Full wwPDB EM Validation Report ⓘ

Feb 25, 2024 – 12:07 AM EST

PDB ID	:	7KFR
EMDB ID	:	EMD-22854
Title	:	Adeno-Associated Virus (AAV-DJ) - cryo-EM structure at 1.56 Angstrom Resolution
Authors	:	Xie, Q.; Yoshioka, C.K.; Chapman, M.S.
Deposited on	:	2020-10-14
Resolution	:	1.56 Å(reported)
Based on initial model	:	5UF6

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 ⓘ 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:

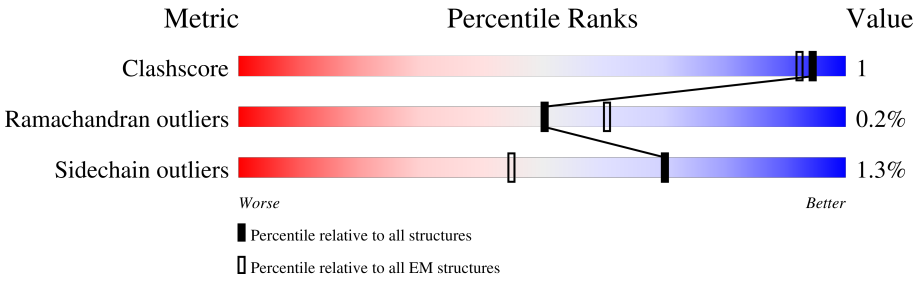
EMDB validation analysis	:	0.0.1.dev70
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.13
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 1.56 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 <=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	521	12% 90% 10%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 8954 atoms, of which 4496 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 Capsid protein VP1.

Mol	Chain	Residues	Atoms						AltConf	Trace
			Total	C	H	N	O	S		
1	A	521	8158	2646	3968	721	809	14	8	0

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	262	ASN	-	insertion	UNP Q6JC41
A	264	THR	GLN	conflict	UNP Q6JC41
A	266	GLY	-	insertion	UNP Q6JC41
A	268	SER	ALA	conflict	UNP Q6JC41
A	273	ALA	HIS	conflict	UNP Q6JC41
A	314	SER	ASN	conflict	UNP Q6JC41
A	329	GLU	ASP	conflict	UNP Q6JC41
A	332	LYS	THR	conflict	UNP Q6JC41
A	342	ILE	VAL	conflict	UNP Q6JC41
A	374	ILE	VAL	conflict	UNP Q6JC41
A	412	GLN	THR	conflict	UNP Q6JC41
A	414	THR	SER	conflict	UNP Q6JC41
A	451	GLN	ASN	conflict	UNP Q6JC41
A	453	THR	PRO	conflict	UNP Q6JC41
A	454	GLY	SER	conflict	UNP Q6JC41
A	458	ASN	THR	conflict	UNP Q6JC41
A	459	THR	GLN	conflict	UNP Q6JC41
A	460	GLN	SER	conflict	UNP Q6JC41
A	461	THR	ARG	conflict	UNP Q6JC41
A	463	GLY	GLN	conflict	UNP Q6JC41
A	467	GLY	ALA	conflict	UNP Q6JC41
A	469	PRO	ALA	conflict	UNP Q6JC41
A	470	ASN	SER	conflict	UNP Q6JC41
A	471	THR	ASP	conflict	UNP Q6JC41
A	472	MET	ILE	conflict	UNP Q6JC41
A	473	ALA	ARG	conflict	UNP Q6JC41
A	474	ASN	ASP	conflict	UNP Q6JC41
A	476	ALA	SER	conflict	UNP Q6JC41

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	477	LYS	ARG	conflict	UNP Q6JC41
A	587	ARG	GLY	conflict	UNP Q6JC41
A	590	ARG	THR	conflict	UNP Q6JC41
A	658	ASP	ASN	conflict	UNP Q6JC41
A	660	PRO	SER	conflict	UNP Q6JC41
A	664	ASN	SER	conflict	UNP Q6JC41
A	665	GLN	ALA	conflict	UNP Q6JC41
A	666	SER	ALA	conflict	UNP Q6JC41
A	668	LEU	PHE	conflict	UNP Q6JC41
A	669	ASN	ALA	conflict	UNP Q6JC41
A	707	TYR	ASN	conflict	UNP Q6JC41
A	710	THR	VAL	conflict	UNP Q6JC41
A	711	SER	ASN	conflict	UNP Q6JC41
A	715	ALA	THR	conflict	UNP Q6JC41
A	717	ASN	ASP	conflict	UNP Q6JC41
A	719	GLU	ASN	conflict	UNP Q6JC41

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		AltConf
2	A	3	Total	Mg	0
			3	3	

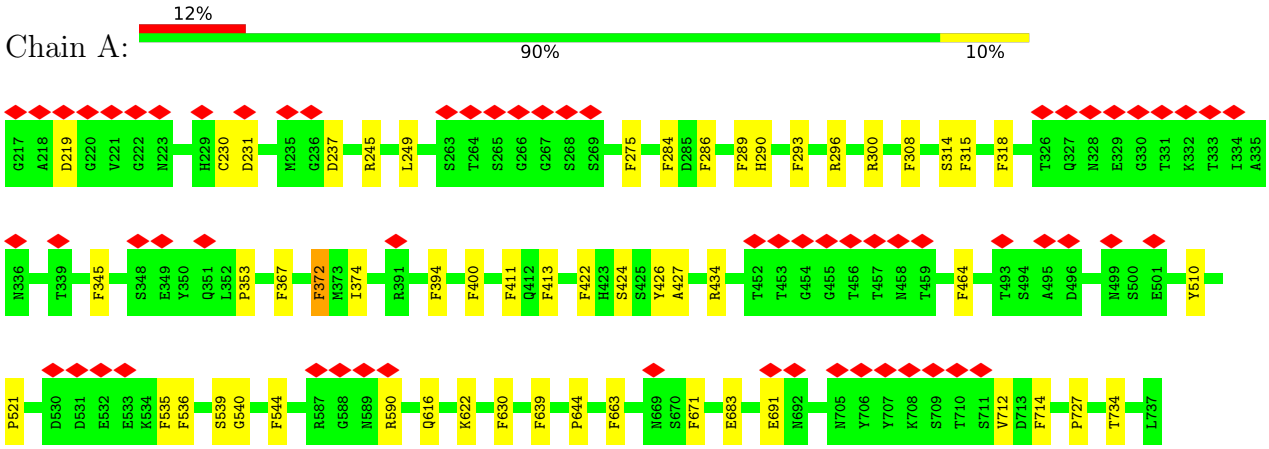
- Molecule 3 is water.

Mol	Chain	Residues	Atoms			AltConf
3	A	265	Total	H	O	0
			793	528	265	

3 Residue-property plots (i)

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: Capsid protein VP1



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, I	Depositor
Number of particles used	48209	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 ($e^-/Å^2$)	30	Depositor
Minimum defocus (nm)	-800	Depositor
Maximum defocus (nm)	-2600	Depositor
Magnification	154400	Depositor
Image detector	FEI FALCON III (4k x 4k)	Depositor
Maximum map value	0.372	Depositor
Minimum map value	-0.177	Depositor
Average map value	-0.000	Depositor
Map value standard deviation	0.018	Depositor
Recommended contour level	0.064	Depositor
Map size (Å)	306.30002, 306.30002, 306.30002	wwPDB
Map dimensions	600, 600, 600	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	0.5105, 0.5105, 0.5105	Depositor

5 Model quality

5.1 Standard geometry

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

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	1.36	41/4337 (0.9%)	1.09	16/5911 (0.3%)

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	1

All (41) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	639	PHE	CG-CD1	8.74	1.51	1.38
1	A	345	PHE	CG-CD1	8.12	1.50	1.38
1	A	536	PHE	CG-CD2	7.78	1.50	1.38
1	A	422	PHE	CG-CD1	7.68	1.50	1.38
1	A	663	PHE	CG-CD2	7.45	1.50	1.38
1	A	413	PHE	CG-CD1	7.34	1.49	1.38
1	A	367	PHE	CG-CD1	6.84	1.49	1.38
1	A	284	PHE	CG-CD2	6.82	1.49	1.38
1	A	318	PHE	CG-CD1	6.77	1.49	1.38
1	A	535	PHE	CG-CD2	6.72	1.48	1.38
1	A	394	PHE	CG-CD2	6.67	1.48	1.38
1	A	464	PHE	CG-CD2	6.63	1.48	1.38
1	A	284	PHE	CG-CD1	6.62	1.48	1.38
1	A	400	PHE	CG-CD1	6.60	1.48	1.38
1	A	413	PHE	CG-CD2	6.36	1.48	1.38
1	A	671	PHE	CG-CD1	6.24	1.48	1.38
1	A	535	PHE	CG-CD1	6.24	1.48	1.38
1	A	411	PHE	CG-CD1	6.20	1.48	1.38

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	286	PHE	CG-CD2	6.15	1.48	1.38
1	A	411	PHE	CG-CD2	6.04	1.47	1.38
1	A	714	PHE	CG-CD2	5.88	1.47	1.38
1	A	345	PHE	CG-CD2	5.76	1.47	1.38
1	A	536	PHE	CG-CD1	5.67	1.47	1.38
1	A	308	PHE	CG-CD2	5.63	1.47	1.38
1	A	315	PHE	CG-CD1	5.56	1.47	1.38
1	A	464	PHE	CG-CD1	5.51	1.47	1.38
1	A	630	PHE	CB-CG	5.49	1.60	1.51
1	A	275	PHE	CG-CD2	5.49	1.47	1.38
1	A	544	PHE	CG-CD2	5.45	1.47	1.38
1	A	372	PHE	CG-CD2	5.35	1.46	1.38
1	A	293	PHE	CG-CD2	5.32	1.46	1.38
1	A	630	PHE	CG-CD2	5.30	1.46	1.38
1	A	671	PHE	CG-CD2	5.27	1.46	1.38
1	A	353	PRO	N-CD	5.25	1.55	1.47
1	A	714	PHE	CG-CD1	5.24	1.46	1.38
1	A	318	PHE	CG-CD2	5.18	1.46	1.38
1	A	289	PHE	CG-CD2	5.16	1.46	1.38
1	A	308	PHE	CG-CD1	5.16	1.46	1.38
1	A	422	PHE	CB-CG	5.03	1.59	1.51
1	A	275	PHE	CG-CD1	5.01	1.46	1.38
1	A	400	PHE	CG-CD2	5.00	1.46	1.38

All (16) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	639	PHE	CB-CG-CD2	12.95	129.87	120.80
1	A	639	PHE	CB-CG-CD1	-11.90	112.47	120.80
1	A	394	PHE	CB-CG-CD1	7.07	125.75	120.80
1	A	422	PHE	CB-CG-CD1	-6.82	116.03	120.80
1	A	422	PHE	CB-CG-CD2	6.61	125.42	120.80
1	A	289	PHE	CB-CG-CD2	-6.30	116.39	120.80
1	A	245	ARG	NE-CZ-NH2	-6.23	117.18	120.30
1	A	300	ARG	NE-CZ-NH1	6.13	123.37	120.30
1	A	394	PHE	CB-CG-CD2	-5.76	116.77	120.80
1	A	296	ARG	NE-CZ-NH2	-5.67	117.47	120.30
1	A	345	PHE	CB-CG-CD2	5.55	124.68	120.80
1	A	663	PHE	CB-CG-CD1	5.46	124.62	120.80
1	A	434	ARG	NE-CZ-NH2	-5.45	117.57	120.30
1	A	434	ARG	NE-CZ-NH1	5.33	122.97	120.30
1	A	245	ARG	NE-CZ-NH1	5.14	122.87	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	590	ARG	NE-CZ-NH1	5.04	122.82	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	510	TYR	Sidechain

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	4190	3968	3980	9	0
2	A	3	0	0	0	0
3	A	265	528	0	0	0
All	All	4458	4496	3980	9	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 1.

All (9) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:372:PHE:CE2	1:A:374[B]:ILE:HD13	2.43	0.53
1:A:314:SER:OG	1:A:683:GLU:HB3	2.12	0.49
1:A:521:PRO:O	1:A:540:GLY:HA2	2.15	0.47
1:A:622[B]:LYS:HB2	1:A:644:PRO:HG3	1.96	0.47
1:A:616:GLN:HE22	1:A:727:PRO:HA	1.81	0.45
1:A:521:PRO:HA	1:A:539:SER:O	2.18	0.43
1:A:249:LEU:HB2	1:A:374[A]:ILE:HD12	2.01	0.42
1:A:424:SER:HB2	1:A:426:TYR:CE2	2.56	0.41
1:A:427:ALA:O	1:A:734:THR:HA	2.21	0.40

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	527/521 (101%)	511 (97%)	15 (3%)	1 (0%)	47 23

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	231	ASP

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	467/459 (102%)	461 (99%)	6 (1%)	69 44

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	219	ASP
1	A	230	CYS
1	A	237	ASP
1	A	290	HIS
1	A	691	GLU
1	A	712	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	351	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 3 ligands modelled in this entry, 3 are 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.

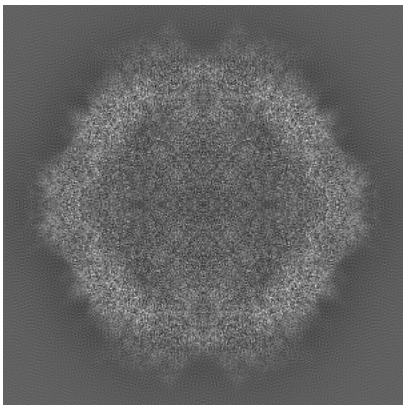
6 Map visualisation

This section contains visualisations of the EMDB entry EMD-22854. These allow visual inspection of the internal detail of the map and identification of artifacts.

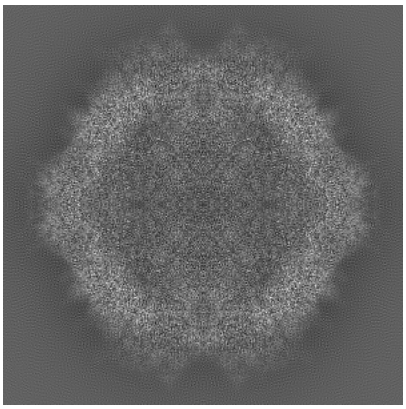
Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

6.1 Orthogonal projections

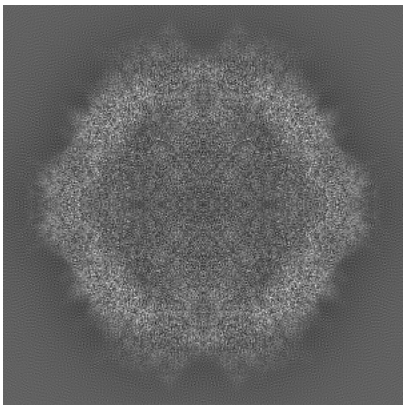
6.1.1 Primary map



X

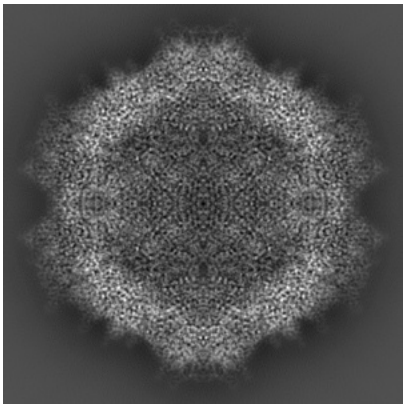


Y

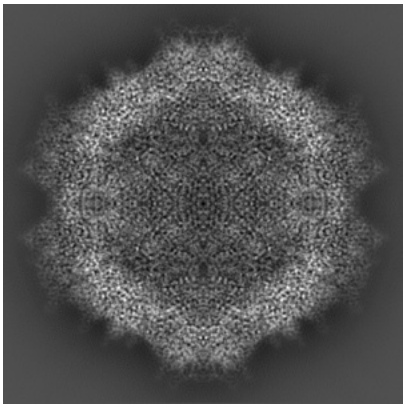


Z

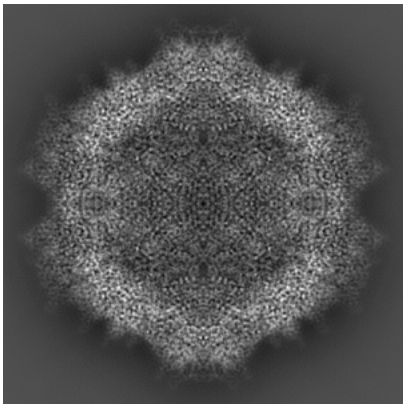
6.1.2 Raw map



X



Y

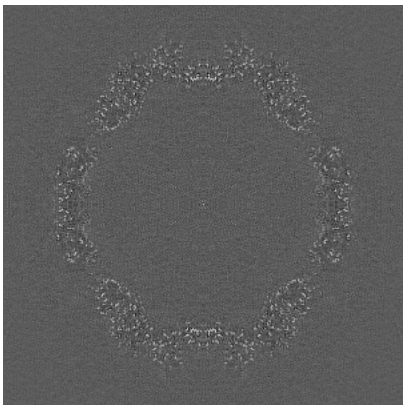


Z

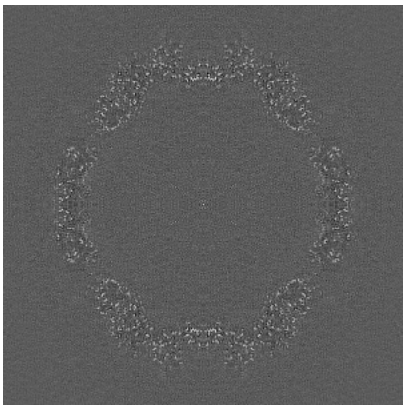
The images above show the map projected in three orthogonal directions.

6.2 Central slices (i)

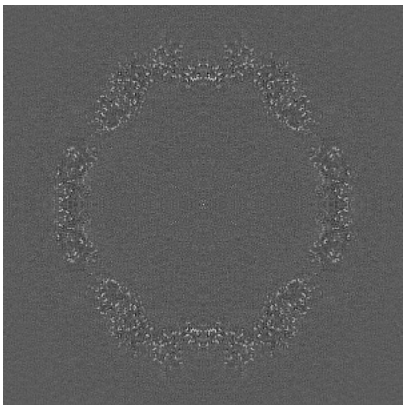
6.2.1 Primary map



X Index: 300

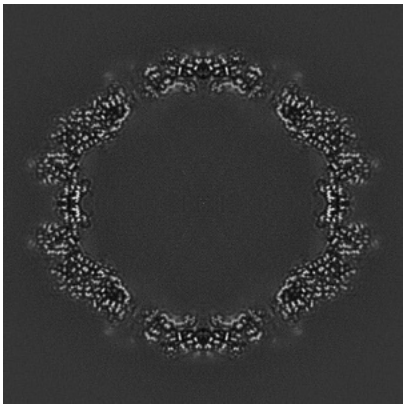


Y Index: 300

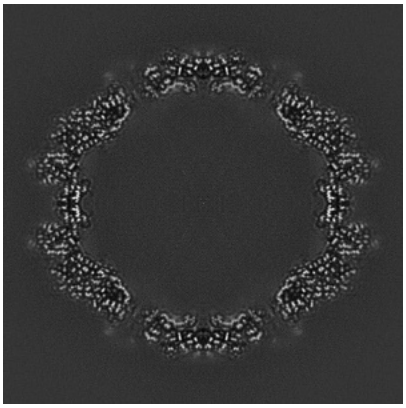


Z Index: 300

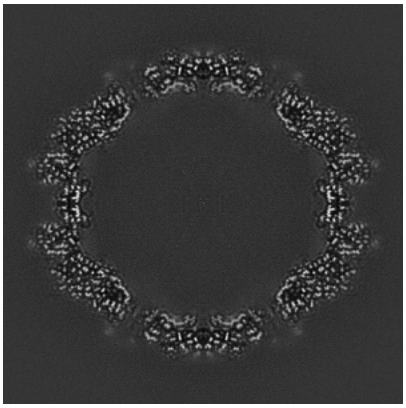
6.2.2 Raw map



X Index: 300



Y Index: 300

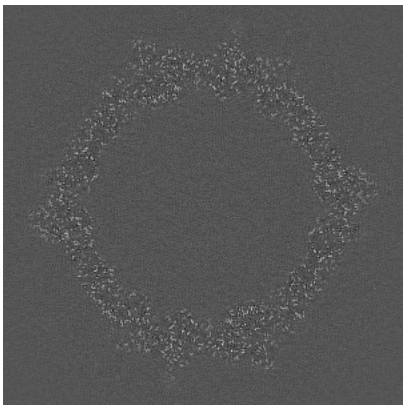


Z Index: 300

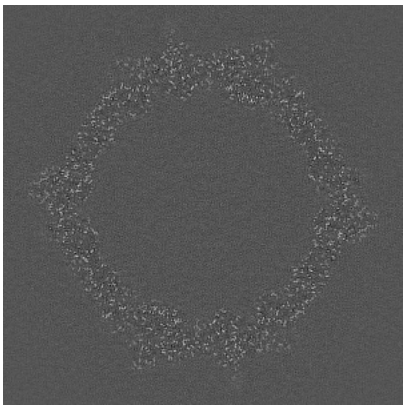
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

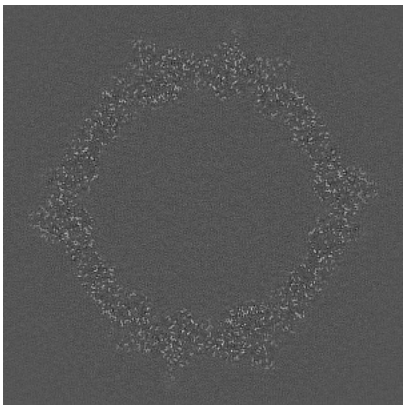
6.3.1 Primary map



X Index: 259

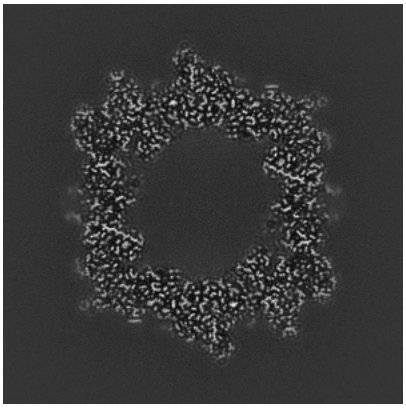


Y Index: 341

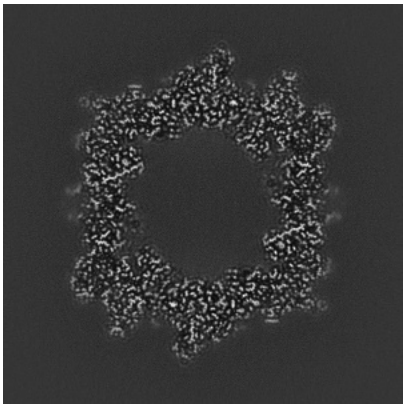


Z Index: 259

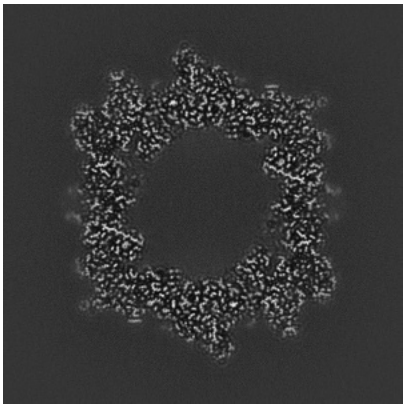
6.3.2 Raw map



X Index: 429



Y Index: 171

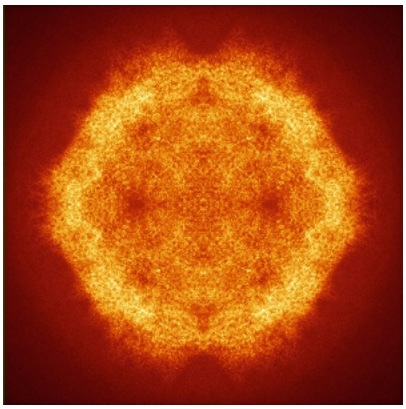


Z Index: 429

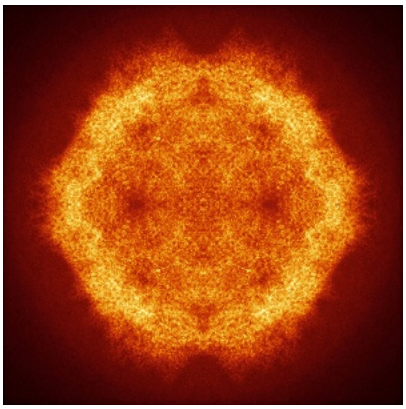
The images above show the largest variance slices of the map in three orthogonal directions.

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

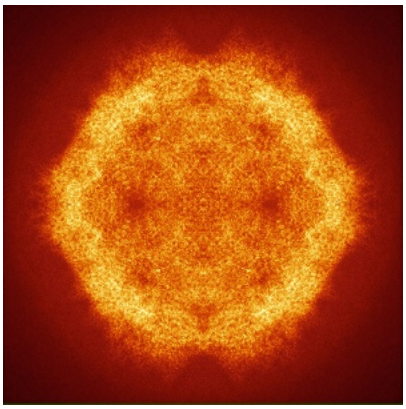
6.4.1 Primary map



X

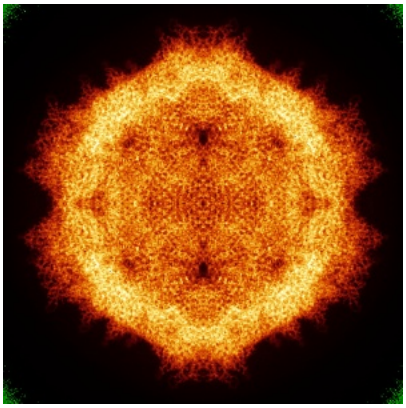


Y

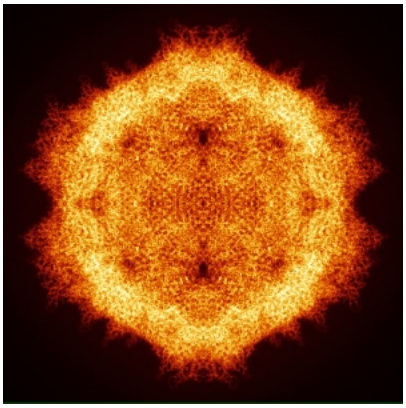


Z

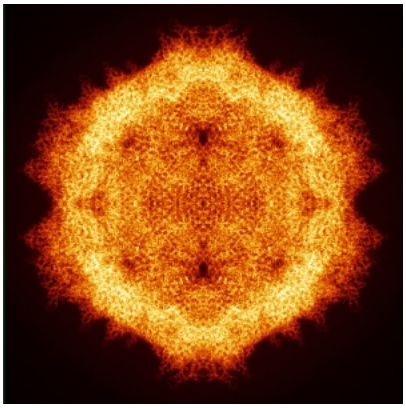
6.4.2 Raw map



X



Y

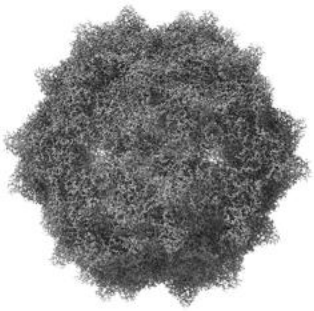


Z

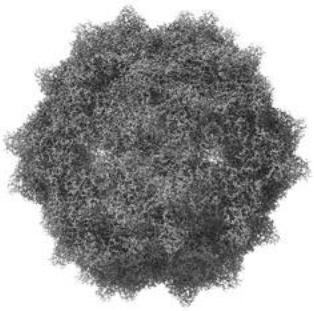
The images above show the map standard deviation projections with false color in three orthogonal directions. Minimum values are shown in green, max in blue, and dark to light orange shades represent small to large values respectively.

6.5 Orthogonal surface views (i)

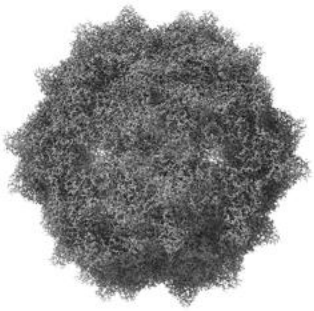
6.5.1 Primary map



X



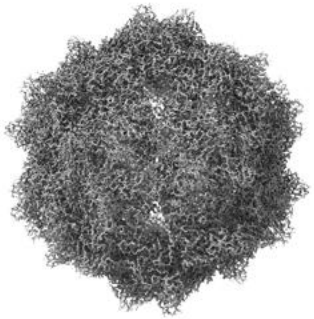
Y



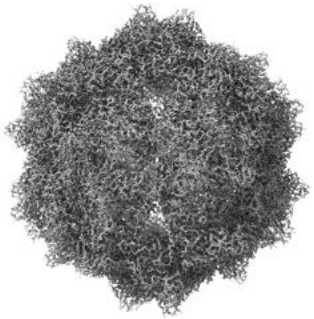
Z

The images above show the 3D surface view of the map at the recommended contour level 0.064. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

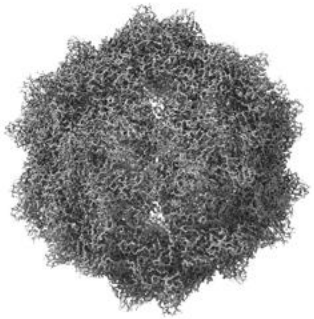
6.5.2 Raw map



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.

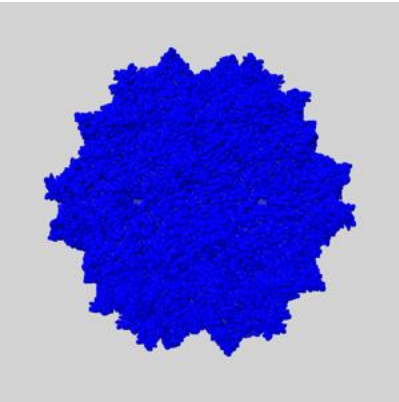
6.6 Mask visualisation (i)

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

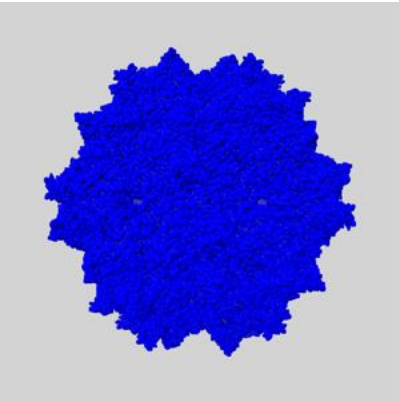
A mask typically either:

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

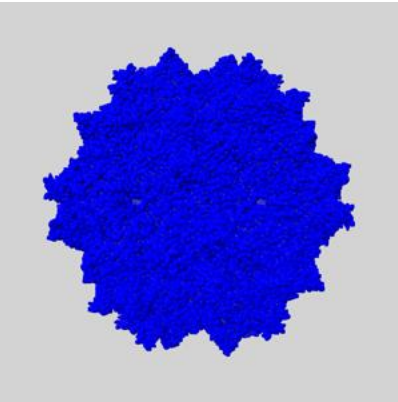
6.6.1 emd_22854_msk_1.map (i)



X



Y

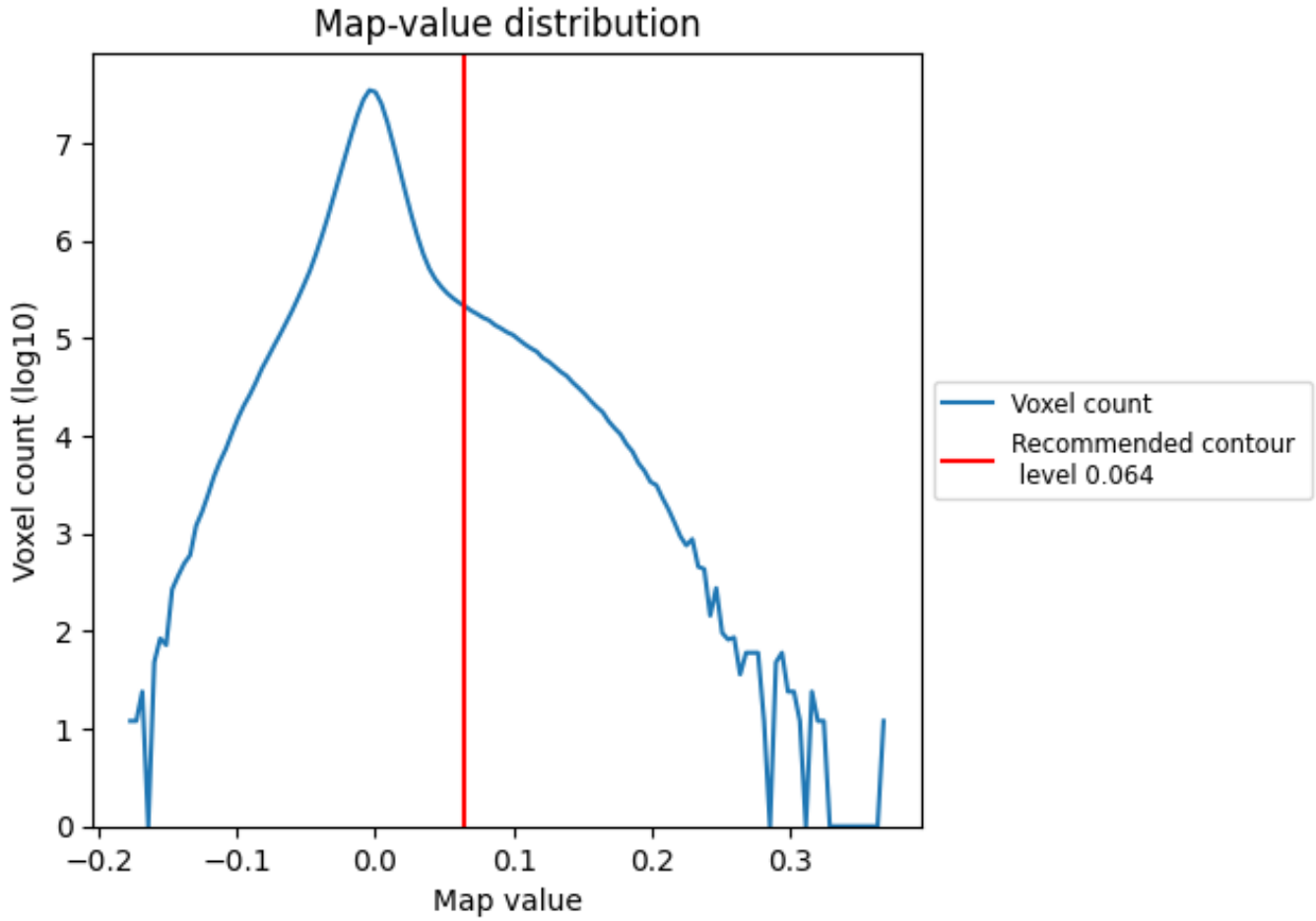


Z

7 Map analysis

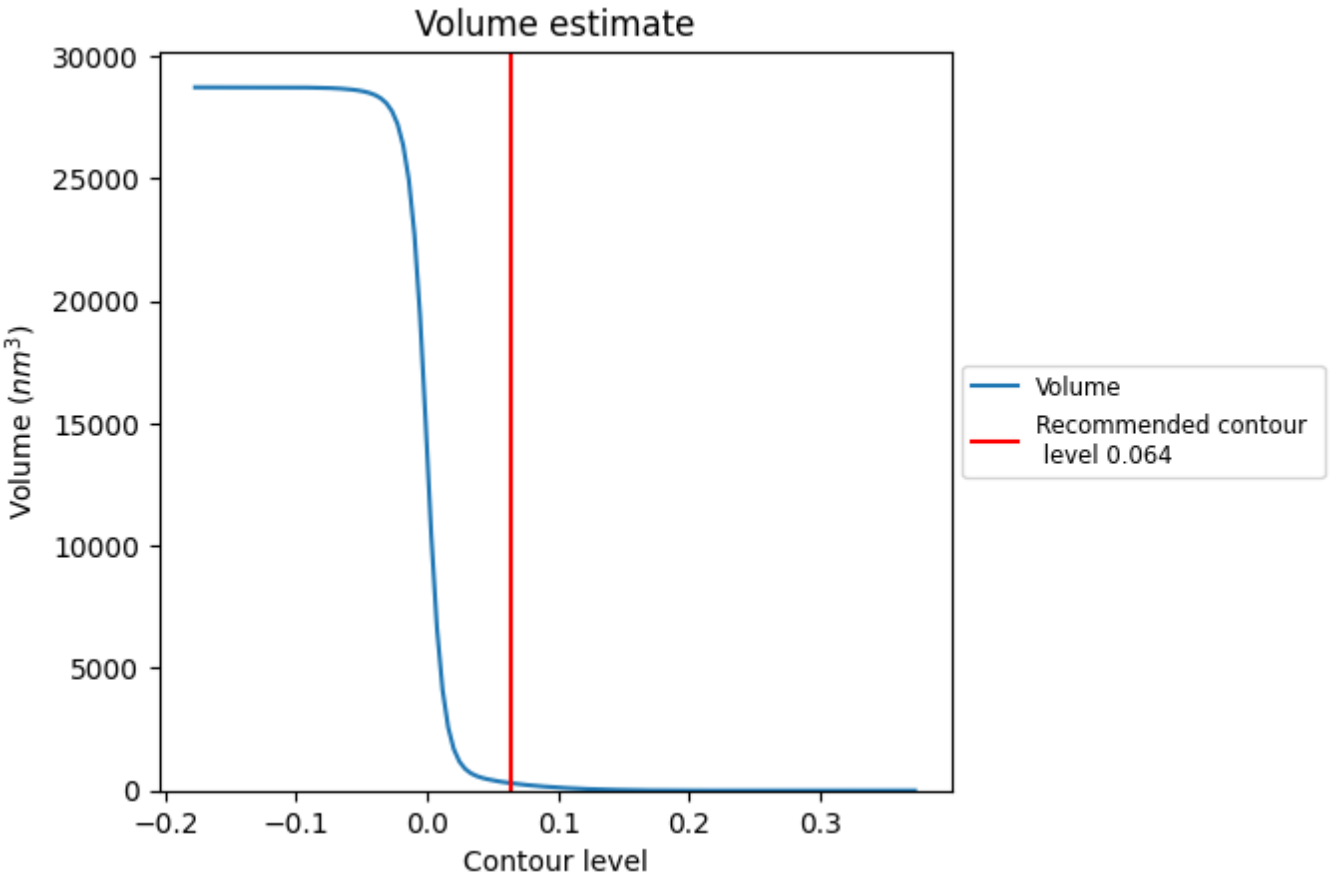
This section contains the results of statistical analysis of the map.

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.

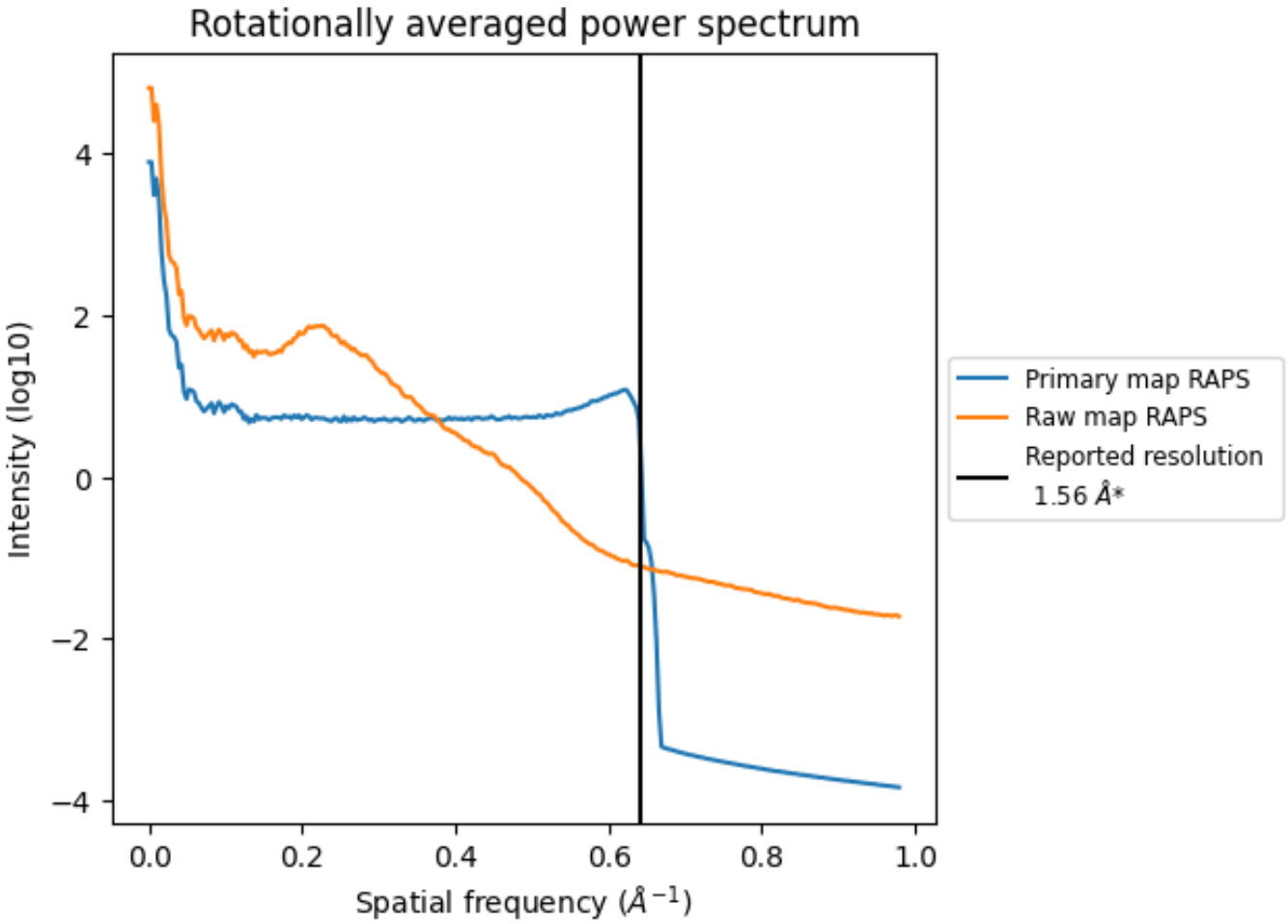
7.2 Volume estimate



The volume at the recommended contour level is 301 nm^3; this corresponds to an approximate mass of 272 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)

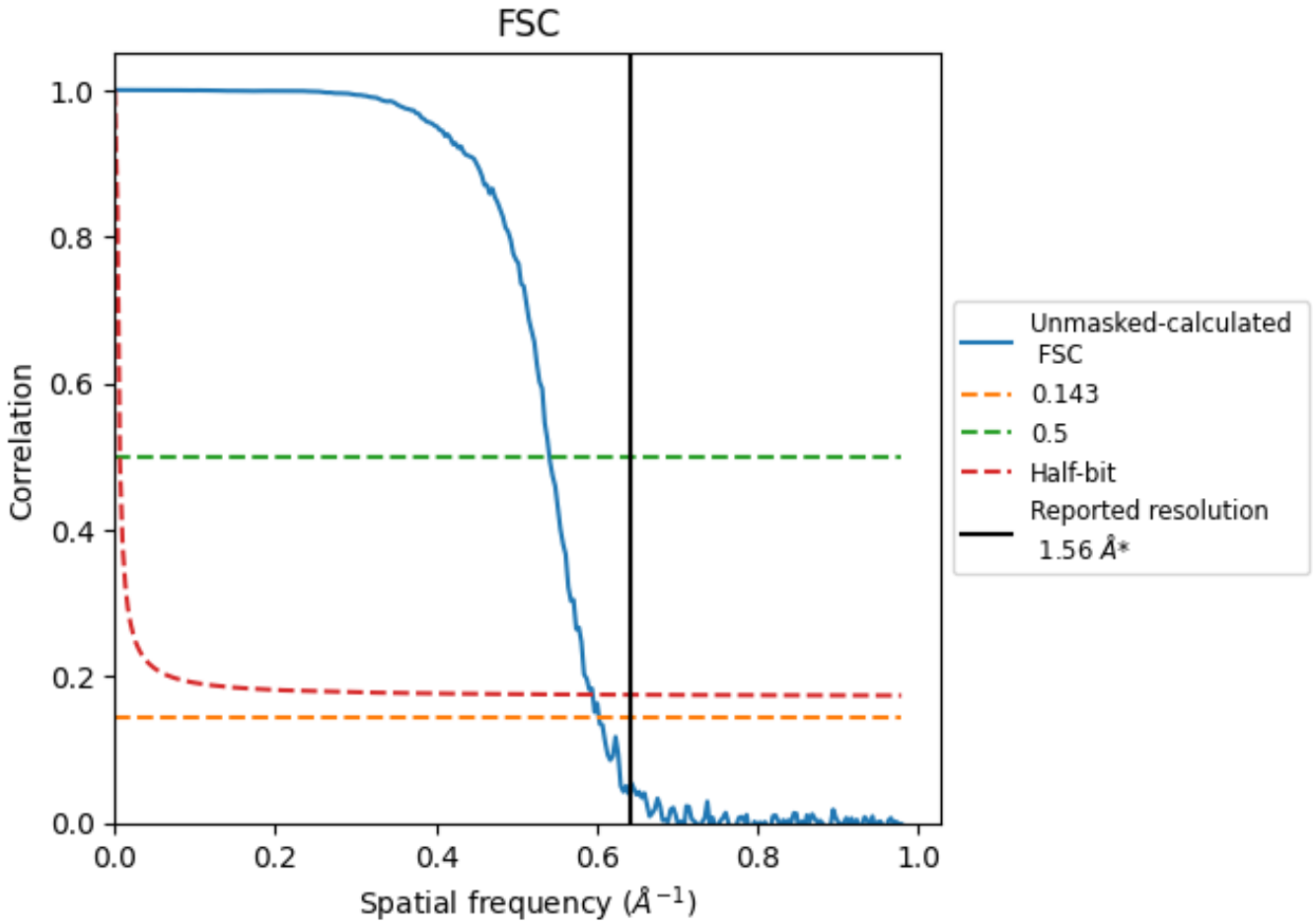


*Reported resolution corresponds to spatial frequency of 0.641 Å^{-1}

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 0.641 Å^{-1}

8.2 Resolution estimates (i)

Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	1.56	-	-
Author-provided FSC curve	-	-	-
Unmasked-calculated*	1.66	1.85	1.68

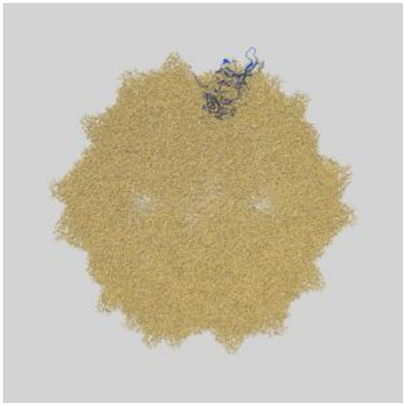
*Resolution estimate based on FSC curve calculated by comparison of deposited half-maps.

9 Map-model fit

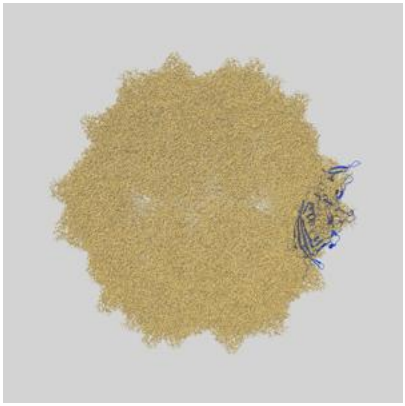
This section contains information regarding the fit between EMDB map EMD-22854 and PDB model 7KFR. Per-residue inclusion information can be found in section 3 on page 5.

9.1 Map-model overlays

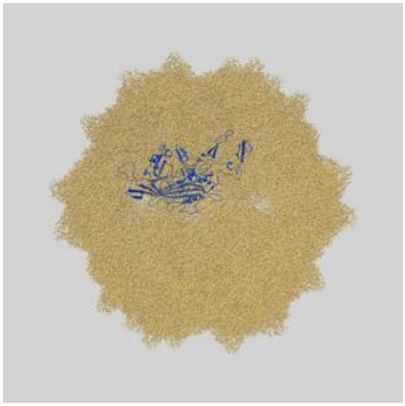
9.1.1 Map-model overlay



X

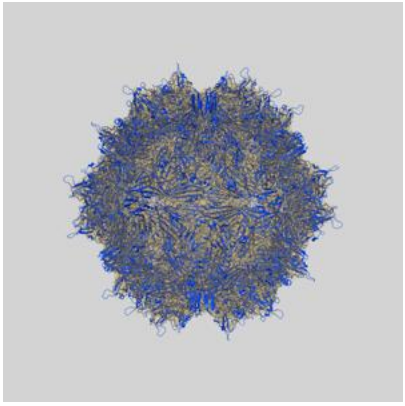


Y

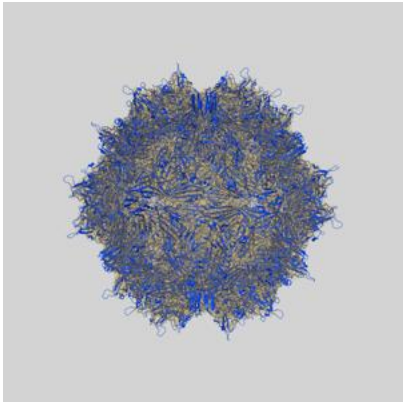


Z

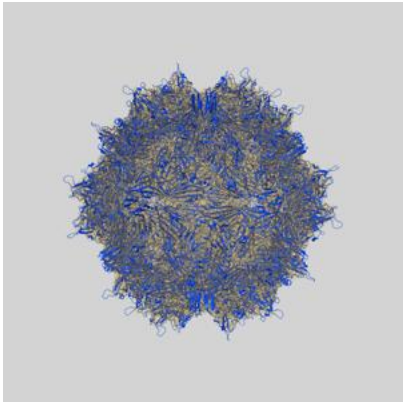
9.1.2 Map-model assembly overlay



X



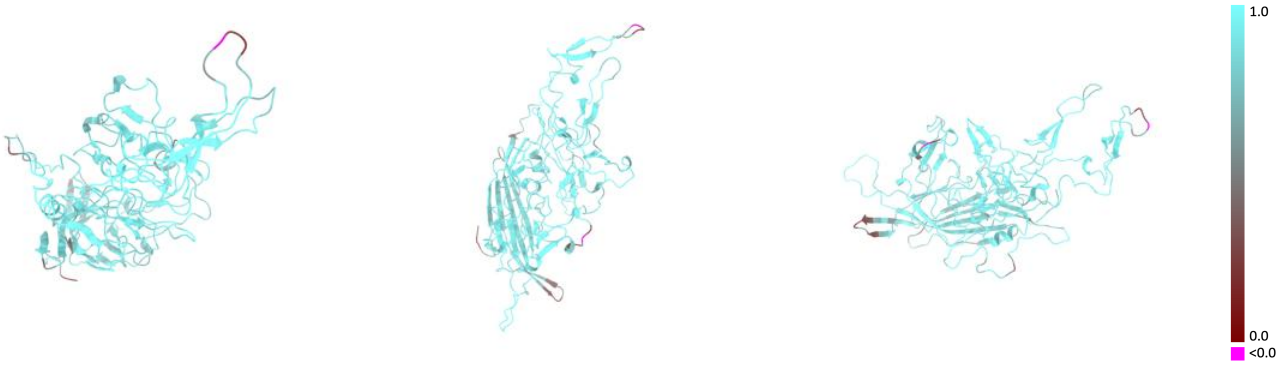
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.064 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.

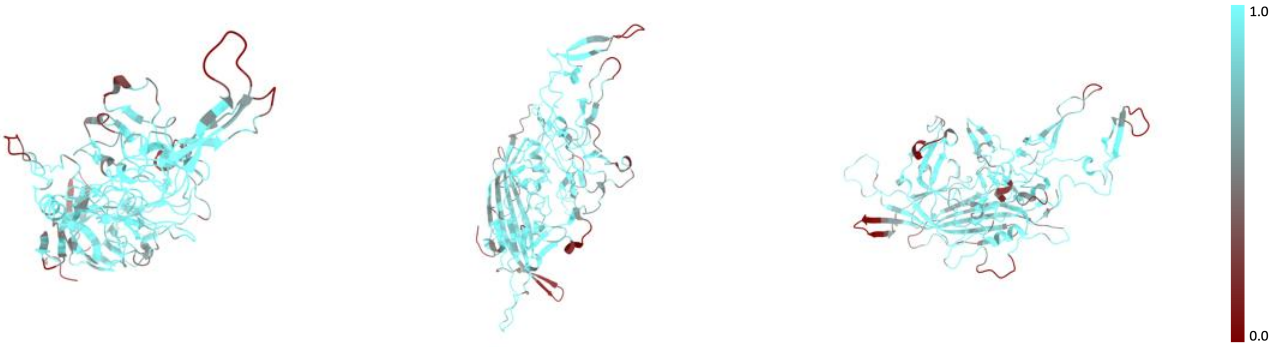
9.2 Q-score mapped to coordinate model (i)



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.

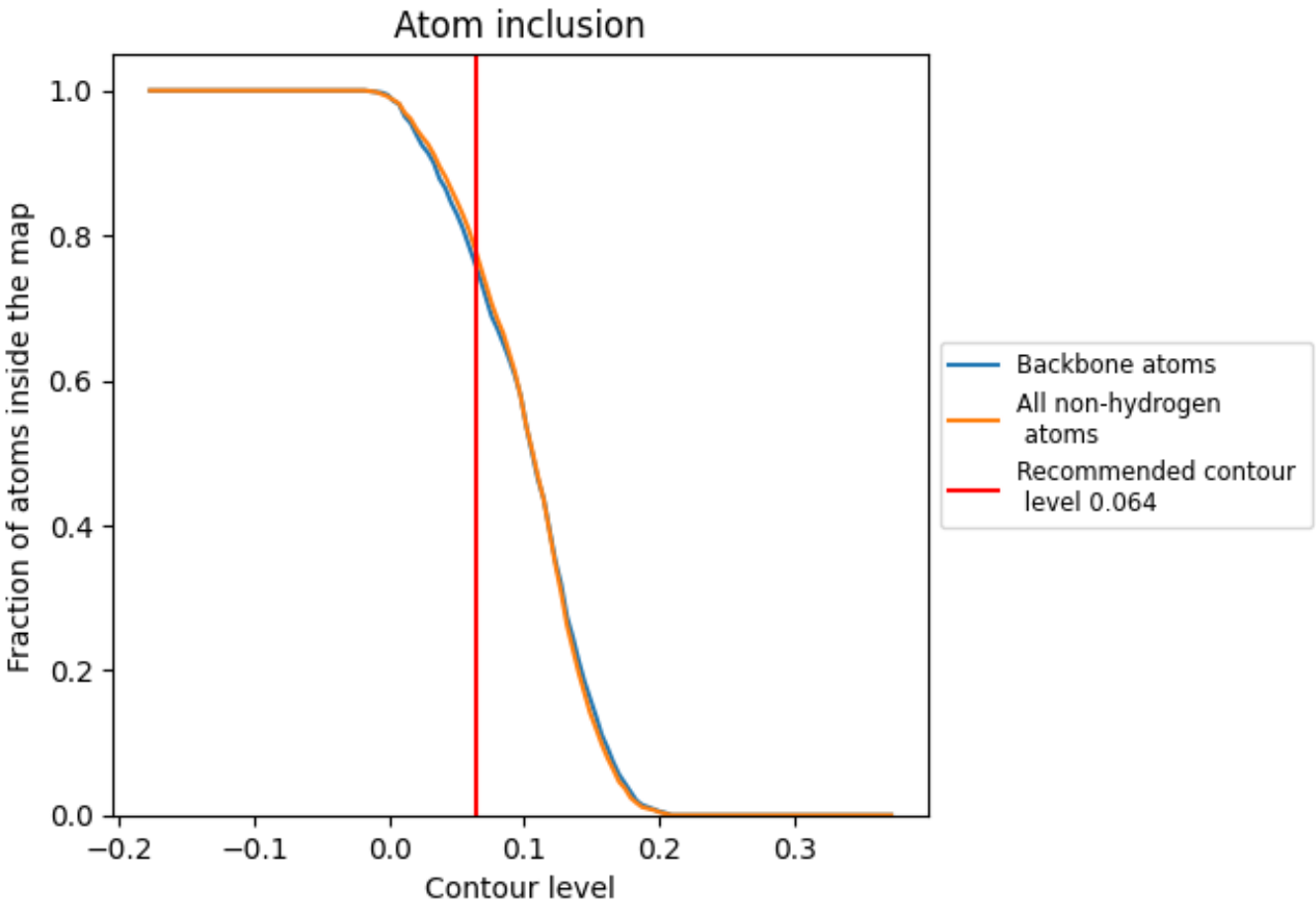
9.3 Atom inclusion mapped to coordinate model (i)



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 (0.064).

9.4 Atom inclusion (i)



At the recommended contour level, 76% of all backbone atoms, 78% 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 (0.064) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.7800	0.8380
A	0.7920	0.8380