Full wwPDB EM Validation Report (i)

Feb 5, 2024 – 02:25 PM EST

PDB ID	:	7UM1
EMDB ID	:	EMD-24765
Title	:	Structure of bacteriophage AR9 non-virion RNAP polymerase holoenzyme determined by cryo-EM
Authors	:	Leiman, P.G.; Fraser, A.; Sokolova, M.L.
Deposited on	:	2022-04-05
Resolution	:	4.20 Å(reported)
Based on initial model	:	7S01

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

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

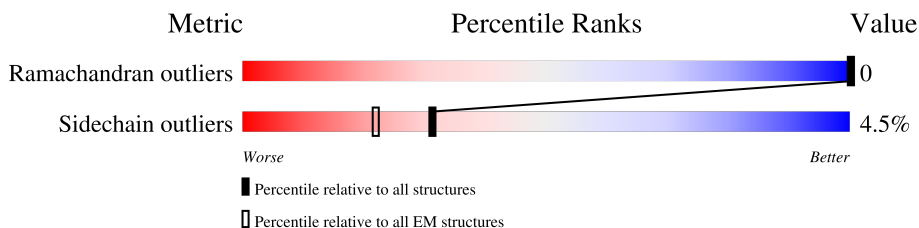|		
EMDB validation analysis	:	0.0.1.dev70
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.9
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 4.20 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	Percentile Ranks	Value
Ramachandran outliers		0
Sidechain outliers		4.5%

Worse   *Better*

■ Percentile relative to all structures
❑ Percentile relative to all EM structures

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

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the EM map (all-atom inclusion < 40%). The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	464	16% / 34% · 63%
2	d	448	16% / 91% · 5%
3	c	496	26% / 93% · ·
4	D	631	· / 78% · 18%
5	C	665	· / 94% · ·

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 18356 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called DNA-directed RNA polymerase subunit.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	173	Total 1421	C 909	N 238	O 264	S 10	0	0

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

Mol	Chain	Residues	Atoms					AltConf	Trace
2	d	426	Total 3488	C 2256	N 561	O 663	S 8	0	0

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
d	-21	MET	-	expression tag	UNP A0A172JIH0
d	-20	GLY	-	expression tag	UNP A0A172JIH0
d	-19	SER	-	expression tag	UNP A0A172JIH0
d	-18	SER	-	expression tag	UNP A0A172JIH0
d	-17	HIS	-	expression tag	UNP A0A172JIH0
d	-16	HIS	-	expression tag	UNP A0A172JIH0
d	-15	HIS	-	expression tag	UNP A0A172JIH0
d	-14	HIS	-	expression tag	UNP A0A172JIH0
d	-13	HIS	-	expression tag	UNP A0A172JIH0
d	-12	HIS	-	expression tag	UNP A0A172JIH0
d	-11	SER	-	expression tag	UNP A0A172JIH0
d	-10	SER	-	expression tag	UNP A0A172JIH0
d	-9	GLY	-	expression tag	UNP A0A172JIH0
d	-8	GLU	-	expression tag	UNP A0A172JIH0
d	-7	ASN	-	expression tag	UNP A0A172JIH0
d	-6	LEU	-	expression tag	UNP A0A172JIH0
d	-5	TYR	-	expression tag	UNP A0A172JIH0
d	-4	PHE	-	expression tag	UNP A0A172JIH0
d	-3	GLN	-	expression tag	UNP A0A172JIH0
d	-2	GLY	-	expression tag	UNP A0A172JIH0
d	-1	HIS	-	expression tag	UNP A0A172JIH0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
d	0	HIS	-	expression tag	UNP A0A172JIH0

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

Mol	Chain	Residues	Atoms					AltConf	Trace
3	c	484	Total	C	N	O	S	0	0
			4003	2580	658	754	11		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
4	D	515	Total	C	N	O	S	0	0
			4172	2659	685	813	15		

- Molecule 5 is a protein called DNA-directed RNA polymerase.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	C	649	Total	C	N	O	S	0	0
			5271	3337	868	1041	25		

- Molecule 6 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		AltConf
6	D	1	Total	Zn	0
			1	1	

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and atom inclusion in map density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red diamond above a residue indicates a poor fit to the EM map for this residue (all-atom inclusion < 40%). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: DNA-directed RNA polymerase subunit

• Molecule 2: DNA-directed RNA polymerase beta' subunit

F176 Y182 Y183 N184 I187 G196 V197 V198 S199 D200 D201 S202 I203 D204 L205 L206 L207 L208 L211 Y212 N213 M214 Y217 A218 N219 R231 R238 Y269 R331 D374 D410 Q414 N426

• Molecule 3: DNA-directed RNA polymerase beta subunit

Chain c:

26%

93%

M1 E42 I43 I44 P45 E46 I47 N64 K65 E66 F84 L85 I86 E87 K88 D89 D90 K91 K92 E93 E94 Y109 A134 I144 V145 E148 K149 K150 E151 T152 F153 D154 D155 I156 N157 G158 E159 T160 H161 T162 M163 L164 N165 V166 D167 L168 D169 L170 F171 K172 V175 P176

F177 L178 F181 F185 E188 E192 Y193 L196 Q197 D198 L199 I200 H201 V202 L203 M204 K205 E206 D207 L208 D209 Q210 L211 D212 E213 D214 E215 I216 N217 D218 N219 V220 I221 F222 M223 I224 T225 K226 L230 V231 V232 D233 F236 F237 S238 N239 K240 N241 N242 Q243 L249 L250

N251 C252 F253 N254 T255 R256 I257 K258 I259 Y263 E264 K265 L272 G273 G274 Y275 F276 T277 S281 N282 K283 Q284 E285 K286 G287 E288 G289 I290 I291 F294 E295 R296 K313 Y317 M323 F357 S358 V364 L365 N366 N367 R368 S370 N371 K372 F373 E374 K375 I376

I379 N395 F410 Q418 S426 S427 N428 S429 T430 N431 L447 G453 N482 L483 N484 ILE ASP ASP ILE SER ILE ASP GLU VAL ILE GLU SER

• Molecule 4: DNA-directed RNA polymerase

Chain D:

78%

18%

M1 L12 K20 D56 L57 V58 V59 D60 Y68 G69 N70 H112 N138 S146 K158 N186 M200 Y204 G210 L222 F242 N254 R270 L327 Y361 L389 L390 S391 A392 K393 H394 LEU LEU GLU THR ARG SER SER LYS ILE ASP

TRP GLY THR ASN PHE GLU GLU ASN PHE ILE VAL ASN ARG ASN LEU ILE TYR PRO LYS VAL TYR ASN GLY THR VAL ILE ILE LYS GLU ASP ASP PHE LYS GLU ASP GLU MET GLN PHE GLU GLU ILE PRO LEU ASN LYS LEU ASP GLU GLY ASP SER PHE ALA THR PHE ILE MET ASP ASN ASN

MET ARG LEU PHE LEU ASN LYS ASP LEU LYS LYS GLN LEU ASP GLU SER PHE TYR ASN ILE GLU GLU MET GLN GLU GLU ILE PRO LEU ASN LYS LEU ASP GLY SER PHE ALA THR PHE ILE MET ASP ASN ASN E511 K514 N526 K527 Y528 I529 M565 D568 N577

M580 Y583 N592 G595 L598 S599 R600 F604 T612 F619 N620 K621 S624 D628 L631

• Molecule 5: DNA-directed RNA polymerase

Chain C:

94%

M1 D2 D3 A20 N63 D90 C113 W124 D136 F167 N171 L174 Y239 Y244 E245 L246 D267 S272 K313 D319 L322 H323 M330 F352 D382 N482 Y528 H534 S546 F547 M548 N559 Y567 S583

M592 R606 E607 L608 I609 M610 L644 S645 I646 V649 GLU GLU GLU TRP GLU ASN LYS LEU ASN GLY LYS VAL GLU ASP GLU LYS

4   Experimental information   ⓘ

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	104471	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 $(e^-/\text{Å}^2)$	43.2	Depositor
Minimum defocus (nm)	1200	Depositor
Maximum defocus (nm)	2400	Depositor
Magnification	Not provided	
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	22.295	Depositor
Minimum map value	-17.974	Depositor
Average map value	0.000	Depositor
Map value standard deviation	1.000	Depositor
Recommended contour level	4.0	Depositor
Map size (Å)	216.00002, 216.00002, 216.00002	wwPDB
Map dimensions	200, 200, 200	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.08, 1.08, 1.08	Depositor

5 Model quality ⓘ

5.1 Standard geometry ⓘ

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

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	0.30	0/1440	0.66	5/1929 (0.3%)
2	d	0.33	0/3558	0.52	0/4804
3	c	0.30	0/4082	0.54	2/5502 (0.0%)
4	D	0.29	0/4234	0.54	1/5705 (0.0%)
5	C	0.33	0/5370	0.53	1/7241 (0.0%)
All	All	0.31	0/18684	0.54	9/25181 (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
2	d	0	1
3	c	0	1
All	All	0	2

There are no bond length outliers.

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	386	PRO	CA-N-CD	-7.84	100.52	111.50
1	A	461	LEU	CA-CB-CG	6.85	131.06	115.30
3	c	483	LEU	CA-CB-CG	6.50	130.24	115.30
5	C	174	LEU	CA-CB-CG	6.24	129.64	115.30
4	D	12	LEU	CA-CB-CG	6.12	129.38	115.30
1	A	361	LEU	CA-CB-CG	5.94	128.96	115.30
1	A	453	LEU	CA-CB-CG	5.31	127.52	115.30
1	A	364	MET	CA-CB-CG	5.20	122.14	113.30
3	c	223	MET	CA-CB-CG	5.20	122.14	113.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	c	162	THR	Peptide
2	d	41	ILE	Peptide

5.2 Too-close contacts ⓘ

Due to software issues we are unable to calculate clashes - this section is therefore empty.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	171/464 (37%)	156 (91%)	15 (9%)	0	100	100
2	d	424/448 (95%)	401 (95%)	23 (5%)	0	100	100
3	c	482/496 (97%)	453 (94%)	29 (6%)	0	100	100
4	D	511/631 (81%)	491 (96%)	20 (4%)	0	100	100
5	C	647/665 (97%)	607 (94%)	40 (6%)	0	100	100
All	All	2235/2704 (83%)	2108 (94%)	127 (6%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

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	164/445 (37%)	153 (93%)	11 (7%)	16	43
2	d	390/409 (95%)	371 (95%)	19 (5%)	25	52
3	c	457/469 (97%)	438 (96%)	19 (4%)	30	55
4	D	479/590 (81%)	458 (96%)	21 (4%)	28	54
5	C	593/608 (98%)	569 (96%)	24 (4%)	31	57
All	All	2083/2521 (83%)	1989 (96%)	94 (4%)	31	54

All (94) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	294	ASP
1	A	297	ASN
1	A	316	TYR
1	A	320	ASP
1	A	321	GLU
1	A	338	TYR
1	A	359	TYR
1	A	364	MET
1	A	422	ASP
1	A	439	MET
1	A	460	MET
2	d	14	TYR
2	d	42	PHE
2	d	45	TYR
2	d	65	TYR
2	d	91	GLN
2	d	111	TYR
2	d	117	PHE
2	d	138	TYR
2	d	145	HIS
2	d	176	PHE
2	d	182	TYR
2	d	201	ASP
2	d	211	LEU
2	d	213	ASN
2	d	217	TYR
2	d	238	ARG
2	d	269	TYR
2	d	331	ARG
2	d	410	ASP
3	c	1	MET

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	c	64	ASN
3	c	109	TYR
3	c	150	LYS
3	c	161	HIS
3	c	188	GLU
3	c	223	MET
3	c	226	LYS
3	c	236	PHE
3	c	239	ASN
3	c	243	GLN
3	c	294	PHE
3	c	323	MET
3	c	357	PHE
3	c	373	PHE
3	c	410	PHE
3	c	418	GLN
3	c	428	ASN
3	c	447	LEU
4	D	1	MET
4	D	12	LEU
4	D	70	ASN
4	D	112	HIS
4	D	138	ASN
4	D	146	SER
4	D	158	LYS
4	D	200	MET
4	D	204	TYR
4	D	222	LEU
4	D	242	PHE
4	D	254	ASN
4	D	270	ARG
4	D	361	TYR
4	D	528	TYR
4	D	565	MET
4	D	580	MET
4	D	598	LEU
4	D	600	ARG
4	D	604	PHE
4	D	619	PHE
5	C	63	ASN
5	C	90	ASP
5	C	113	CYS

Continued from previous page...

Mol	Chain	Res	Type
5	C	124	TRP
5	C	167	PHE
5	C	244	TYR
5	C	267	ASP
5	C	272	SER
5	C	319	ASP
5	C	323	HIS
5	C	352	PHE
5	C	382	ASP
5	C	482	ASN
5	C	528	TYR
5	C	546	SER
5	C	548	MET
5	C	559	ASN
5	C	567	TYR
5	C	583	SER
5	C	592	MET
5	C	606	ARG
5	C	608	LEU
5	C	610	MET
5	C	644	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (17) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	419	ASN
1	A	429	GLN
1	A	435	ASN
2	d	213	ASN
2	d	252	ASN
2	d	425	ASN
3	c	165	ASN
3	c	251	ASN
3	c	279	ASN
3	c	312	ASN
3	c	367	ASN
4	D	254	ASN
4	D	553	GLN
5	C	42	HIS
5	C	155	ASN
5	C	412	ASN
5	C	415	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6 Map visualisation ⓘ

This section contains visualisations of the EMDB entry EMD-24765. 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 ⓘ
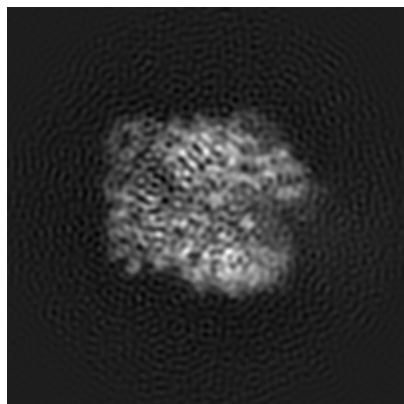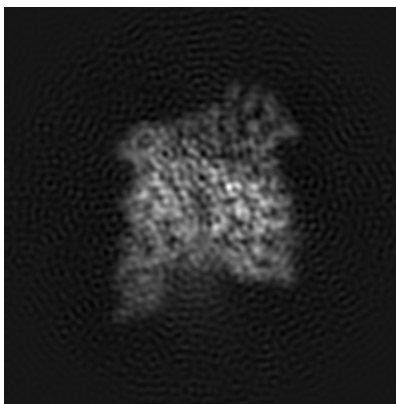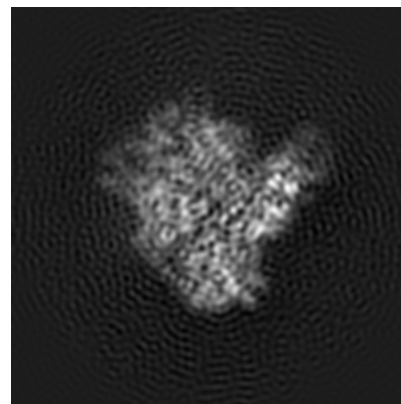
6.1.1 Primary map

X　　　　　　　　　Y　　　　　　　　　Z

6.1.2 Raw map

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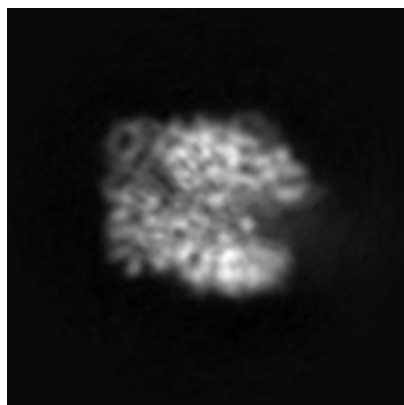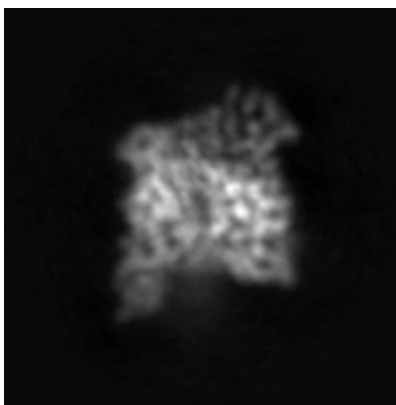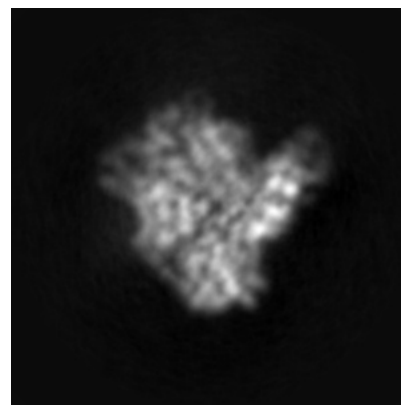
The images above show the map projected in three orthogonal directions.

6.2 Central slices ⓘ
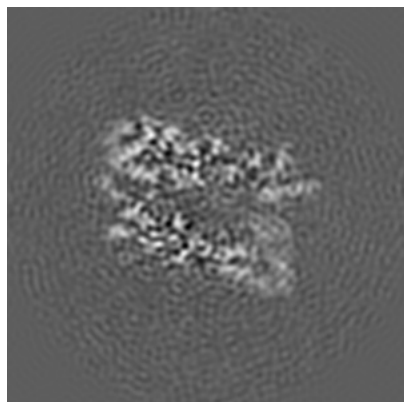
6.2.1 Primary map
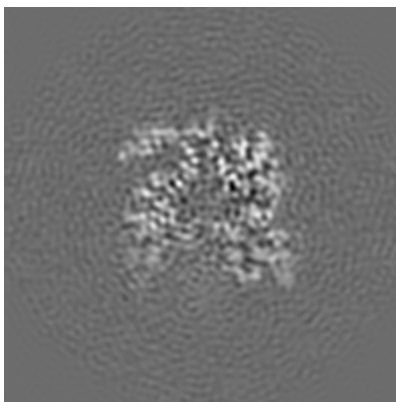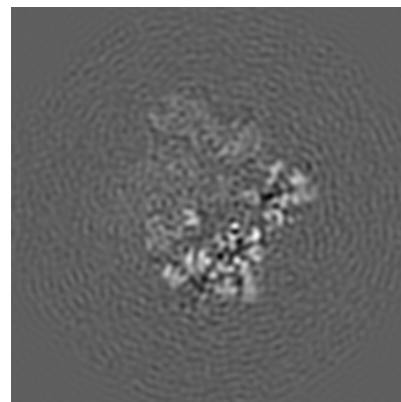
X Index: 100   Y Index: 100   Z Index: 100

6.2.2 Raw map
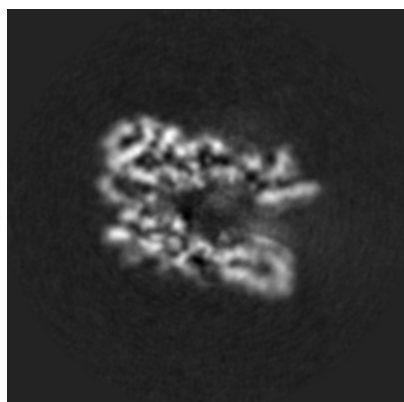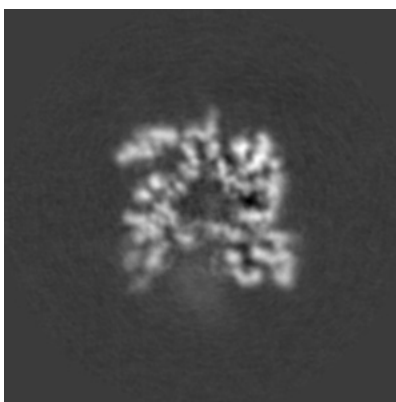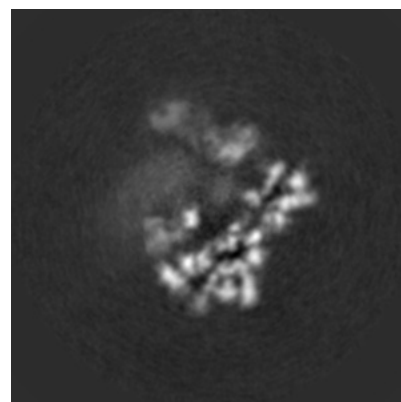
X Index: 100   Y Index: 100   Z Index: 100

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

6.3 Largest variance slices (i)

6.3.1 Primary map
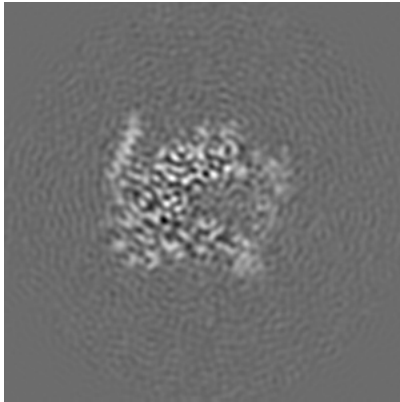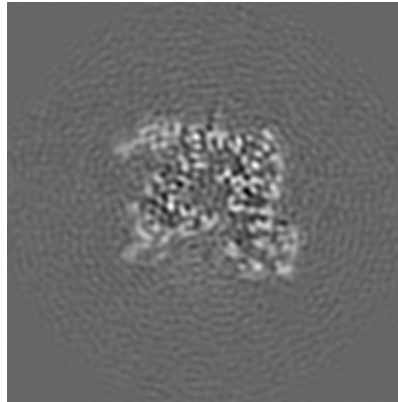
| X Index: 110 | Y Index: 97 | Z Index: 116 |
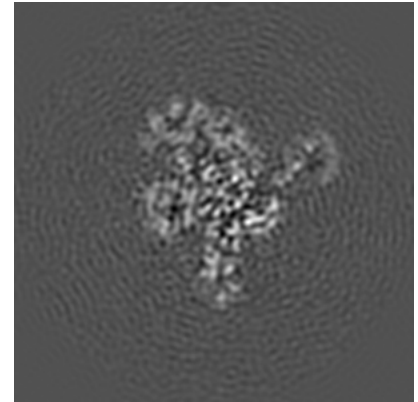
6.3.2 Raw map
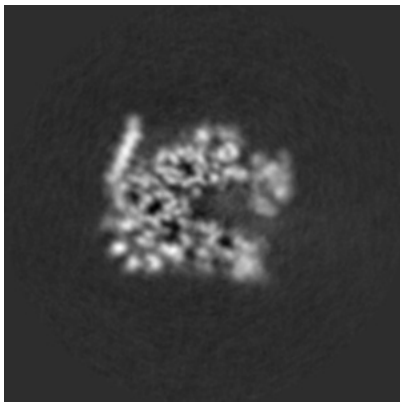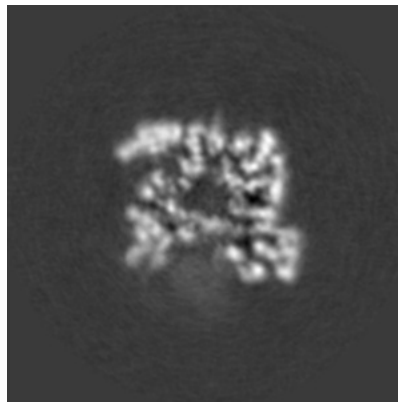
| X Index: 110 | Y Index: 98 | Z Index: 126 |
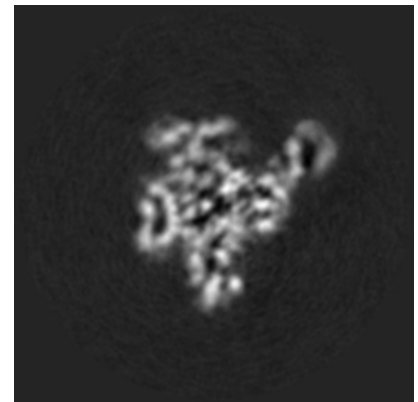
The images above show the largest variance slices of the map in three orthogonal directions.

6.4 Orthogonal standard-deviation projections (False-color) ⓘ

6.4.1 Primary map
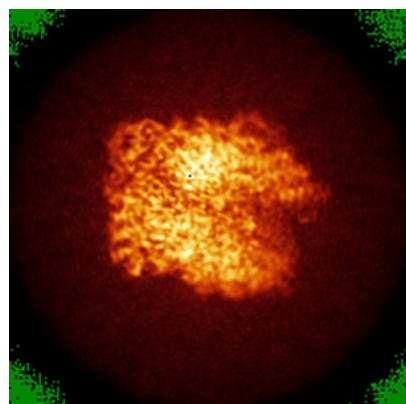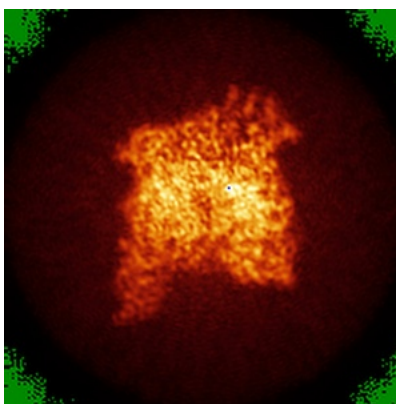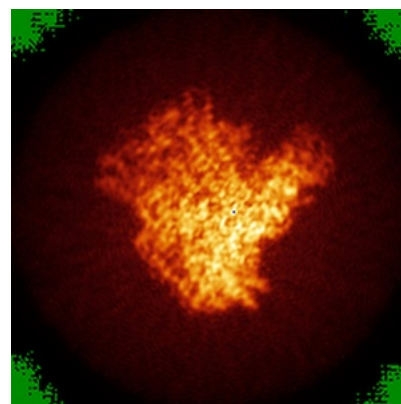
| X | Y | Z |

6.4.2 Raw map

| X | Y | Z |
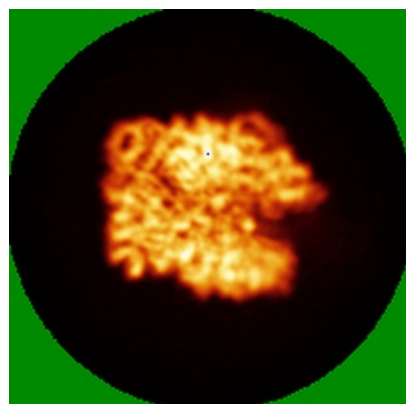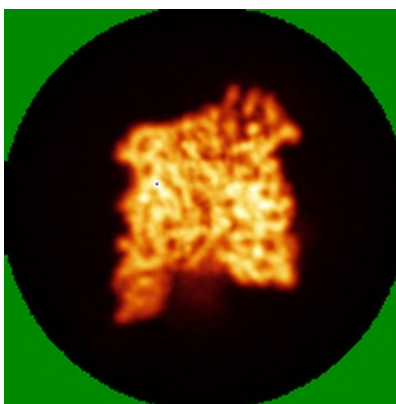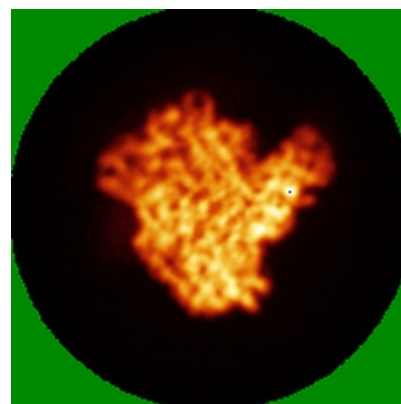
The images above show the map standard deviation projections with false color in three orthogonal directions. Minimum values are shown in green, max in blue, and dark to light orange shades represent small to large values respectively.

6.5 Orthogonal surface views (i)

6.5.1 Primary map

X   Y   Z
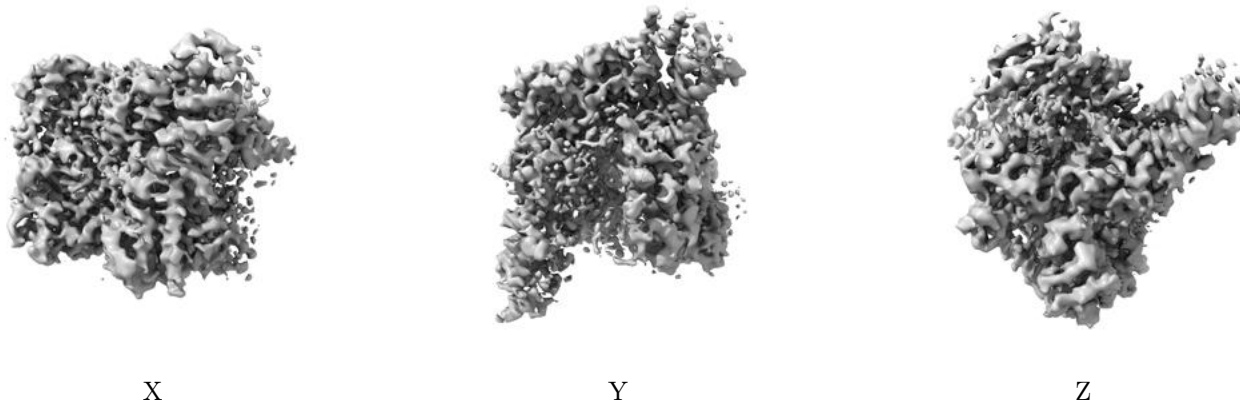
The images above show the 3D surface view of the map at the recommended contour level 4.0. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

6.5.2 Raw map

X   Y   Z
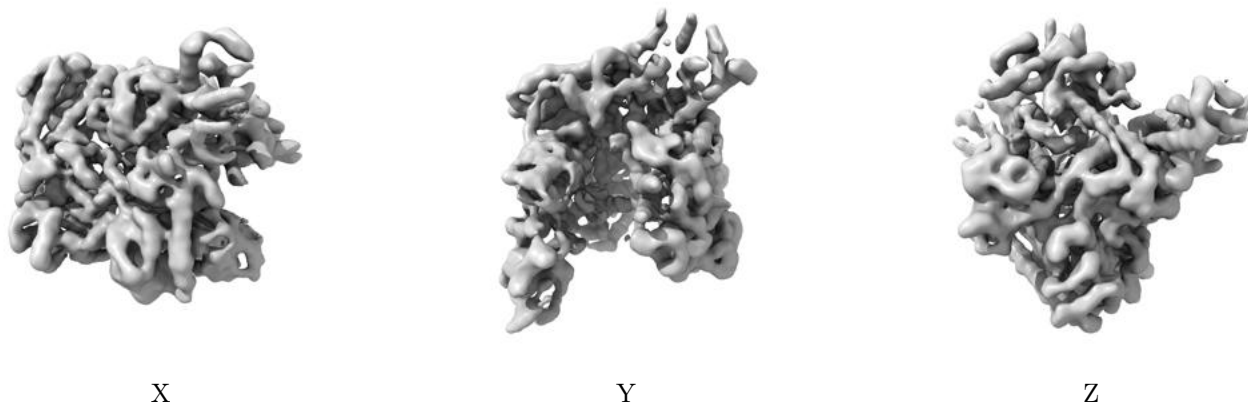
These images show the 3D surface of the raw map. The raw map's contour level was selected so that its surface encloses the same volume as the primary map does at its recommended contour level.

6.6 Mask visualisation (i)

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

7 Map analysis ⓘ

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

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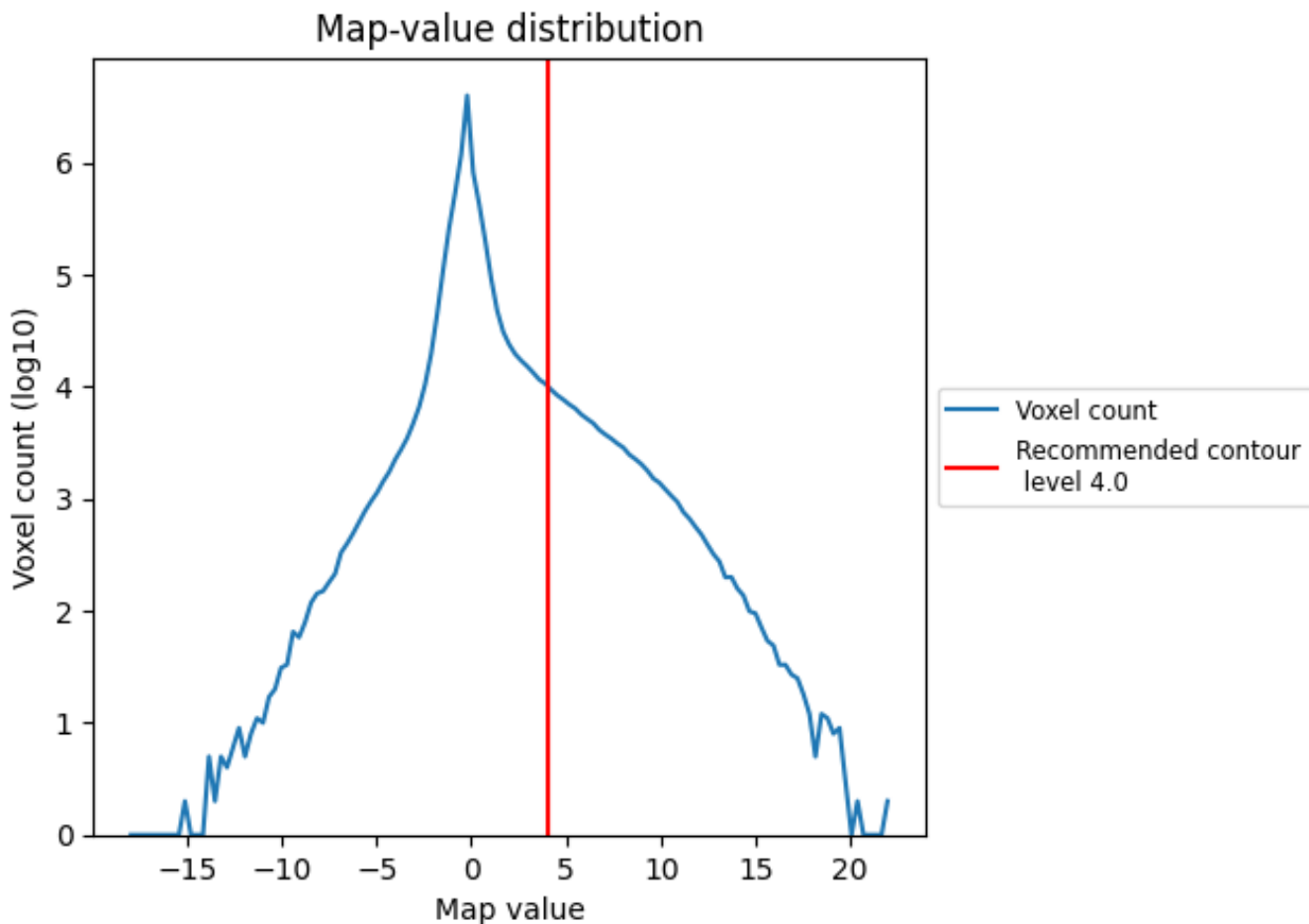
The map-value distribution is plotted in 128 intervals along the x-axis. The y-axis is logarithmic. A spike in this graph at zero usually indicates that the volume has been masked.

7.2   Volume estimate ⓘ
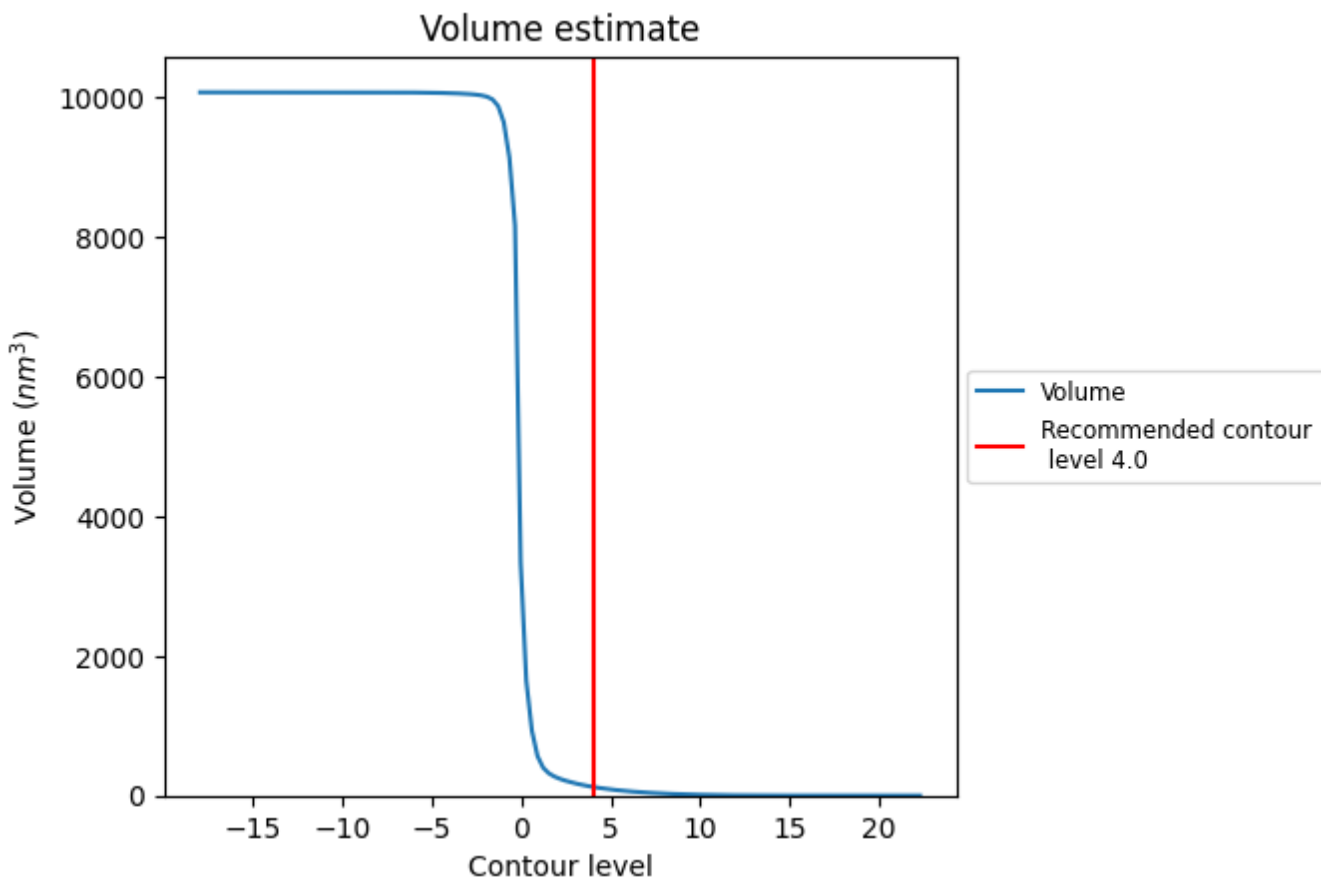
The volume at the recommended contour level is 124 nm^3; this corresponds to an approximate mass of 112 kDa.

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)
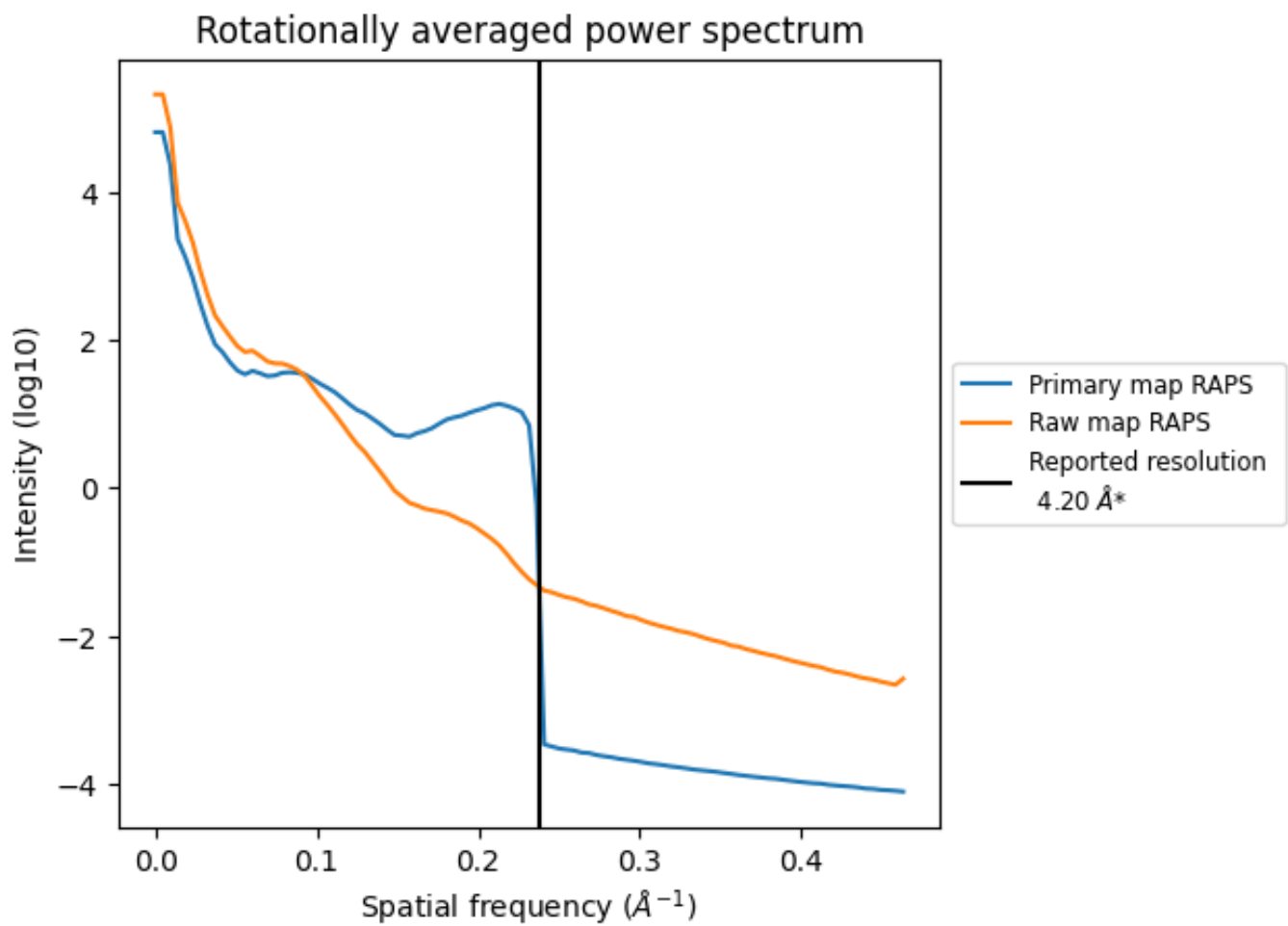
*Reported resolution corresponds to spatial frequency of 0.238 Å^{-1}

8 Fourier-Shell correlation ⓘ

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 ⓘ
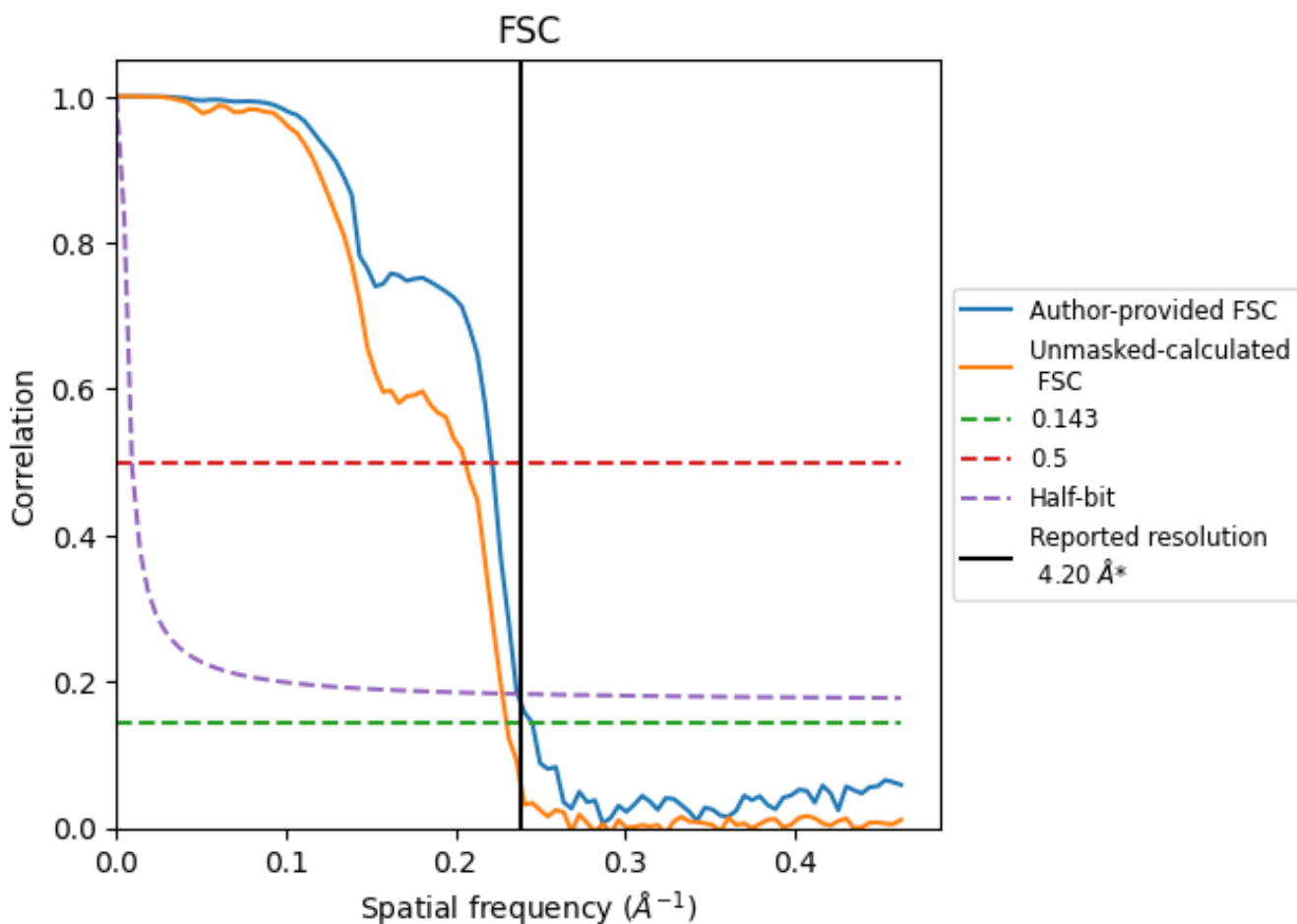
*Reported resolution corresponds to spatial frequency of 0.238 Å^{-1}

8.2 Resolution estimates ⓘ

Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	4.20	-	-
Author-provided FSC curve	4.07	4.51	4.22
Unmasked-calculated*	4.34	4.86	4.39

*Resolution estimate based on FSC curve calculated by comparison of deposited half-maps.

9 Map-model fit (i)

This section contains information regarding the fit between EMDB map EMD-24765 and PDB model 7UM1. Per-residue inclusion information can be found in section 3 on page 5.

9.1 Map-model overlay (i)
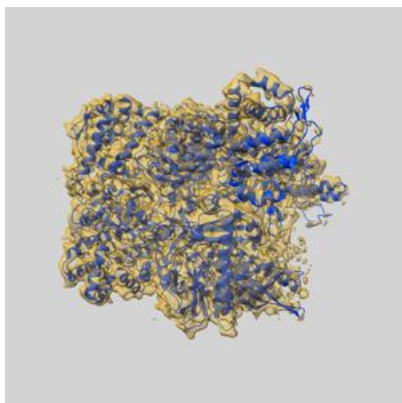
X   Y   Z
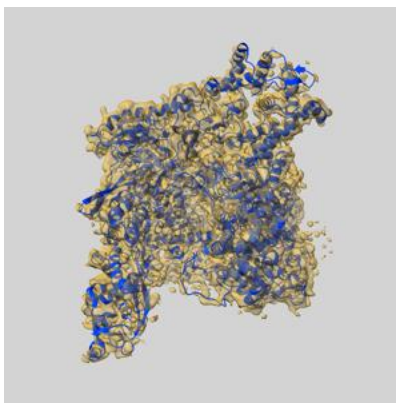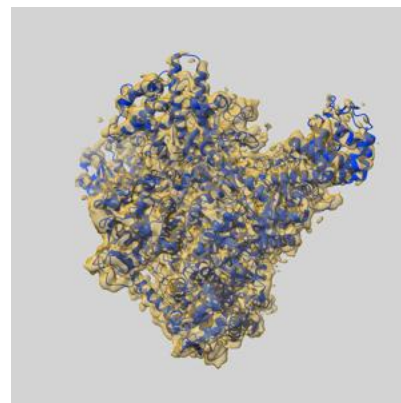
The images above show the 3D surface view of the map at the recommended contour level 4.0 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

9.2   Q-score mapped to coordinate model ⓘ

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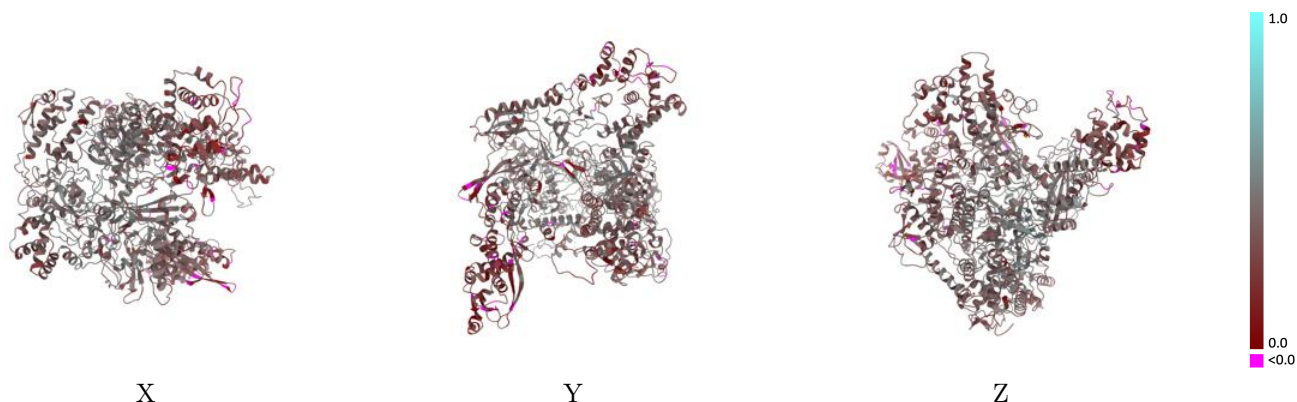
The images above show the model with each residue coloured according its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions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

X   Y   Z
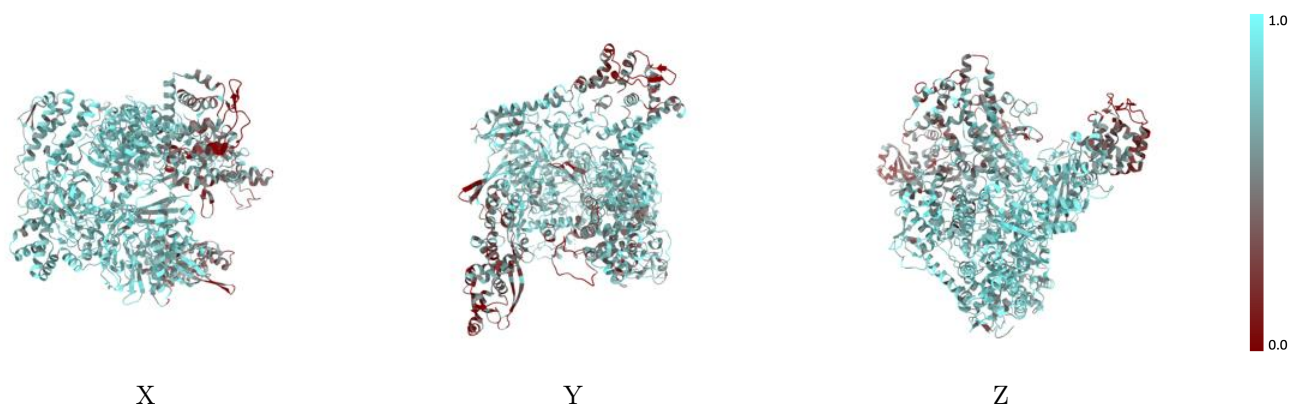
The images above show the model with each residue coloured according to its atom inclusion. This shows to what extent they are inside the map at the recommended contour level (4.0).

9.4　Atom inclusion ⓘ
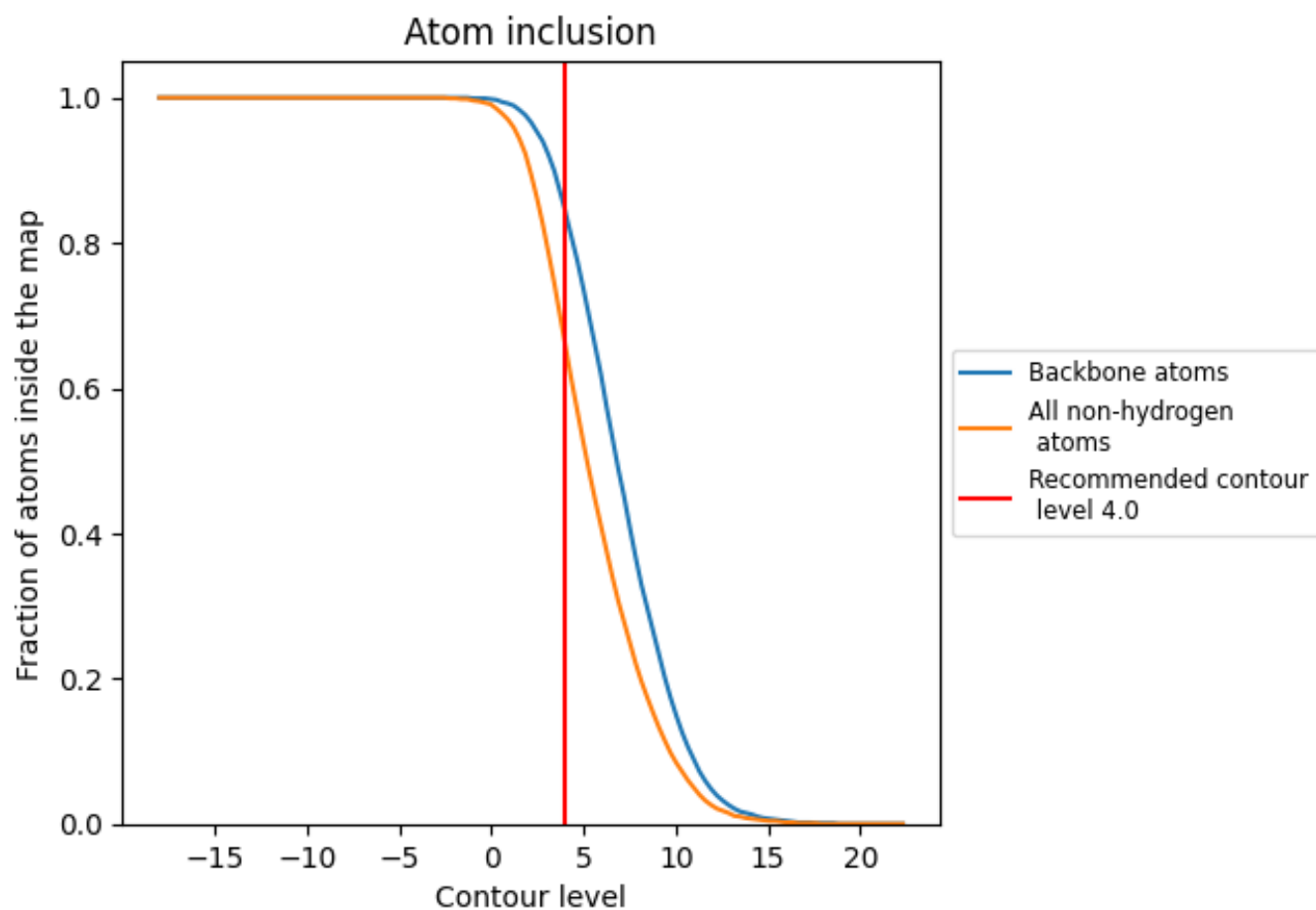
At the recommended contour level, 84% of all backbone atoms, 66% of all non-hydrogen atoms, are inside the map.

9.5 Map-model fit summary ⓘ

The table lists the average atom inclusion at the recommended contour level (4.0) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.6610	0.3570
A	0.4260	0.2670
C	0.7590	0.4030
D	0.7150	0.3700
c	0.5680	0.3100
d	0.6510	0.3630

1.0

0.0
<0.0