



wwPDB EM Validation Summary Report ⓘ

Dec 18, 2022 – 05:10 pm GMT

PDB ID : 7AQO
EMDB ID : EMD-11871
Title : yeast THO-Sub2 complex dimer
Authors : Schuller, S.K.; Schuller, J.M.; Prabu, R.J.; Baumgartner, M.; Bonneau, F.;
basquin, J.; Conti, E.
Deposited on : 2020-10-22
Resolution : 4.50 Å(reported)
Based on initial model : 7APX

This is a wwPDB EM Validation Summary Report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/EMValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

EMDB validation analysis : 0.0.1.dev43
MolProbity : 4.02b-467
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ : 1.9.9
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.31.3

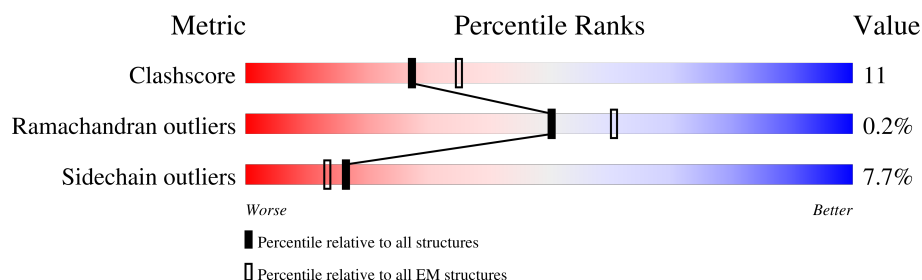
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 4.50 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	1601	
1	G	1601	
2	B	720	
2	H	720	
3	E	380	
3	K	380	
4	F	400	
4	L	400	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	C	261	<div><div></div><div>62%19%5%13%</div></div>
5	J	261	<div><div>20%</div><div>64%19%5%12%</div></div>
6	D	392	<div><div></div><div>47%8%43%</div></div>
6	I	392	<div><div>18%</div><div>47%8%42%</div></div>

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 40261 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 THO complex subunit 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	G	1079	Total	C	N	O	S	0	0
			7785	5074	1304	1379	28		
1	A	1079	Total	C	N	O	S	0	0
			7785	5074	1304	1379	28		

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	-3	GLY	-	expression tag	UNP A0A6A5Q535
G	-2	ALA	-	expression tag	UNP A0A6A5Q535
G	-1	ALA	-	expression tag	UNP A0A6A5Q535
G	0	SER	-	expression tag	UNP A0A6A5Q535
A	-3	GLY	-	expression tag	UNP A0A6A5Q535
A	-2	ALA	-	expression tag	UNP A0A6A5Q535
A	-1	ALA	-	expression tag	UNP A0A6A5Q535
A	0	SER	-	expression tag	UNP A0A6A5Q535

- Molecule 2 is a protein called THO complex subunit HPR1.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	H	563	Total	C	N	O	S	0	0
			4143	2713	693	721	16		
2	B	563	Total	C	N	O	S	0	0
			4143	2713	693	721	16		

- Molecule 3 is a protein called TEX1 isoform 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	K	348	Total	C	N	O	S	0	0
			2357	1530	402	414	11		
3	E	348	Total	C	N	O	S	0	0
			2357	1530	402	414	11		

- Molecule 4 is a protein called EM14S01-3B_G0007820.mRNA.1.CDS.1.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	L	361	Total	C	N	O	S	0	0
			2818	1783	491	535	9		
4	F	361	Total	C	N	O	S	0	0
			2818	1783	491	535	9		

There are 8 discrepancies between the modelled and reference sequences:

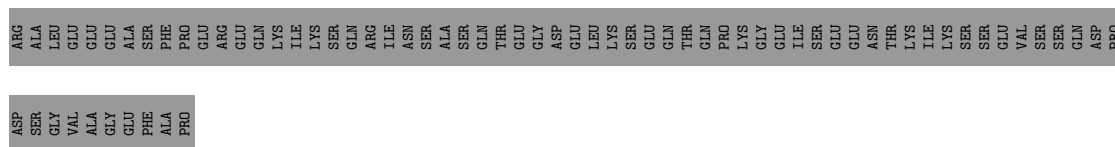
Chain	Residue	Modelled	Actual	Comment	Reference
L	47	GLY	-	expression tag	UNP A0A6A5Q316
L	48	ALA	-	expression tag	UNP A0A6A5Q316
L	49	ALA	-	expression tag	UNP A0A6A5Q316
L	50	SER	-	expression tag	UNP A0A6A5Q316
F	47	GLY	-	expression tag	UNP A0A6A5Q316
F	48	ALA	-	expression tag	UNP A0A6A5Q316
F	49	ALA	-	expression tag	UNP A0A6A5Q316
F	50	SER	-	expression tag	UNP A0A6A5Q316

- Molecule 5 is a protein called BJ4_G0025130.mRNA.1.CDS.1.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	J	230	Total	C	N	O	S	1	0
			1588	997	278	310	3		
5	C	227	Total	C	N	O	S	1	0
			1573	988	275	307	3		

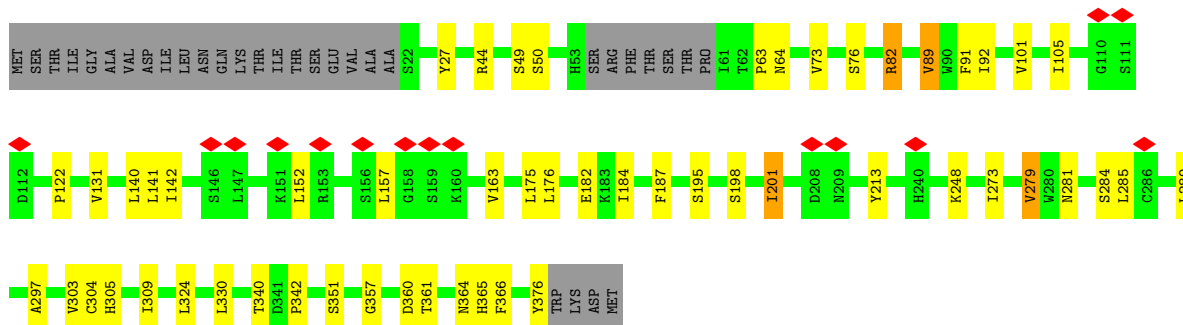
- Molecule 6 is a protein called THO complex subunit MFT1.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	I	227	Total	C	N	O	S	0	0
			1452	902	267	282	1		
6	D	225	Total	C	N	O	S	0	0
			1442	896	265	280	1		



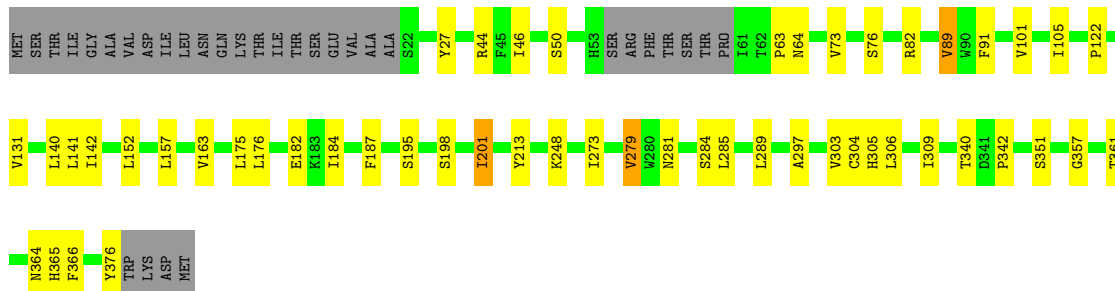
• Molecule 3: TEX1 isoform 1

Chain K:



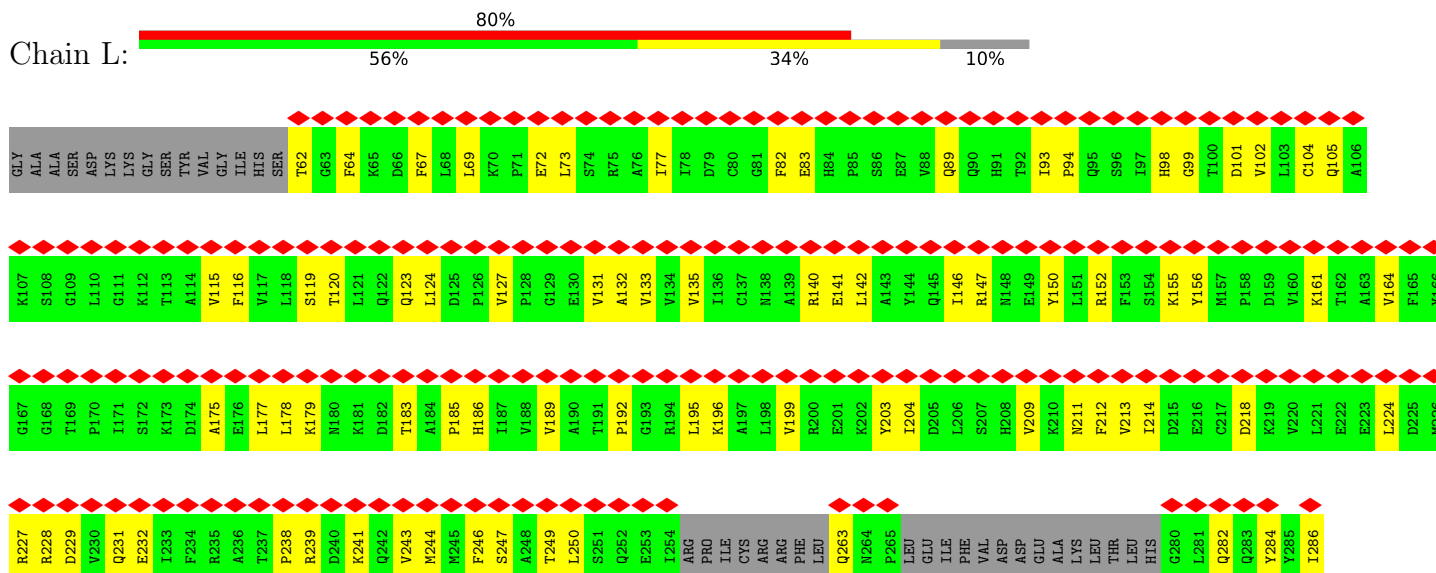
• Molecule 3: TEX1 isoform 1

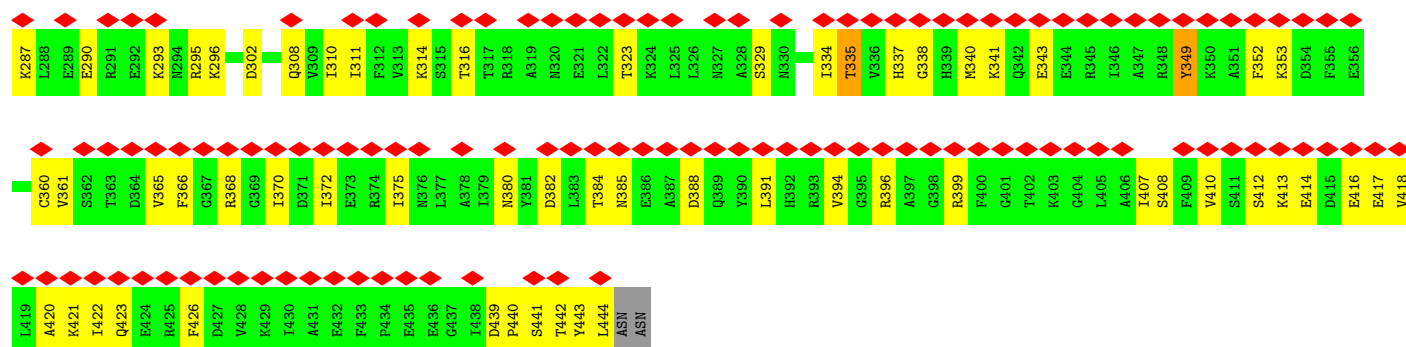
Chain E:



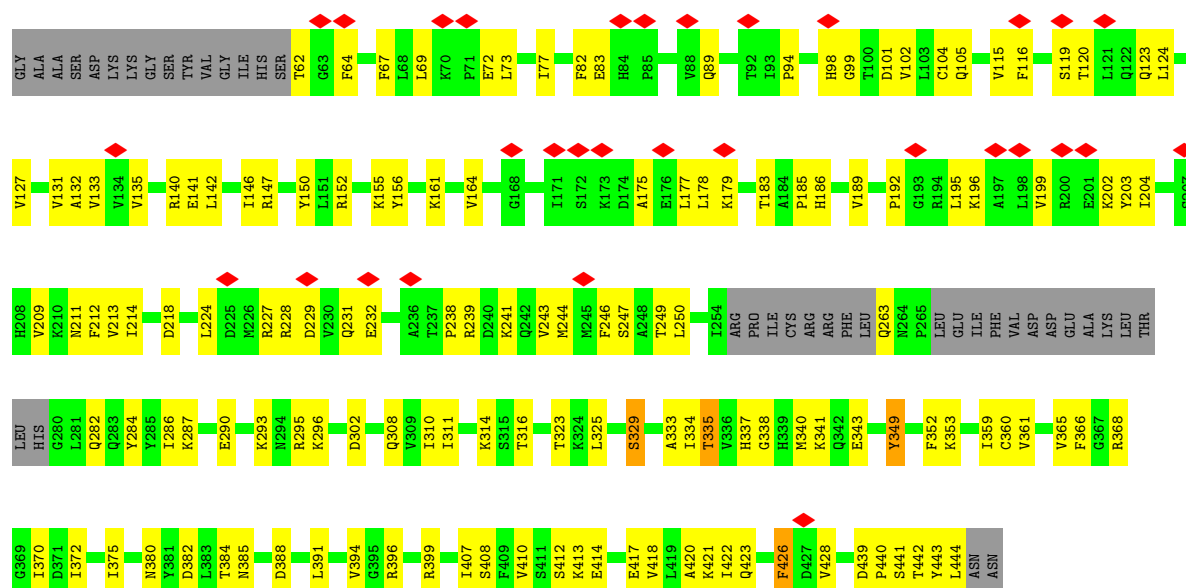
• Molecule 4: EM14S01-3B_G0007820.mRNA.1.CDS.1

Chain L:

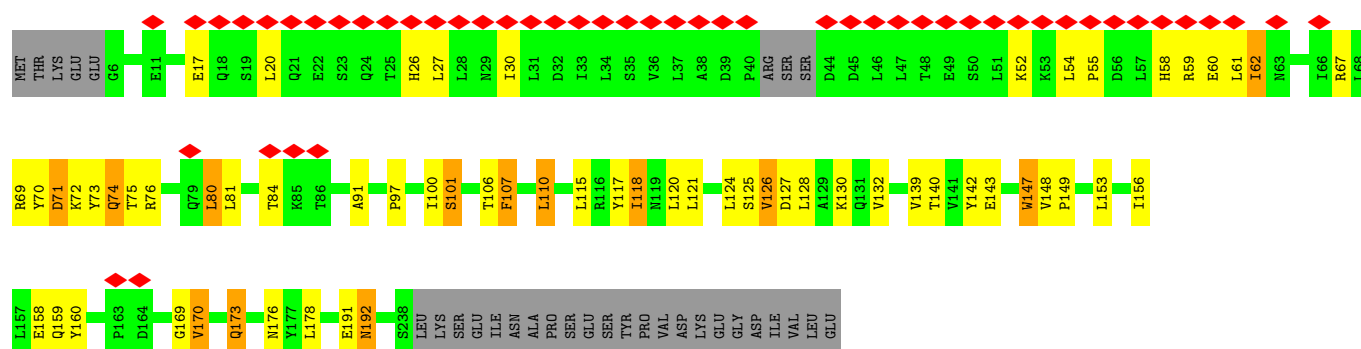




• Molecule 4: EM14S01-3B_G0007820.mRNA.1.CDS.1

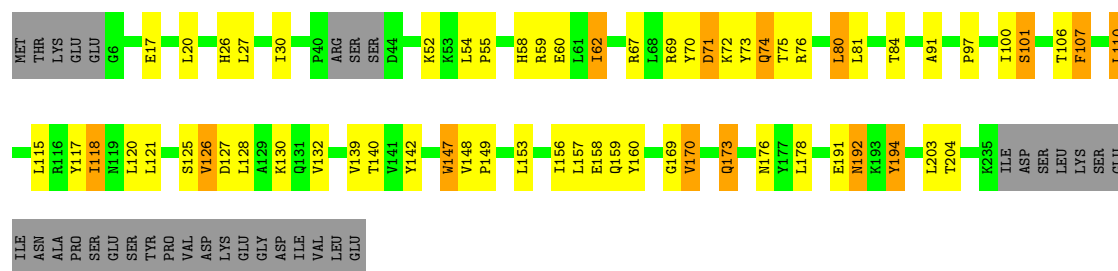


• Molecule 5: BJ4_G0025130.mRNA.1.CDS.1



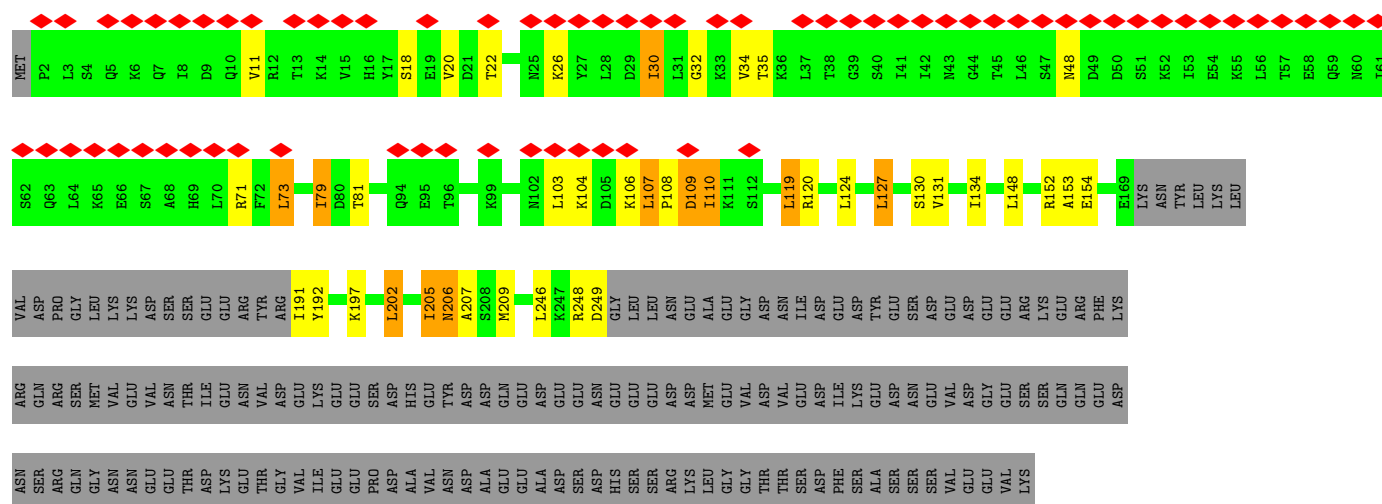
• Molecule 5: BJ4_G0025130.mRNA.1.CDS.1

Chain C:  62% 19% 5% 13%



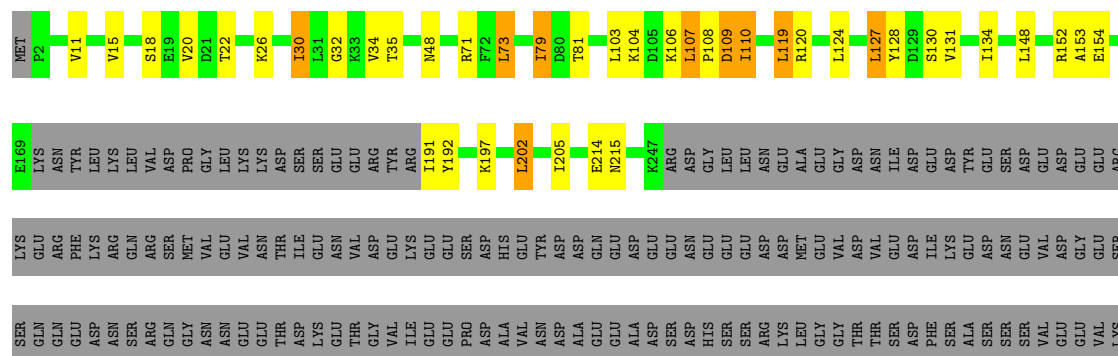
• Molecule 6: THO complex subunit MFT1

Chain I:  18% 47% 8% 27%



• Molecule 6: THO complex subunit MFT1

Chain D:  47% 8% 2% 43%



4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	113076	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	55.0	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K3 (6k x 4k)	Depositor
Maximum map value	1.789	Depositor
Minimum map value	-0.636	Depositor
Average map value	0.004	Depositor
Map value standard deviation	0.060	Depositor
Recommended contour level	0.25	Depositor
Map size (Å)	414.0, 414.0, 414.0	wwPDB
Map dimensions	300, 300, 300	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.38, 1.38, 1.38	Depositor

5 Model quality

5.1 Standard geometry

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	A	0.39	0/7953	0.48	0/10871
1	G	0.39	0/7953	0.48	0/10871
2	B	0.36	0/4246	0.46	0/5811
2	H	0.36	0/4246	0.46	0/5811
3	E	0.31	0/2411	0.47	0/3318
3	K	0.32	0/2411	0.47	0/3318
4	F	0.25	0/2862	0.46	0/3865
4	L	0.26	0/2862	0.46	0/3865
5	C	0.75	0/1595	0.84	0/2184
5	J	0.74	0/1609	0.83	0/2202
6	D	0.71	0/1454	0.79	1/1989 (0.1%)
6	I	0.68	0/1464	0.79	1/2003 (0.0%)
All	All	0.43	0/41066	0.54	2/56108 (0.0%)

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
6	I	0	1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	I	81	THR	CB-CA-C	-6.00	95.40	111.60
6	D	81	THR	CB-CA-C	-5.98	95.47	111.60

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
6	I	206	ASN	Mainchain

5.2 Too-close contacts

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

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	7785	0	7050	140	0
1	G	7785	0	7050	132	0
2	B	4143	0	3661	76	0
2	H	4143	0	3661	76	0
3	E	2357	0	1985	28	0
3	K	2357	0	1985	30	0
4	F	2818	0	2793	102	0
4	L	2818	0	2793	99	0
5	C	1573	0	1319	77	0
5	J	1588	0	1324	77	0
6	D	1442	0	1153	46	0
6	I	1452	0	1157	48	0
All	All	40261	0	35931	824	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 824 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:C:67:ARG:O	5:C:71:ASP:OD1	1.58	1.21
5:J:67:ARG:O	5:J:71:ASP:OD1	1.58	1.20
3:K:50:SER:CB	3:E:46:ILE:HD12	1.71	1.20
5:C:117:TYR:CZ	6:D:120:ARG:HD3	1.84	1.12
5:J:117:TYR:CZ	6:I:120:ARG:HD3	1.84	1.11

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	1069/1601 (67%)	1002 (94%)	66 (6%)	1 (0%)	51	85
1	G	1069/1601 (67%)	1002 (94%)	66 (6%)	1 (0%)	51	85
2	B	555/720 (77%)	503 (91%)	51 (9%)	1 (0%)	47	81
2	H	555/720 (77%)	503 (91%)	51 (9%)	1 (0%)	47	81
3	E	344/380 (90%)	320 (93%)	24 (7%)	0	100	100
3	K	344/380 (90%)	320 (93%)	24 (7%)	0	100	100
4	F	355/400 (89%)	346 (98%)	9 (2%)	0	100	100
4	L	355/400 (89%)	346 (98%)	9 (2%)	0	100	100
5	C	224/261 (86%)	211 (94%)	9 (4%)	4 (2%)	8	42
5	J	225/261 (86%)	213 (95%)	9 (4%)	3 (1%)	12	48
6	D	221/392 (56%)	208 (94%)	12 (5%)	1 (0%)	29	68
6	I	223/392 (57%)	210 (94%)	12 (5%)	1 (0%)	34	72
All	All	5539/7508 (74%)	5184 (94%)	342 (6%)	13 (0%)	50	81

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	G	993	ILE
1	A	993	ILE
5	C	194	TYR
2	H	442	GLU
2	B	442	GLU

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	702/1468 (48%)	640 (91%)	62 (9%)	10	33
1	G	702/1468 (48%)	640 (91%)	62 (9%)	10	33
2	B	367/670 (55%)	343 (94%)	24 (6%)	17	44
2	H	367/670 (55%)	343 (94%)	24 (6%)	17	44
3	E	186/344 (54%)	172 (92%)	14 (8%)	13	40
3	K	186/344 (54%)	171 (92%)	15 (8%)	11	37
4	F	299/351 (85%)	293 (98%)	6 (2%)	55	73
4	L	299/351 (85%)	293 (98%)	6 (2%)	55	73
5	C	129/242 (53%)	112 (87%)	17 (13%)	4	20
5	J	129/242 (53%)	112 (87%)	17 (13%)	4	20
6	D	102/363 (28%)	88 (86%)	14 (14%)	3	19
6	I	102/363 (28%)	88 (86%)	14 (14%)	3	19
All	All	3570/6876 (52%)	3295 (92%)	275 (8%)	16	39

5 of 275 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
6	I	22	THR
6	I	109	ASP
5	C	176	ASN
3	K	309	ILE
3	K	248	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 79 such sidechains are listed below:

Mol	Chain	Res	Type
4	F	186	HIS
5	C	109	HIS
4	F	385	ASN
5	J	180	GLN
6	D	25	ASN

5.3.3 RNA ⓘ

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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
5	J	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	J	235:LYS	C	236:ILE	N	3.04

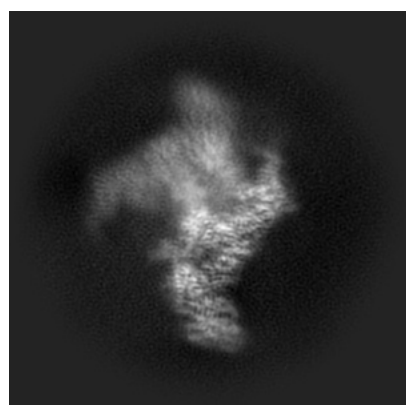
6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-11871. These allow visual inspection of the internal detail of the map and identification of artifacts.

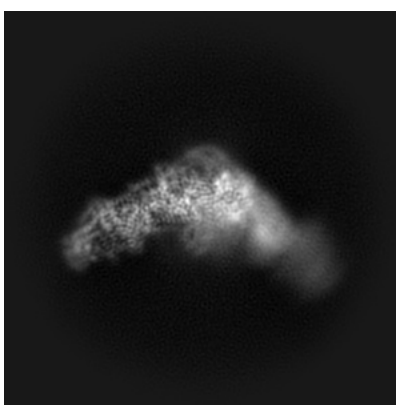
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections [i](#)

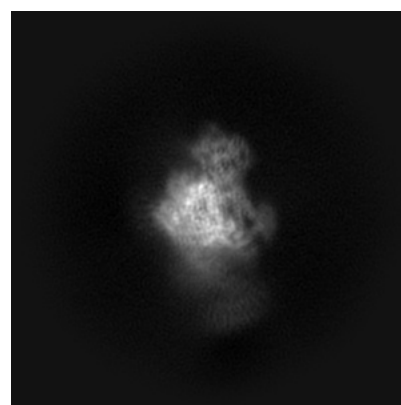
6.1.1 Primary map



X



Y

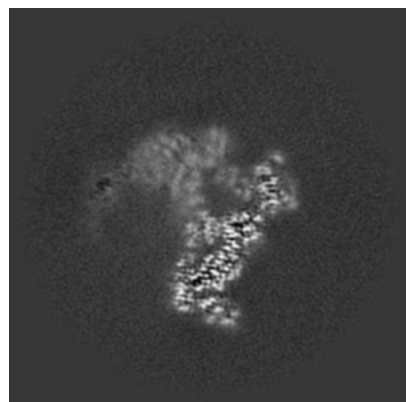


Z

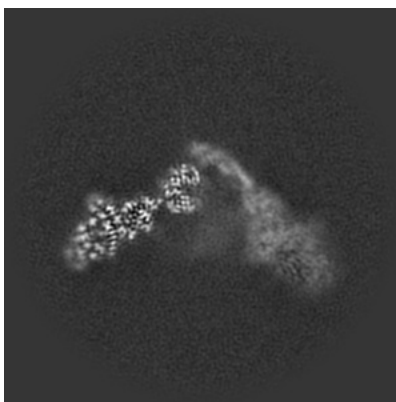
The images above show the map projected in three orthogonal directions.

6.2 Central slices [i](#)

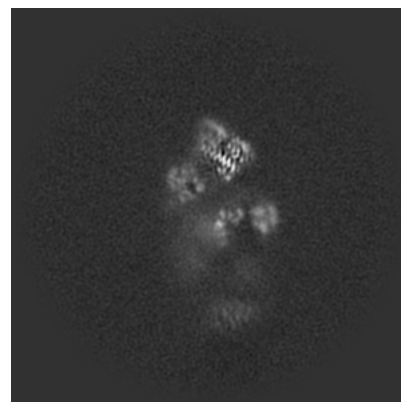
6.2.1 Primary map



X Index: 150



Y Index: 150

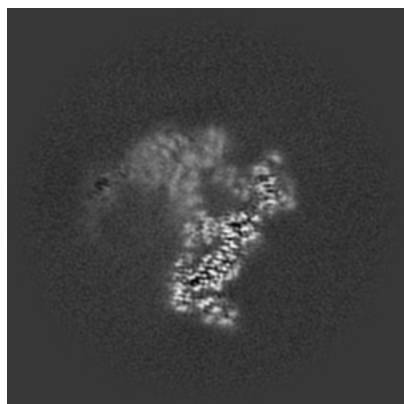


Z Index: 150

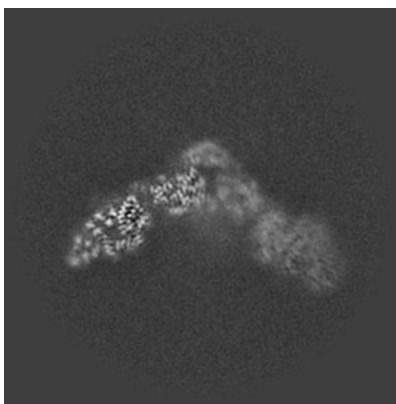
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices [i](#)

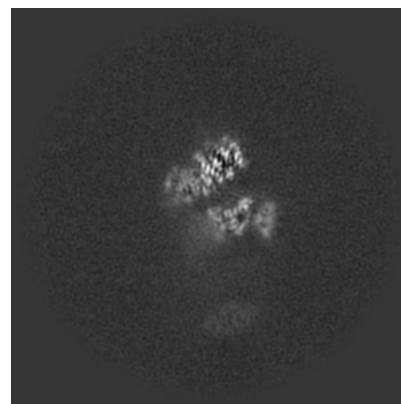
6.3.1 Primary map



X Index: 150



Y Index: 141



Z Index: 144

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

6.4 Orthogonal surface views [i](#)

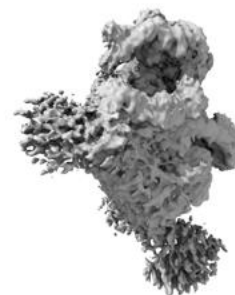
6.4.1 Primary map



X



Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.25. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

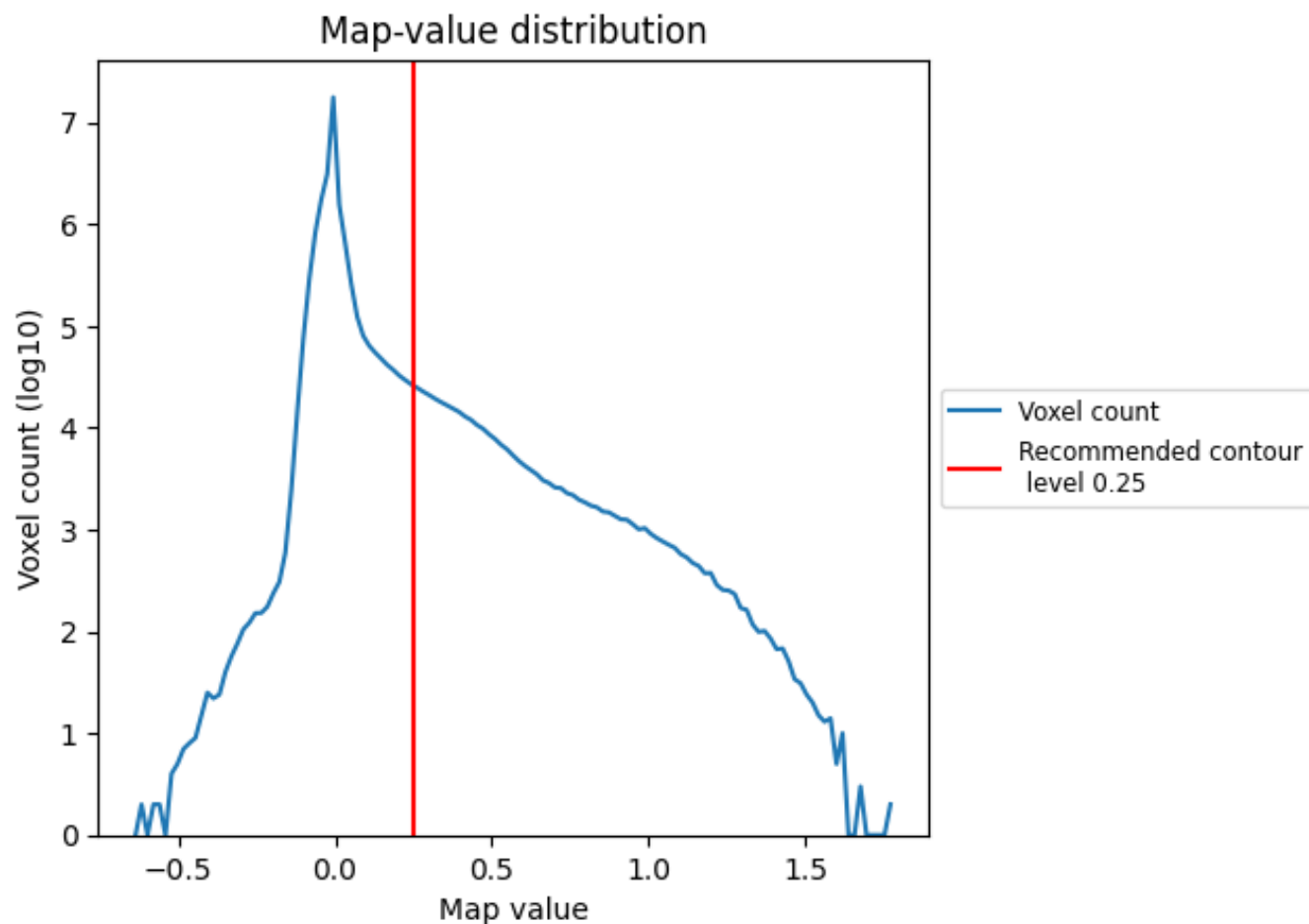
6.5 Mask visualisation

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

7 Map analysis [i](#)

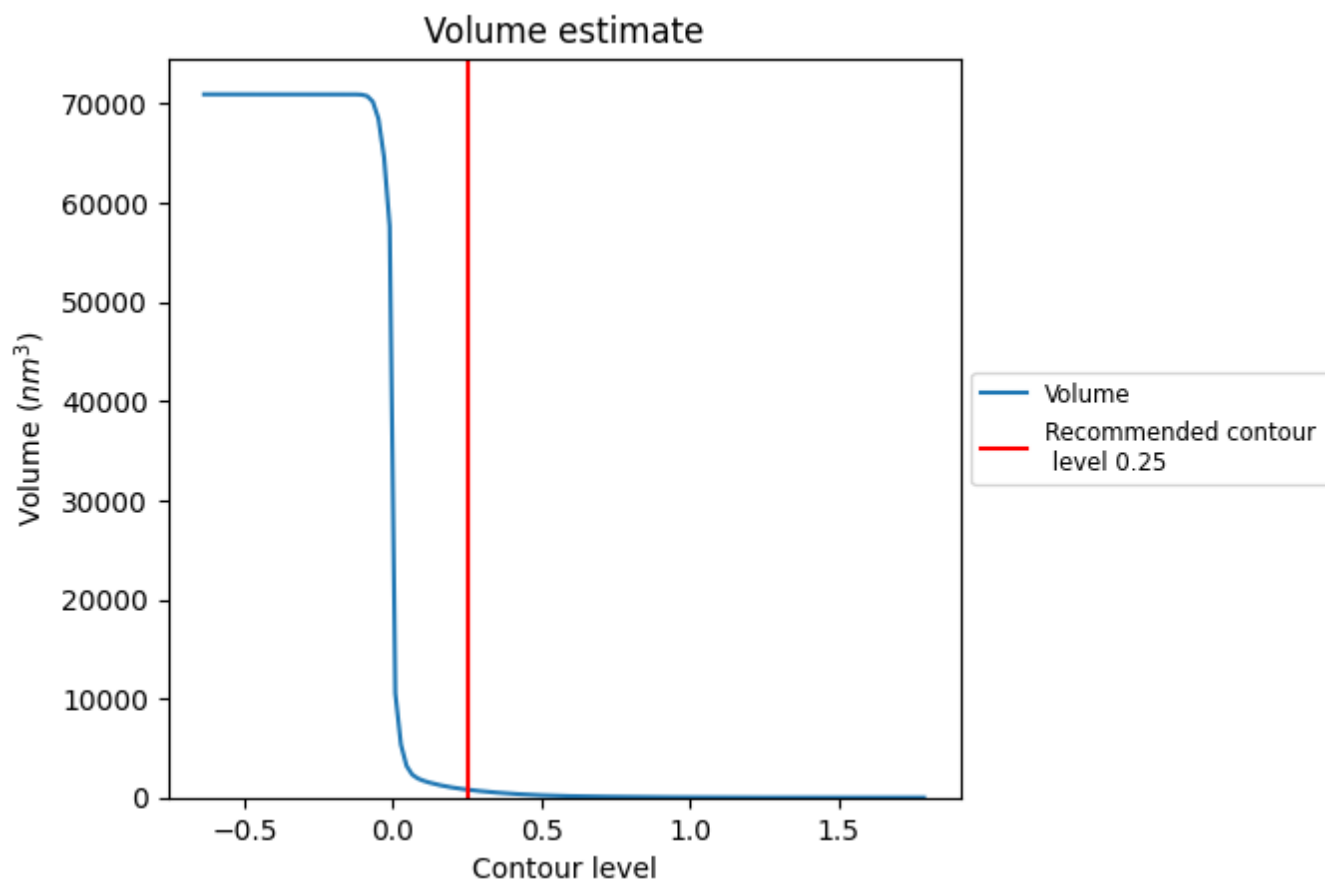
This section contains the results of statistical analysis of the map.

7.1 Map-value distribution [i](#)



The map-value distribution is plotted in 128 intervals along the x-axis. The y-axis is logarithmic. A spike in this graph at zero usually indicates that the volume has been masked.

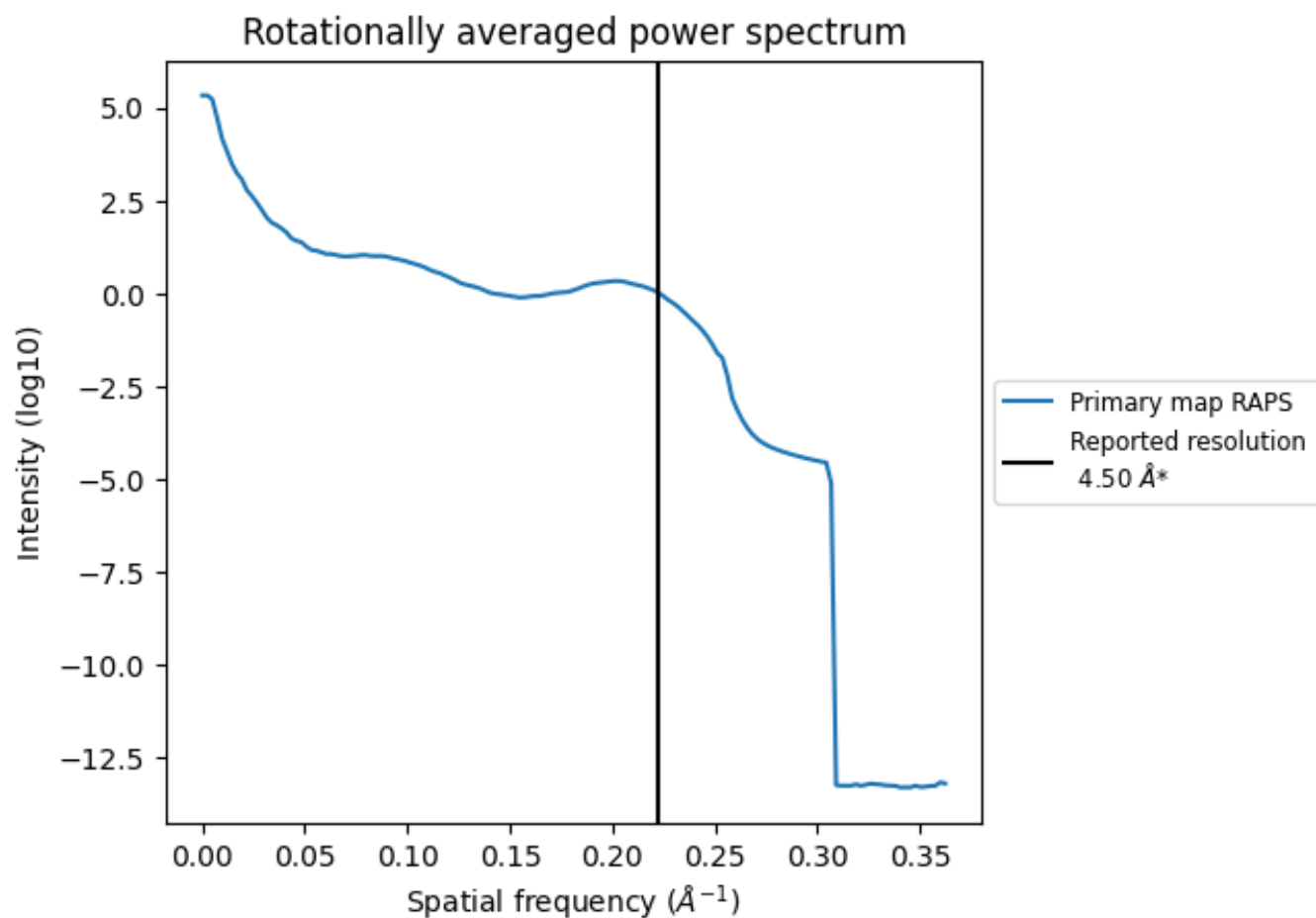
7.2 Volume estimate [i](#)



The volume at the recommended contour level is 804 nm³; this corresponds to an approximate mass of 726 kDa.

The volume estimate graph shows how the enclosed volume varies with the contour level. The recommended contour level is shown as a vertical line and the intersection between the line and the curve gives the volume of the enclosed surface at the given level.

7.3 Rotationally averaged power spectrum ⓘ



*Reported resolution corresponds to spatial frequency of 0.222 Å⁻¹

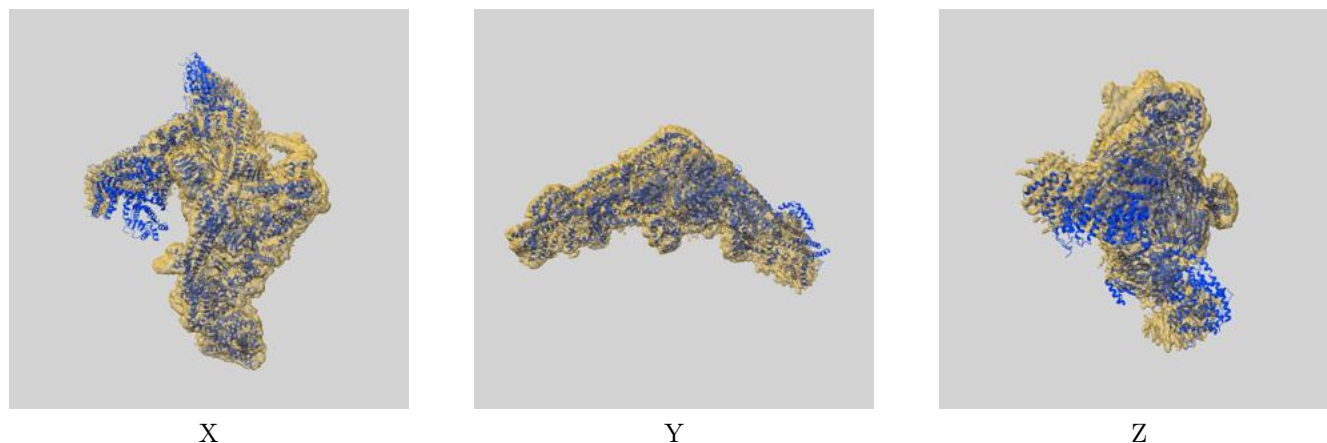
8 Fourier-Shell correlation ⓘ

This section was not generated. No FSC curve or half-maps provided.

9 Map-model fit [i](#)

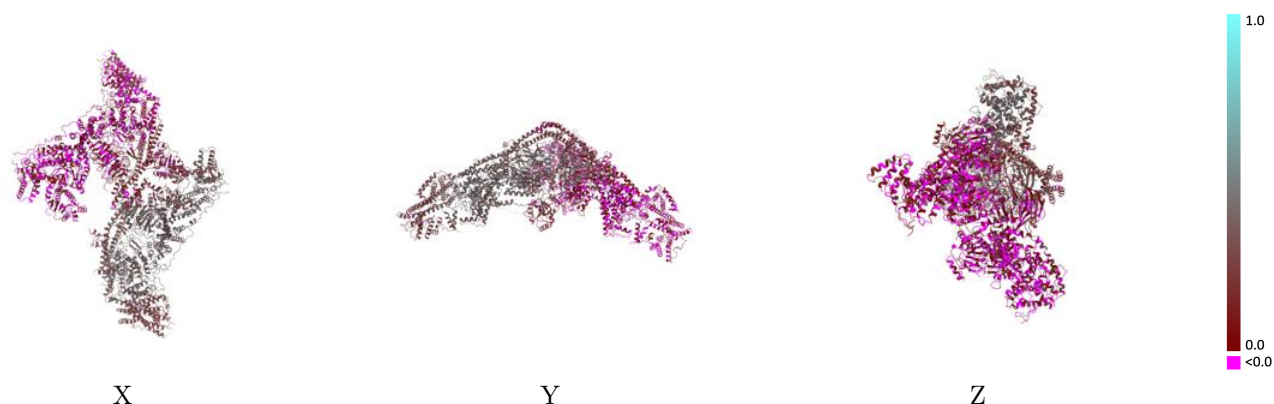
This section contains information regarding the fit between EMDB map EMD-11871 and PDB model 7AQO. Per-residue inclusion information can be found in [section 3](#) on [page 6](#).

9.1 Map-model overlay [i](#)



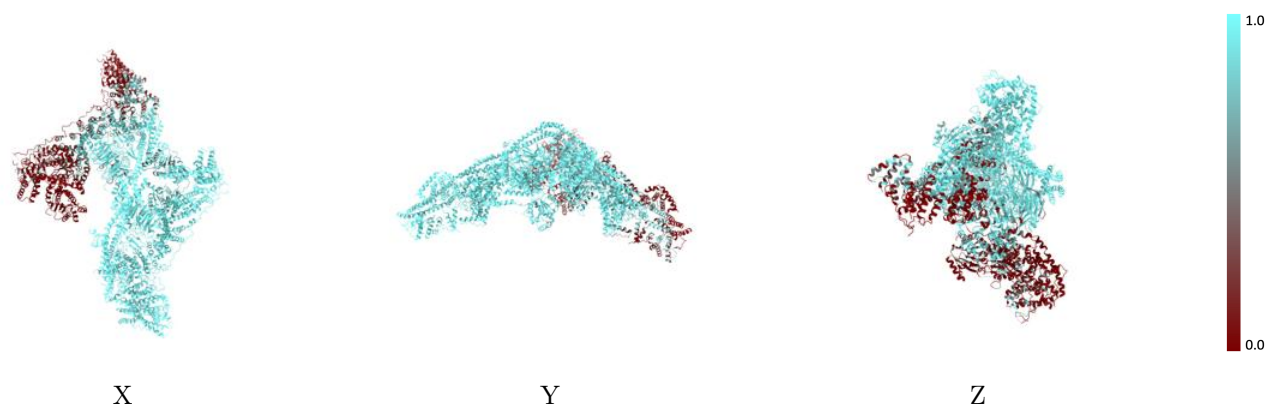
The images above show the 3D surface view of the map at the recommended contour level 0.25 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

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



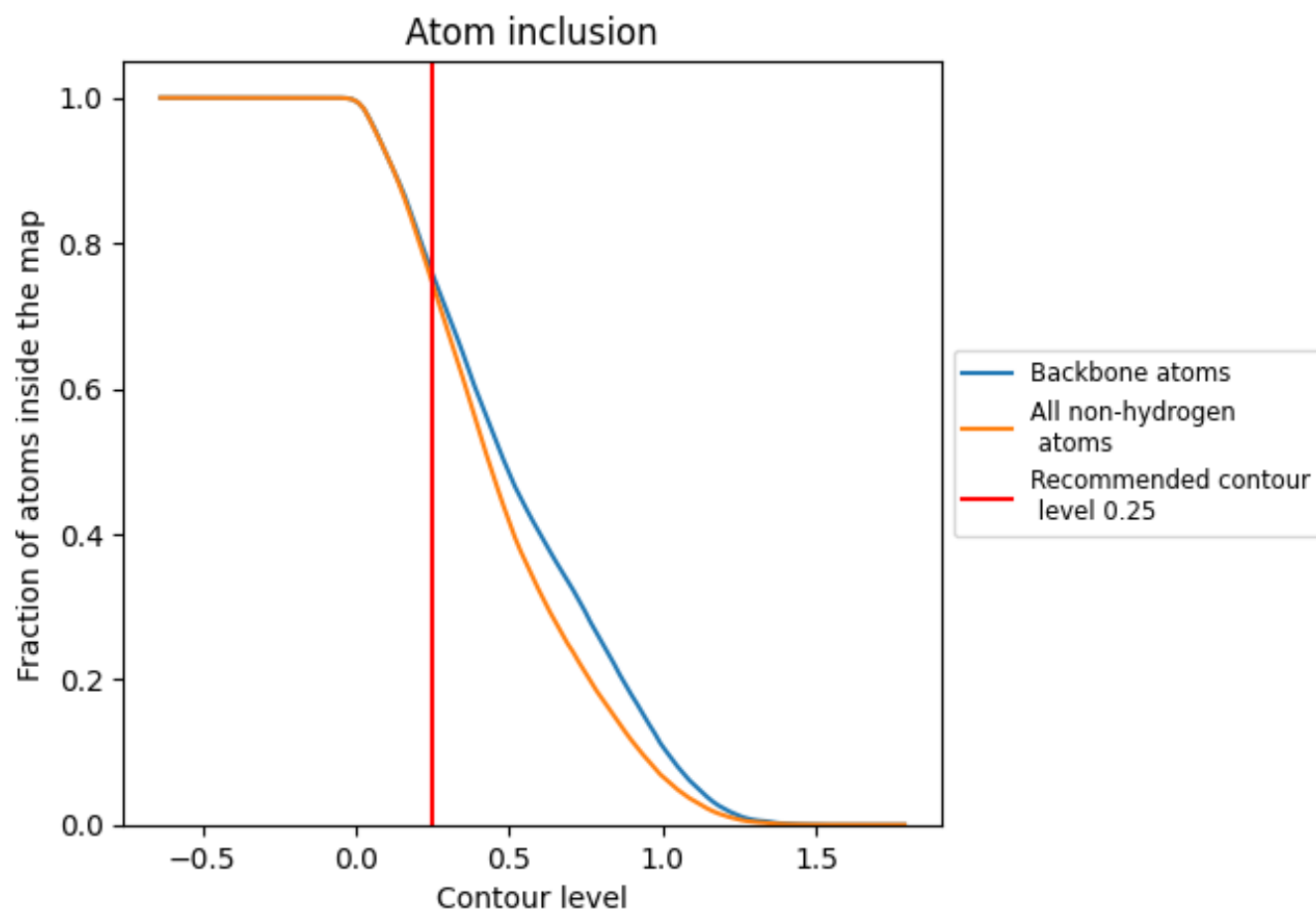
The images above show the model with each residue coloured according its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions 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 (0.25).

9.4 Atom inclusion [i](#)



At the recommended contour level, 76% of all backbone atoms, 75% 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 (0.25) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	<div></div> 0.7452	<div></div> 0.1890
A	<div></div> 0.9870	<div></div> 0.3920
B	<div></div> 0.9823	<div></div> 0.3790
C	<div></div> 0.9695	<div></div> 0.3400
D	<div></div> 0.9735	<div></div> 0.3130
E	<div></div> 0.9931	<div></div> 0.4000
F	<div></div> 0.8172	<div></div> 0.1320
G	<div></div> 0.4695	<div></div> 0.0170
H	<div></div> 0.5790	<div></div> 0.0200
I	<div></div> 0.6436	<div></div> 0.0790
J	<div></div> 0.7465	<div></div> 0.0910
K	<div></div> 0.9343	<div></div> 0.0960
L	<div></div> 0.1044	<div></div> 0.0020

1.0

0.0

<0.0