



wwPDB EM Validation Summary Report ⓘ

Nov 7, 2022 – 06:29 AM EST

PDB ID : 6C23
EMDB ID : EMD-7334
Title : Cryo-EM structure of PRC2 bound to cofactors AEBP2 and JARID2 in the Compact Active State
Authors : Kasinath, V.; Faini, M.; Poepsel, S.; Reif, D.; Feng, A.; Stjepanovic, G.; Aebersold, R.; Nogales, E.
Deposited on : 2018-01-05
Resolution : 3.90 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.dev43
Mogul : 1.8.5 (274361), CSD as541be (2020)
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

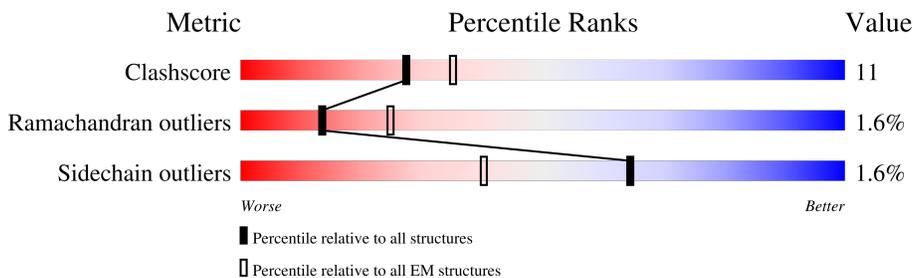
1 Overall quality at a glance

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.



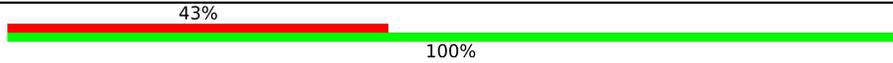
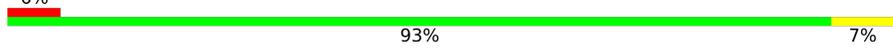
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

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	739	14% 14% 7% 61% 83%
1	M	739	14% 14% 7% 61% 83%
1	Q	739	7% 7% 7% 77% 91%
2	E	348	6% 6% 6% 82% 92%
3	C	746	22% 22% 22% 34% 74%
3	K	746	36% 36% 8% 20% 56%
4	L	441	61% 20% 18% 1% 18%
5	N	425	71% 19% 9% 1% 9%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	O	7	 43% 100%
7	P	295	 19% 78% 6%
8	Z	135	 6% 93% 7%
9	B	345	 6% 96% 6%

2 Entry composition [i](#)

There are 9 unique types of molecules in this entry. The entry contains 13440 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 Polycomb protein SUZ12.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	123	Total	C	N	O	S	0	0
			860	549	157	148	6		
1	M	122	Total	C	N	O	S	0	0
			962	609	169	175	9		
1	Q	66	Total	C	N	O	S	0	0
			496	315	94	85	2		

- Molecule 2 is a protein called Protein Jumonji.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	E	27	Total	C	N	O	S	0	0
			181	111	32	37	1		

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	103	SER	-	expression tag	UNP Q92833
E	104	ASN	-	expression tag	UNP Q92833
E	105	ALA	-	expression tag	UNP Q92833

- Molecule 3 is a protein called Histone-lysine N-methyltransferase EZH2.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	K	325	Total	C	N	O	S	0	0
			2406	1514	426	436	30		
3	C	193	Total	C	N	O	S	0	0
			1307	820	230	253	4		

- Molecule 4 is a protein called Polycomb protein EED.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	L	362	Total	C	N	O	S	0	0
			2877	1826	505	526	20		

- Molecule 5 is a protein called Histone-binding protein RBBP4.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
5	N	386	3058	1931	524	593	10	0	0

- Molecule 6 is a protein called JARID2-substrate.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
6	O	7	51	33	11	7	0	0

- Molecule 7 is a protein called Zinc finger protein AEBP2.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
7	P	64	484	304	91	87	2	0	0

- Molecule 8 is a protein called SUZ12.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
8	Z	135	674	404	135	135	0	0

- Molecule 9 is a protein called Protein Jumonji.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
9	B	13	84	55	15	14	0	0

4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	145592	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING ONLY	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	60	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	0.121	Depositor
Minimum map value	-0.061	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.004	Depositor
Recommended contour level	0.0119	Depositor
Map size (\AA)	241.92, 241.92, 241.92	wwPDB
Map dimensions	288, 288, 288	wwPDB
Map angles ($^\circ$)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (\AA)	0.84, 0.84, 0.84	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: M3L

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	0.58	0/882	0.82	1/1208 (0.1%)
1	M	0.58	0/983	0.75	0/1328
1	Q	0.61	0/506	0.86	0/687
2	E	0.40	0/182	0.75	0/246
3	C	0.49	0/1328	0.66	0/1815
3	K	0.49	0/2466	0.66	2/3352 (0.1%)
4	L	0.84	1/2952 (0.0%)	0.92	7/4008 (0.2%)
5	N	0.58	1/3142 (0.0%)	0.73	0/4286
6	O	0.54	0/51	0.74	0/66
7	P	0.53	0/492	0.81	0/666
9	B	0.44	0/73	0.43	0/100
All	All	0.62	2/13057 (0.0%)	0.77	10/17762 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	3
1	M	0	1
1	Q	0	1
3	C	0	1
All	All	0	6

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	L	234	GLY	CA-C	-22.83	1.15	1.51
5	N	24	TRP	CB-CG	-6.19	1.39	1.50

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	L	234	GLY	CA-C-O	-12.45	98.19	120.60
4	L	234	GLY	CA-C-N	9.64	138.40	117.20
4	L	234	GLY	N-CA-C	8.16	133.51	113.10
4	L	302	ARG	NE-CZ-NH2	-7.22	116.69	120.30
3	K	685	ARG	NE-CZ-NH2	5.79	123.20	120.30

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	441	GLN	Peptide
1	A	478	TYR	Peptide
1	A	480	TYR	Peptide
1	M	579	MET	Peptide
1	Q	129	ARG	Peptide

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	860	0	729	20	0
1	M	962	0	898	13	0
1	Q	496	0	450	11	0
2	E	181	0	170	3	0
3	C	1307	0	1048	23	0
3	K	2406	0	2122	65	0
4	L	2877	0	2749	68	0
5	N	3058	0	2904	81	0
6	O	51	0	54	0	0
7	P	484	0	476	8	0
8	Z	674	0	167	5	0
9	B	84	0	65	3	0
All	All	13440	0	11832	274	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 274 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)
5:N:152:PHE:CD1	5:N:170:ASP:OD2	1.71	1.43
3:K:443:LEU:HB3	3:K:447:TYR:CE2	1.61	1.36
3:K:443:LEU:CB	3:K:447:TYR:CE2	2.09	1.34
3:K:443:LEU:CB	3:K:447:TYR:HE2	1.41	1.32
3:K:662:MET:SD	3:K:732:ASP:OD1	1.92	1.28

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	121/739 (16%)	89 (74%)	28 (23%)	4 (3%)	4	31
1	M	120/739 (16%)	105 (88%)	13 (11%)	2 (2%)	9	43
1	Q	64/739 (9%)	48 (75%)	15 (23%)	1 (2%)	9	44
2	E	25/348 (7%)	15 (60%)	8 (32%)	2 (8%)	1	15
3	C	189/746 (25%)	160 (85%)	24 (13%)	5 (3%)	5	35
3	K	319/746 (43%)	270 (85%)	46 (14%)	3 (1%)	17	54
4	L	360/441 (82%)	305 (85%)	51 (14%)	4 (1%)	14	51
5	N	382/425 (90%)	335 (88%)	42 (11%)	5 (1%)	12	48
6	O	5/7 (71%)	4 (80%)	1 (20%)	0	100	100
7	P	62/295 (21%)	48 (77%)	14 (23%)	0	100	100
9	B	10/345 (3%)	9 (90%)	1 (10%)	0	100	100
All	All	1657/5570 (30%)	1388 (84%)	243 (15%)	26 (2%)	13	44

5 of 26 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	481	HIS
4	L	400	LYS
1	M	581	VAL
5	N	171	LEU
3	C	127	VAL

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	72/646 (11%)	72 (100%)	0	100	100
1	M	101/646 (16%)	100 (99%)	1 (1%)	76	86
1	Q	45/646 (7%)	45 (100%)	0	100	100
2	E	18/297 (6%)	17 (94%)	1 (6%)	21	51
3	C	100/667 (15%)	99 (99%)	1 (1%)	76	86
3	K	233/667 (35%)	227 (97%)	6 (3%)	46	68
4	L	310/392 (79%)	305 (98%)	5 (2%)	62	79
5	N	340/375 (91%)	335 (98%)	5 (2%)	65	80
6	O	3/3 (100%)	3 (100%)	0	100	100
7	P	50/263 (19%)	48 (96%)	2 (4%)	31	58
9	B	3/294 (1%)	3 (100%)	0	100	100
All	All	1275/4896 (26%)	1254 (98%)	21 (2%)	64	79

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	N	79	LEU
5	N	387	PRO
3	C	45	ASN
7	P	236	LYS
5	N	261	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
7	P	290	GLN
1	Q	83	HIS
5	N	168	ASN
5	N	188	ASN
5	N	226	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](#)

1 non-standard protein/DNA/RNA residue is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
9	M3L	B	116	9,3	10,11,12	0.47	0	9,14,16	0.58	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
9	M3L	B	116	9,3	-	1/9/10/12	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
9	B	116	M3L	CE-CD-CG-CB

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
9	B	116	M3L	2	0

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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
8	Z	6

The worst 5 of 6 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	Z	182:UNK	C	247:UNK	N	39.79
1	Z	45:UNK	C	55:UNK	N	32.97
1	Z	67:UNK	C	77:UNK	N	29.83
1	Z	138:UNK	C	143:UNK	N	25.66
1	Z	99:UNK	C	120:UNK	N	10.57

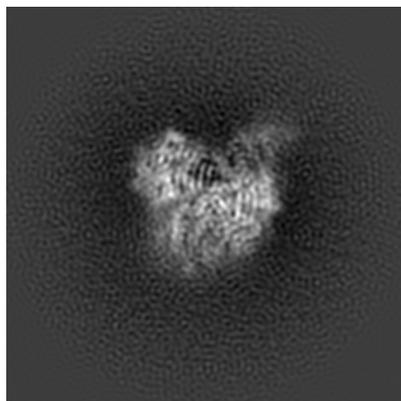
6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-7334. These allow visual inspection of the internal detail of the map and identification of artifacts.

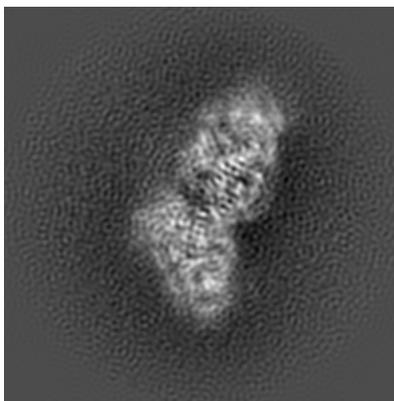
Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

6.1 Orthogonal projections [i](#)

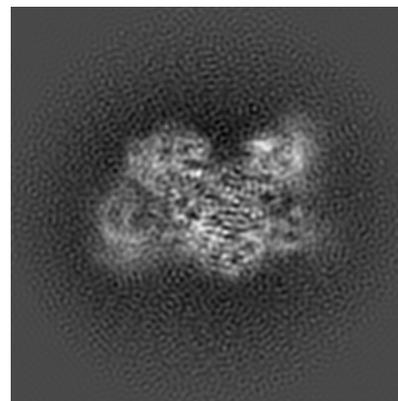
6.1.1 Primary map



X

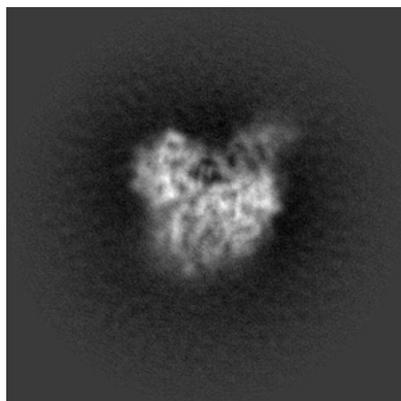


Y

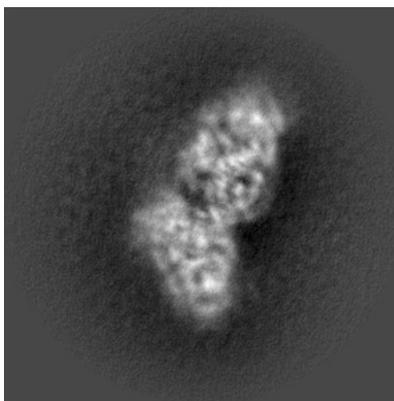


Z

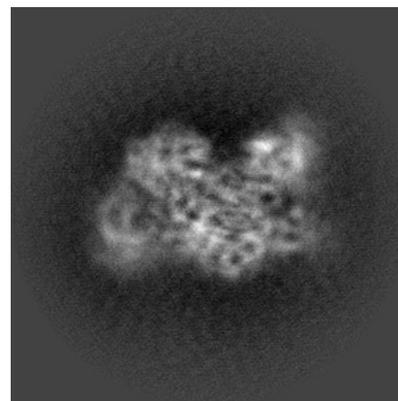
6.1.2 Raw map



X



Y

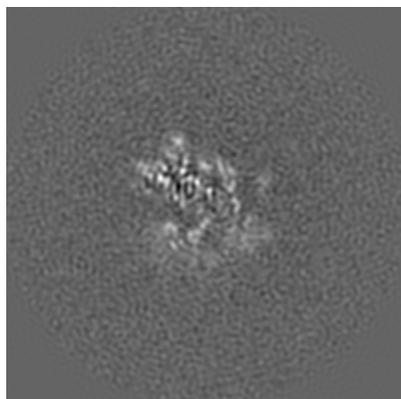


Z

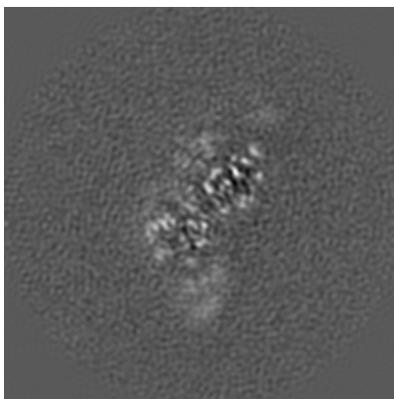
The images above show the map projected in three orthogonal directions.

6.2 Central slices [i](#)

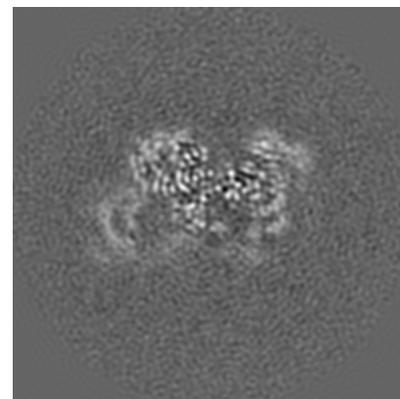
6.2.1 Primary map



X Index: 144

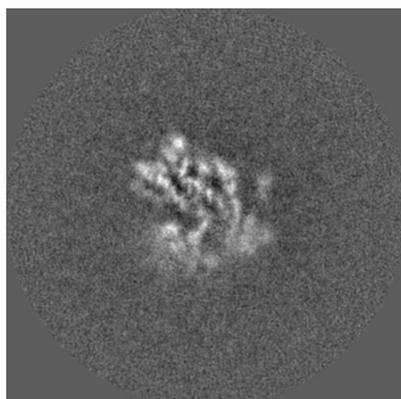


Y Index: 144

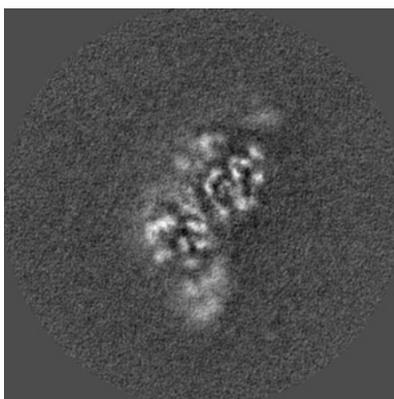


Z Index: 144

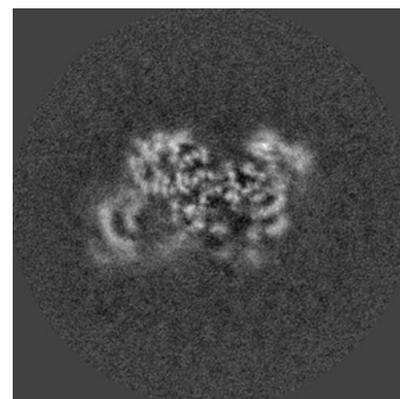
6.2.2 Raw map



X Index: 144



Y Index: 144

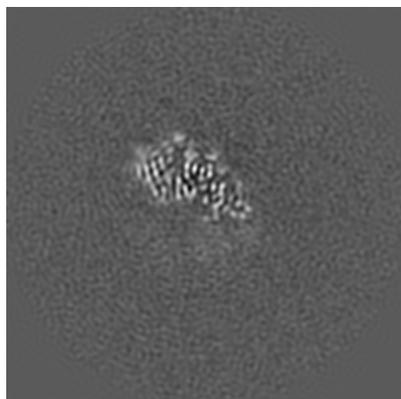


Z Index: 144

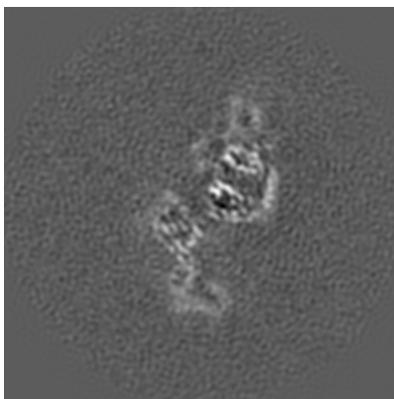
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices [i](#)

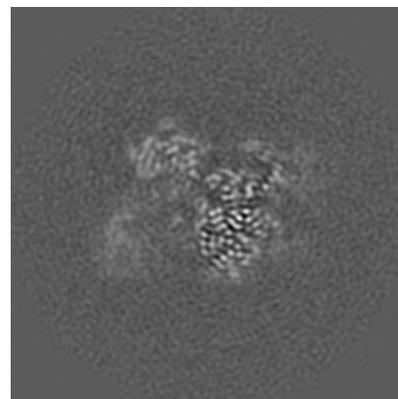
6.3.1 Primary map



X Index: 156

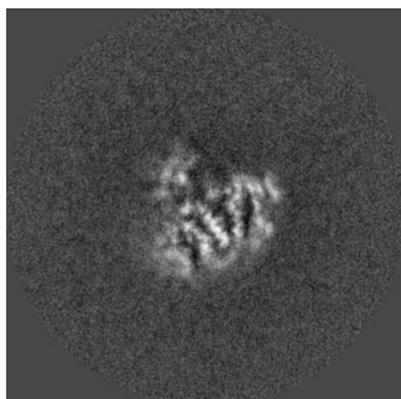


Y Index: 124

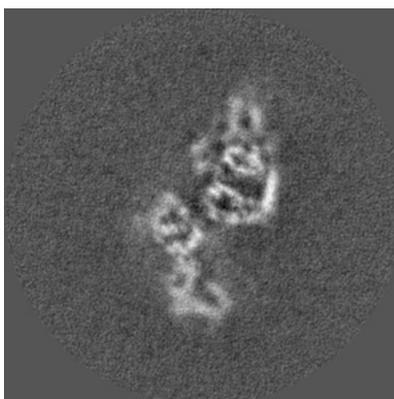


Z Index: 159

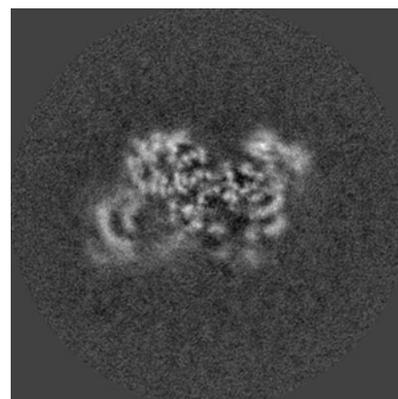
6.3.2 Raw map



X Index: 132



Y Index: 124



Z Index: 144

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

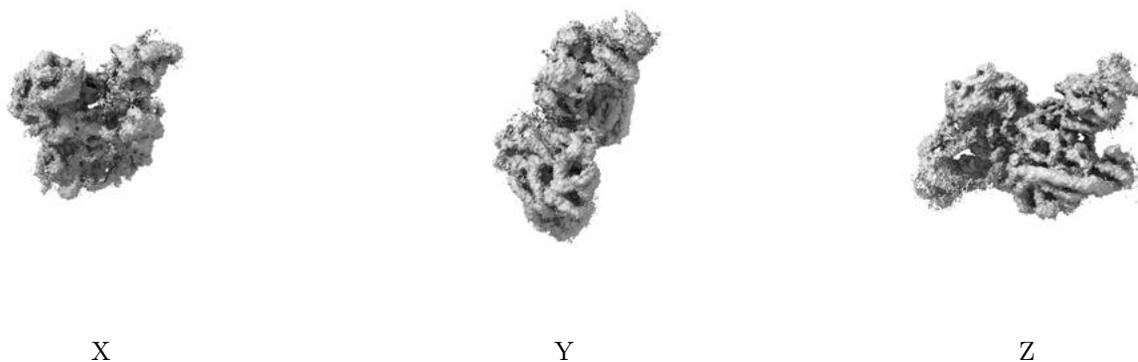
6.4 Orthogonal surface views [i](#)

6.4.1 Primary map



The images above show the 3D surface view of the map at the recommended contour level 0.0119. 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



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.

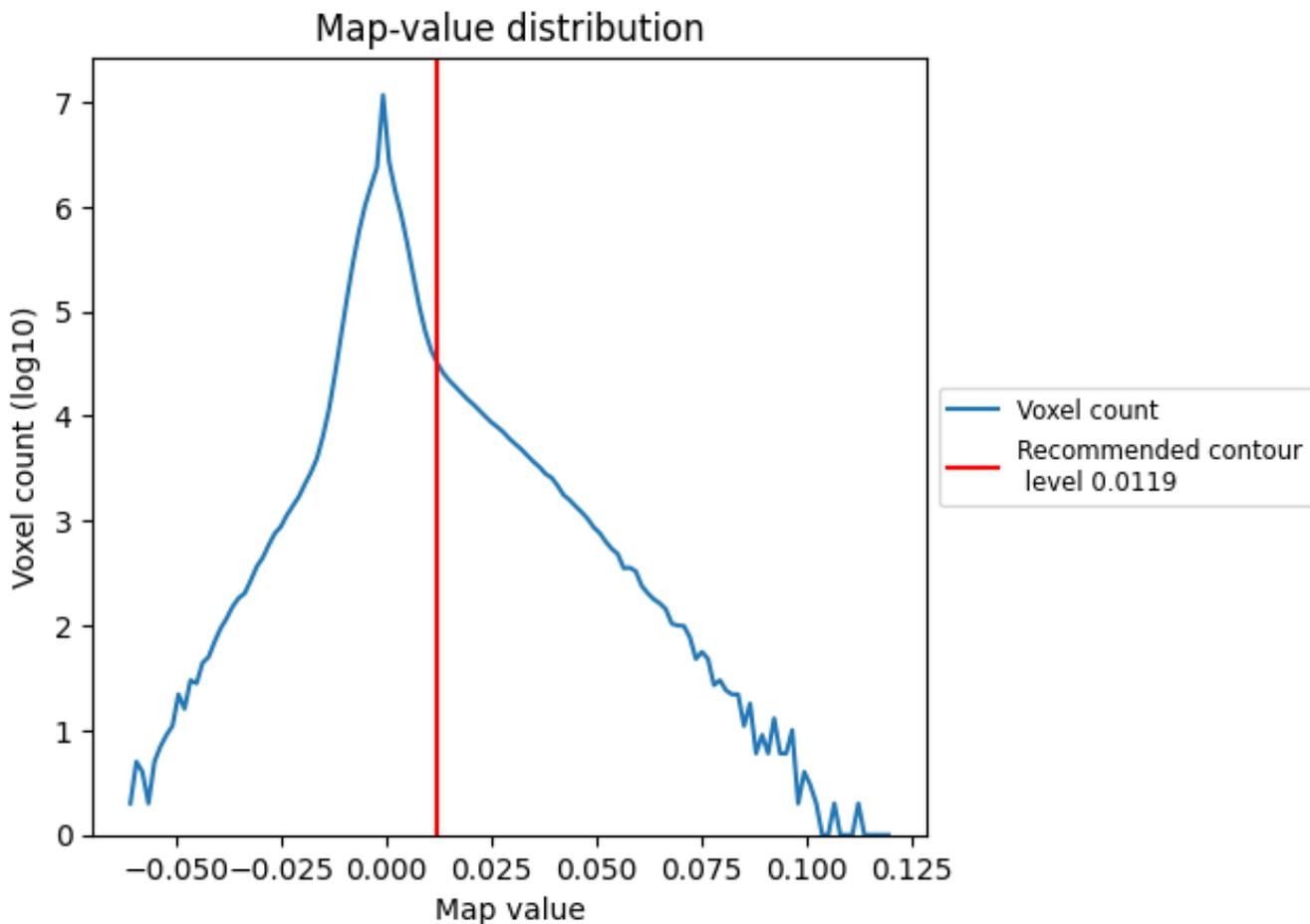
6.5 Mask visualisation [i](#)

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

7 Map analysis [i](#)

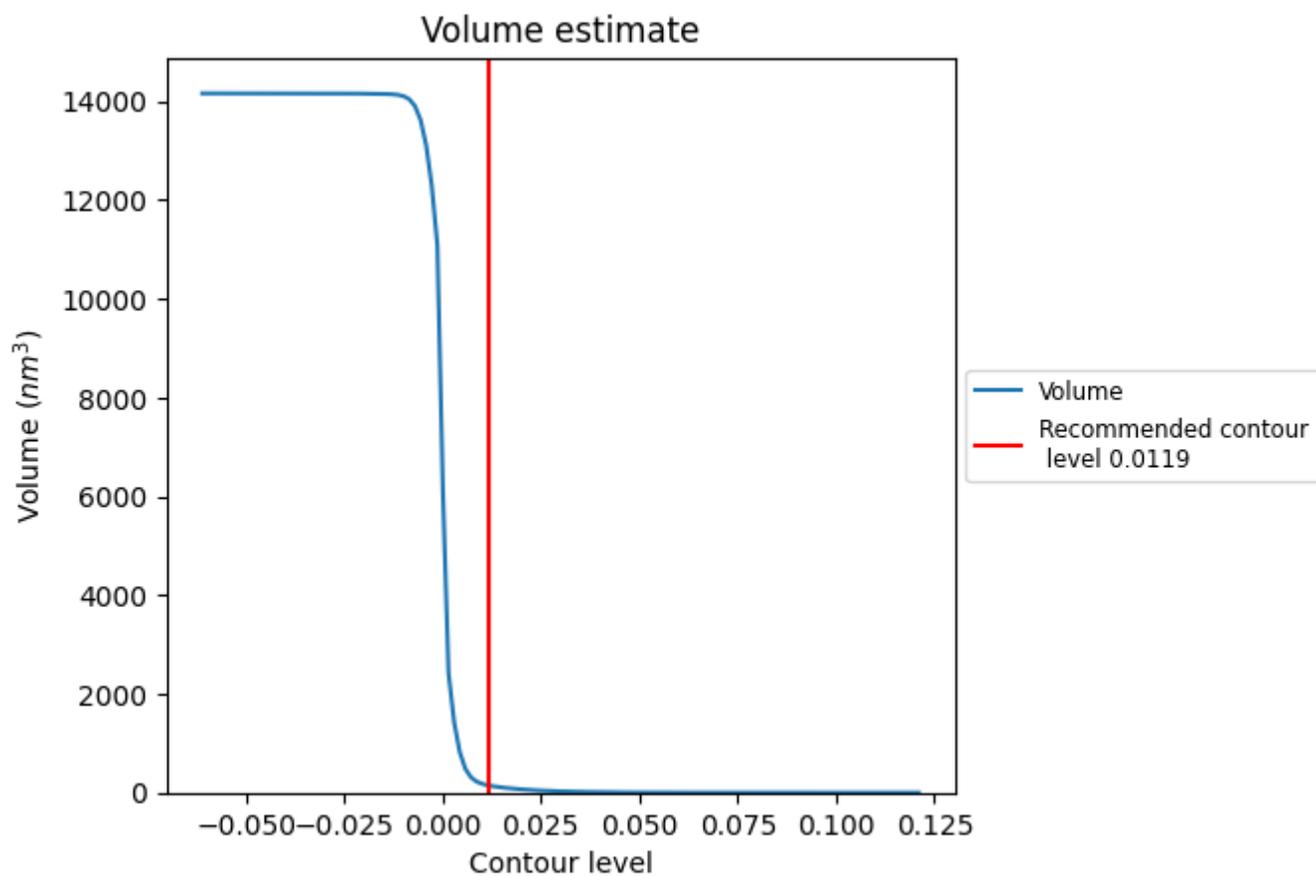
This section contains the results of statistical analysis of the map.

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.

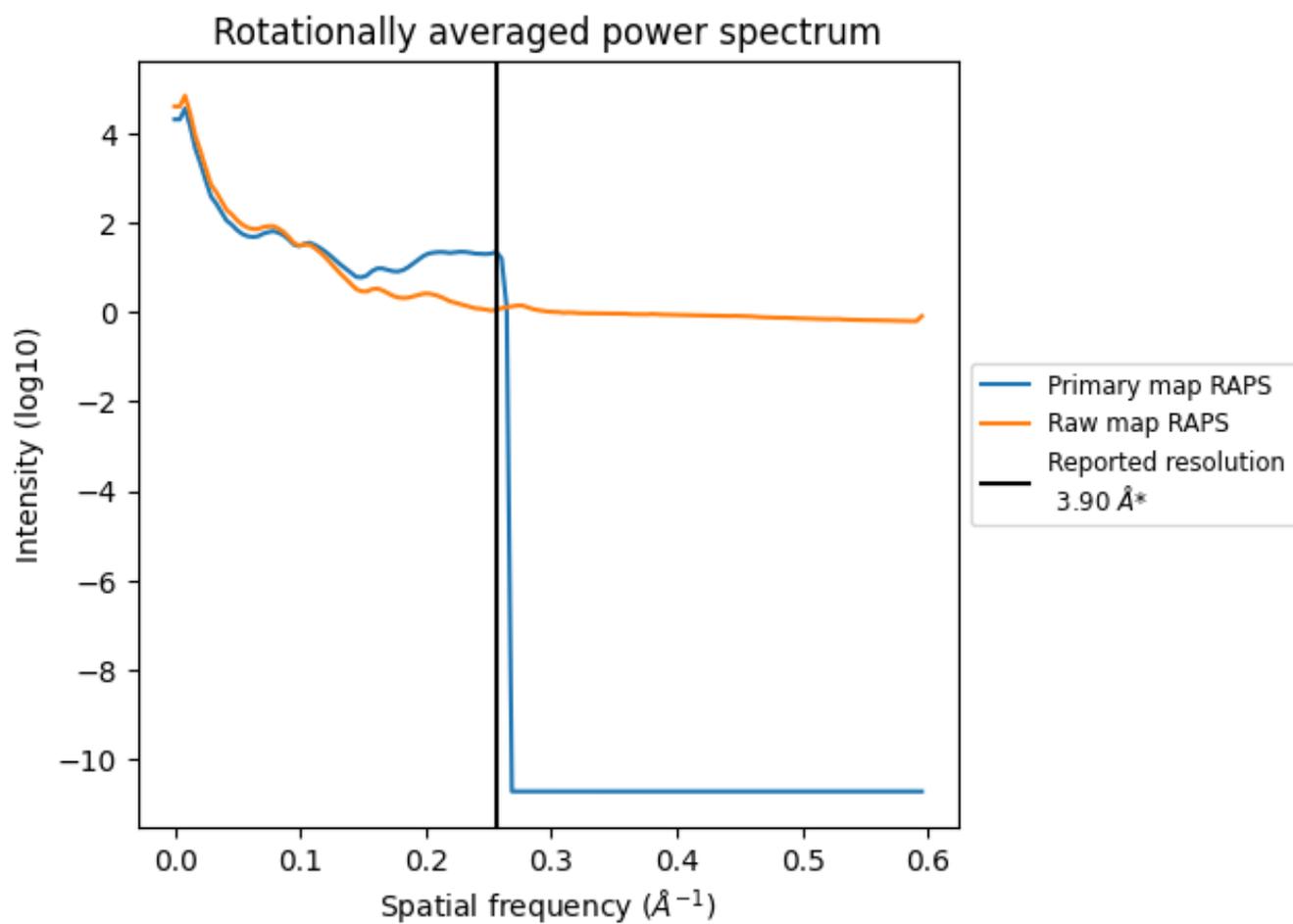
7.2 Volume estimate [\(i\)](#)



The volume at the recommended contour level is 143 nm³; this corresponds to an approximate mass of 129 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](#)

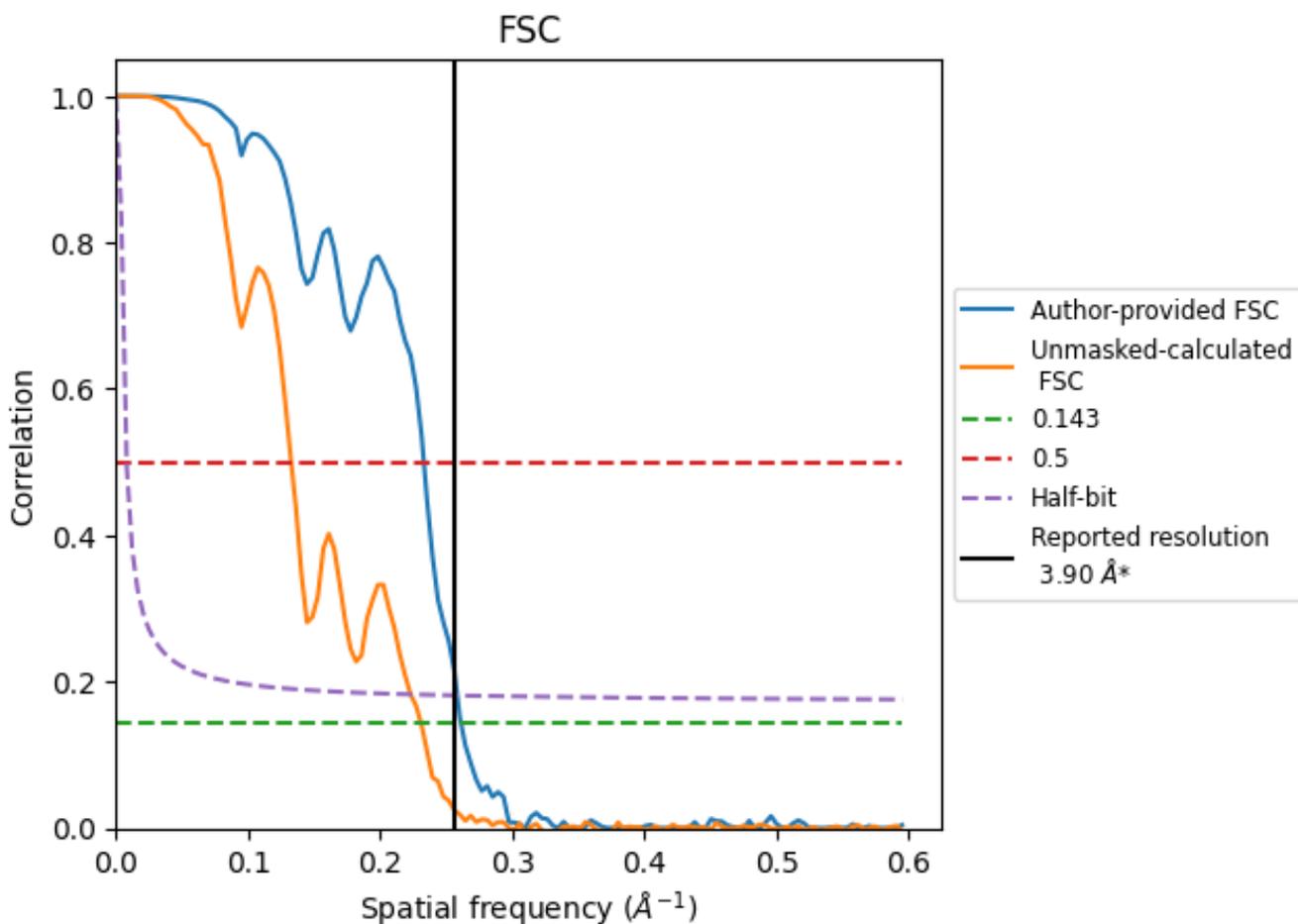


*Reported resolution corresponds to spatial frequency of 0.256 Å⁻¹

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](#)

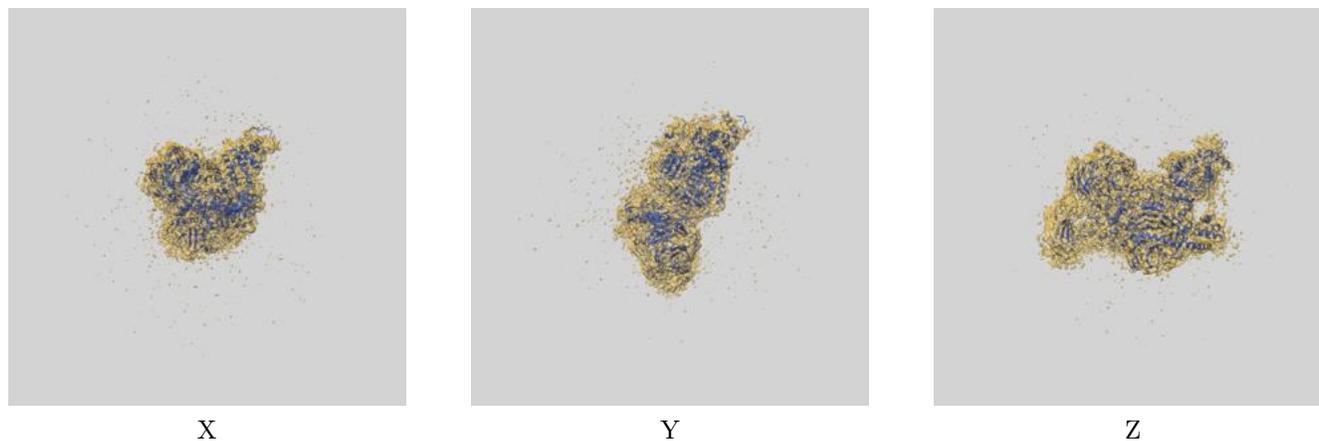
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.90	-	-
Author-provided FSC curve	3.83	4.29	3.87
Unmasked-calculated*	4.33	7.54	4.48

*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 4.33 differs from the reported value 3.9 by more than 10 %

9 Map-model fit [i](#)

This section contains information regarding the fit between EMDB map EMD-7334 and PDB model 6C23. Per-residue inclusion information can be found in section 3 on page 6.

9.1 Map-model overlay [i](#)



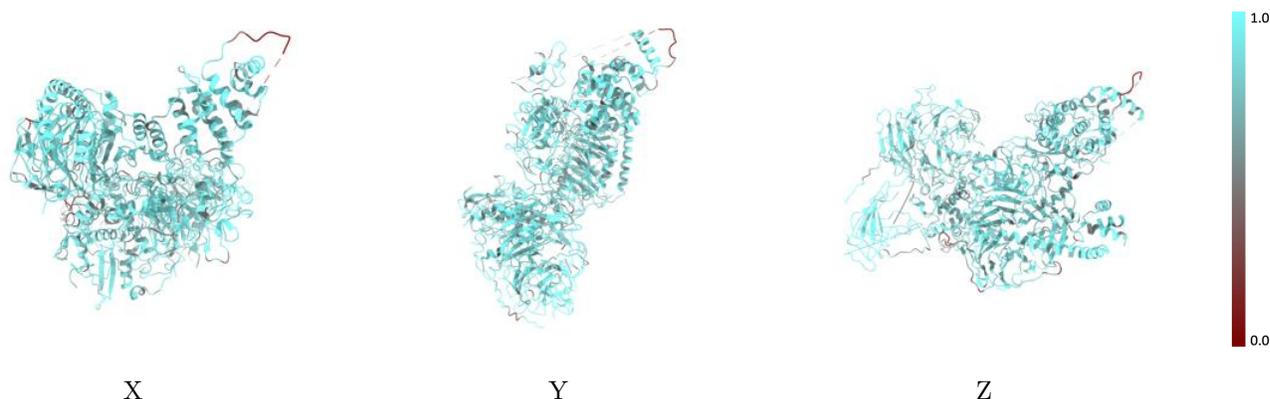
The images above show the 3D surface view of the map at the recommended contour level 0.0119 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

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



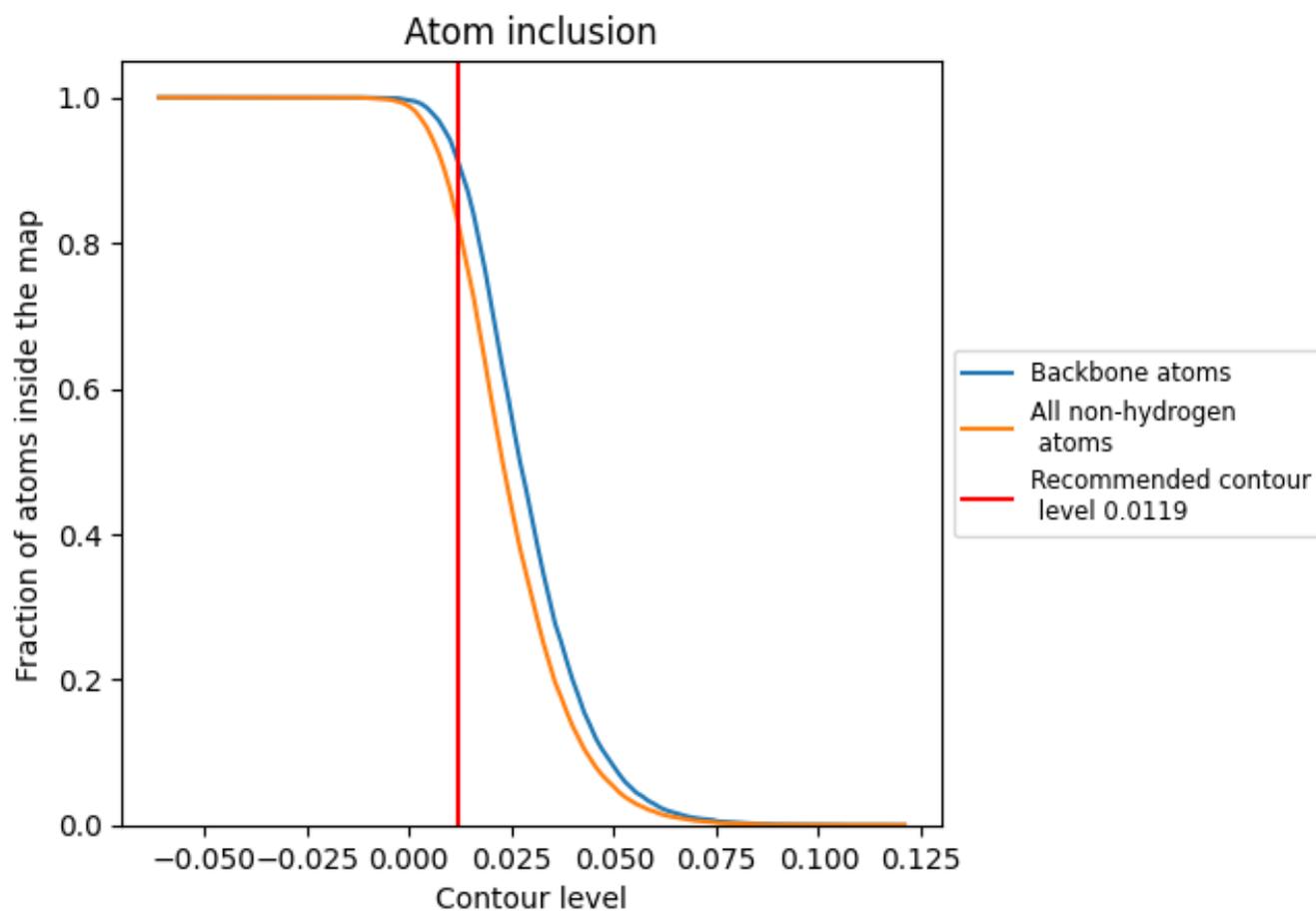
The images above show the model with each residue coloured according to its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions 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 (0.0119).

9.4 Atom inclusion [i](#)



At the recommended contour level, 91% of all backbone atoms, 83% 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 (0.0119) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	 0.8298	 0.3940
A	 0.8504	 0.3980
B	 0.8810	 0.3300
C	 0.7705	 0.3300
E	 0.7374	 0.3230
K	 0.8120	 0.3620
L	 0.8470	 0.4380
M	 0.8189	 0.3960
N	 0.8418	 0.4160
O	 0.5102	 0.3390
P	 0.8000	 0.3890
Q	 0.8610	 0.4180
Z	 0.9095	 0.3590

