wwPDB EM Validation Summary Report

May 20, 2023 – 09:24 am BST

PDB ID	:	7QOJ
EMDB ID	:	EMD-14092
Title	:	Tail barrel assembly of the phicrAss001 virion with C12 symmetry imposed
Authors	:	Bayfield, O.W.; Shkoporov, A.N.; Yutin, N.; Khokhlova, E.V.; Smith, J.L.R.; Hawkins, D.E.D.P.; Koonin, E.V.; Hill, C.; Antson, A.A.
Deposited on	:	2021-12-24
Resolution	:	3.21 Å(reported)

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

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

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

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

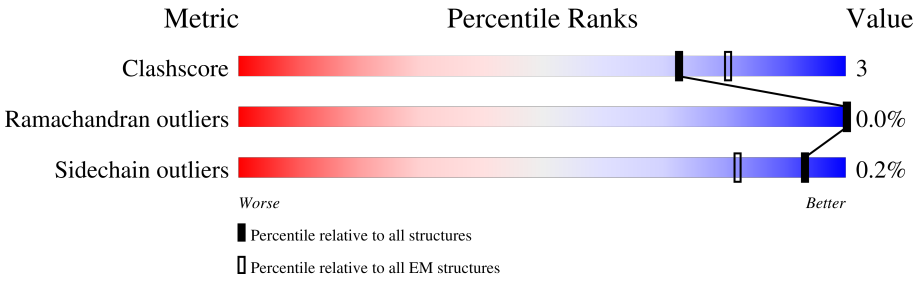
EMDB validation analysis	:	0.0.1.dev50
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.9
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.32.2

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 3.21 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.



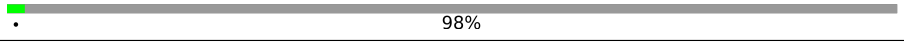
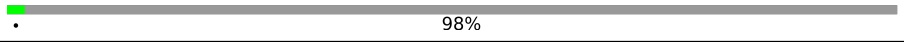
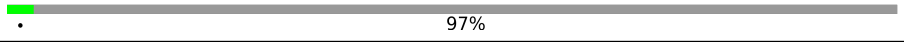
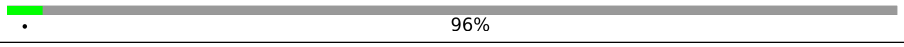
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

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	806	5% 73% 9% 18%
2	B	236	87% 9% .
3	C	225	92% 8%
4	D	230	90% 9% .
5	E	238	. 89% 11%
5	F	238	. 89% 9% .
6	G	215	. 76% 6% 18%
6	H	215	6% 80% 7% 13%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
7	I	114	
8	J	832	
8	K	832	
8	L	832	
9	M	842	
9	N	842	

2 Entry composition (i)

There are 11 unique types of molecules in this entry. The entry contains 20218 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 Portal protein gp20.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	657	5371	3404	917	1027	23	0	0

- Molecule 2 is a protein called Ring protein 1 gp43.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	B	226	1841	1188	308	336	9	0	0

- Molecule 3 is a protein called Ring protein 2 gp40.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	C	225	1841	1180	292	357	12	0	0

- Molecule 4 is a protein called Ring protein 3 gp35.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	D	226	1805	1151	293	357	4	0	0

- Molecule 5 is a protein called Ring protein 4/5 gp34.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
5	E	238	1878	1185	312	377	4	0	0
			Total	C	N	O	S		
5	F	232	1843	1165	305	369	4	0	0

- Molecule 6 is a protein called Tail hub protein A gp38.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	G	177	Total	C	N	O	S	0	0
			1441	924	230	276	11		
6	H	188	Total	C	N	O	S	0	0
			1516	968	243	292	13		

- Molecule 7 is a protein called Tail hub protein B gp39.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	I	107	Total	C	N	O	S	0	0
			878	560	134	175	9		

- Molecule 8 is a protein called Tail fiber protein gp22.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	J	19	Total	C	N	O	S	0	0
			170	112	27	30	1		
8	K	19	Total	C	N	O	S	0	0
			170	112	27	30	1		
8	L	25	Total	C	N	O	S	0	0
			224	145	37	41	1		

- Molecule 9 is a protein called Cargo protein 1 gp45.

Mol	Chain	Residues	Atoms					AltConf	Trace
9	M	126	Total	C	N	O	S	0	0
			987	605	179	201	2		
9	N	32	Total	C	N	O	S	0	0
			251	157	44	49	1		

- Molecule 10 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		AltConf
10	A	1	Total	Mg	0
			1	1	

- Molecule 11 is water.

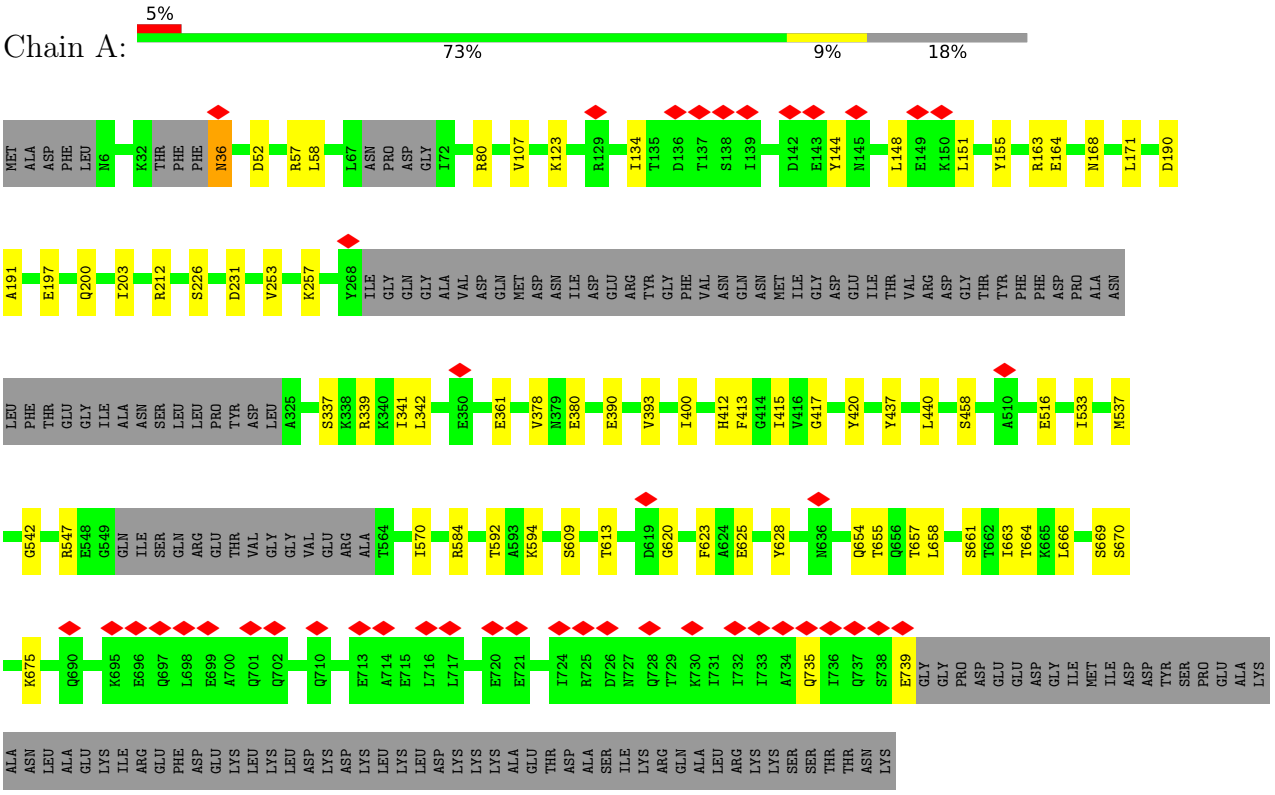
Mol	Chain	Residues	Atoms		AltConf
11	A	1	Total	O	0
			1	1	

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.

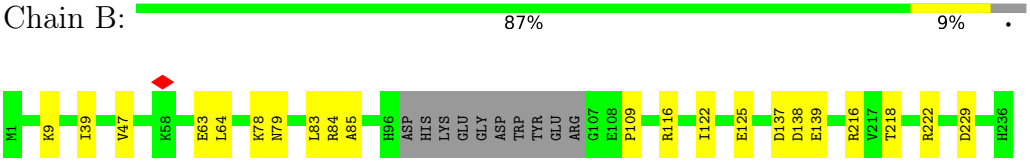
• Molecule 1: Portal protein gp20

Chain A:



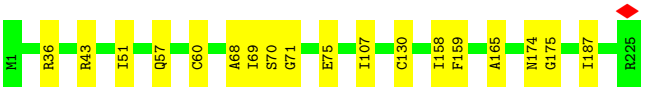
• Molecule 2: Ring protein 1 gp43

Chain B:



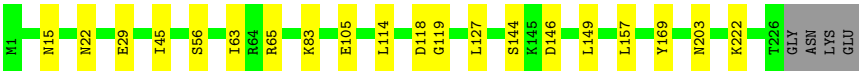
• Molecule 3: Ring protein 2 gp40

Chain C:



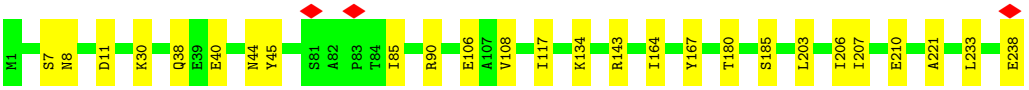
• Molecule 4: Ring protein 3 gp35

Chain D:



• Molecule 5: Ring protein 4/5 gp34

Chain E:

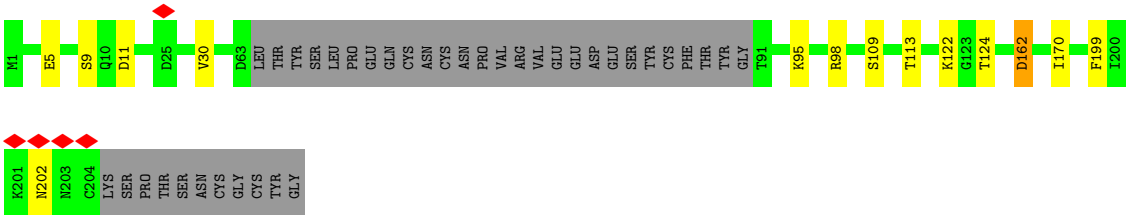


• Molecule 5: Ring protein 4/5 gp34

Chain F:

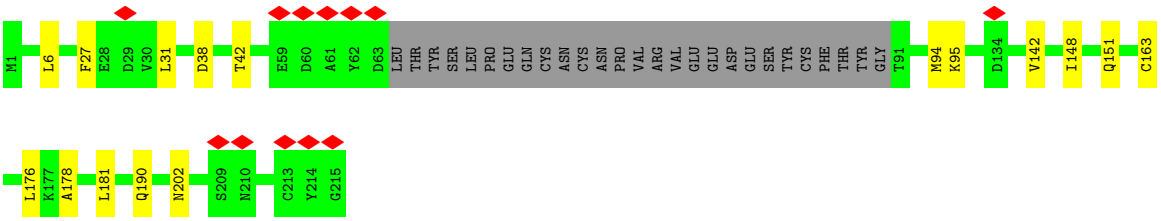
• Molecule 6: Tail hub protein A gp38

Chain G:



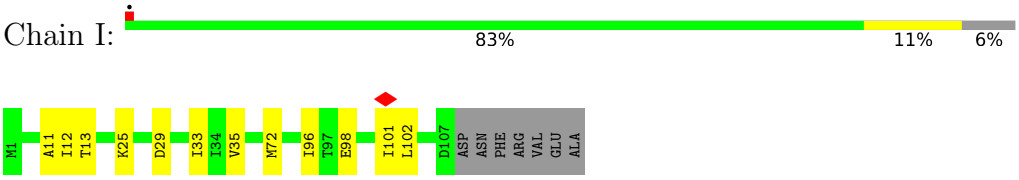
• Molecule 6: Tail hub protein A gp38

Chain H:



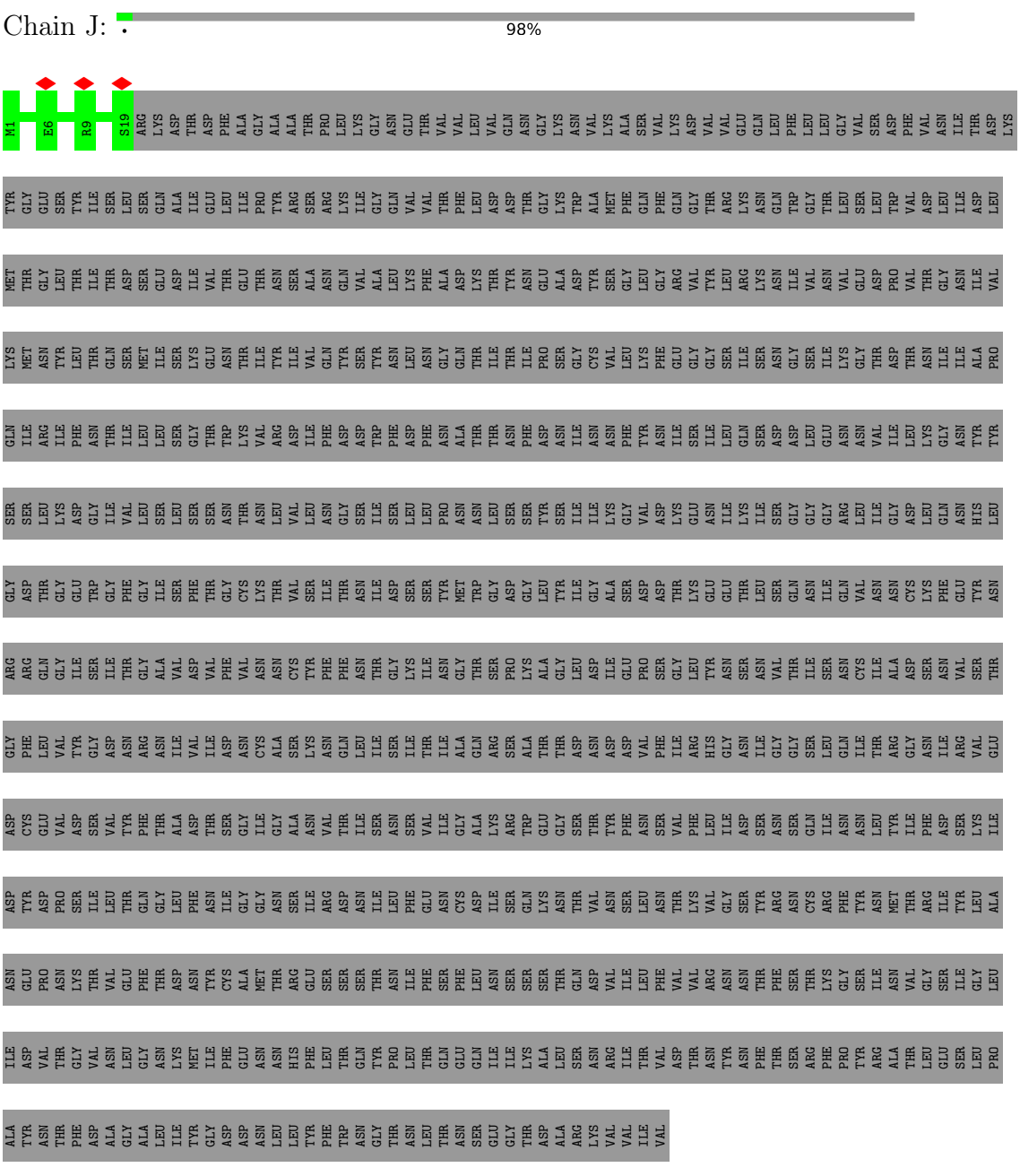
• Molecule 7: Tail hub protein B gp39

Chain I:



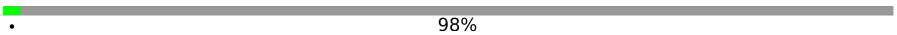
• Molecule 8: Tail fiber protein gp22

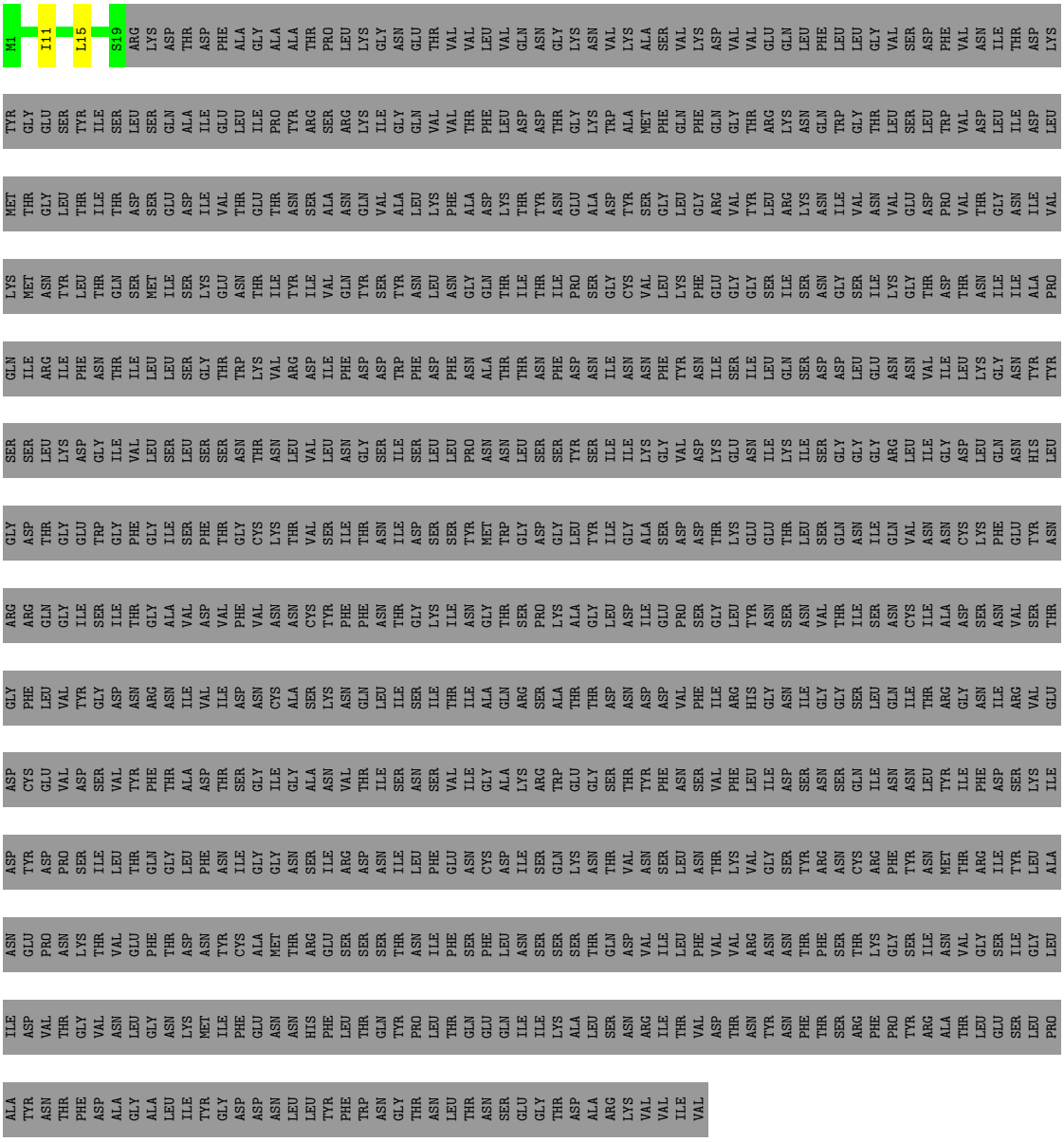
Chain J:



• Molecule 8: Tail fiber protein gp22

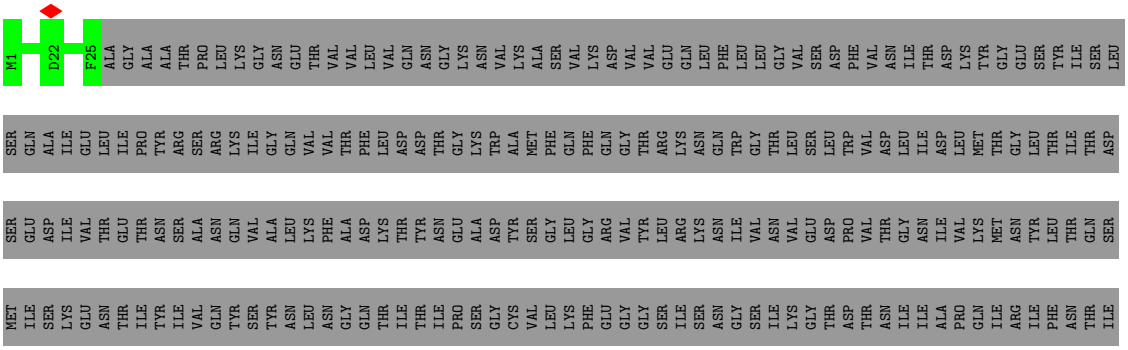
Chain K:

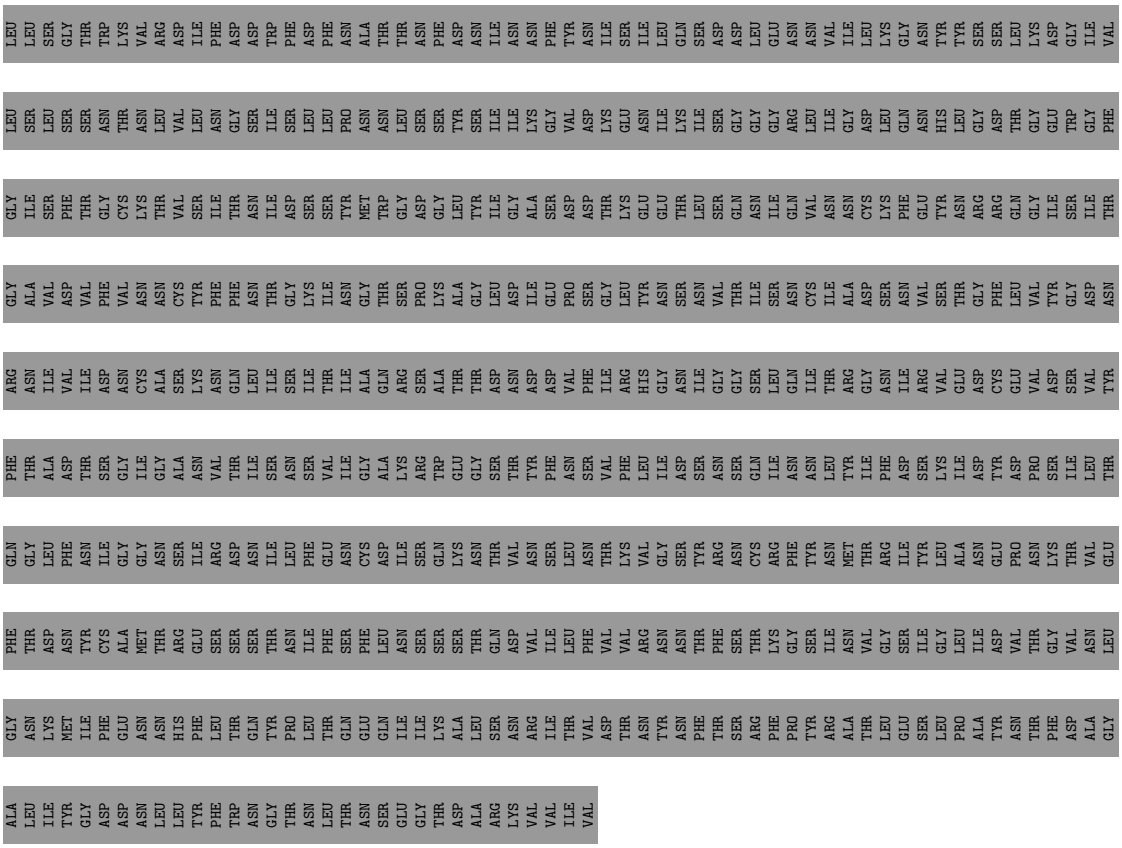




• Molecule 8: Tail fiber protein gp22

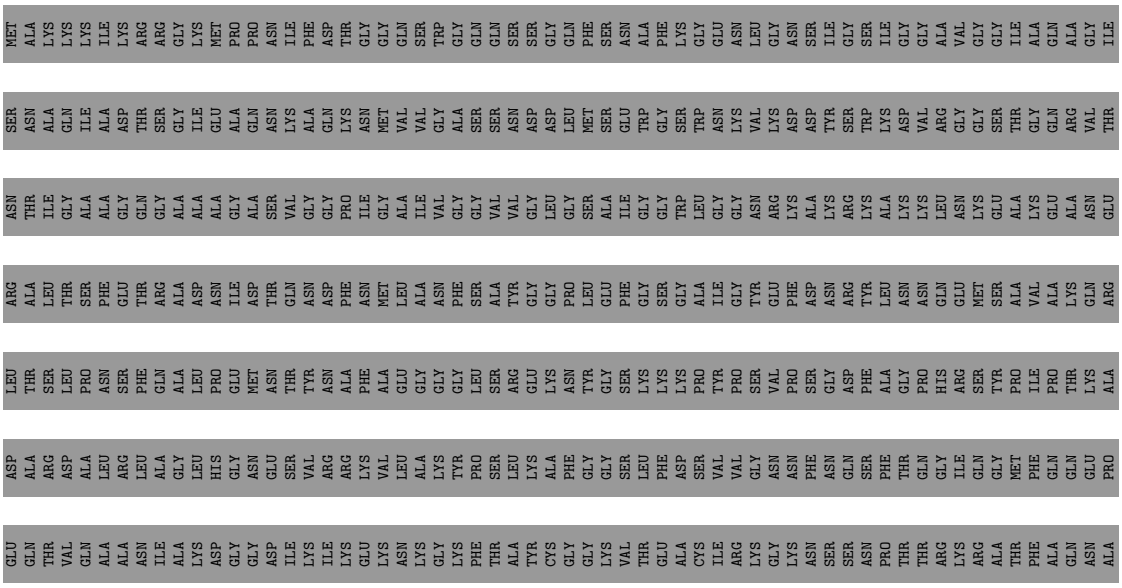
Chain L: 97%

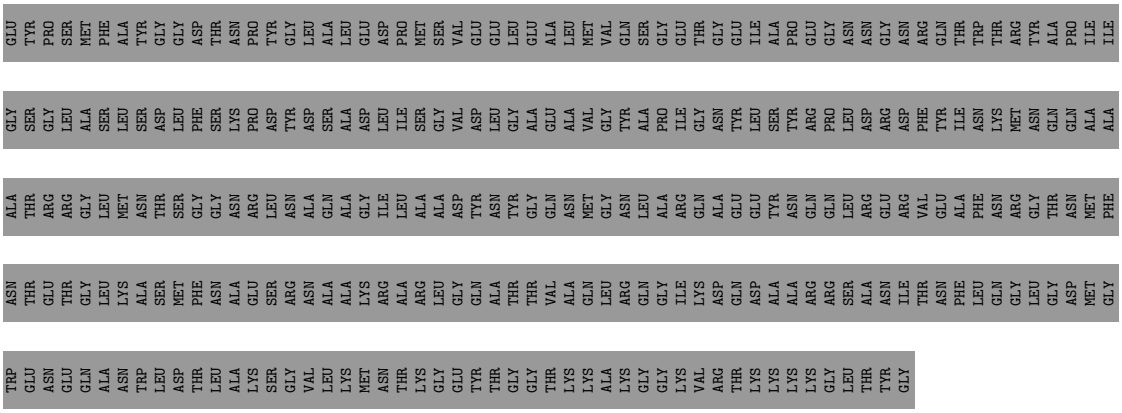




• Molecule 9: Cargo protein 1 gp45

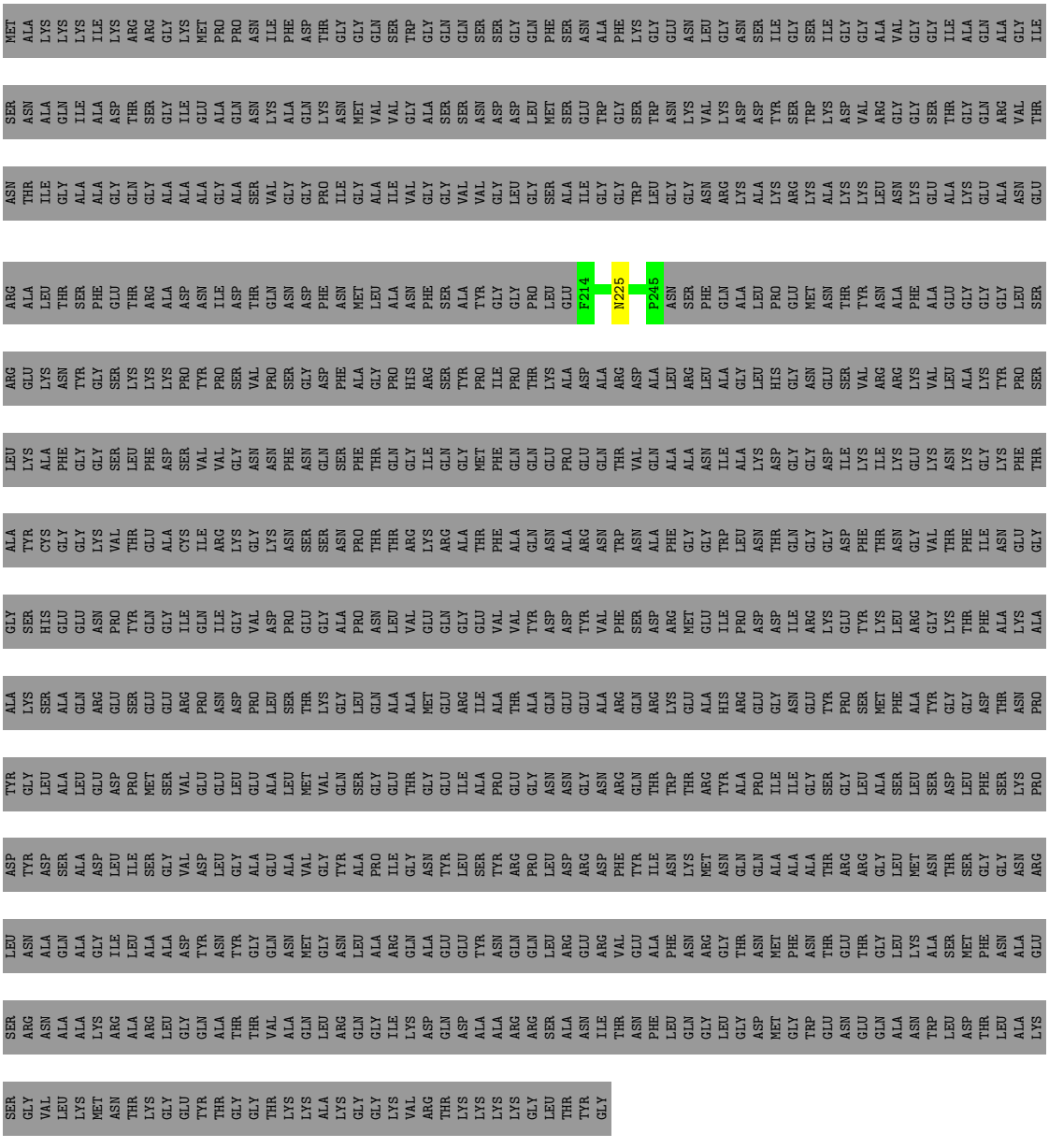
Chain M:





• Molecule 9: Cargo protein 1 gp45

Chain N: 96%



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C12	Depositor
Number of particles used	122709	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 ($e^-/Å^2$)	51	Depositor
Minimum defocus (nm)	300	Depositor
Maximum defocus (nm)	1500	Depositor
Magnification	Not provided	
Image detector	FEI FALCON III (4k x 4k)	Depositor
Maximum map value	0.762	Depositor
Minimum map value	-0.553	Depositor
Average map value	0.001	Depositor
Map value standard deviation	0.029	Depositor
Recommended contour level	0.06	Depositor
Map size (Å)	535.3336, 535.3336, 535.3336	wwPDB
Map dimensions	384, 384, 384	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.3940979, 1.3940979, 1.3940979	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	0.34	0/5468	0.57	0/7374
2	B	0.35	0/1881	0.59	0/2548
3	C	0.42	1/1886 (0.1%)	0.58	0/2552
4	D	0.35	0/1837	0.58	0/2491
5	E	0.36	0/1910	0.53	0/2594
5	F	0.32	0/1873	0.53	0/2541
6	G	0.33	0/1469	0.57	1/1989 (0.1%)
6	H	0.31	0/1546	0.53	0/2093
7	I	0.37	0/895	0.58	0/1208
8	J	0.39	0/174	0.66	0/232
8	K	0.39	0/174	0.55	0/232
8	L	0.28	0/229	0.53	0/305
9	M	0.29	0/1003	0.58	0/1347
9	N	0.35	0/255	0.59	0/342
All	All	0.35	1/20600 (0.0%)	0.57	1/27848 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	60	CYS	CB-SG	-5.56	1.72	1.81

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
6	G	162	ASP	CB-CG-OD1	7.47	125.02	118.30

There are no chirality outliers.

There are no planarity outliers.

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	5371	0	5333	40	0
2	B	1841	0	1863	12	0
3	C	1841	0	1800	11	0
4	D	1805	0	1813	11	0
5	E	1878	0	1870	15	0
5	F	1843	0	1838	13	0
6	G	1441	0	1425	8	0
6	H	1516	0	1493	16	0
7	I	878	0	844	13	0
8	J	170	0	167	0	0
8	K	170	0	167	1	0
8	L	224	0	217	0	0
9	M	987	0	930	7	0
9	N	251	0	239	1	0
10	A	1	0	0	0	0
11	A	1	0	0	0	0
All	All	20218	0	19999	135	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 3.

The worst 5 of 135 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:H:190:GLN:HE22	7:I:13:THR:HG22	1.06	1.18
6:H:190:GLN:HE22	7:I:13:THR:CG2	1.61	1.13
6:H:190:GLN:NE2	7:I:13:THR:HG22	1.76	1.01
2:B:116:ARG:NH1	6:H:163:CYS:SG	2.51	0.84
1:A:458:SER:OG	1:A:516:GLU:OE1	1.96	0.84

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	647/806 (80%)	636 (98%)	11 (2%)	0	100	100
2	B	222/236 (94%)	217 (98%)	5 (2%)	0	100	100
3	C	223/225 (99%)	214 (96%)	9 (4%)	0	100	100
4	D	224/230 (97%)	213 (95%)	11 (5%)	0	100	100
5	E	236/238 (99%)	225 (95%)	11 (5%)	0	100	100
5	F	228/238 (96%)	216 (95%)	12 (5%)	0	100	100
6	G	173/215 (80%)	168 (97%)	4 (2%)	1 (1%)	25	63
6	H	184/215 (86%)	180 (98%)	4 (2%)	0	100	100
7	I	105/114 (92%)	101 (96%)	4 (4%)	0	100	100
8	J	17/832 (2%)	16 (94%)	1 (6%)	0	100	100
8	K	17/832 (2%)	16 (94%)	1 (6%)	0	100	100
8	L	23/832 (3%)	23 (100%)	0	0	100	100
9	M	122/842 (14%)	116 (95%)	6 (5%)	0	100	100
9	N	30/842 (4%)	28 (93%)	2 (7%)	0	100	100
All	All	2451/6697 (37%)	2369 (97%)	81 (3%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
6	G	162	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	588/714 (82%)	585 (100%)	3 (0%)	88 94
2	B	206/215 (96%)	205 (100%)	1 (0%)	88 94
3	C	208/208 (100%)	208 (100%)	0	100 100
4	D	207/210 (99%)	207 (100%)	0	100 100
5	E	211/211 (100%)	211 (100%)	0	100 100
5	F	208/211 (99%)	208 (100%)	0	100 100
6	G	168/203 (83%)	168 (100%)	0	100 100
6	H	177/203 (87%)	177 (100%)	0	100 100
7	I	100/106 (94%)	100 (100%)	0	100 100
8	J	18/736 (2%)	18 (100%)	0	100 100
8	K	18/736 (2%)	18 (100%)	0	100 100
8	L	24/736 (3%)	24 (100%)	0	100 100
9	M	100/644 (16%)	100 (100%)	0	100 100
9	N	26/644 (4%)	26 (100%)	0	100 100
All	All	2259/5777 (39%)	2255 (100%)	4 (0%)	93 97

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	36	ASN
1	A	200	GLN
1	A	547	ARG
2	B	222	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	200	GLN
5	F	15	ASN
5	F	16	ASN
6	G	190	GLN
6	H	190	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.

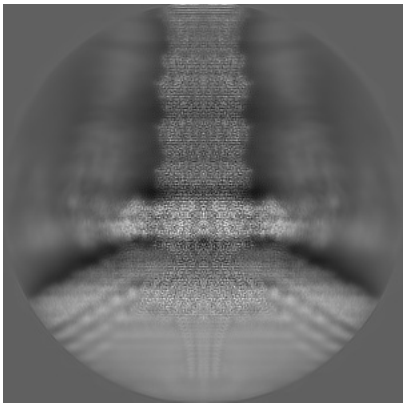
6 Map visualisation

This section contains visualisations of the EMDB entry EMD-14092. These allow visual inspection of the internal detail of the map and identification of artifacts.

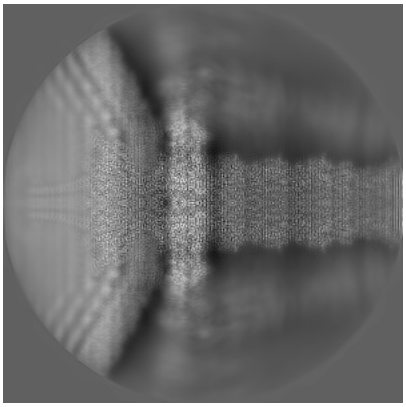
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections

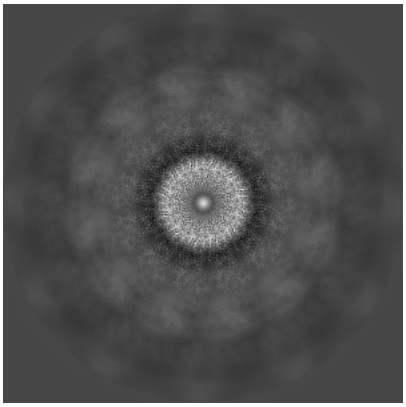
6.1.1 Primary map



X



Y

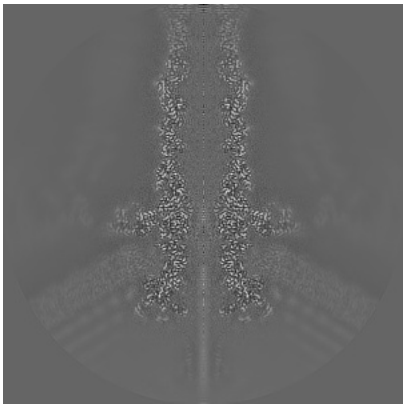


Z

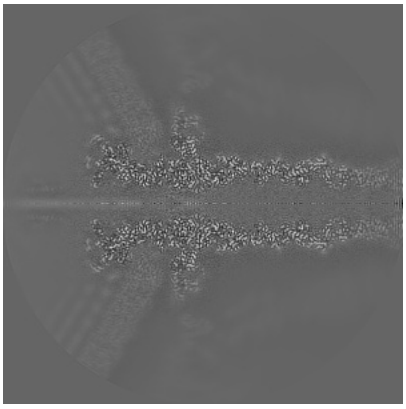
The images above show the map projected in three orthogonal directions.

6.2 Central slices

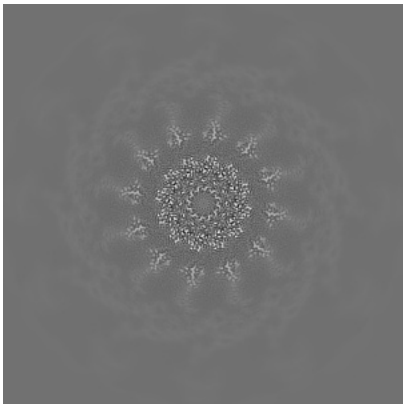
6.2.1 Primary map



X Index: 192



Y Index: 192

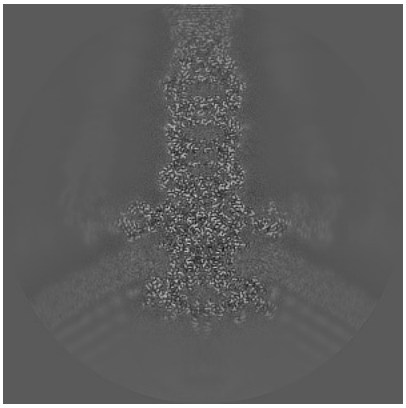


Z Index: 192

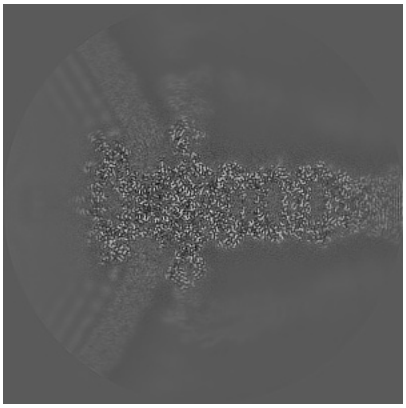
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

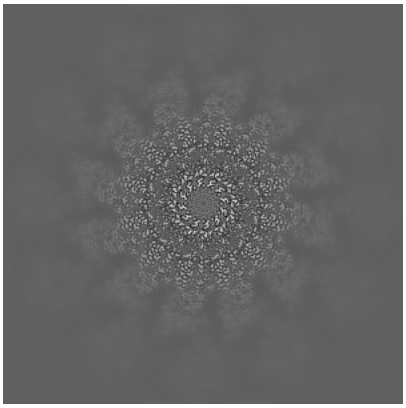
6.3.1 Primary map



X Index: 213



Y Index: 171

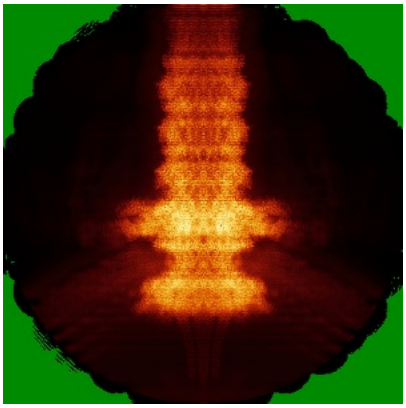


Z Index: 171

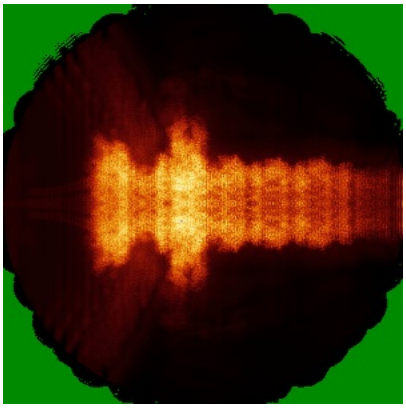
The images above show the largest variance slices of the map in three orthogonal directions.

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

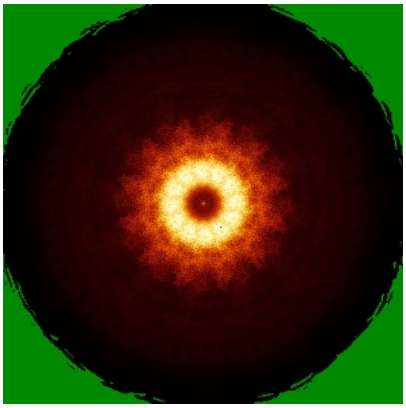
6.4.1 Primary map



X



Y

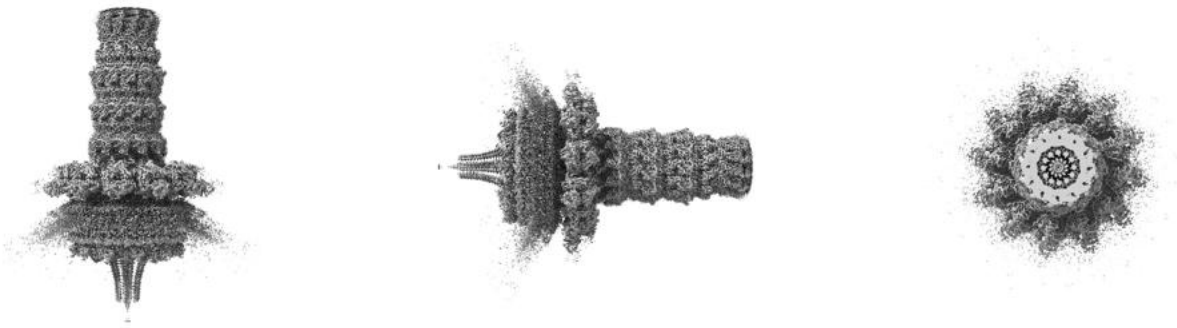


Z

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 0.06. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

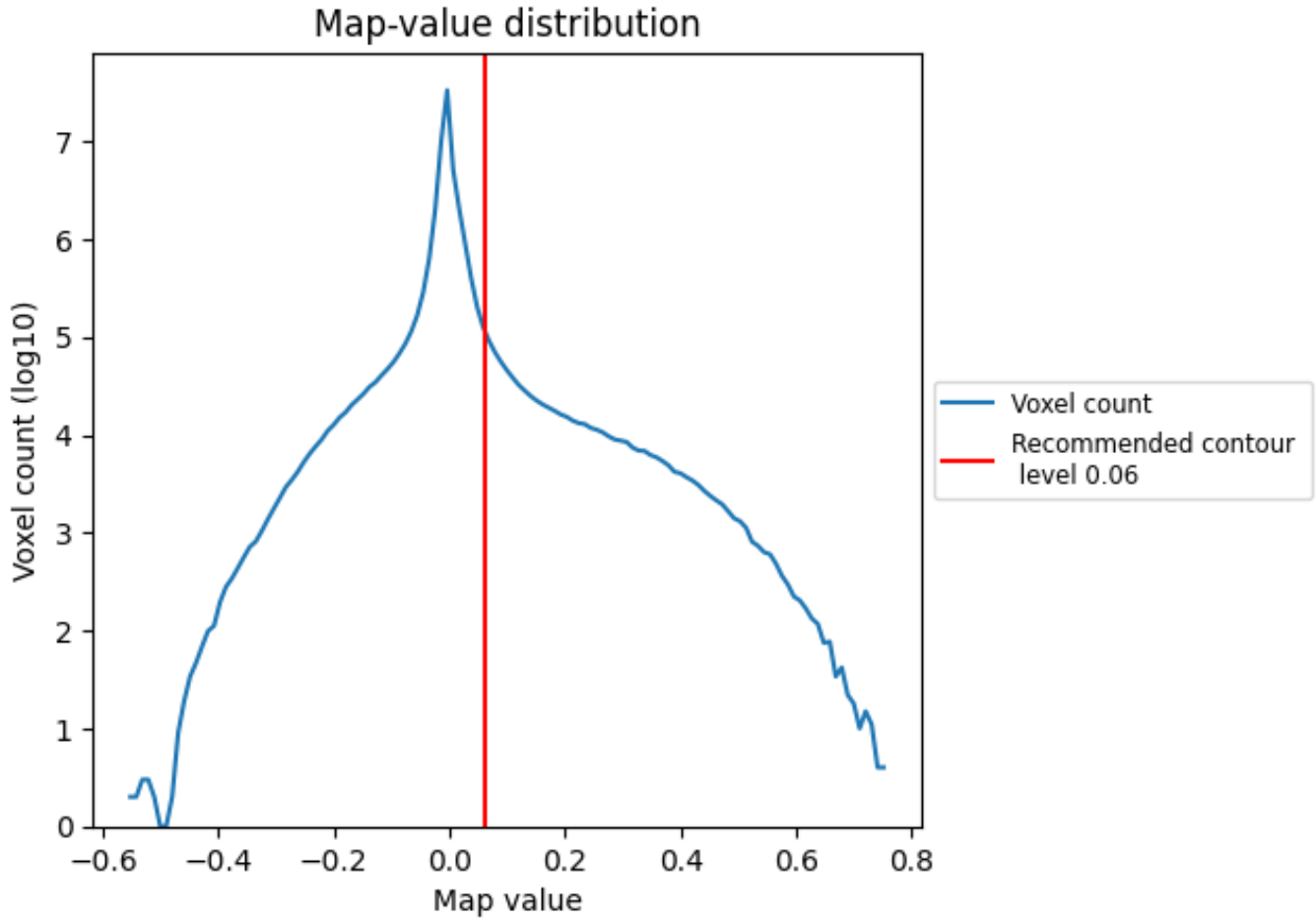
6.6 Mask visualisation (i)

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

7 Map analysis

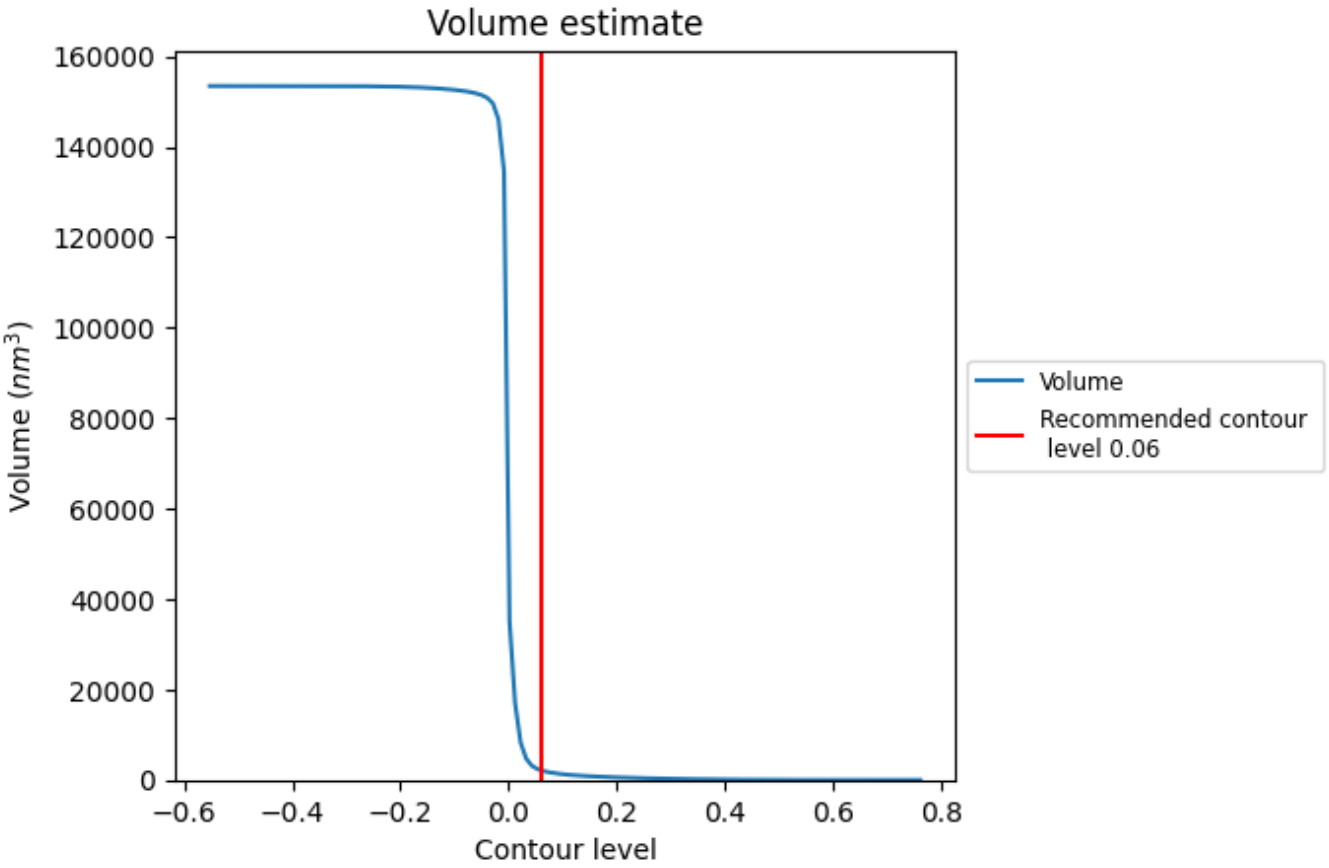
This section contains the results of statistical analysis of the map.

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.

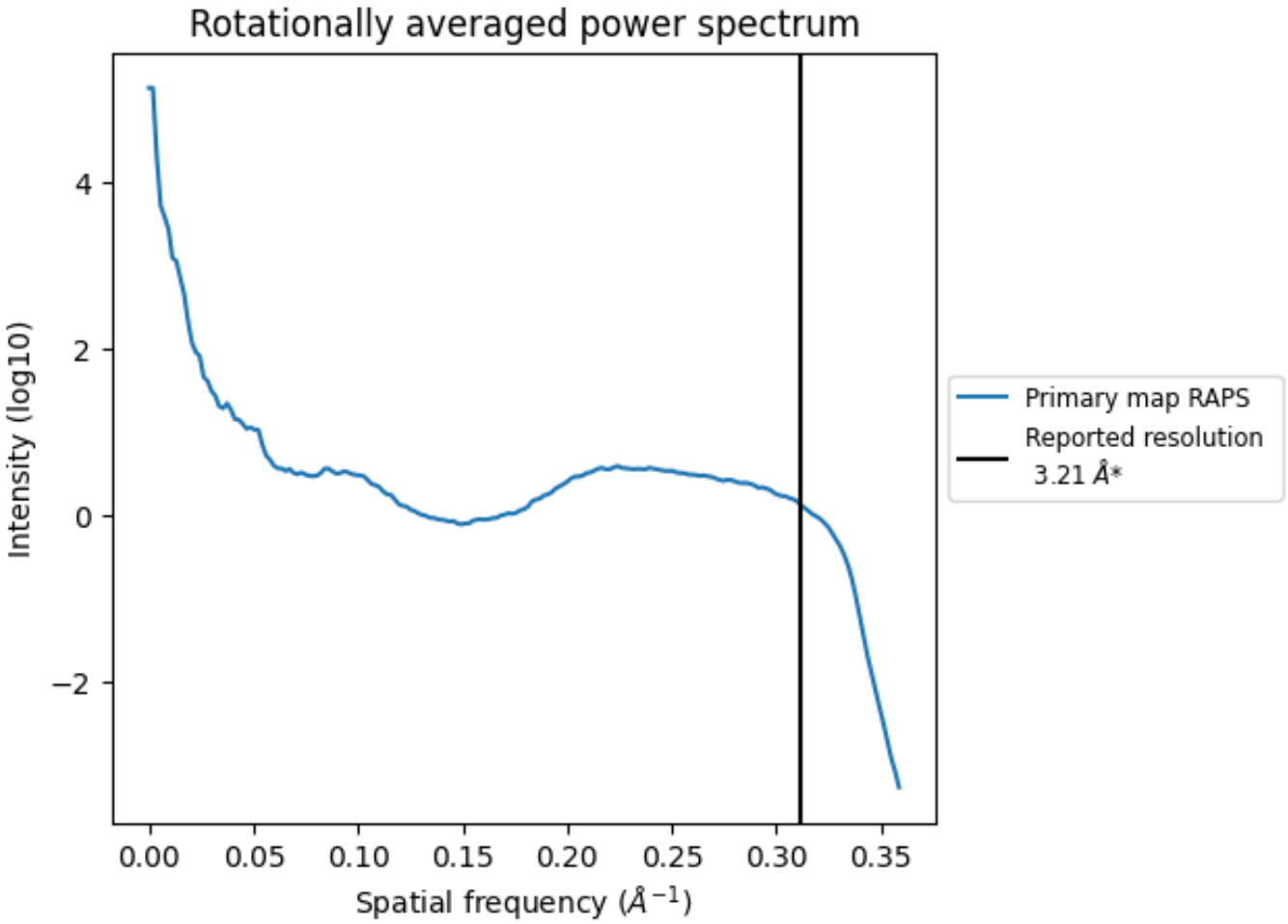
7.2 Volume estimate



The volume at the recommended contour level is 2153 nm^3; this corresponds to an approximate mass of 1945 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)

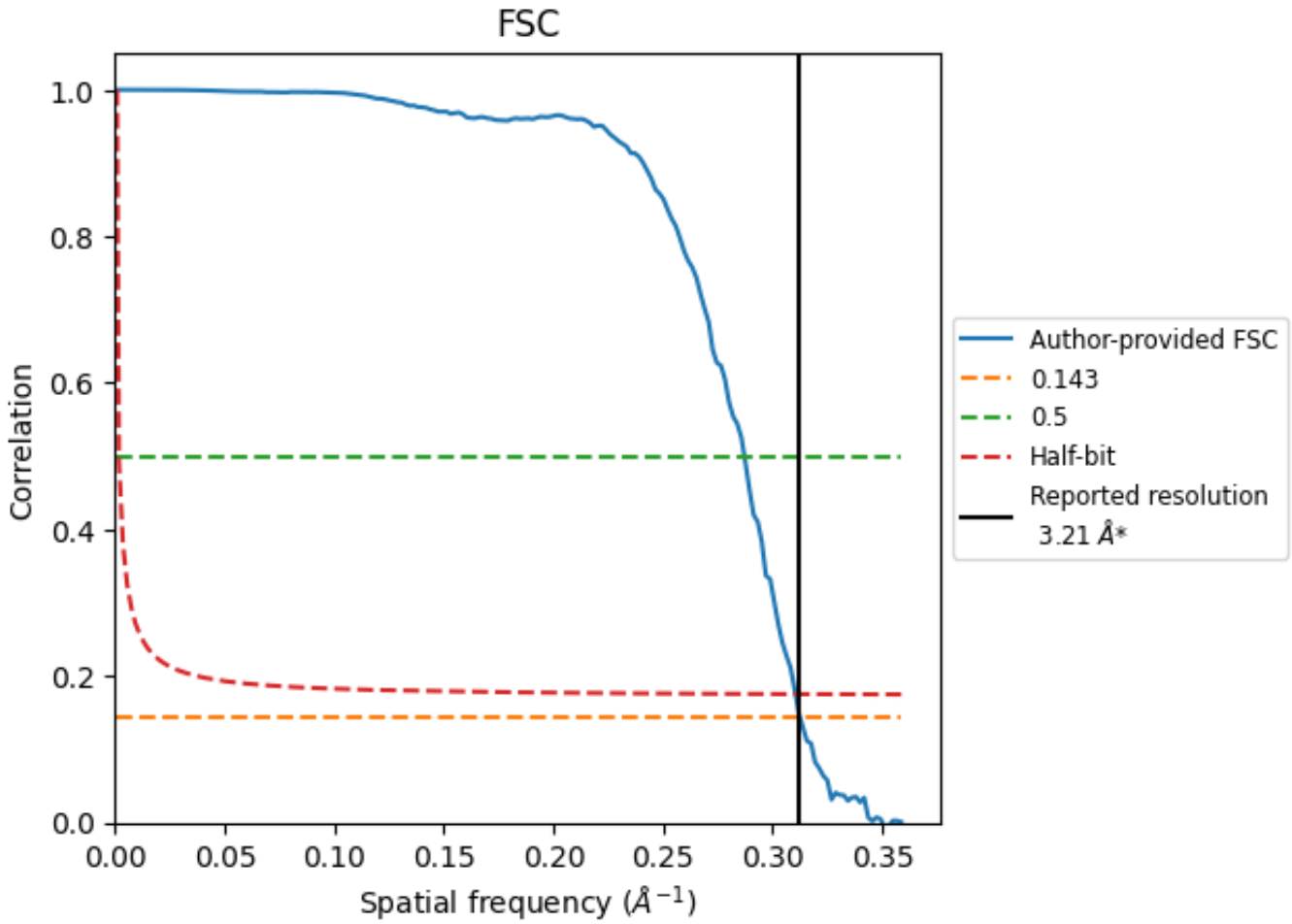


*Reported resolution corresponds to spatial frequency of 0.312 Å^{-1}

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 0.312 Å^{-1}

8.2 Resolution estimates (i)

Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.21	-	-
Author-provided FSC curve	3.20	3.48	3.22
Unmasked-calculated*	-	-	-

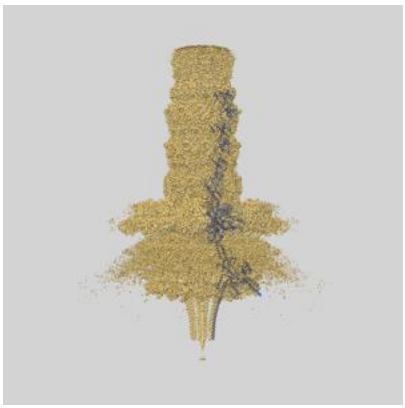
*Resolution estimate based on FSC curve calculated by comparison of deposited half-maps.

9 Map-model fit

This section contains information regarding the fit between EMDB map EMD-14092 and PDB model 7QOJ. Per-residue inclusion information can be found in section 3 on page 6.

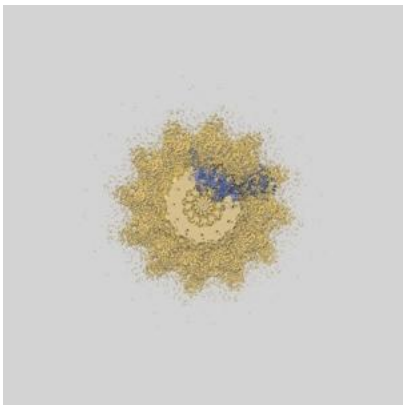
9.1 Map-model overlays

9.1.1 Map-model overlay



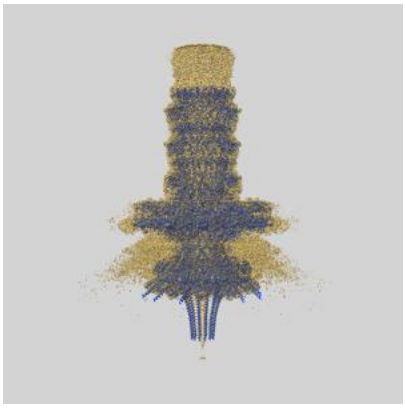
X

Y

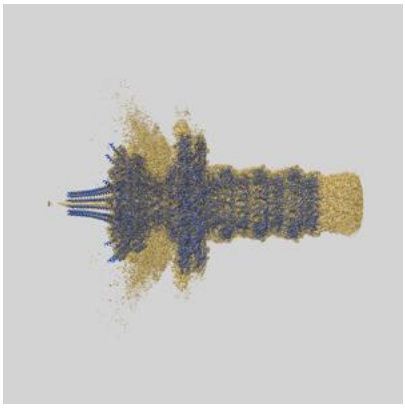


Z

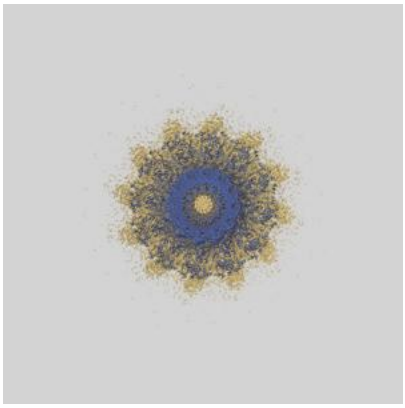
9.1.2 Map-model assembly overlay



X



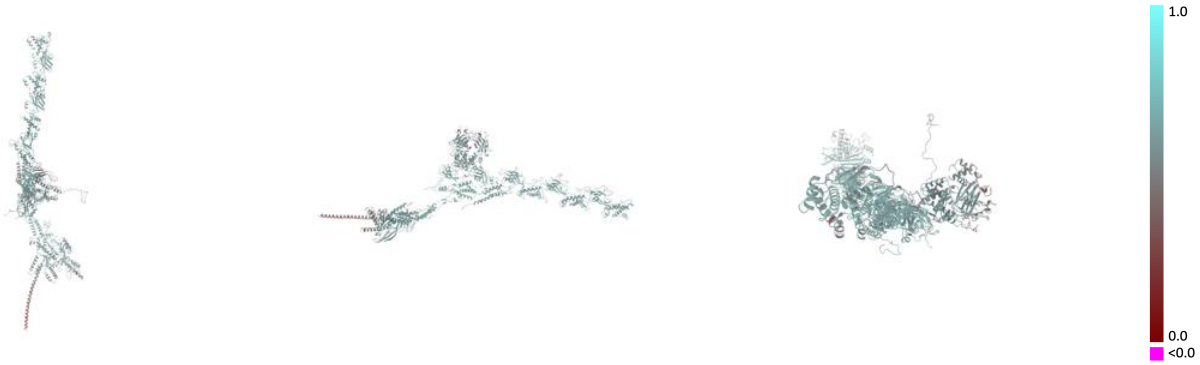
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.06 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.

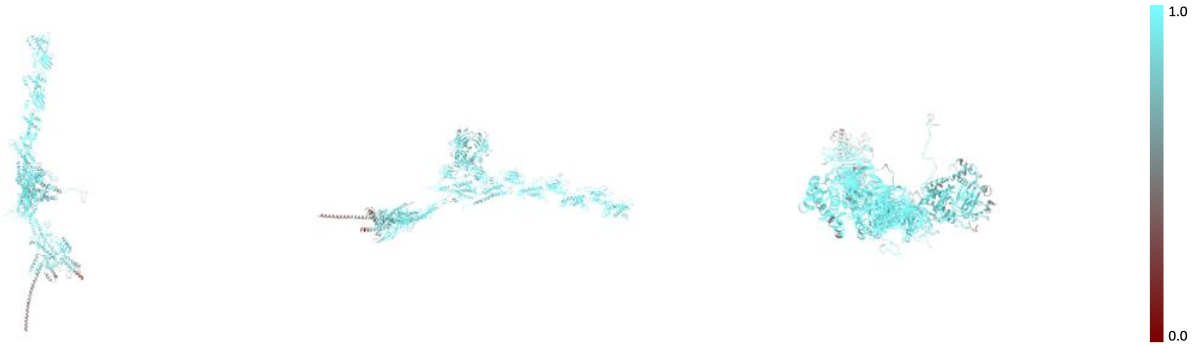
9.2 Q-score mapped to coordinate model (i)



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.

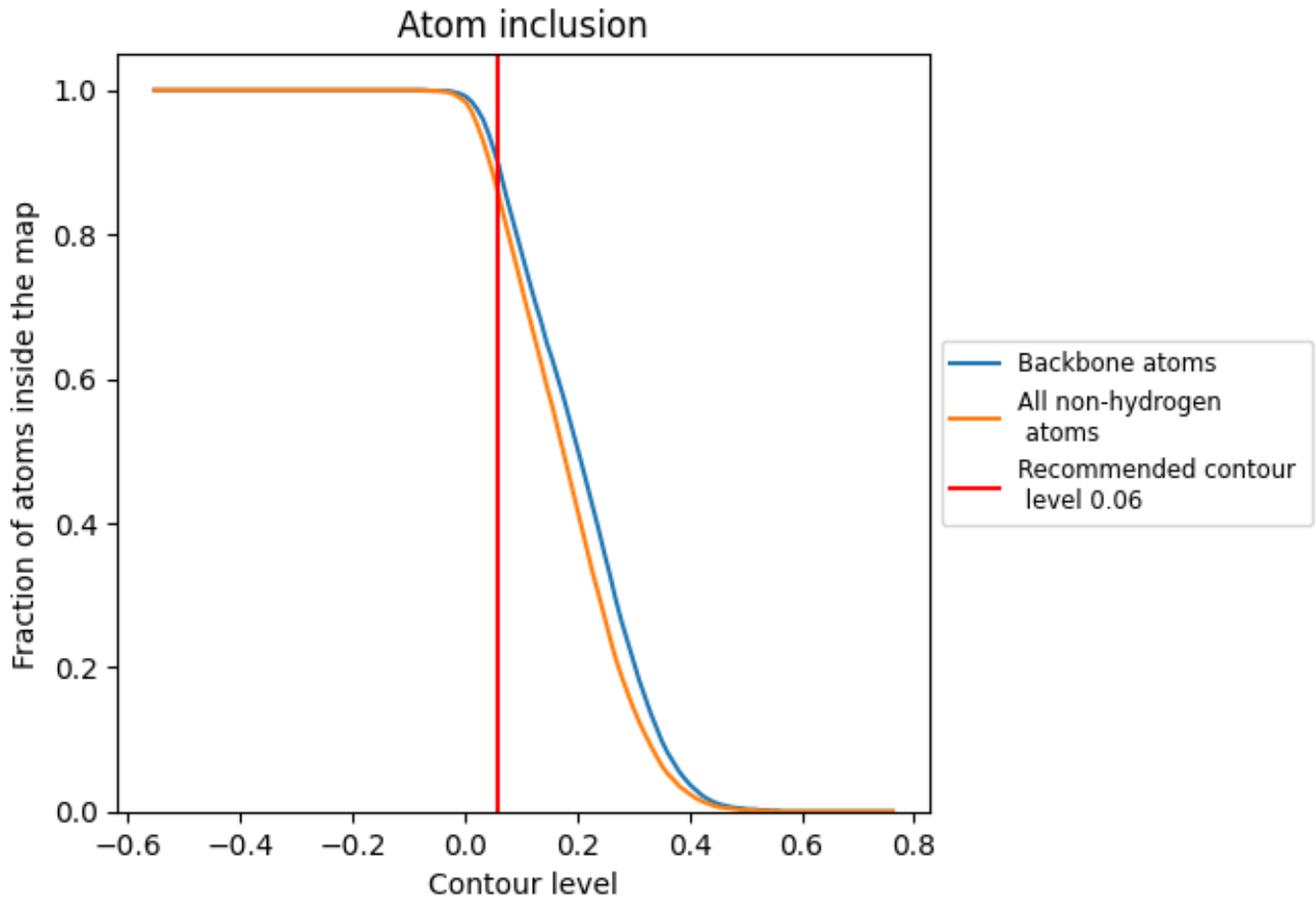
9.3 Atom inclusion mapped to coordinate model (i)



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 (0.06).

9.4 Atom inclusion (i)



At the recommended contour level, 90% of all backbone atoms, 86% 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 (0.06) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.8560	0.5620
A	0.8180	0.5510
B	0.9250	0.5980
C	0.9310	0.5950
D	0.9420	0.5930
E	0.9250	0.5880
F	0.9020	0.5760
G	0.8380	0.5430
H	0.8320	0.5270
I	0.8560	0.5440
J	0.7170	0.4820
K	0.7950	0.5120
L	0.7250	0.5230
M	0.6130	0.4970
N	0.7710	0.5300