



wwPDB EM Validation Summary Report ⓘ

Nov 16, 2022 – 07:28 AM EST

PDB ID : 7KTS
EMDB ID : EMD-23028
Title : Negative stain EM structure of the human SAGA coactivator complex (TR-RAP, core, splicing module)
Authors : Herbst, D.A.; Esbin, M.N.; Nogales, E.
Deposited on : 2020-11-24
Resolution : 19.09 Å (reported)
Based on initial model : 5IFE

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) : 2.31.2

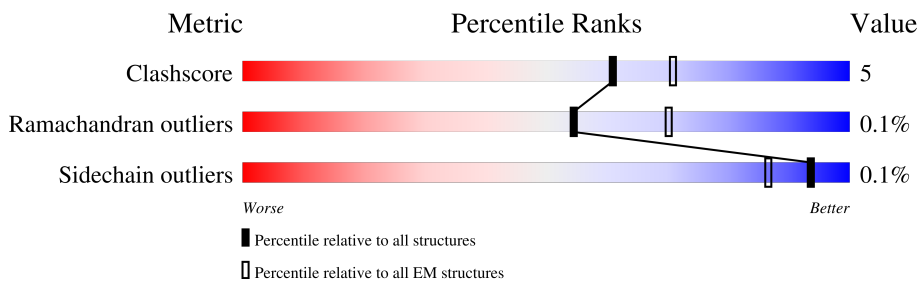
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 19.09 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	3848	
2	B	589	
3	C	811	
4	D	749	
5	E	251	
6	F	622	
7	G	161	
8	H	218	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
9	I	317	
10	J	335	
11	N	892	
12	S	1217	
13	T	86	

2 Entry composition [i](#)

There are 13 unique types of molecules in this entry. The entry contains 100130 atoms, of which 48957 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 Transformation/transcription domain-associated protein, Transformation/transcription domain-associated protein, Transformation/transcription domain-associated protein, Transformation/transcription domain-associated protein, Transformation/transcription domain-associated protein.

Mol	Chain	Residues	Atoms						AltConf	Trace
			Total	C	H	N	O	S		
1	A	3042	45856	14890	22462	4097	4249	158	3	0

- Molecule 2 is a protein called TAF5-like RNA polymerase II p300/CBP-associated factor-associated factor 65 kDa subunit 5L.

Mol	Chain	Residues	Atoms						AltConf	Trace
			Total	C	H	N	O	S		
2	B	534	8338	2657	4113	727	819	22	0	0

- Molecule 3 is a protein called Isoform 3 of Transcription factor SPT20 homolog.

Mol	Chain	Residues	Atoms						AltConf	Trace
			Total	C	H	N	O	S		
3	C	407	6609	2066	3316	580	624	23	1	0

- Molecule 4 is a protein called STAGA complex 65 subunit gamma, DhaA, STAGA complex 65 subunit gamma, STAGA complex 65 subunit gamma, DhaA, STAGA complex 65 subunit gamma, STAGA complex 65 subunit gamma, DhaA.

Mol	Chain	Residues	Atoms						AltConf	Trace
			Total	C	H	N	O	S		
4	D	219	3149	1045	1485	299	312	8	0	0

- Molecule 5 is a protein called Transcription initiation factor TFIID subunit 9B.

Mol	Chain	Residues	Atoms						AltConf	Trace
			Total	C	H	N	O	S		
5	E	124	2013	635	1018	175	180	5	0	0

- Molecule 6 is a protein called TAF6-like RNA polymerase II p300/CBP-associated factor-associated factor 65 kDa subunit 6L,TAF6-like RNA polymerase II p300/CBP-associated factor-associated factor 65 kDa subunit 6L,TAF6-like RNA polymerase II p300/CBP-associated factor-associated factor 65 kDa subunit 6L.

Mol	Chain	Residues	Atoms						AltConf	Trace
			Total	C	H	N	O	S		
6	F	351	3398	1308	1254	409	421	6	0	0

- Molecule 7 is a protein called Transcription initiation factor TFIID subunit 12.

Mol	Chain	Residues	Atoms						AltConf	Trace
			Total	C	H	N	O	S		
7	G	103	1704	527	859	152	161	5	0	0

- Molecule 8 is a protein called Transcription initiation factor TFIID subunit 10.

Mol	Chain	Residues	Atoms						AltConf	Trace
			Total	C	H	N	O	S		
8	H	99	1566	501	782	128	151	4	0	0

- Molecule 9 is a protein called Transcription initiation protein SPT3 homolog,Transcription initiation protein SPT3 homolog,Transcription initiation protein SPT3 homolog.

Mol	Chain	Residues	Atoms						AltConf	Trace
			Total	C	H	N	O	S		
9	I	243	3823	1207	1911	343	350	12	0	0

- Molecule 10 is a protein called Transcriptional adapter 1.

Mol	Chain	Residues	Atoms						AltConf	Trace
			Total	C	H	N	O	S		
10	J	214	3380	1073	1675	305	321	6	0	0

- Molecule 11 is a protein called Ataxin-7.

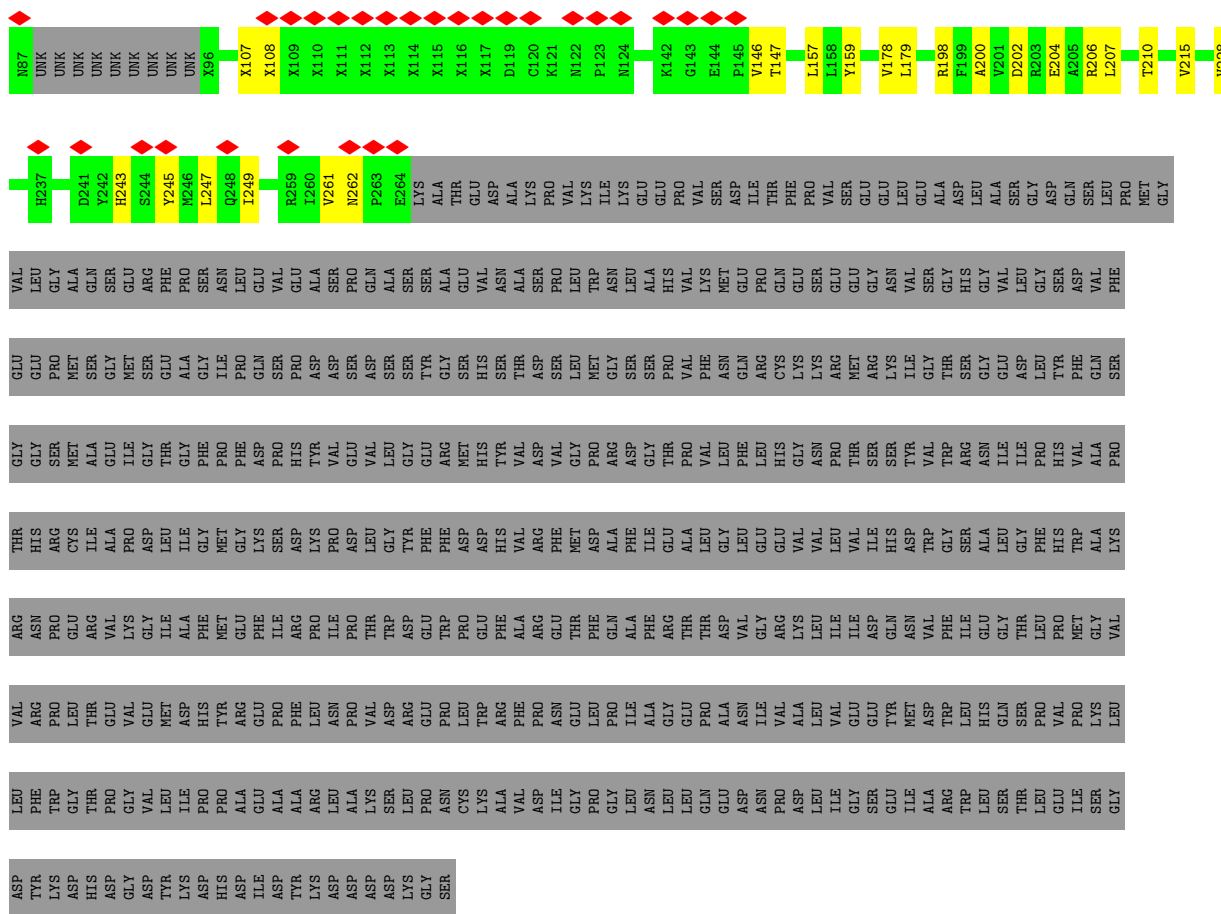
Mol	Chain	Residues	Atoms						AltConf	Trace
			Total	C	H	N	O	S		
11	N	53	864	280	419	87	74	4	1	0

- Molecule 12 is a protein called Splicing factor 3B subunit 3.

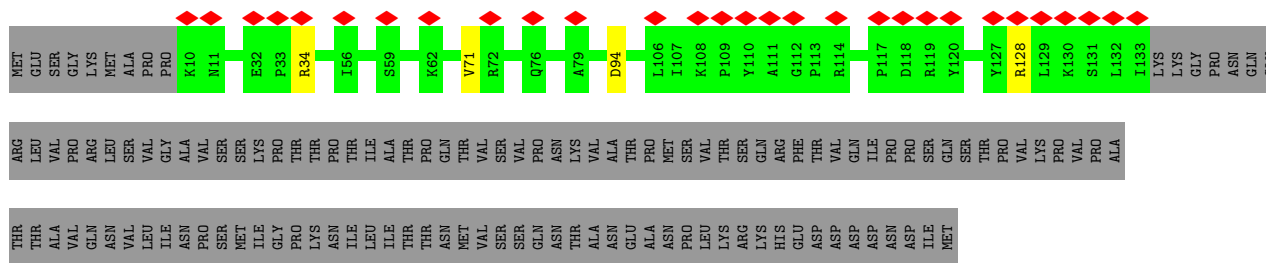
Mol	Chain	Residues	Atoms					AltConf	Trace	
			Total	C	H	N	O			S
12	S	1177	18381	5858	9154	1568	1756	45	0	0

- Molecule 13 is a protein called Splicing factor 3B subunit 5.

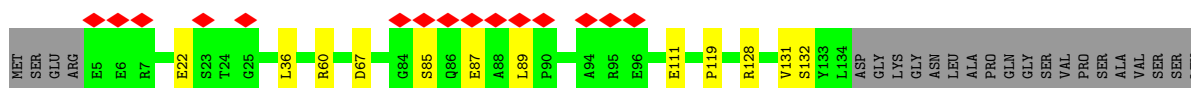
Mol	Chain	Residues	Atoms					AltConf	Trace	
			Total	C	H	N	O			S
13	T	66	1049	343	509	94	98	5	0	0

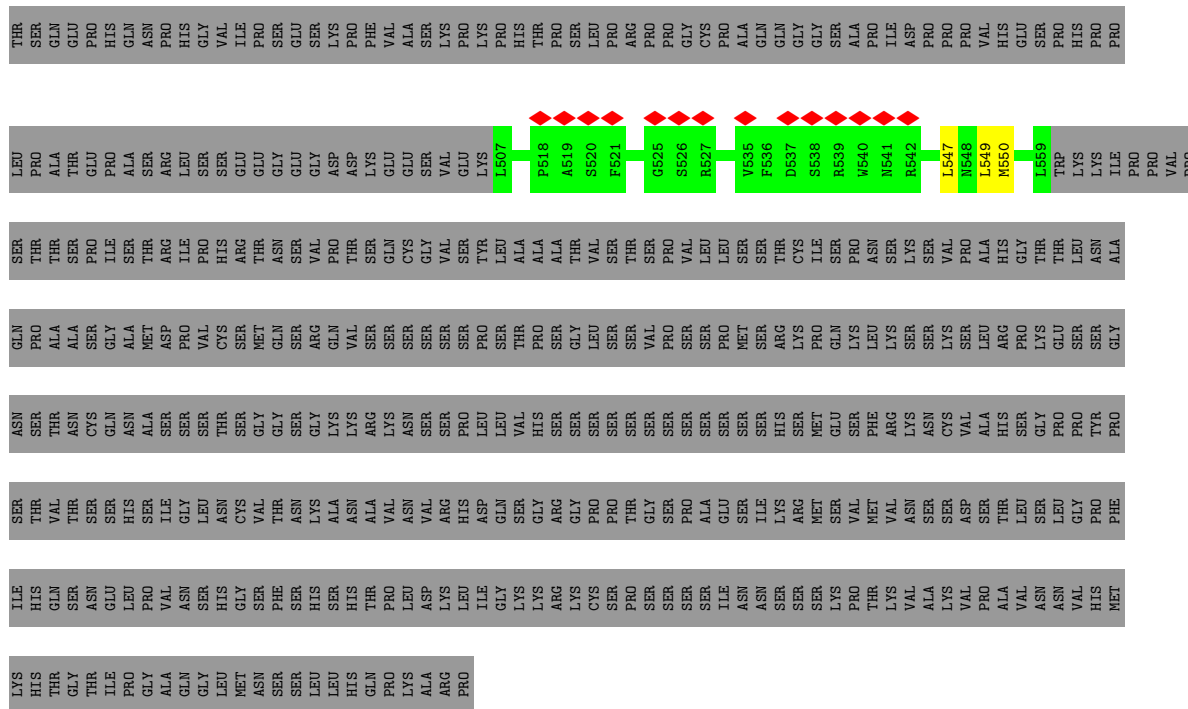


• Molecule 5: Transcription initiation factor TFIID subunit 9B

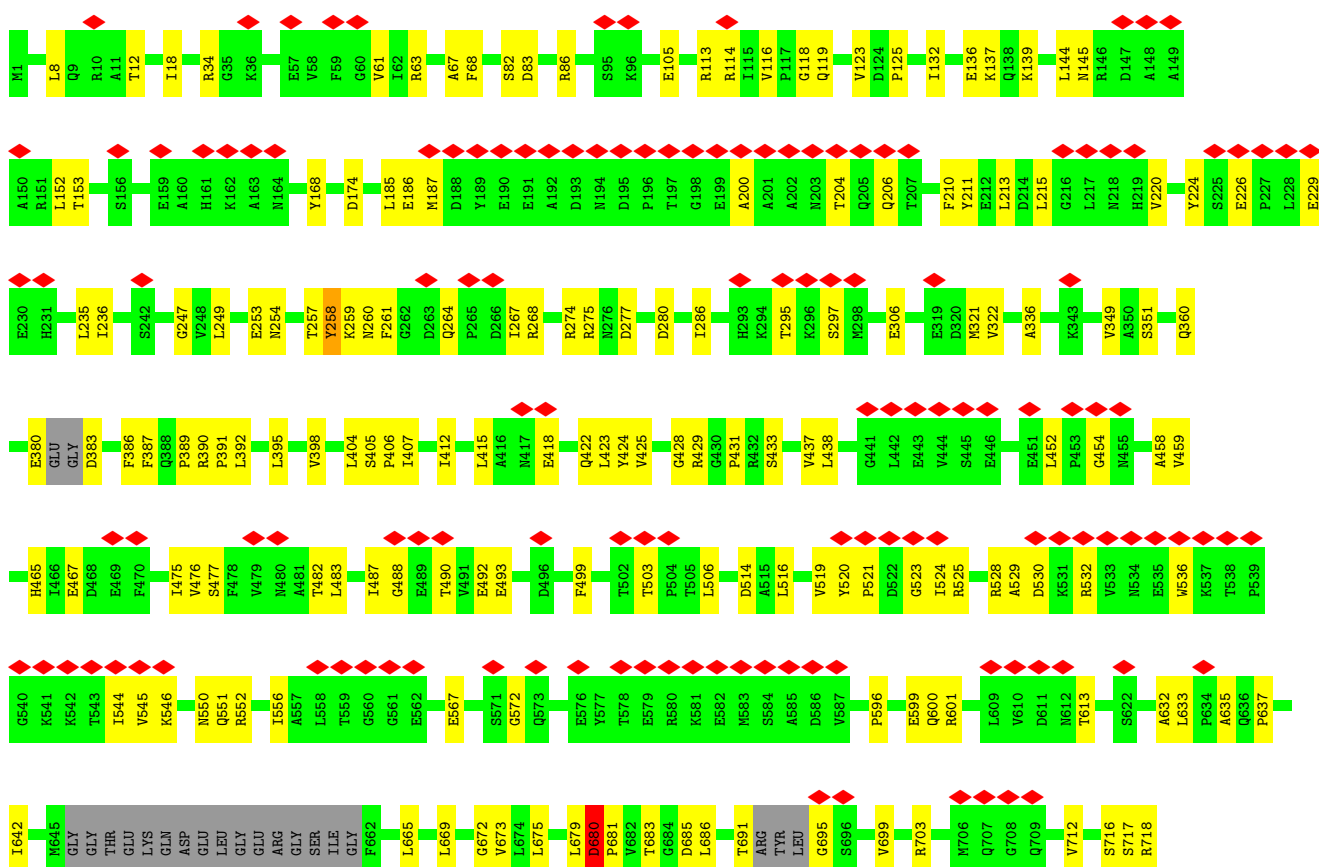


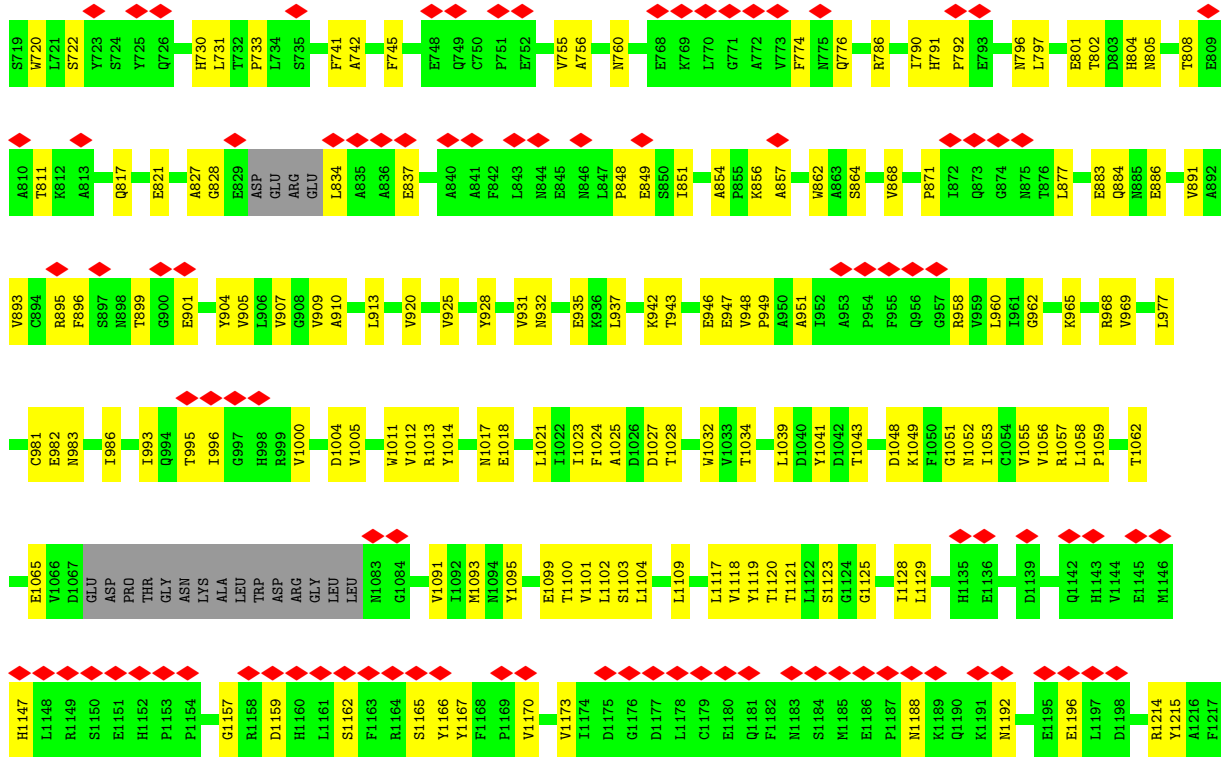
• Molecule 6: TAF6-like RNA polymerase II p300/CBP-associated factor-associated factor 65 kDa subunit 6L, TAF6-like RNA polymerase II p300/CBP-associated factor-associated factor 65 kDa subunit 6L, TAF6-like RNA polymerase II p300/CBP-associated factor-associated factor 65 kDa subunit 6L



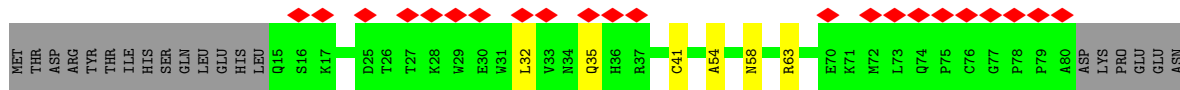


• Molecule 12: Splicing factor 3B subunit 3





• Molecule 13: Splicing factor 3B subunit 5



4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	3157	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TECNAI 20	Depositor
Voltage (kV)	120	Depositor
Electron dose ($e^-/\text{\AA}^2$)	35	Depositor
Minimum defocus (nm)	400	Depositor
Maximum defocus (nm)	3900	Depositor
Magnification	80000	Depositor
Image detector	GATAN ULTRASCAN 4000 (4k x 4k)	Depositor
Maximum map value	0.066	Depositor
Minimum map value	-0.028	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.006	Depositor
Recommended contour level	0.0302	Depositor
Map size (Å)	420.0, 420.0, 420.0	wwPDB
Map dimensions	300, 300, 300	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.4, 1.4, 1.4	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	0.28	0/22230	0.41	0/30095
2	B	0.28	0/4315	0.43	0/5856
3	C	0.28	0/3355	0.40	0/4527
4	D	0.29	0/1513	0.41	0/2050
5	E	0.29	0/1018	0.40	0/1384
6	F	0.28	0/1055	0.43	0/1424
7	G	0.26	0/857	0.40	0/1150
8	H	0.30	0/800	0.41	0/1082
9	I	0.27	0/1879	0.39	0/2523
10	J	0.28	0/1745	0.40	0/2373
11	N	0.29	0/459	0.39	0/619
12	S	0.25	0/9415	0.46	0/12775
13	T	0.24	0/556	0.42	0/751
All	All	0.28	0/49197	0.42	0/66609

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	F	0	4

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
6	F	193	UNK	Peptide
6	F	199	UNK	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
6	F	226	UNK	Peptide
6	F	247	UNK	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	23394	22462	22469	152	0
2	B	4225	4113	4113	31	0
3	C	3293	3316	3313	23	0
4	D	1664	1485	1488	20	0
5	E	995	1018	1018	5	0
6	F	2144	1254	1255	12	0
7	G	845	859	859	2	0
8	H	784	782	781	6	0
9	I	1912	1911	1910	20	0
10	J	1705	1675	1675	19	0
11	N	445	419	418	3	0
12	S	9227	9154	9154	196	0
13	T	540	509	509	5	0
All	All	51173	48957	48962	460	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.

The worst 5 of 460 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)
6:F:317:UNK:O	6:F:318:UNK:CB	2.26	0.83
1:A:2920:ILE:HD11	1:A:2941:VAL:HG11	1.61	0.82
2:B:144:GLN:N	2:B:148:ASP:OD2	2.14	0.80
6:F:22:GLU:OE2	7:G:145:THR:OG1	1.99	0.80
10:J:135:ASP:OD2	10:J:137:LYS:NZ	2.16	0.78

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	2703/3848 (70%)	2570 (95%)	131 (5%)	2 (0%)	51	86
2	B	530/589 (90%)	494 (93%)	36 (7%)	0	100	100
3	C	402/811 (50%)	371 (92%)	31 (8%)	0	100	100
4	D	178/749 (24%)	171 (96%)	7 (4%)	0	100	100
5	E	122/251 (49%)	115 (94%)	7 (6%)	0	100	100
6	F	128/622 (21%)	124 (97%)	4 (3%)	0	100	100
7	G	101/161 (63%)	97 (96%)	4 (4%)	0	100	100
8	H	95/218 (44%)	93 (98%)	2 (2%)	0	100	100
9	I	223/317 (70%)	211 (95%)	12 (5%)	0	100	100
10	J	210/335 (63%)	196 (93%)	14 (7%)	0	100	100
11	N	52/892 (6%)	50 (96%)	2 (4%)	0	100	100
12	S	1165/1217 (96%)	1098 (94%)	66 (6%)	1 (0%)	51	86
13	T	64/86 (74%)	59 (92%)	5 (8%)	0	100	100
All	All	5973/10096 (59%)	5649 (95%)	321 (5%)	3 (0%)	54	86

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
12	S	680	ASP
1	A	3719	THR
1	A	1117	LYS

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	2409/2847 (85%)	2408 (100%)	1 (0%)	100	100
2	B	472/521 (91%)	472 (100%)	0	100	100
3	C	375/716 (52%)	373 (100%)	2 (0%)	88	93
4	D	164/610 (27%)	164 (100%)	0	100	100
5	E	110/224 (49%)	110 (100%)	0	100	100
6	F	112/315 (36%)	112 (100%)	0	100	100
7	G	94/141 (67%)	93 (99%)	1 (1%)	73	84
8	H	86/154 (56%)	86 (100%)	0	100	100
9	I	202/263 (77%)	201 (100%)	1 (0%)	88	93
10	J	188/287 (66%)	188 (100%)	0	100	100
11	N	47/779 (6%)	47 (100%)	0	100	100
12	S	1020/1051 (97%)	1017 (100%)	3 (0%)	92	95
13	T	57/77 (74%)	57 (100%)	0	100	100
All	All	5336/7985 (67%)	5328 (100%)	8 (0%)	93	97

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
12	S	1166	TYR
12	S	680	ASP
9	I	109	LYS
7	G	151	ARG
12	S	258	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	15
4	D	2

The worst 5 of 17 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	2089:UNK	C	2090:UNK	N	31.46
1	A	2048:UNK	C	2049:UNK	N	19.99
1	A	2122:UNK	C	2123:UNK	N	18.57
1	A	2078:UNK	C	2079:UNK	N	16.12
1	A	357:UNK	C	358:GLU	N	14.95

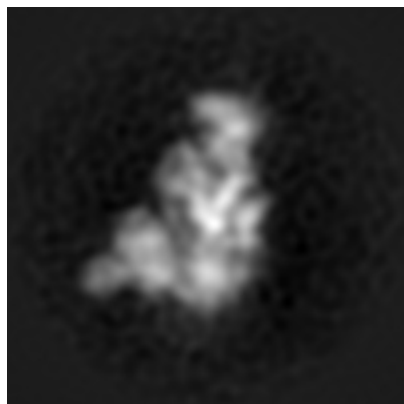
6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-23028. These allow visual inspection of the internal detail of the map and identification of artifacts.

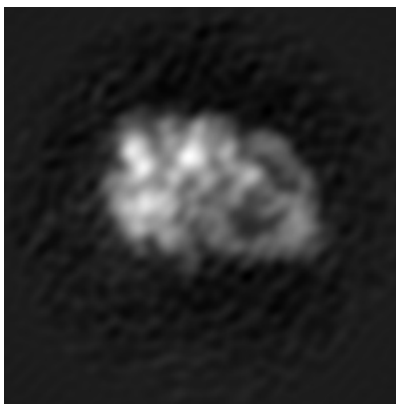
Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

6.1 Orthogonal projections [i](#)

6.1.1 Primary map



X

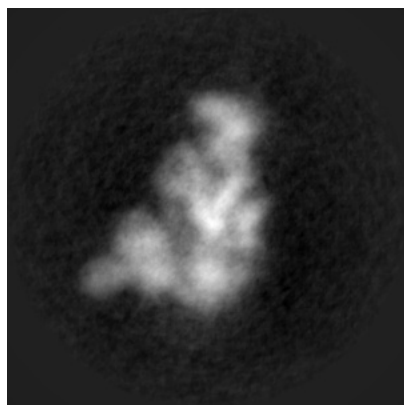


Y

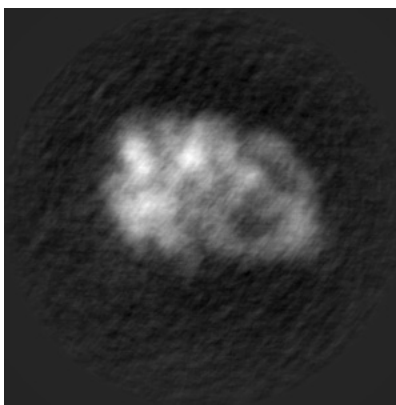


Z

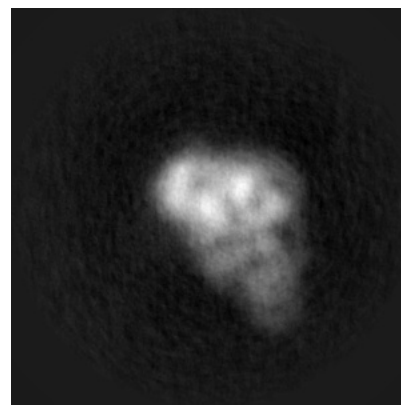
6.1.2 Raw map



X



Y

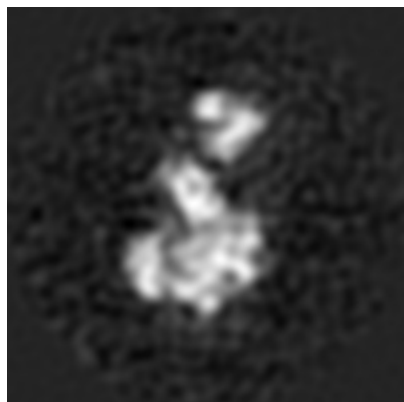


Z

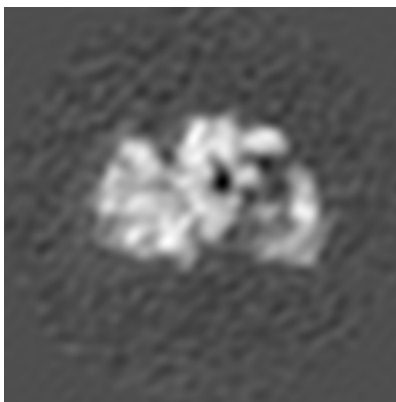
The images above show the map projected in three orthogonal directions.

6.2 Central slices [i](#)

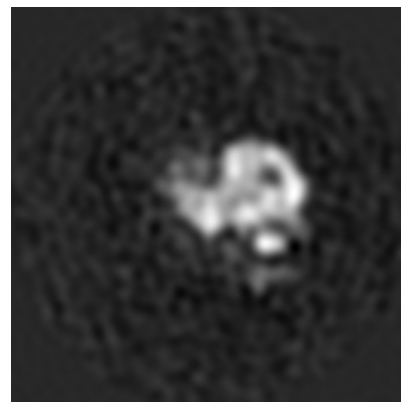
6.2.1 Primary map



X Index: 150

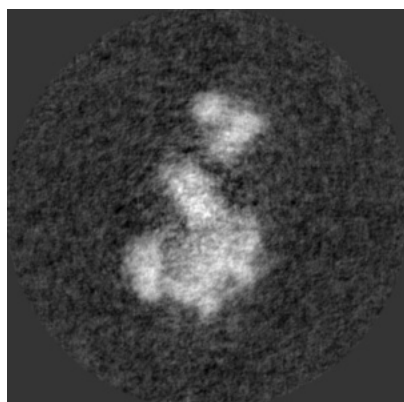


Y Index: 150

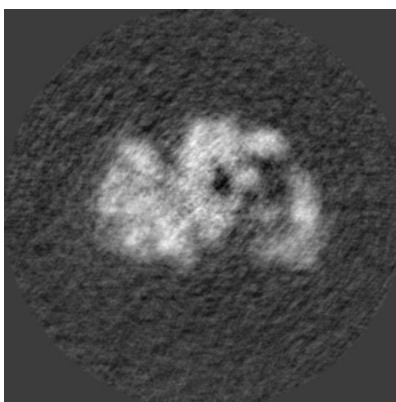


Z Index: 150

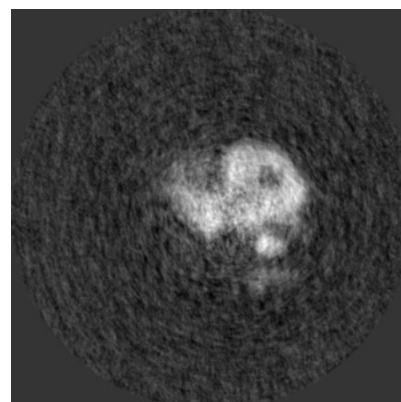
6.2.2 Raw map



X Index: 150



Y Index: 150

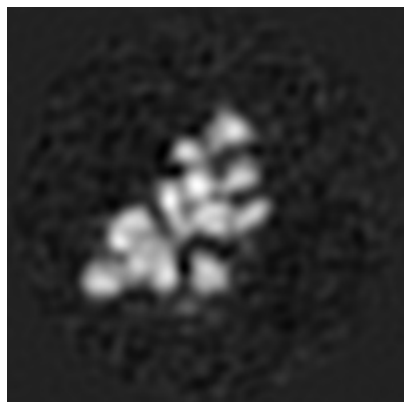


Z Index: 150

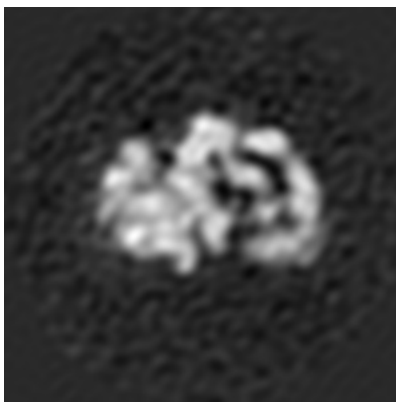
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices [i](#)

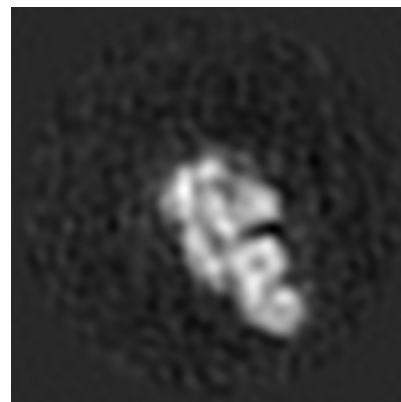
6.3.1 Primary map



X Index: 190

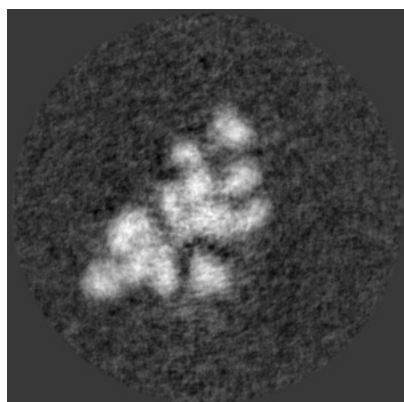


Y Index: 156

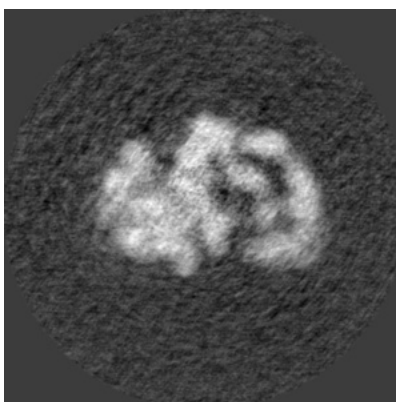


Z Index: 101

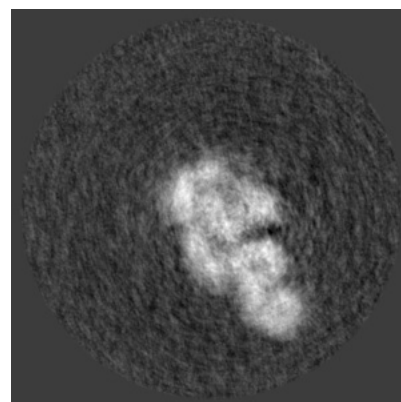
6.3.2 Raw map



X Index: 188



Y Index: 156

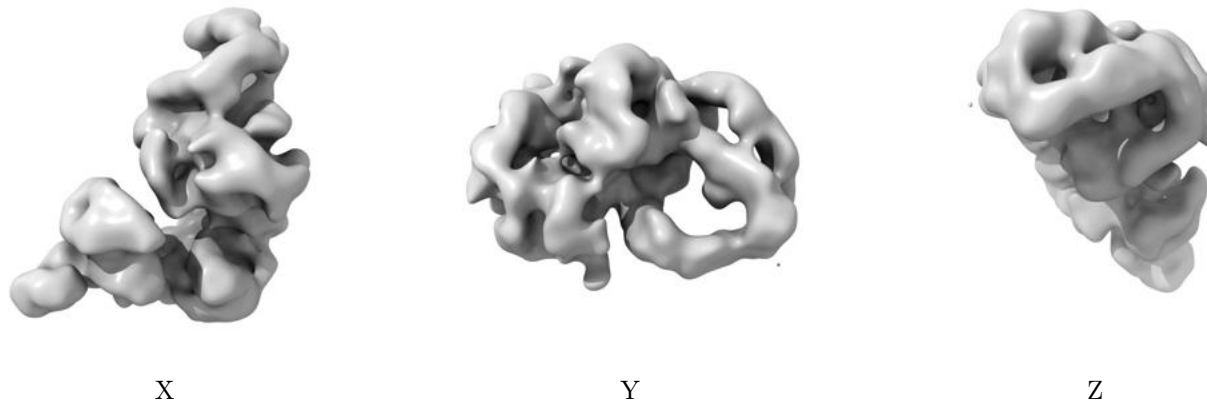


Z Index: 100

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

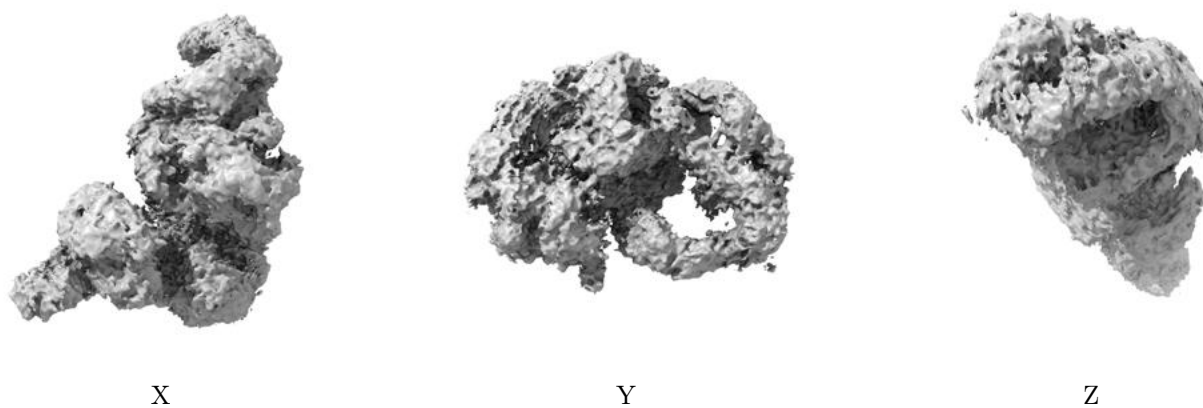
6.4 Orthogonal surface views [i](#)

6.4.1 Primary map



The images above show the 3D surface view of the map at the recommended contour level 0.0302. 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.

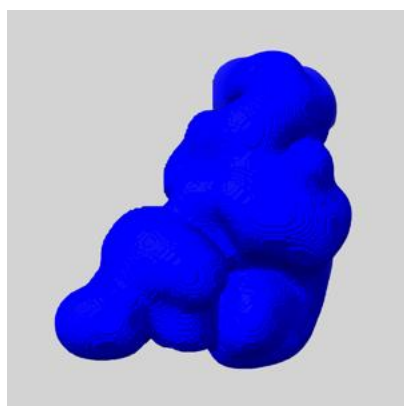
6.5 Mask visualisation [i](#)

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

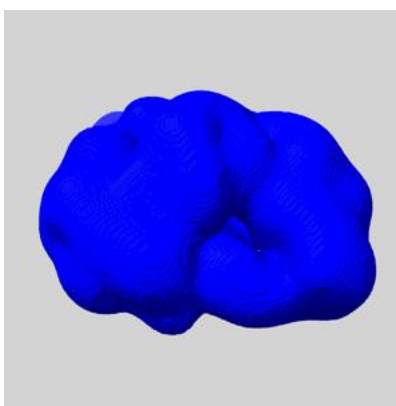
A mask typically either:

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

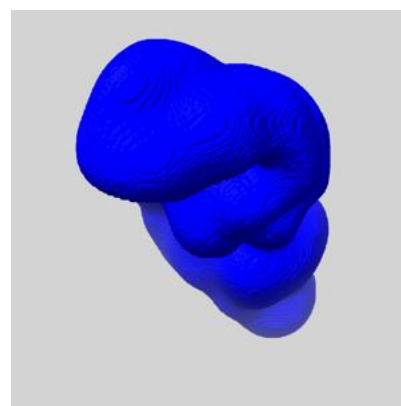
6.5.1 emd_23028_msk_1.map [i](#)



X



Y

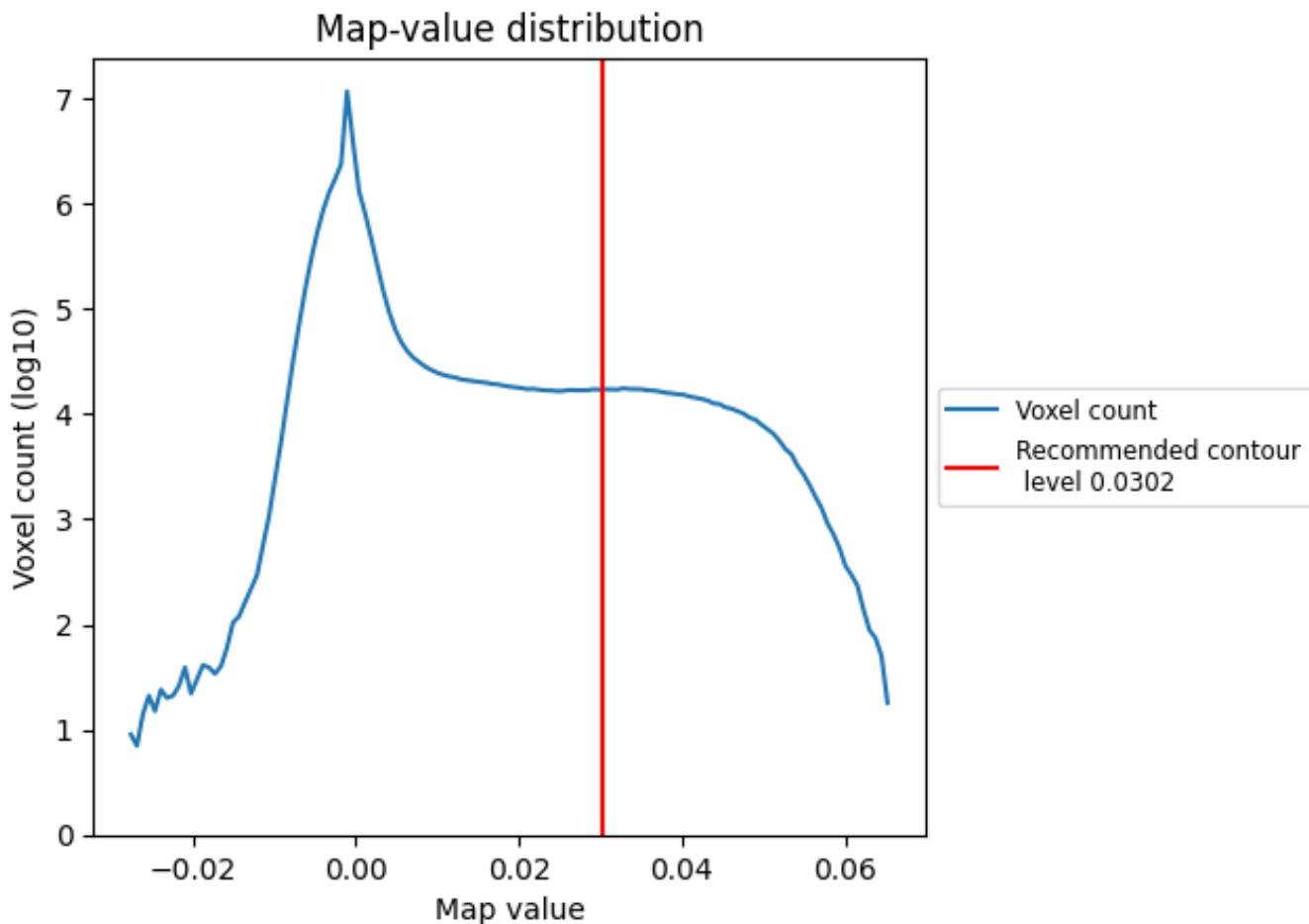


Z

7 Map analysis [i](#)

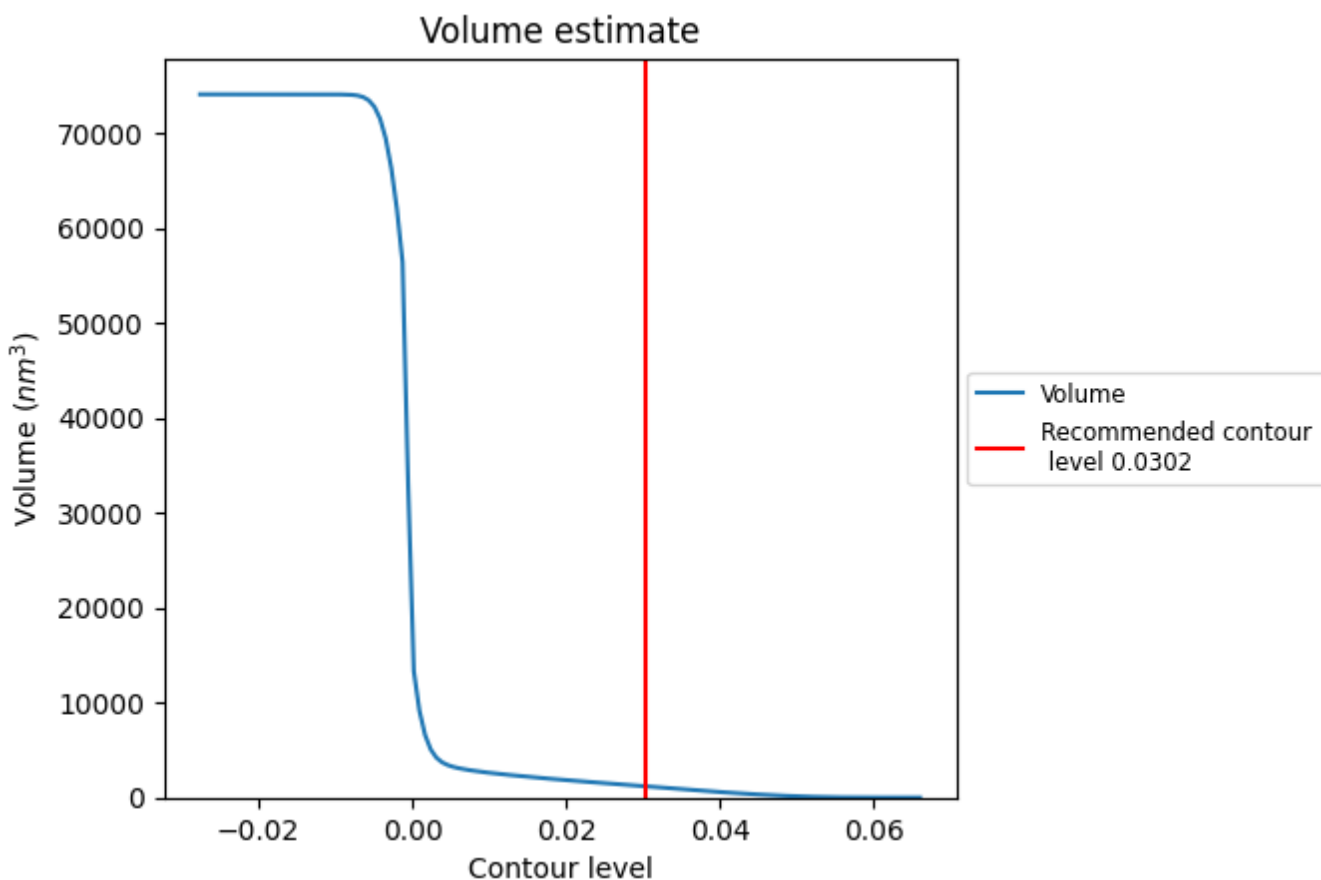
This section contains the results of statistical analysis of the map.

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.

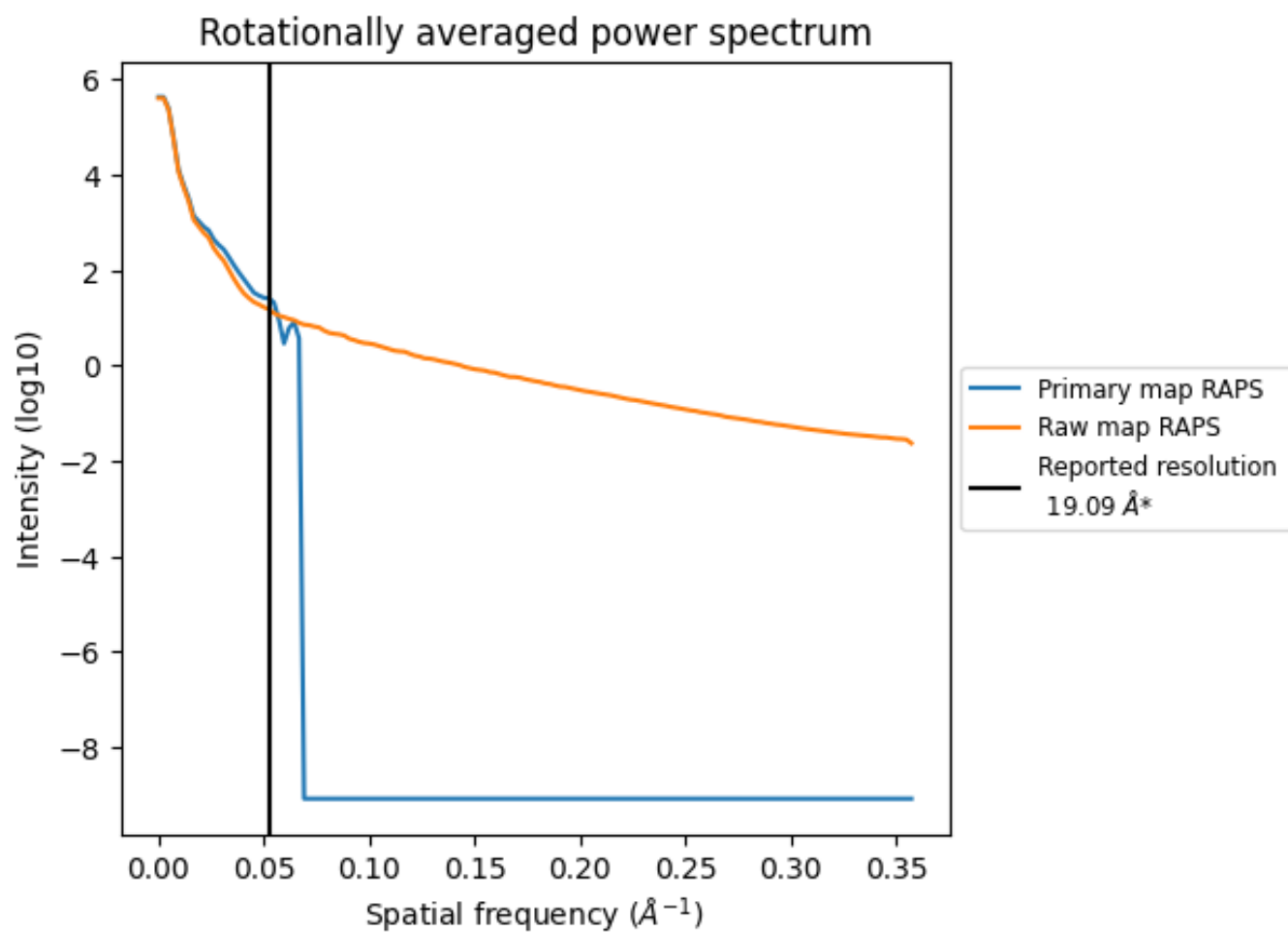
7.2 Volume estimate [\(i\)](#)



The volume at the recommended contour level is 1192 nm³; this corresponds to an approximate mass of 1077 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](#)

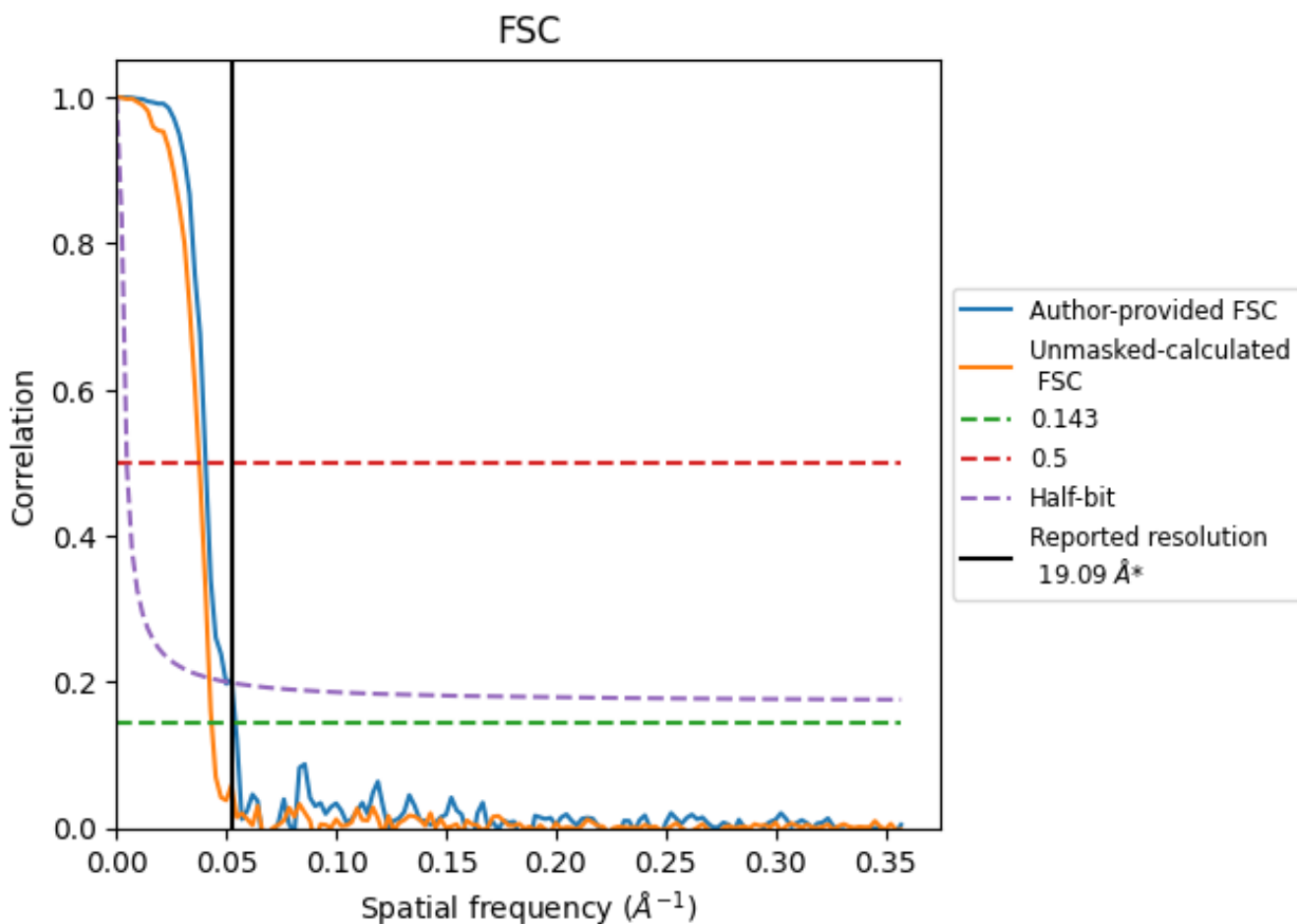


*Reported resolution corresponds to spatial frequency of 0.052 Å⁻¹

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 0.052 Å⁻¹

8.2 Resolution estimates [i](#)

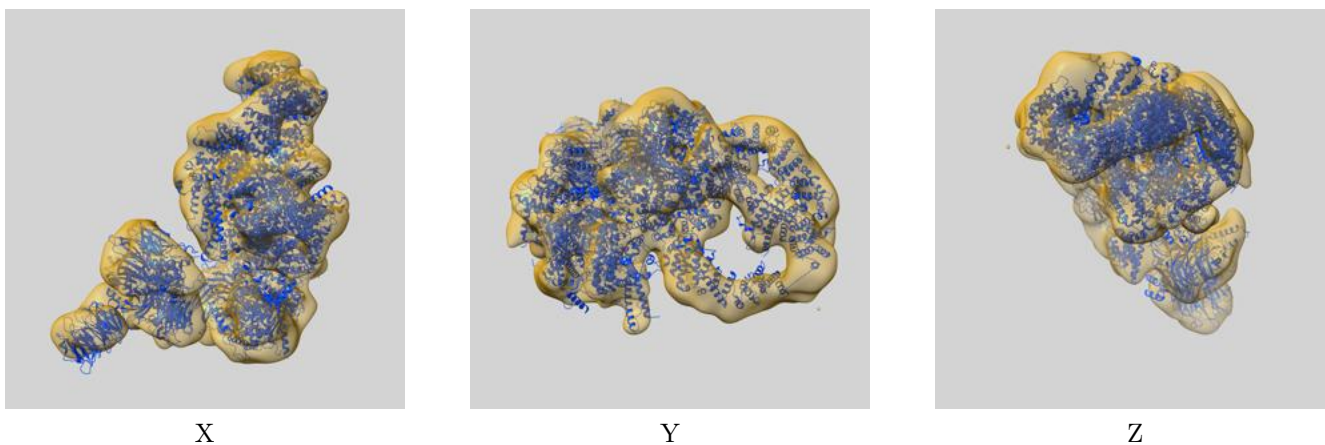
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	19.09	-	-
Author-provided FSC curve	18.48	24.63	20.04
Unmasked-calculated*	23.09	26.53	23.70

*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 23.09 differs from the reported value 19.09 by more than 10 %

9 Map-model fit [i](#)

This section contains information regarding the fit between EMDB map EMD-23028 and PDB model 7KTS. Per-residue inclusion information can be found in section 3 on page 7.

9.1 Map-model overlay [i](#)



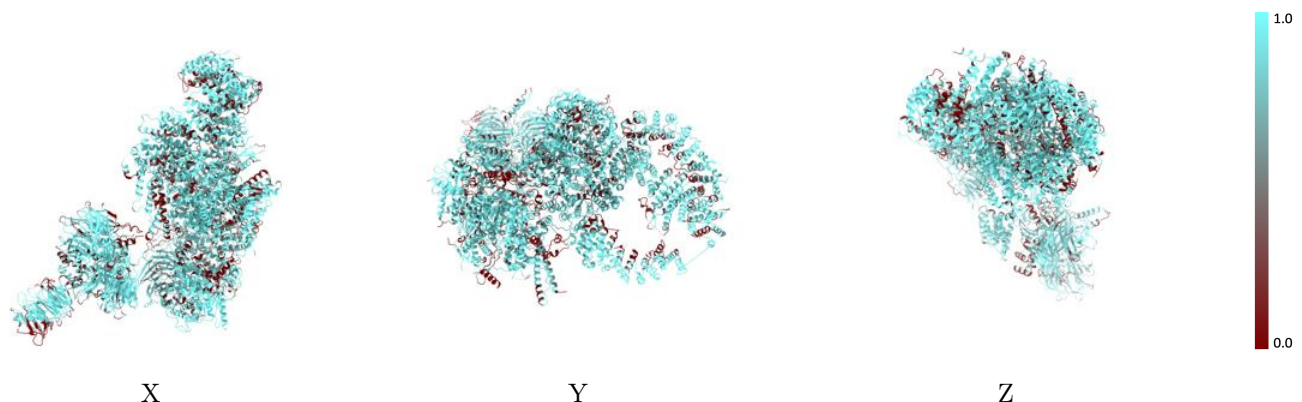
The images above show the 3D surface view of the map at the recommended contour level 0.0302 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

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



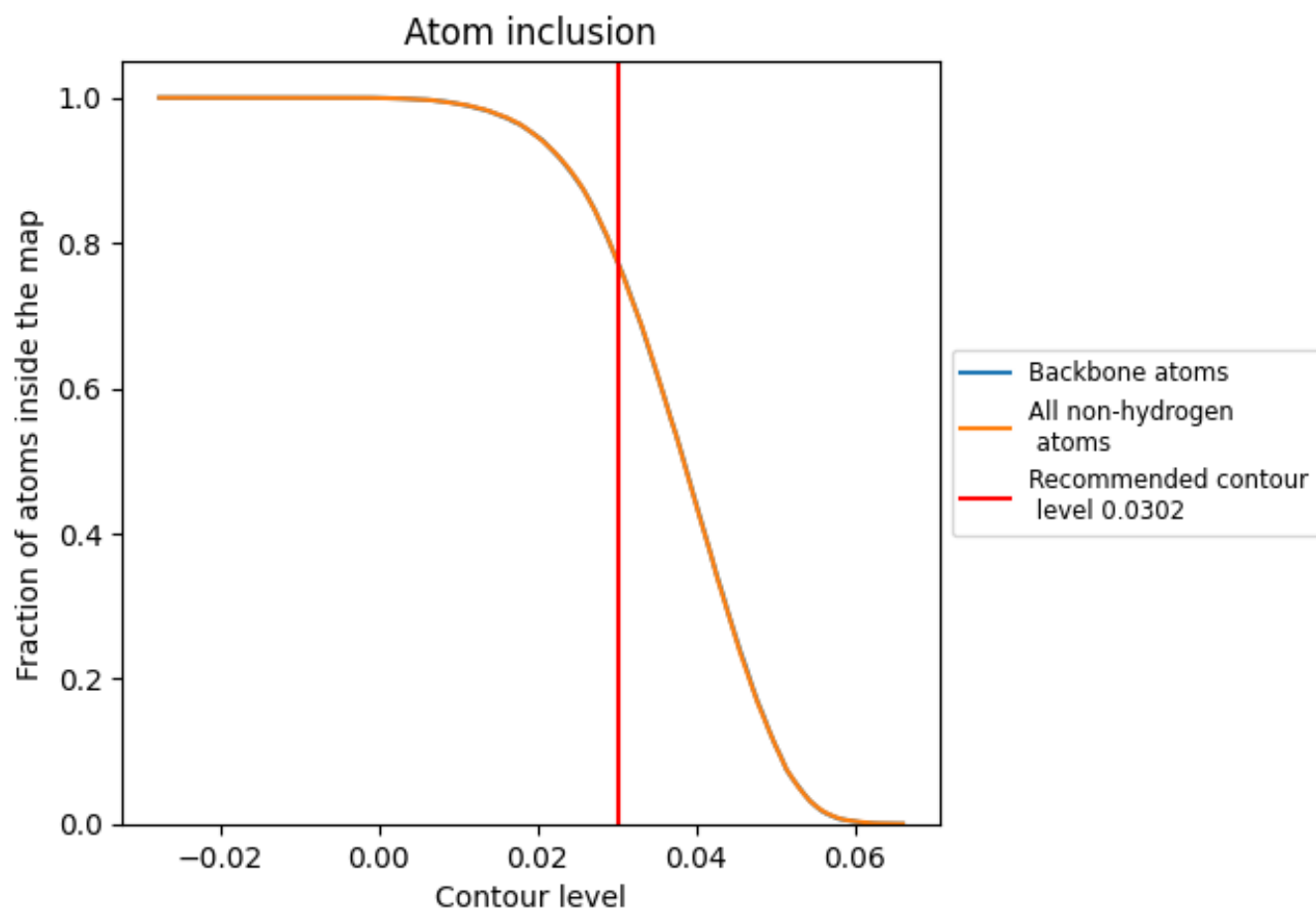
The images above show the model with each residue coloured according to its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions 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 (0.0302).





























9.4 Atom inclusion [i](#)



At the recommended contour level, 77% 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 (0.0302) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	 0.7699	 0.0480
A	 0.7848	 0.0550
B	 0.8248	 0.0470
C	 0.7255	 0.0330
D	 0.7968	 0.0420
E	 0.7523	 0.0200
F	 0.8702	 0.0470
G	 0.6449	 0.0240
H	 0.9260	 0.0550
I	 0.6413	 0.0300
J	 0.6944	 0.0280
N	 0.7102	 0.0180
S	 0.7754	 0.0510
T	 0.6250	 0.0180

