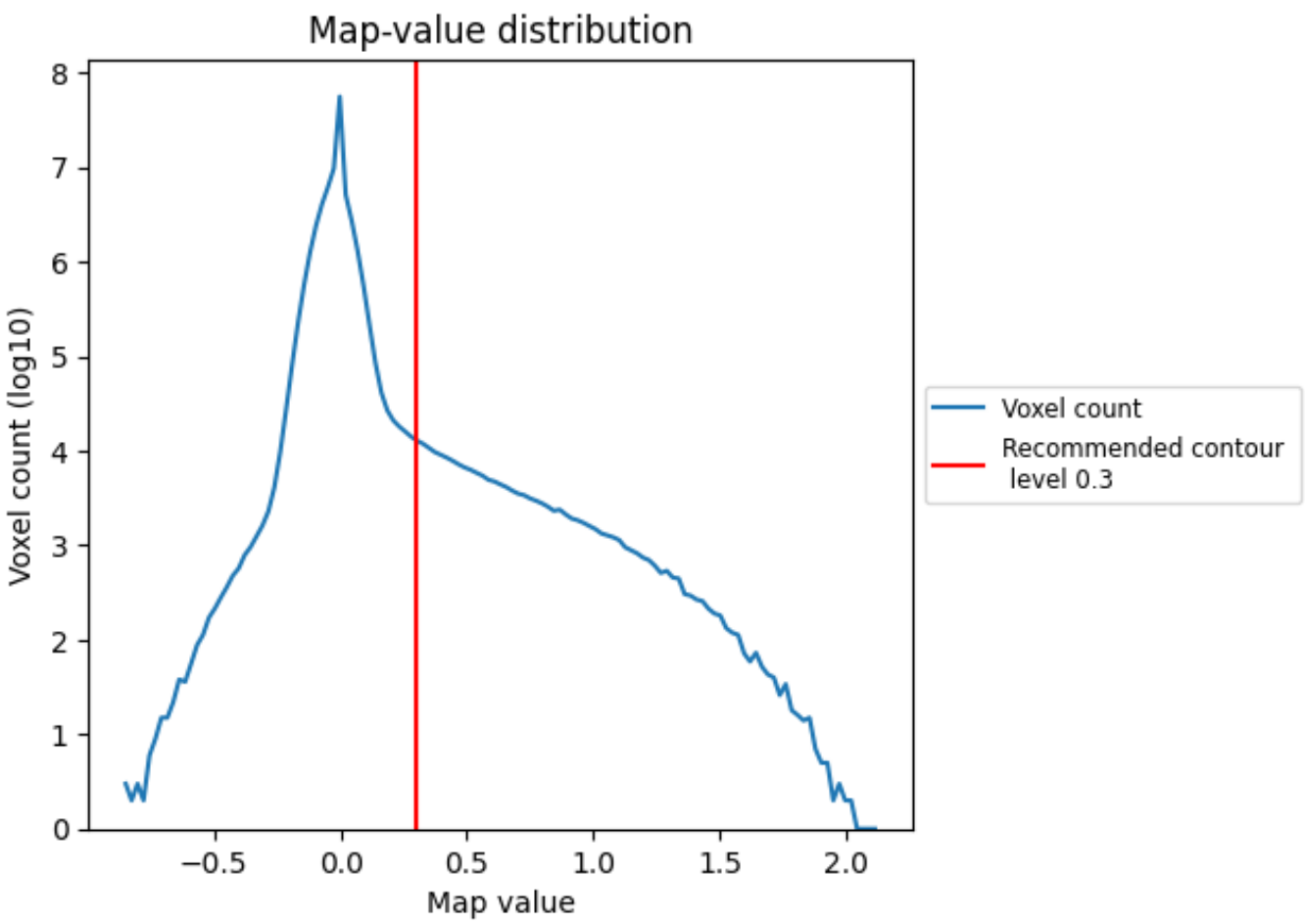


Z

7 Map analysis (i)

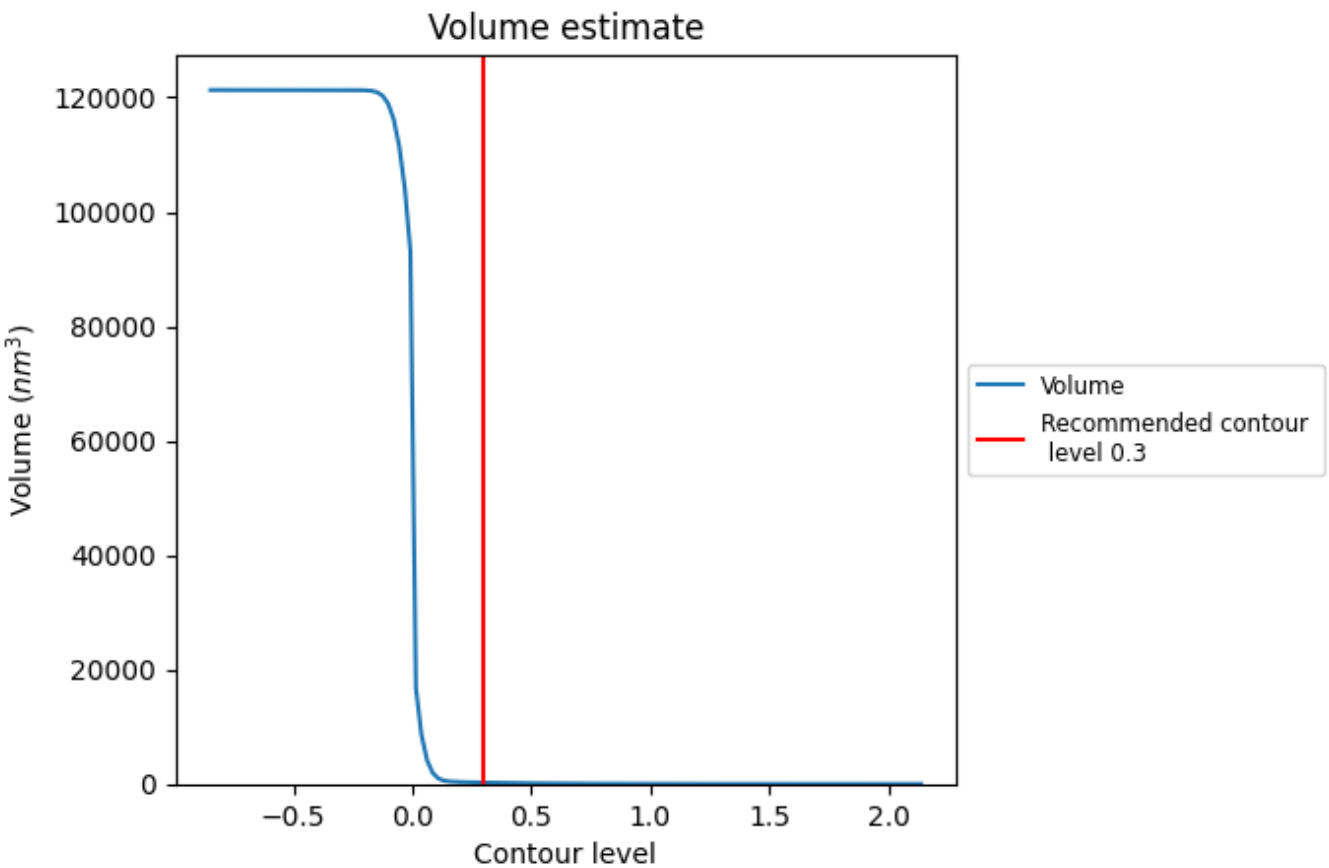
This section contains the results of statistical analysis of the map.

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.

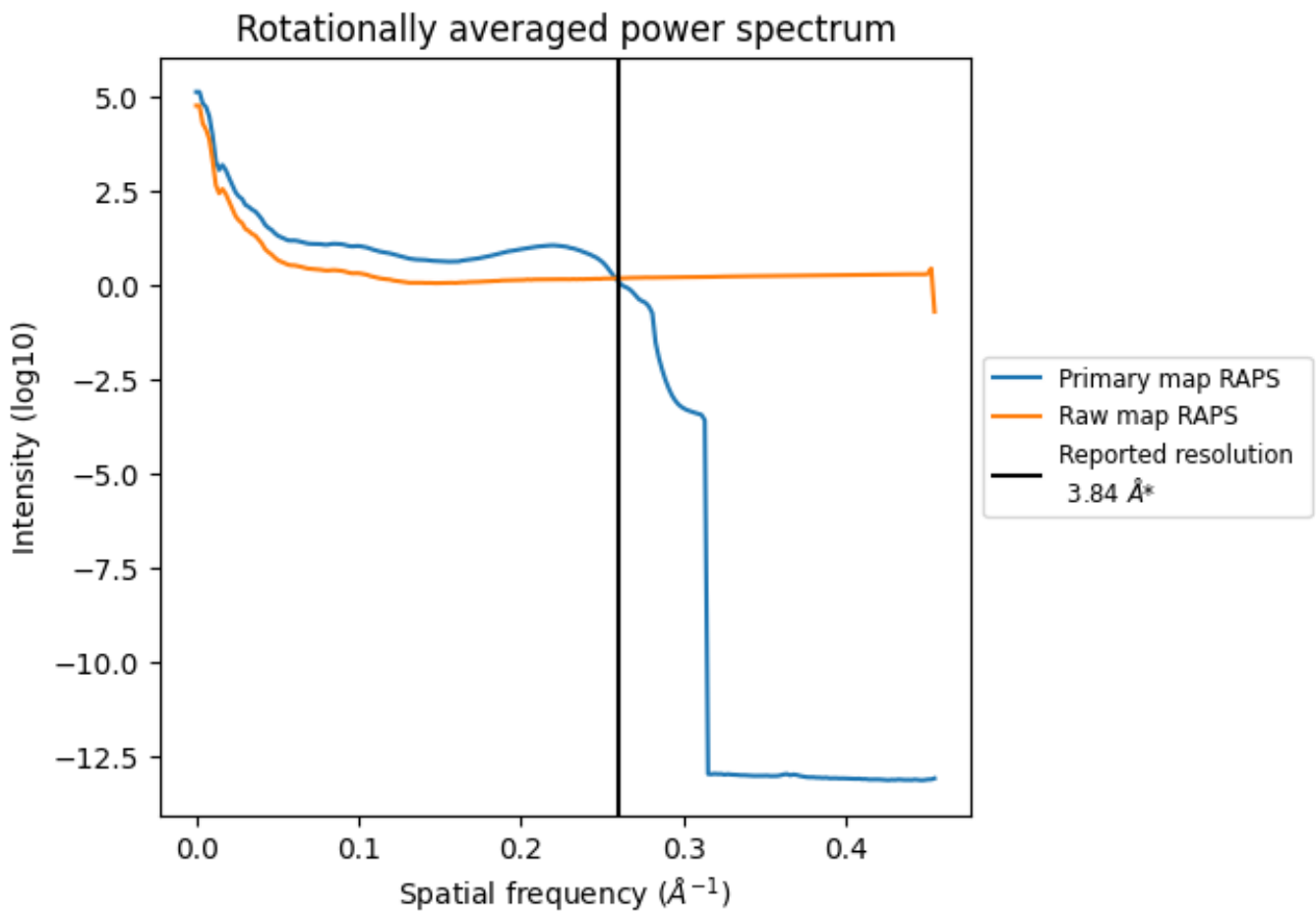
7.2 Volume estimate (i)



The volume at the recommended contour level is 232 nm^3; this corresponds to an approximate mass of 209 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)

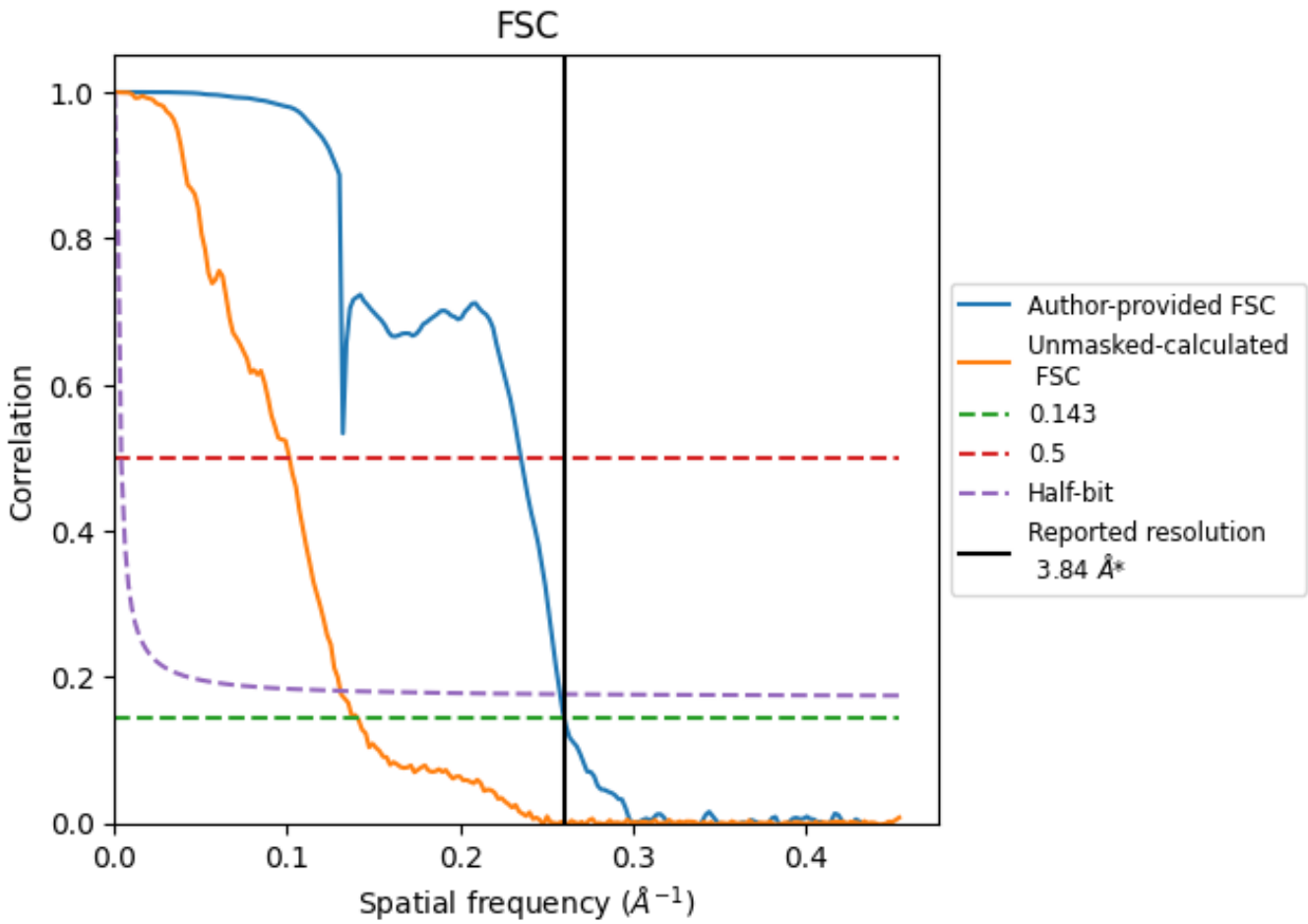


*Reported resolution corresponds to spatial frequency of 0.260 Å^{-1}

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 0.260 Å^{-1}

8.2 Resolution estimates (i)

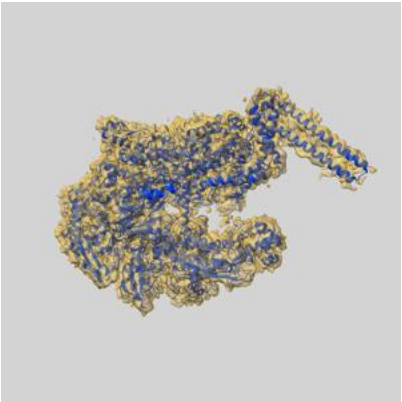
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.84	-	-
Author-provided FSC curve	3.84	4.25	3.88
Unmasked-calculated*	7.10	9.87	7.64

*Resolution estimate based on FSC curve calculated by comparison of deposited half-maps. The value from deposited half-maps intersecting FSC 0.143 CUT-OFF 7.10 differs from the reported value 3.84 by more than 10 %

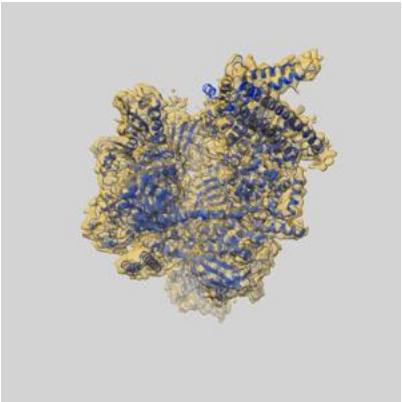
9 Map-model fit (i)

This section contains information regarding the fit between EMDB map EMD-13579 and PDB model 7PP4. Per-residue inclusion information can be found in section 3 on page 6.

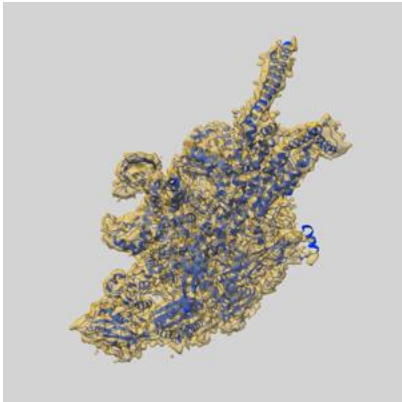
9.1 Map-model overlay (i)



X



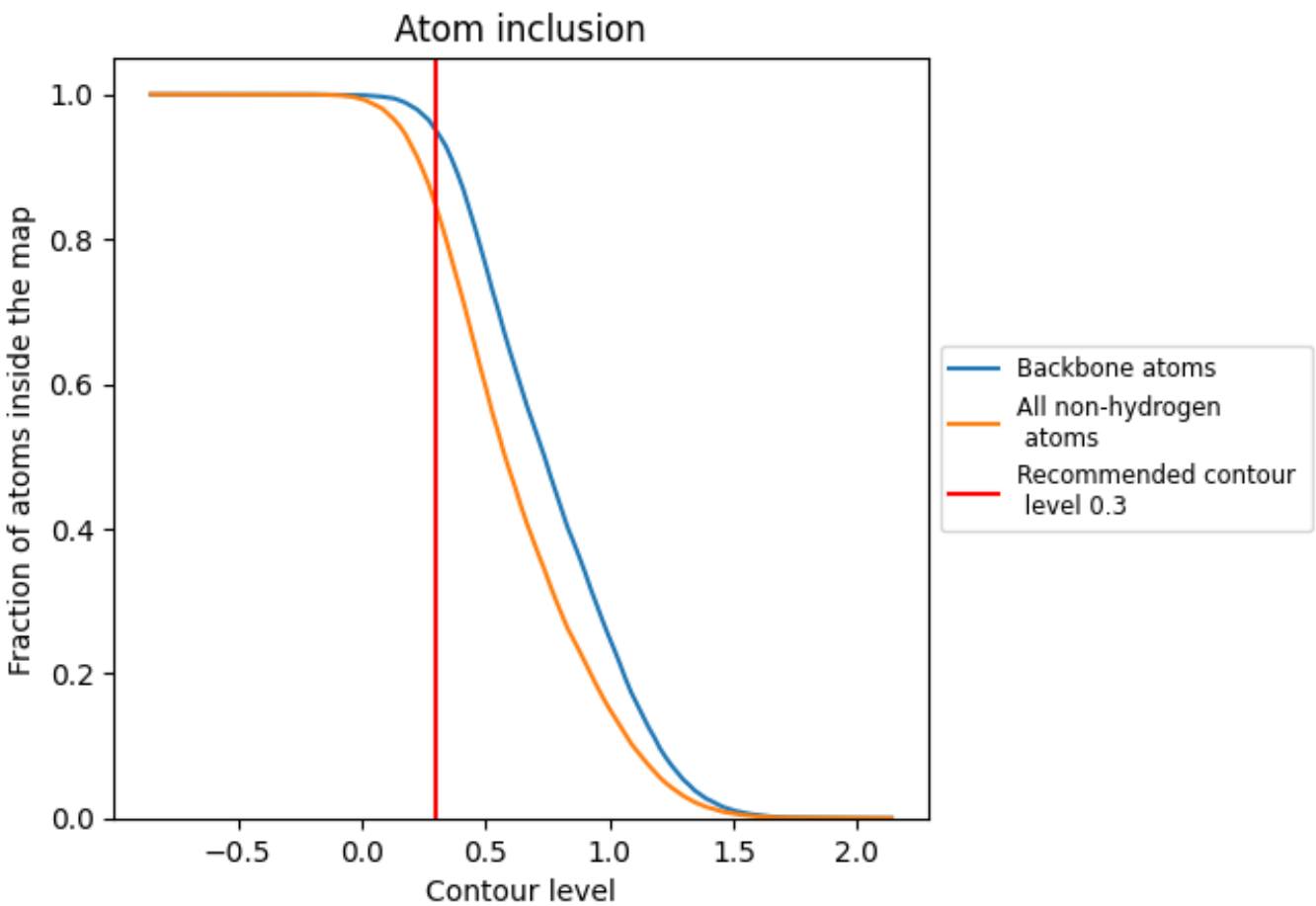
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.3 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 Atom inclusion (i)



At the recommended contour level, 95% of all backbone atoms, 84% of all non-hydrogen atoms, are inside the map.