



wwPDB EM Validation Summary Report ⓘ

Sep 10, 2023 – 12:53 AM JST

PDB ID : 8HI1
EMDB ID : EMD-34809
Title : Streptococcus thermophilus Cas1-Cas2- prespacer ternary complex
Authors : Chen, Q.; Luo, Y.
Deposited on : 2022-11-18
Resolution : 3.09 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.dev50
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) : 2.35

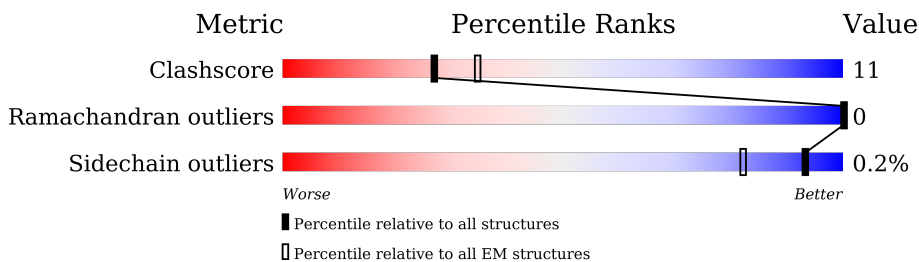
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 3.09 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	318	
1	B	318	
1	C	318	
1	D	318	
2	E	122	
2	F	122	
3	G	38	
4	H	38	

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 20998 atoms, of which 10143 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 CRISPR-associated endonuclease Cas1.

Mol	Chain	Residues	Atoms					AltConf	Trace	
			Total	C	H	N	O			S
1	A	289	Total	C	H	N	O	S	0	0
			4568	1447	2284	402	428	7		
1	B	232	Total	C	H	N	O	S	0	0
			3576	1147	1778	321	324	6		
1	C	285	Total	C	H	N	O	S	0	0
			4379	1424	2131	397	420	7		
1	D	244	Total	C	H	N	O	S	0	0
			3731	1204	1834	347	340	6		

- Molecule 2 is a protein called Type I-E CRISPR-associated endoribonuclease Cas2.

Mol	Chain	Residues	Atoms					AltConf	Trace	
			Total	C	H	N	O			S
2	E	92	Total	C	H	N	O	S	0	0
			1475	476	736	124	137	2		
2	F	92	Total	C	H	N	O	S	0	0
			1483	478	742	124	136	3		

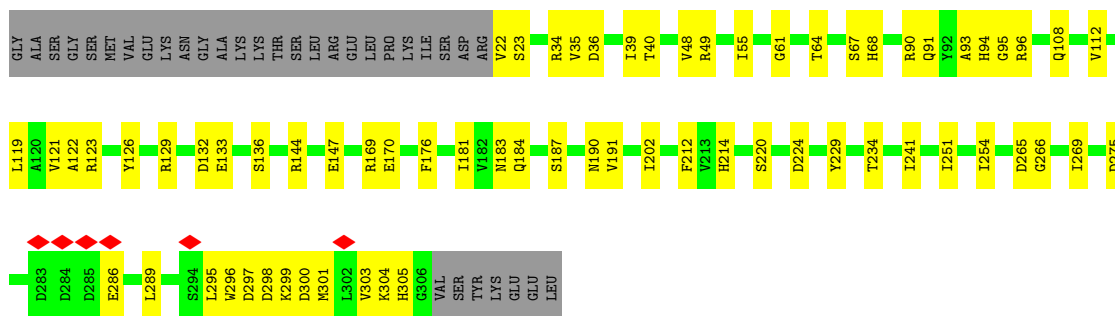
- Molecule 3 is a DNA chain called DNA (26-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace	
			Total	C	H	N	O			P
3	G	26	Total	C	H	N	O	P	0	0
			826	256	290	110	145	25		

- Molecule 4 is a DNA chain called DNA (31-MER).

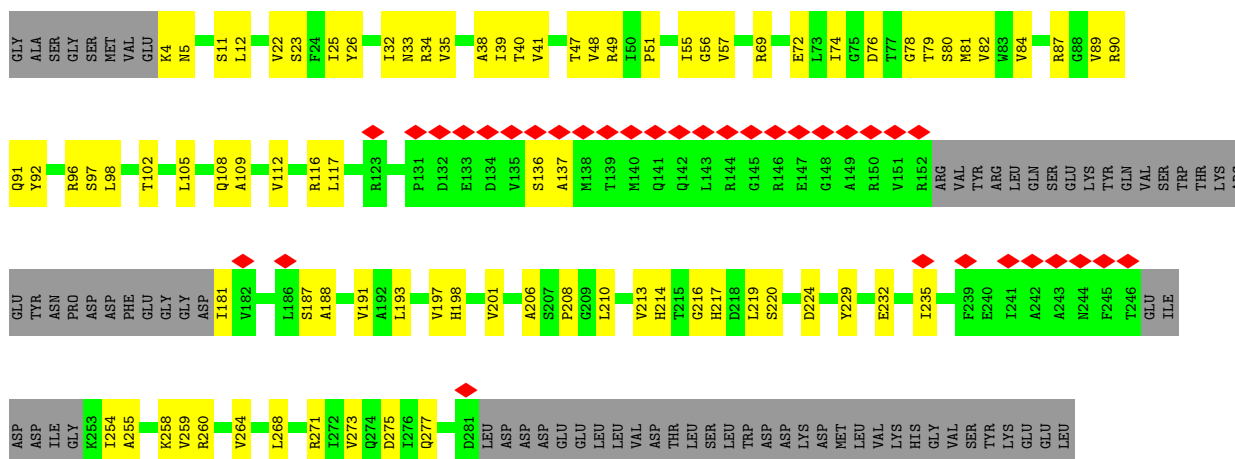
Mol	Chain	Residues	Atoms					AltConf	Trace	
			Total	C	H	N	O			P
4	H	31	Total	C	H	N	O	P	0	0
			960	296	348	91	195	30		

Chain C:  68% 21% 10%



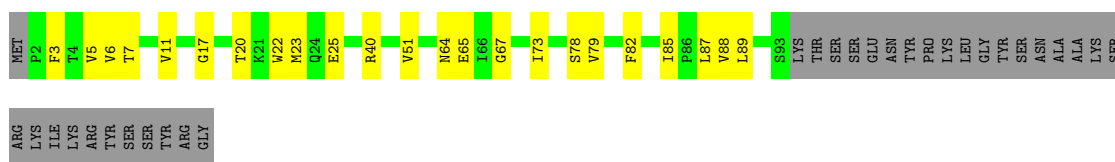
• Molecule 1: CRISPR-associated endonuclease Cas1

Chain D:  11% 51% 26% 23%



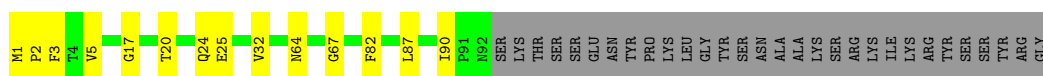
• Molecule 2: Type I-E CRISPR-associated endonuclease Cas2

Chain E:  57% 19% 25%



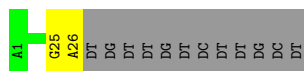
• Molecule 2: Type I-E CRISPR-associated endonuclease Cas2

Chain F:  64% 11% 25%



• Molecule 3: DNA (26-MER)

Chain G:  63% 5% 32%



- Molecule 4: DNA (31-MER)

Chain H: 47% 34% 18%

A horizontal bar chart showing the distribution of nucleotide types for Chain H. The bar is divided into three segments: green (47%), yellow (34%), and grey (18%).



4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	188628	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	49	Depositor
Minimum defocus (nm)	1213	Depositor
Maximum defocus (nm)	2465	Depositor
Magnification	Not provided	
Image detector	GATAN K2 QUANTUM (4k x 4k)	Depositor
Maximum map value	4.064	Depositor
Minimum map value	-3.215	Depositor
Average map value	0.001	Depositor
Map value standard deviation	0.102	Depositor
Recommended contour level	0.282	Depositor
Map size (\AA)	217.6, 217.6, 217.6	wwPDB
Map dimensions	256, 256, 256	wwPDB
Map angles ($^\circ$)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (\AA)	0.85, 0.85, 0.85	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	0.26	0/2324	0.45	0/3141
1	B	0.27	0/1826	0.46	0/2464
1	C	0.31	0/2287	0.48	0/3092
1	D	0.32	0/1925	0.48	0/2594
2	E	0.37	0/756	0.52	0/1026
2	F	0.37	0/758	0.52	0/1029
3	G	0.58	0/605	0.86	0/932
4	H	0.55	0/680	1.12	0/1050
All	All	0.34	0/11161	0.57	0/15328

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	2284	2284	2292	44	0
1	B	1798	1778	1859	41	0
1	C	2248	2131	2259	54	0
1	D	1897	1834	1964	66	0
2	E	739	736	736	19	0
2	F	741	742	742	12	0
3	G	536	290	292	2	0
4	H	612	348	348	12	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	10855	10143	10492	232	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 11.

The worst 5 of 232 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:D:76:ASP:O	1:D:96:ARG:NH2	2.08	0.86
1:B:229:TYR:OH	1:B:275:ASP:OD2	1.95	0.84
1:B:76:ASP:O	1:B:96:ARG:NH2	2.10	0.84
1:D:214:HIS:ND1	1:D:224:ASP:OD1	2.14	0.81
1:C:214:HIS:ND1	1:C:224:ASP:OD1	2.14	0.81

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	287/318 (90%)	258 (90%)	29 (10%)	0	100	100
1	B	226/318 (71%)	209 (92%)	17 (8%)	0	100	100
1	C	283/318 (89%)	251 (89%)	32 (11%)	0	100	100
1	D	238/318 (75%)	219 (92%)	19 (8%)	0	100	100
2	E	90/122 (74%)	82 (91%)	8 (9%)	0	100	100
2	F	90/122 (74%)	81 (90%)	9 (10%)	0	100	100
All	All	1214/1516 (80%)	1100 (91%)	114 (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	244/268 (91%)	243 (100%)	1 (0%)	91	96
1	B	192/268 (72%)	192 (100%)	0	100	100
1	C	240/268 (90%)	239 (100%)	1 (0%)	91	96
1	D	201/268 (75%)	201 (100%)	0	100	100
2	E	82/108 (76%)	82 (100%)	0	100	100
2	F	82/108 (76%)	82 (100%)	0	100	100
All	All	1041/1288 (81%)	1039 (100%)	2 (0%)	93	97

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	118	ARG
1	C	144	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	68	HIS
1	D	91	GLN
2	F	24	GLN
1	D	100	HIS
1	D	277	GLN

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.

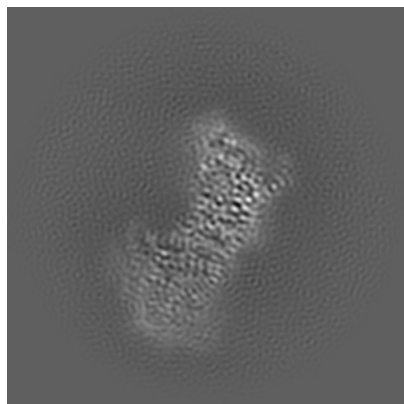
6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-34809. These allow visual inspection of the internal detail of the map and identification of artifacts.

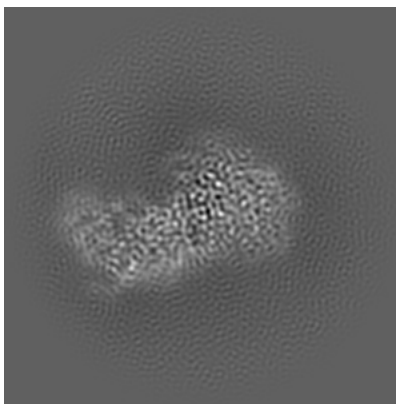
Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

6.1 Orthogonal projections [i](#)

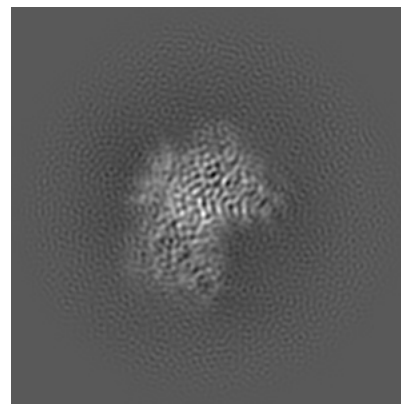
6.1.1 Primary map



X

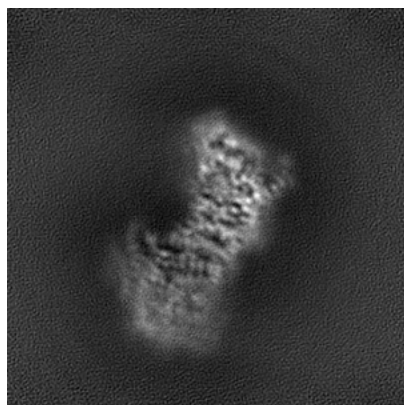


Y

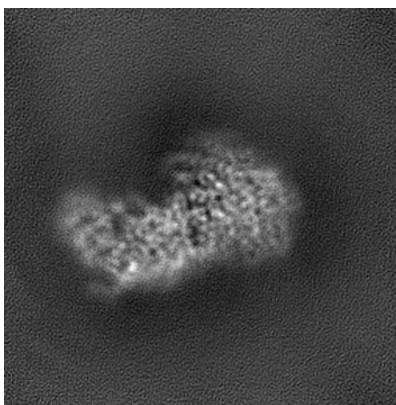


Z

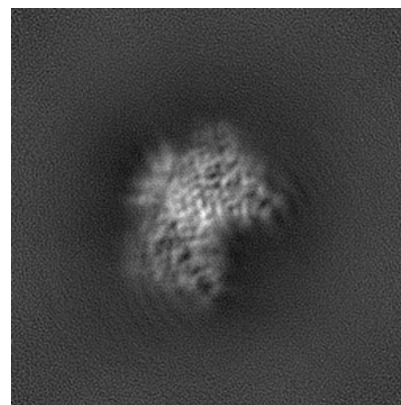
6.1.2 Raw map



X



Y

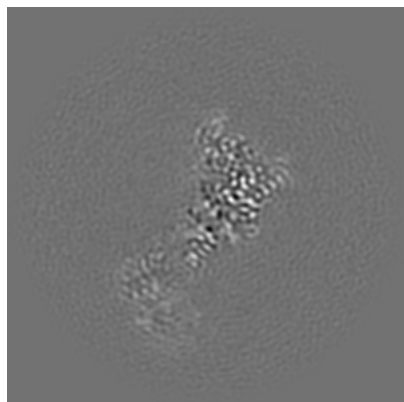


Z

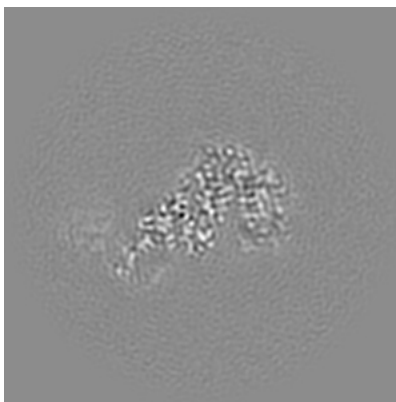
The images above show the map projected in three orthogonal directions.

6.2 Central slices [i](#)

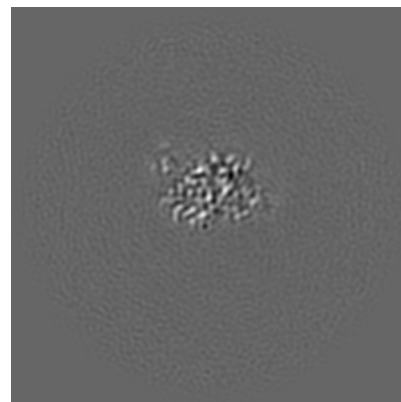
6.2.1 Primary map



X Index: 128

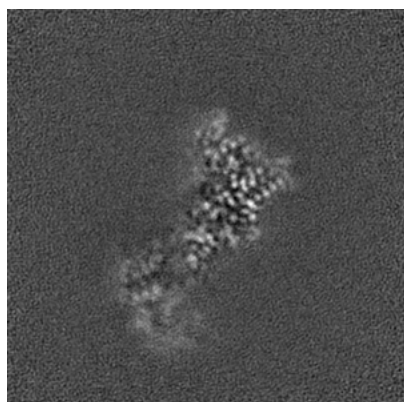


Y Index: 128

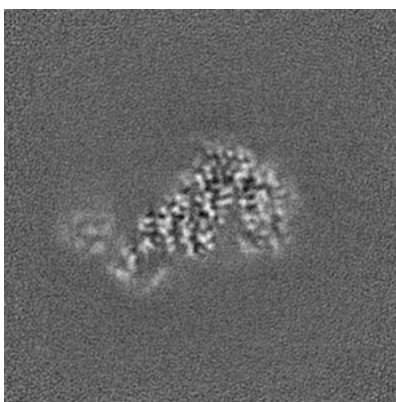


Z Index: 128

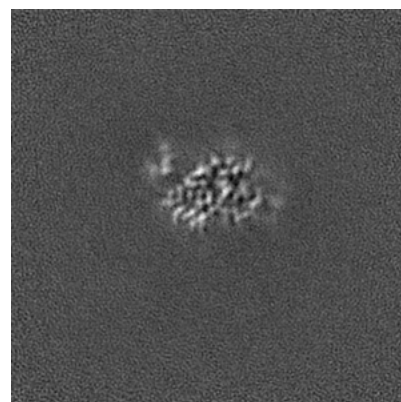
6.2.2 Raw map



X Index: 128



Y Index: 128

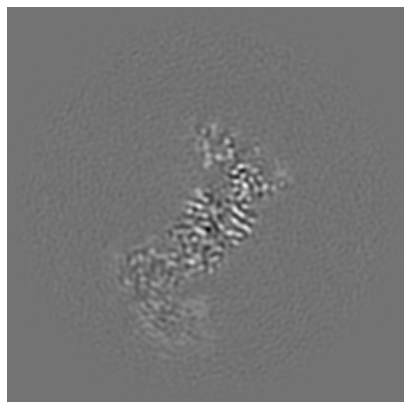


Z Index: 128

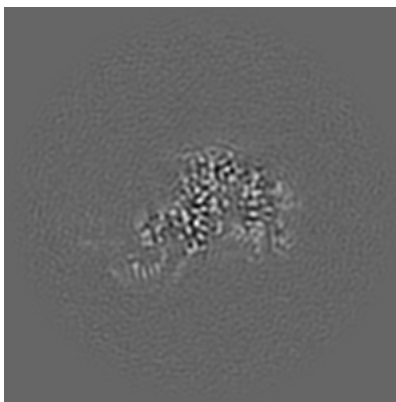
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices [i](#)

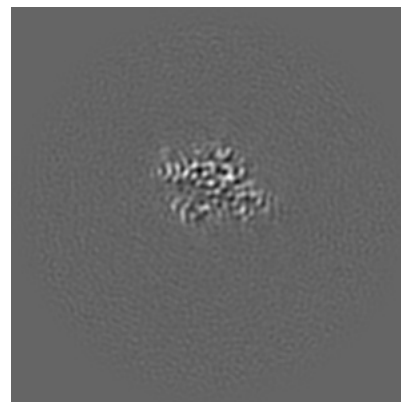
6.3.1 Primary map



X Index: 122

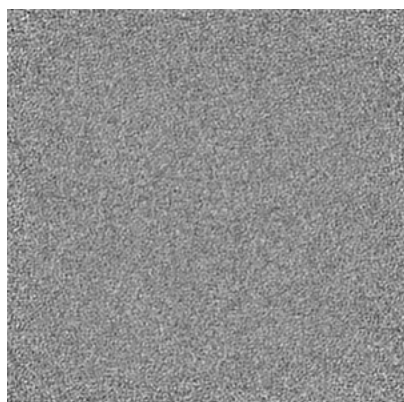


Y Index: 136

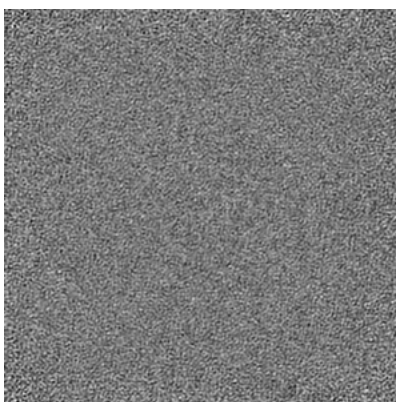


Z Index: 132

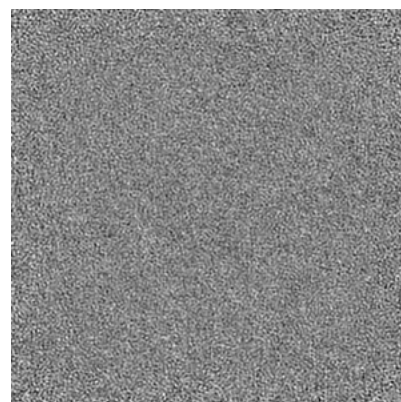
6.3.2 Raw map



X Index: 0



Y Index: 0

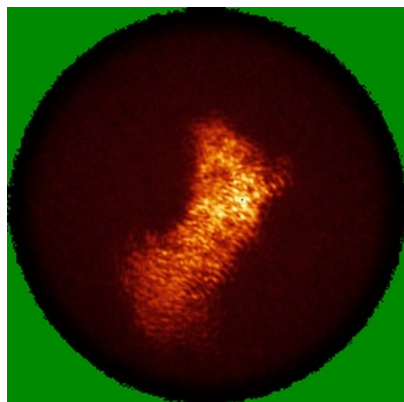


Z Index: 0

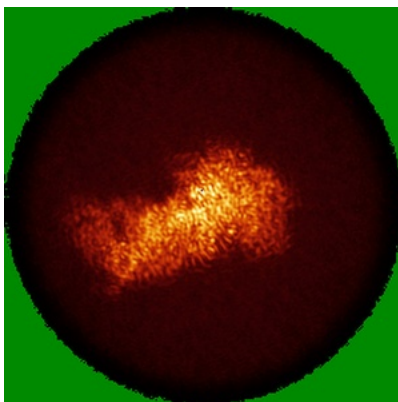
The images above show the largest variance slices of the map in three orthogonal directions.

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

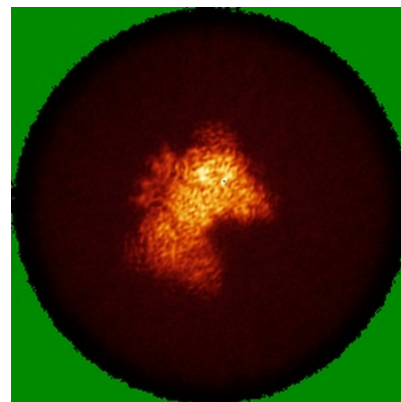
6.4.1 Primary map



X

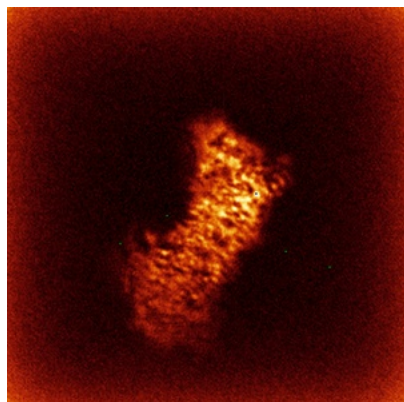


Y

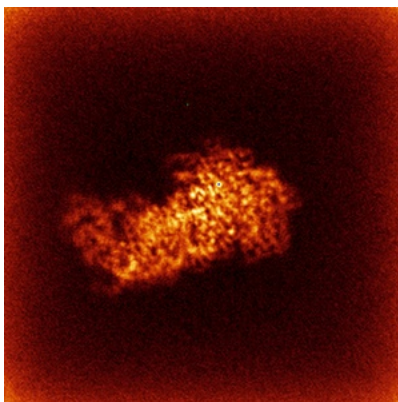


Z

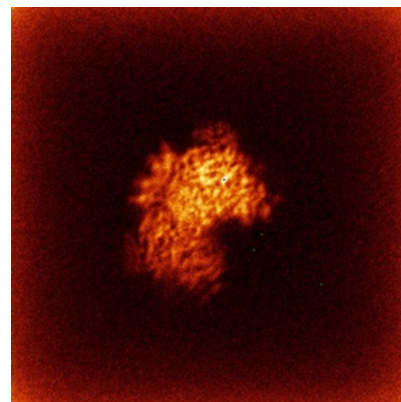
6.4.2 Raw map



X



Y

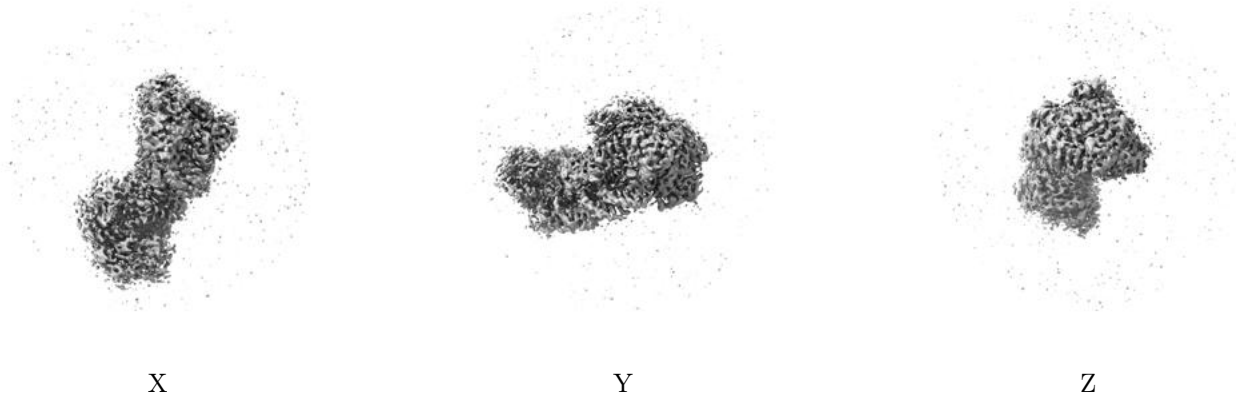


Z

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.

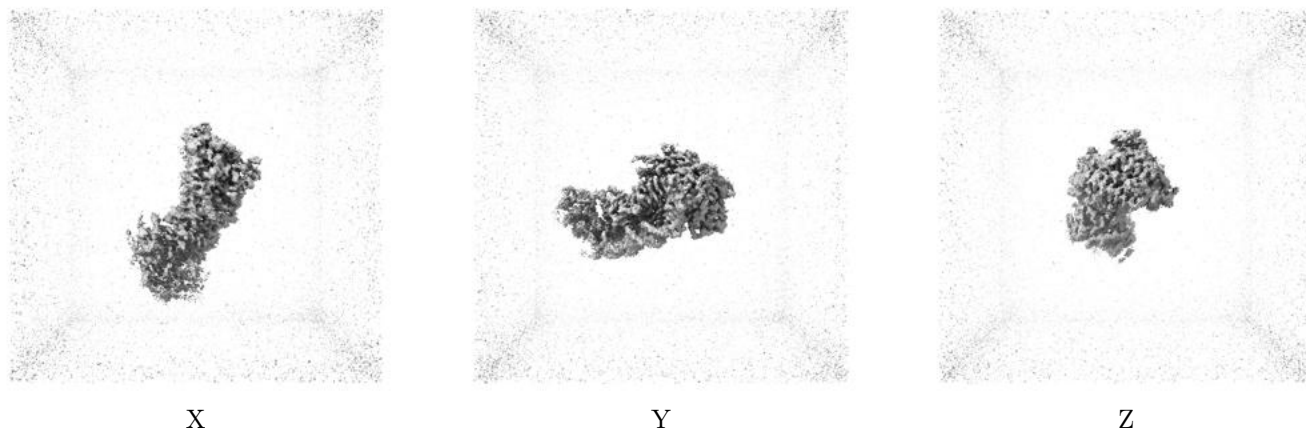
6.5 Orthogonal surface views [i](#)

6.5.1 Primary map



The images above show the 3D surface view of the map at the recommended contour level 0.282. 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



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.

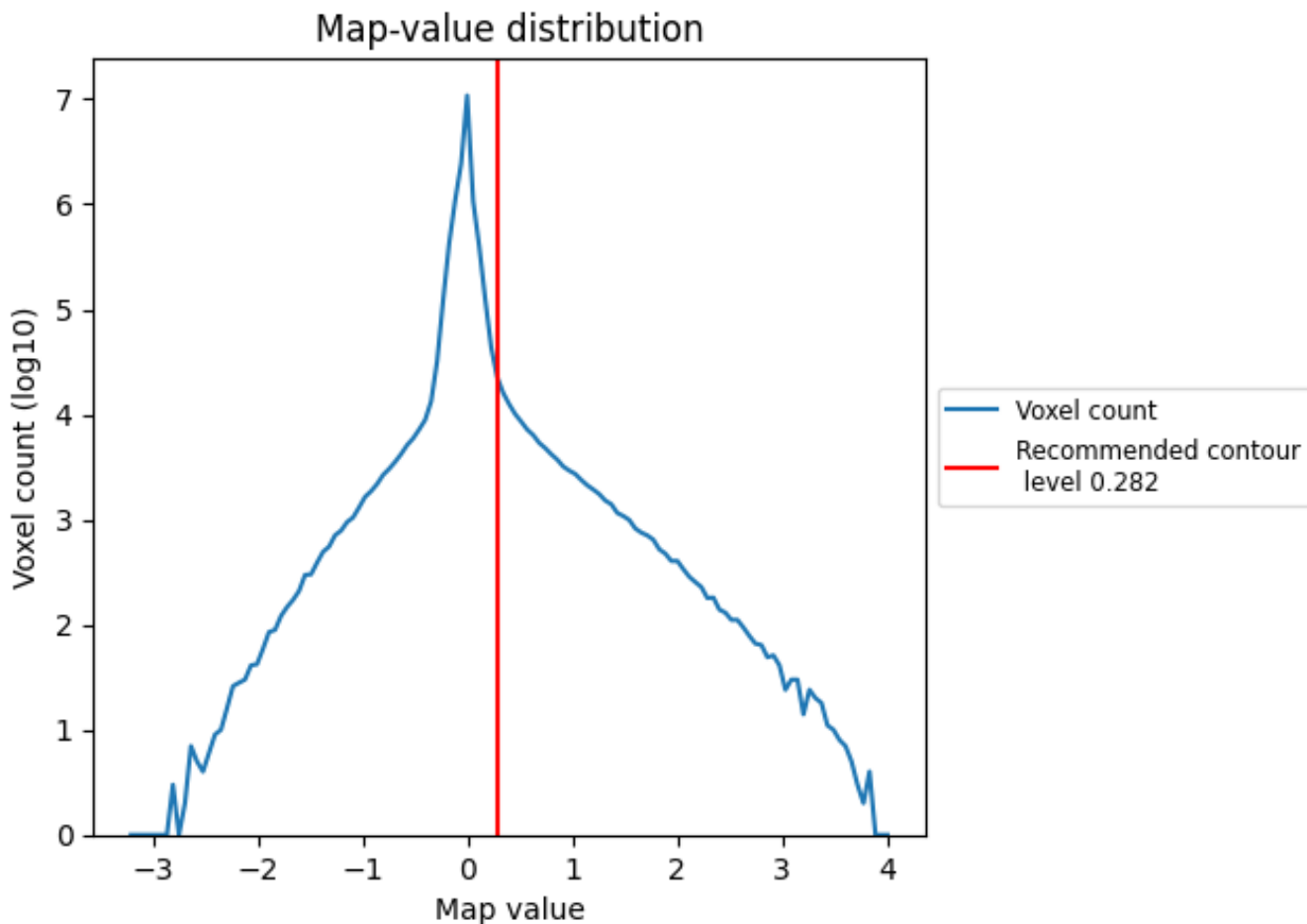
6.6 Mask visualisation [i](#)

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

7 Map analysis [i](#)

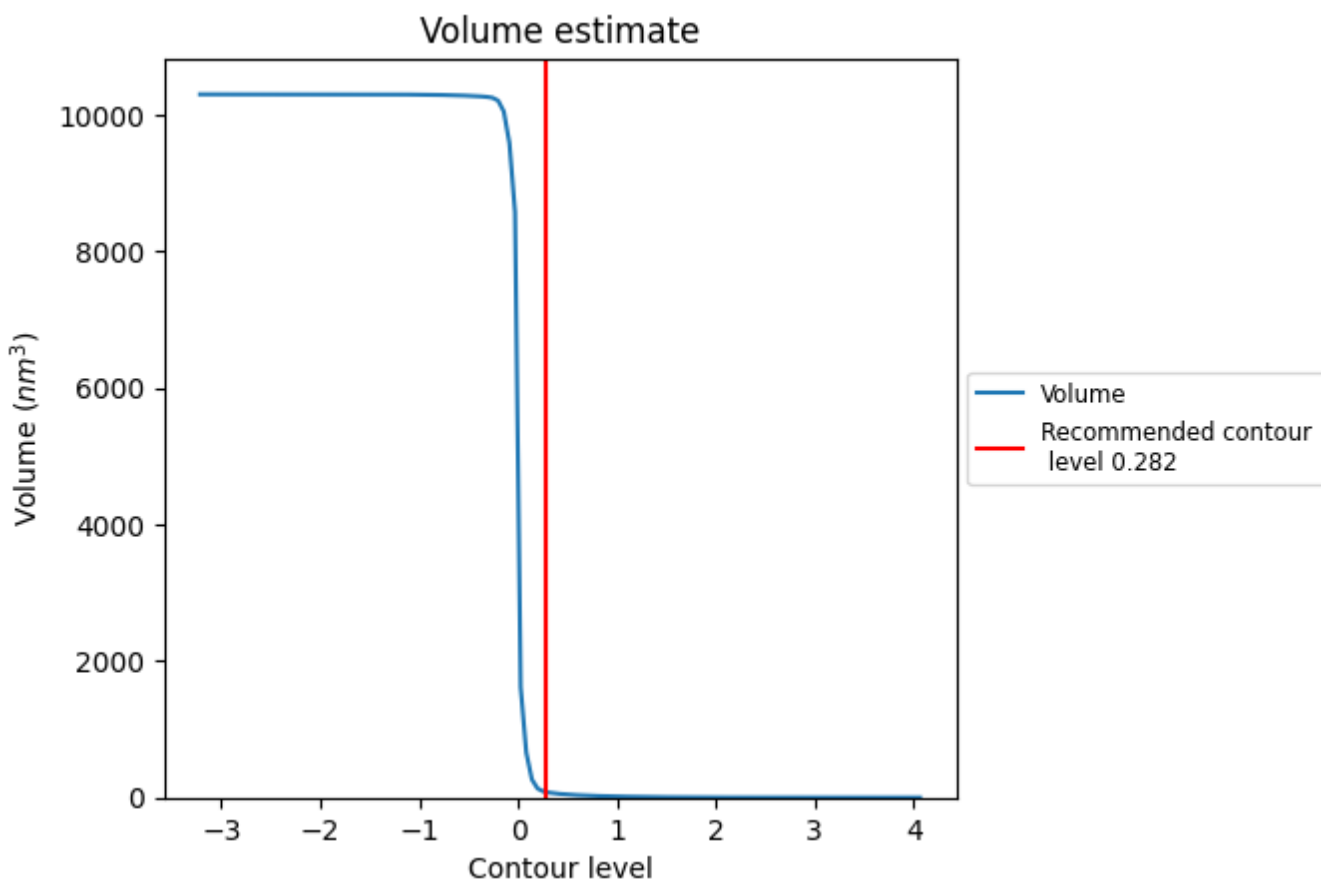
This section contains the results of statistical analysis of the map.

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.

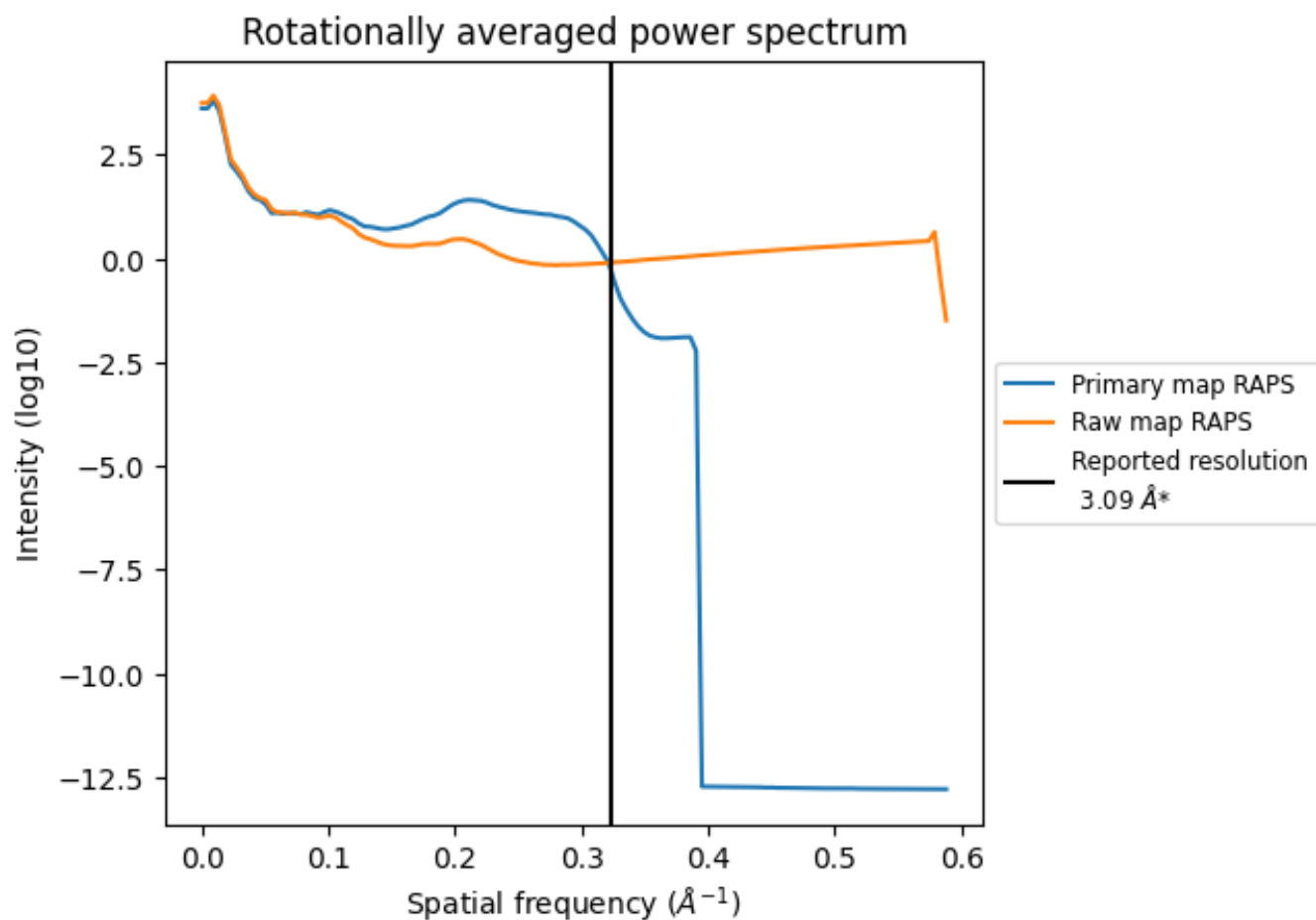
7.2 Volume estimate [i](#)



The volume at the recommended contour level is 82 nm³; this corresponds to an approximate mass of 74 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](#)

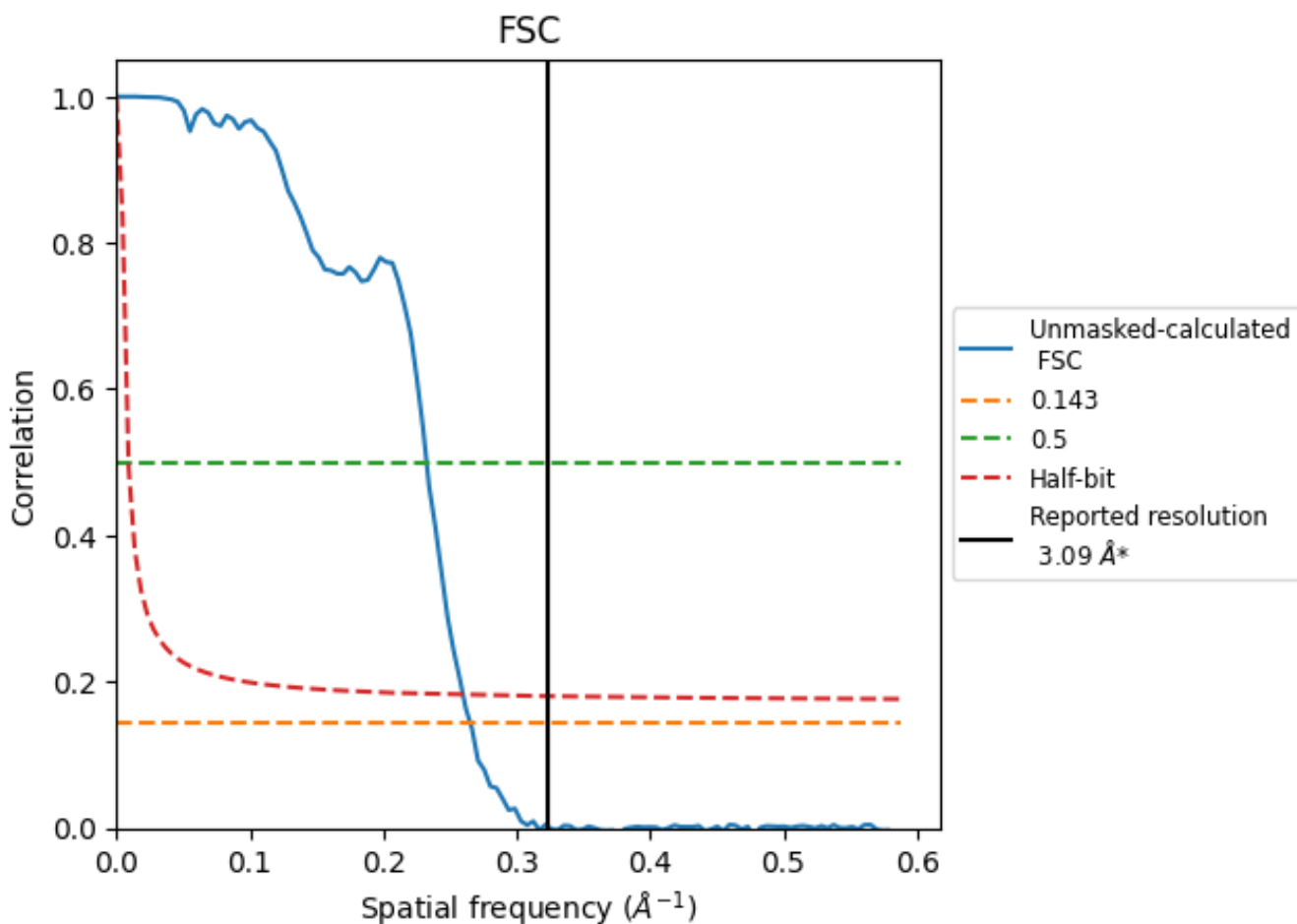


*Reported resolution corresponds to spatial frequency of 0.324 Å⁻¹

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 0.324 Å⁻¹

8.2 Resolution estimates [i](#)

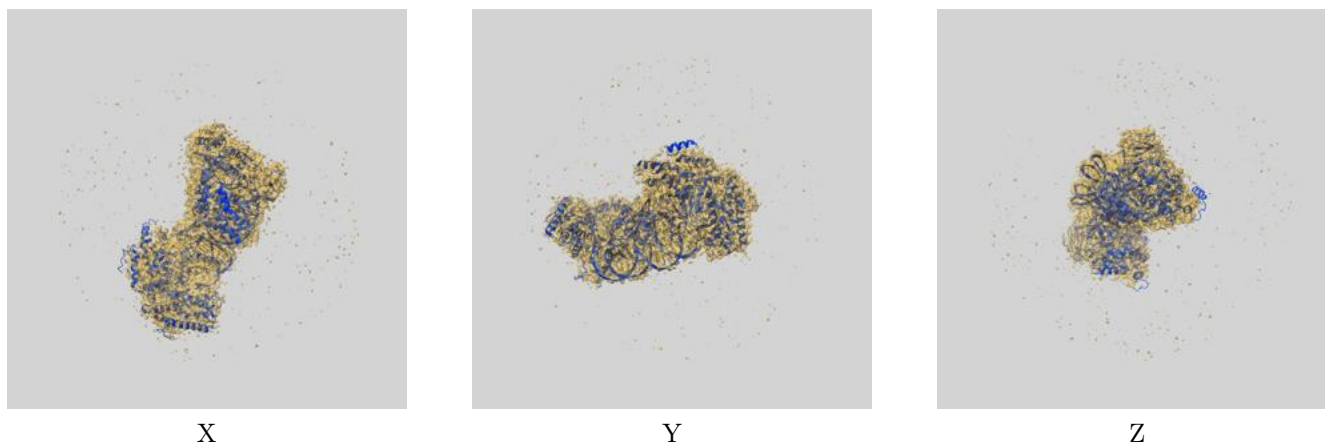
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.09	-	-
Author-provided FSC curve	-	-	-
Unmasked-calculated*	3.77	4.30	3.85

*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 3.77 differs from the reported value 3.09 by more than 10 %

9 Map-model fit [i](#)

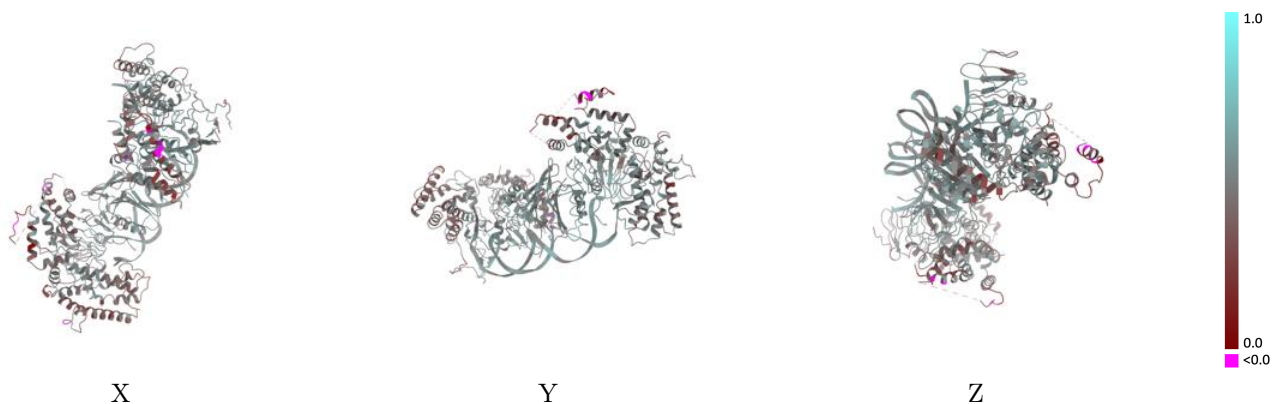
This section contains information regarding the fit between EMDB map EMD-34809 and PDB model 8HI1. Per-residue inclusion information can be found in section 3 on page 4.

9.1 Map-model overlay [i](#)



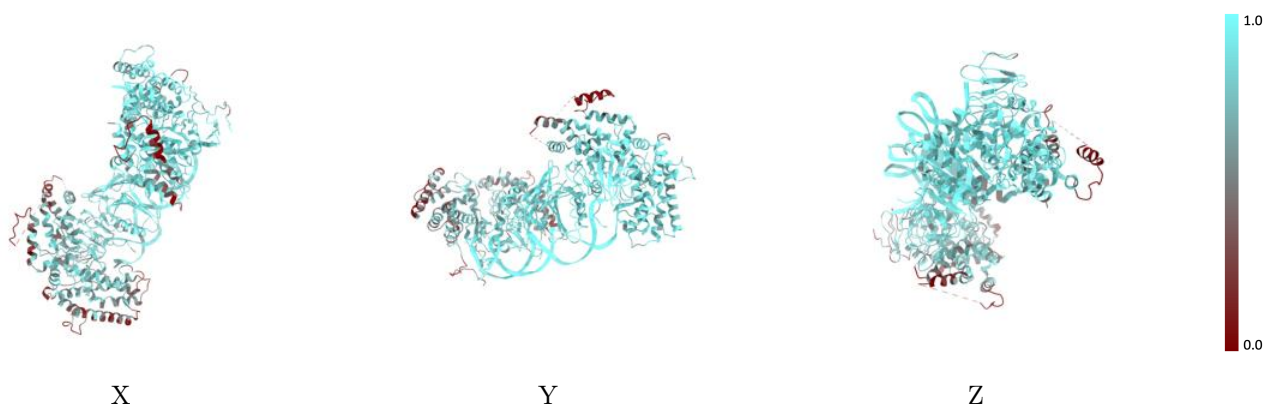
The images above show the 3D surface view of the map at the recommended contour level 0.282 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

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



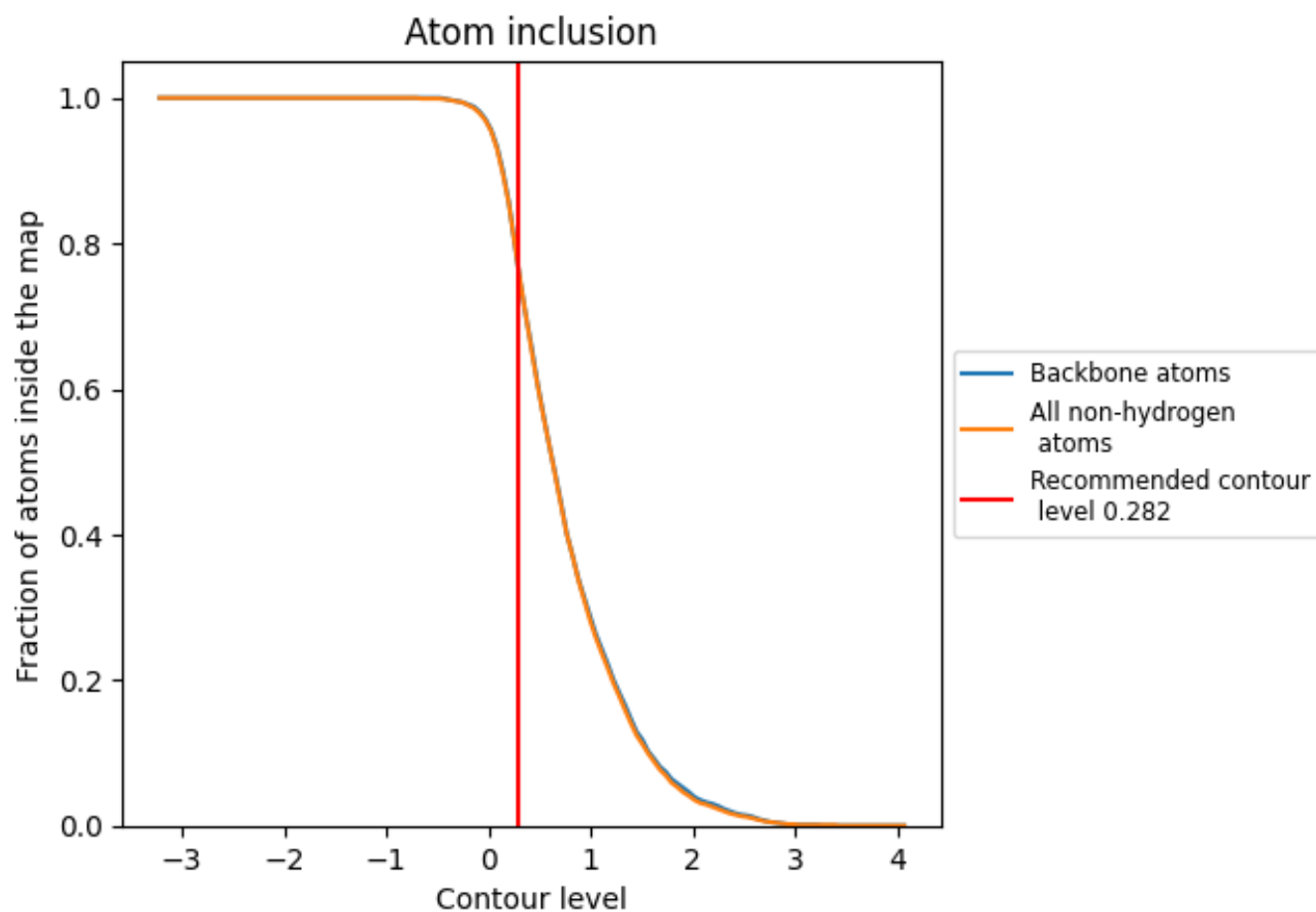
The images above show the model with each residue coloured according to its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions 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 (0.282).



















9.4 Atom inclusion [i](#)



At the recommended contour level, 78% 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 (0.282) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	 0.7740	 0.4730
A	 0.6250	 0.4330
B	 0.6770	 0.4400
C	 0.8400	 0.4880
D	 0.7680	 0.4640
E	 0.9110	 0.5250
F	 0.9020	 0.5200
G	 0.8820	 0.5270
H	 0.9310	 0.5320

