wwPDB EM Validation Summary Report (i)

Oct 28, 2024 – 10:34 pm GMT

PDB ID	:	7YX4
EMDB ID	:	EMD-14354
Title	:	Structure of the Mimivirus genomic fibre in its compact 5-start helix form
Authors	:	Villalta, A.; Schmitt, A.; Estrozi, L.F.; Quemin, E.R.J.; Alempic, J.M.; Lartigue, A.; Prazak, V.; Belmudes, L.; Vasishtan, D.; Colmant, A.M.G.; Honore, F.A.; Coute, Y.; Grunewald, K.; Abergel, C.
Deposited on	:	2022-02-15
Resolution	:	3.70 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 (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:

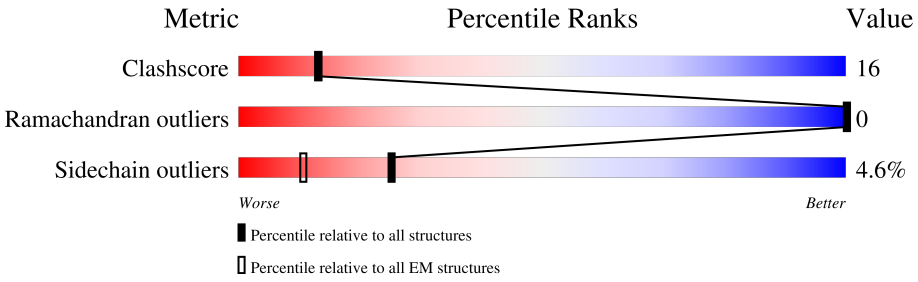
EMDB validation analysis	:	0.0.1.dev113
Mogul	:	1.8.4, CSD as541be (2020)
MolProbity	:	4.02b-467
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
MapQ	:	1.9.13
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

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 3.70 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	210492	15764
Ramachandran outliers	207382	16835
Sidechain outliers	206894	16415

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	702	5% 61% 30% · 7%
1	B	702	6% 60% 30% · 7%

2 Entry composition (i)

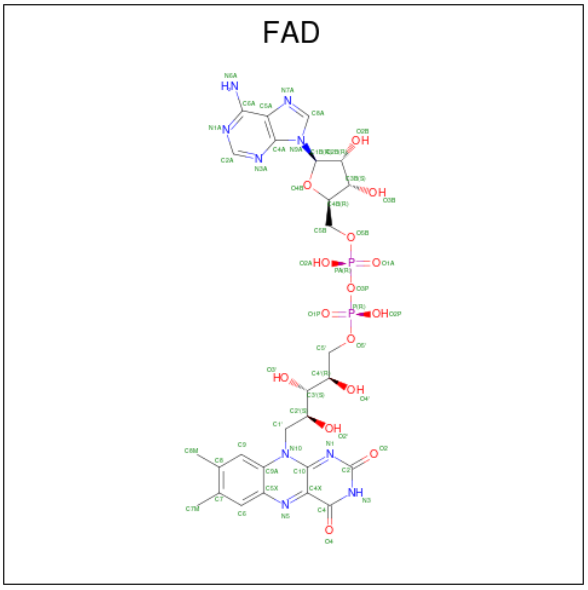
There are 2 unique types of molecules in this entry. The entry contains 10072 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Putative glucose-methanol-choline oxidoreductase protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	650	Total	C	N	O	S	0	0
			4983	3184	836	947	16		
1	B	650	Total	C	N	O	S	0	0
			4983	3184	836	947	16		

- Molecule 2 is FLAVIN-ADENINE DINUCLEOTIDE (three-letter code: FAD) (formula: $C_{27}H_{33}N_9O_{15}P_2$) (labeled as "Ligand of Interest" by depositor).



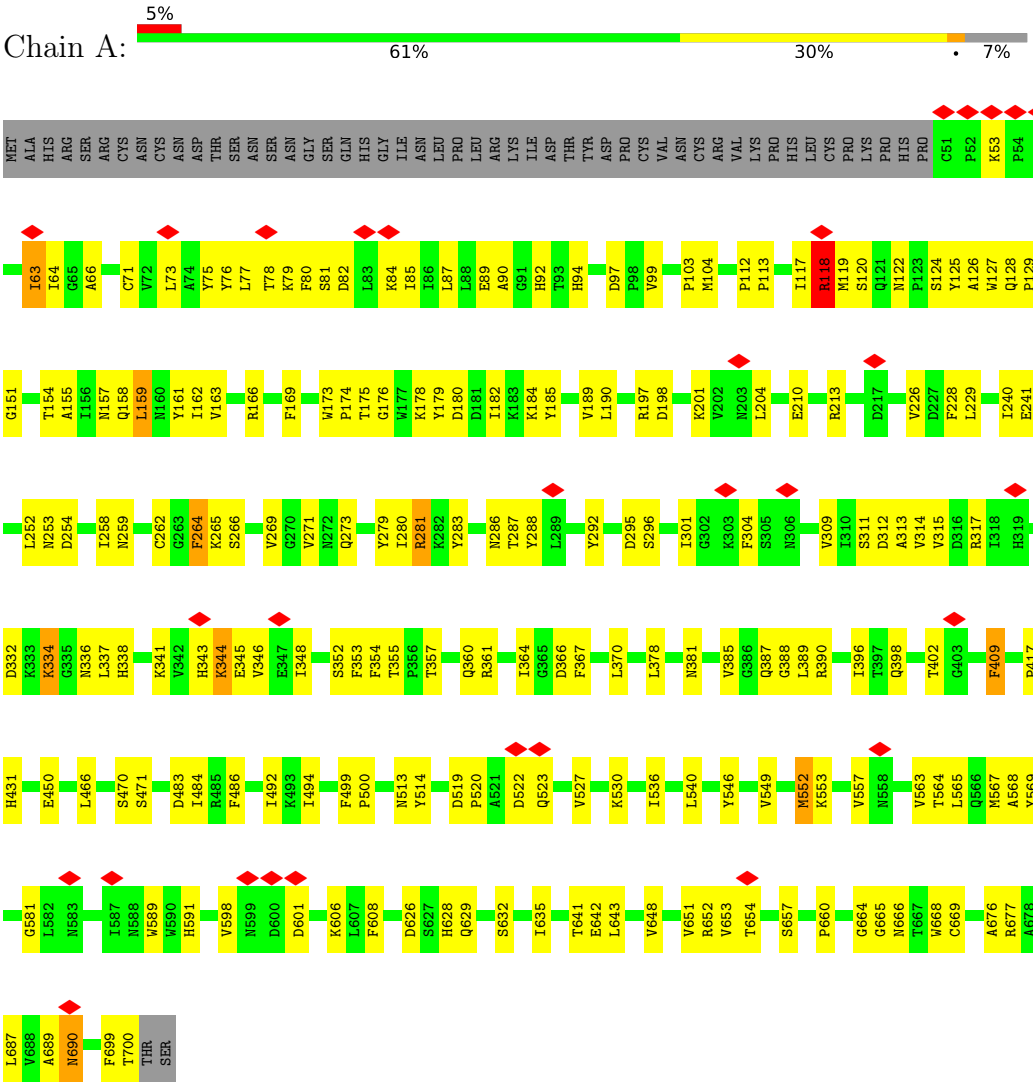
Mol	Chain	Residues	Atoms					AltConf
2	A	1	Total	C	N	O	P	0
			53	27	9	15	2	
2	B	1	Total	C	N	O	P	0
			53	27	9	15	2	

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.

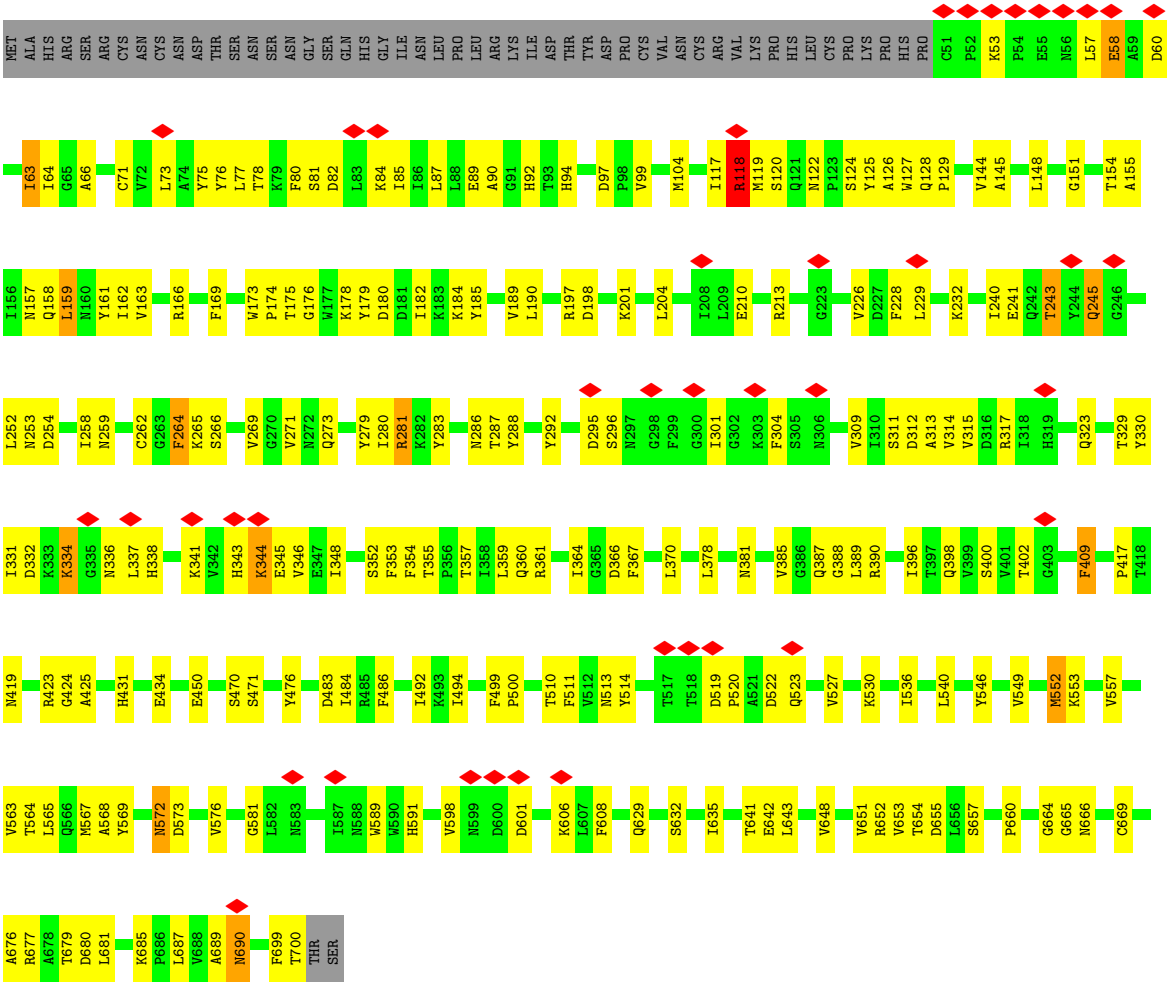
• Molecule 1: Putative glucose-methanol-choline oxidoreductase protein

Chain A:



• Molecule 1: Putative glucose-methanol-choline oxidoreductase protein

Chain B:


MET ALA HIS ARG SER ARG CYS ASN CYS ASN ASP THR SER ASN SER ASN GLY SER GLN HIS GLY ILE ASN LEU PRO LEU ARG LYS ILE ASP THR TYR ASP PRO CYS VAL ASN CYS ARG VAL LYS PRO HIS LEU CYS PRO LYS PRO HIS PRO C51 P52 K53 P54 E55 N56 L57 E58 A59 D60
I63 I64 G65 A66 C71 V72 L73 A74 Y75 Y76 L77 T78 K79 F80 S81 D82 L83 K84 I85 I86 L87 L88 E89 A90 G91 H92 T93 H94 D97 P98 V99 M104 I117 R118 M119 S120 Q121 N122 P123 S124 Y125 A126 W127 Q128 P129 V144 A145 L148 G151 T154 A155
I156 N157 Q158 L159 N160 Y161 I162 V163 R166 F169 W173 P174 T175 G176 W177 K178 Y179 D180 D181 I182 K183 K184 Y185 V189 L190 R197 D198 K201 L204 I208 L209 E210 R213 G223 V226 D227 F228 L229 K232 I240 E241 Q242 T243 Y244 Q245 G246
L252 N253 D254 I258 N259 C262 G263 F264 K265 S266 V269 G270 V271 N272 Q273 Y279 I280 R281 K282 Y283 N286 T287 Y288 Y292 D295 S296 N297 G298 F299 G300 I301 G302 K303 F304 S305 N306 V309 I310 S311 D312 A313 V314 V315 D316 R317 I318 H319 Q323 T329 Y330
I331 D332 K333 K334 G335 N336 L337 H338 K341 V342 H343 K344 E345 V346 E347 I348 S352 F353 F354 T355 P356 T357 I358 L359 Q360 R361 I364 G365 D366 F367 L370 L378 N381 V385 G386 Q387 G388 L389 R390 I396 T397 Q398 V399 S400 V401 T402 G403 F409 P417 T418
N419 R423 G424 A425 H431 E434 E450 S470 S471 Y476 D483 I484 R485 F486 I492 K493 I494 F499 P500 T510 F511 V512 N513 Y514 T517 T518 D519 P520 A521 D522 Q523 V527 K530 I536 L540 Y546 V549 N552 K553 V557
V563 T564 L565 Q566 M567 A568 Y569 N572 D573 V576 G581 L582 N583 I587 N588 W589 W590 H591 V598 N599 D600 D601 K606 L607 F608 Q629 S632 I635 T641 E642 L643 V648 V651 R652 V653 T654 D655 L656 S657 P660 G664 G665 N666 C669
A676 R677 A678 T679 D680 L681 K685 P686 L687 V688 A689 N690 F699 T700 THR SER

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	HELICAL	Depositor
Imposed symmetry	HELICAL, twist=138.925°, rise=7.93 Å, axial sym=C1	Depositor
Number of segments used	121429	Depositor
Resolution determination method	FSC 0.5 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	50.6	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	3000	Depositor
Magnification	Not provided	
Image detector	GATAN K2 QUANTUM (4k x 4k)	Depositor
Maximum map value	0.101	Depositor
Minimum map value	-0.049	Depositor
Average map value	0.001	Depositor
Map value standard deviation	0.005	Depositor
Recommended contour level	0.02	Depositor
Map size (Å)	414.2, 414.2, 414.2	wwPDB
Map dimensions	380, 380, 380	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.09, 1.09, 1.09	Depositor

5 Model quality

5.1 Standard geometry

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

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	0.36	0/5112	0.54	2/6982 (0.0%)
1	B	0.36	0/5112	0.54	2/6982 (0.0%)
All	All	0.36	0/10224	0.54	4/13964 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	118	ARG	CA-CB-CG	6.27	127.20	113.40
1	B	118	ARG	CA-CB-CG	6.26	127.17	113.40
1	A	159	LEU	CB-CG-CD1	-5.18	102.19	111.00
1	B	159	LEU	CB-CG-CD1	-5.15	102.25	111.00

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

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

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	4983	0	4881	160	0
1	B	4983	0	4881	165	0
2	A	53	0	31	4	0
2	B	53	0	31	4	0
All	All	10072	0	9824	321	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 16.

The worst 5 of 321 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:58:GLU:HB3	1:A:341:LYS:HB3	1.53	0.90
1:B:58:GLU:HB3	1:B:341:LYS:HB3	1.53	0.89
1:B:197:ARG:HH22	1:B:265:LYS:HE3	1.41	0.86
1:A:197:ARG:HH22	1:A:265:LYS:HE3	1.41	0.83
1:B:90:ALA:HA	1:B:313:ALA:H	1.49	0.77

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	648/702 (92%)	590 (91%)	58 (9%)	0	100	100
1	B	648/702 (92%)	591 (91%)	57 (9%)	0	100	100
All	All	1296/1404 (92%)	1181 (91%)	115 (9%)	0	100	100

There are no Ramachandran outliers to report.

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	542/591 (92%)	517 (95%)	25 (5%)	23	49
1	B	542/591 (92%)	517 (95%)	25 (5%)	23	49
All	All	1084/1182 (92%)	1034 (95%)	50 (5%)	25	49

5 of 50 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	64	ILE
1	B	245	GLN
1	B	690	ASN
1	B	71	CYS
1	B	175	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	398	GLN
1	B	398	GLN

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 oligosaccharides in this entry.

5.6 Ligand geometry (i)

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	FAD	A	901	-	53,58,58	0.59	1 (1%)	68,89,89	0.54	2 (2%)
2	FAD	B	901	-	53,58,58	0.59	1 (1%)	68,89,89	0.53	2 (2%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	FAD	A	901	-	-	17/30/50/50	0/6/6/6
2	FAD	B	901	-	-	17/30/50/50	0/6/6/6

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	901	FAD	P-O2P	-2.13	1.45	1.55
2	B	901	FAD	P-O2P	-2.13	1.45	1.55

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	901	FAD	C5A-C6A-N6A	2.34	123.90	120.35
2	B	901	FAD	C5A-C6A-N6A	2.30	123.85	120.35
2	B	901	FAD	P-O3P-PA	-2.19	125.30	132.83
2	A	901	FAD	P-O3P-PA	-2.19	125.31	132.83

There are no chirality outliers.

5 of 34 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	901	FAD	C5B-O5B-PA-O1A
2	A	901	FAD	C5B-O5B-PA-O2A
2	A	901	FAD	C1'-C2'-C3'-O3'
2	A	901	FAD	C1'-C2'-C3'-C4'
2	A	901	FAD	O2'-C2'-C3'-O3'

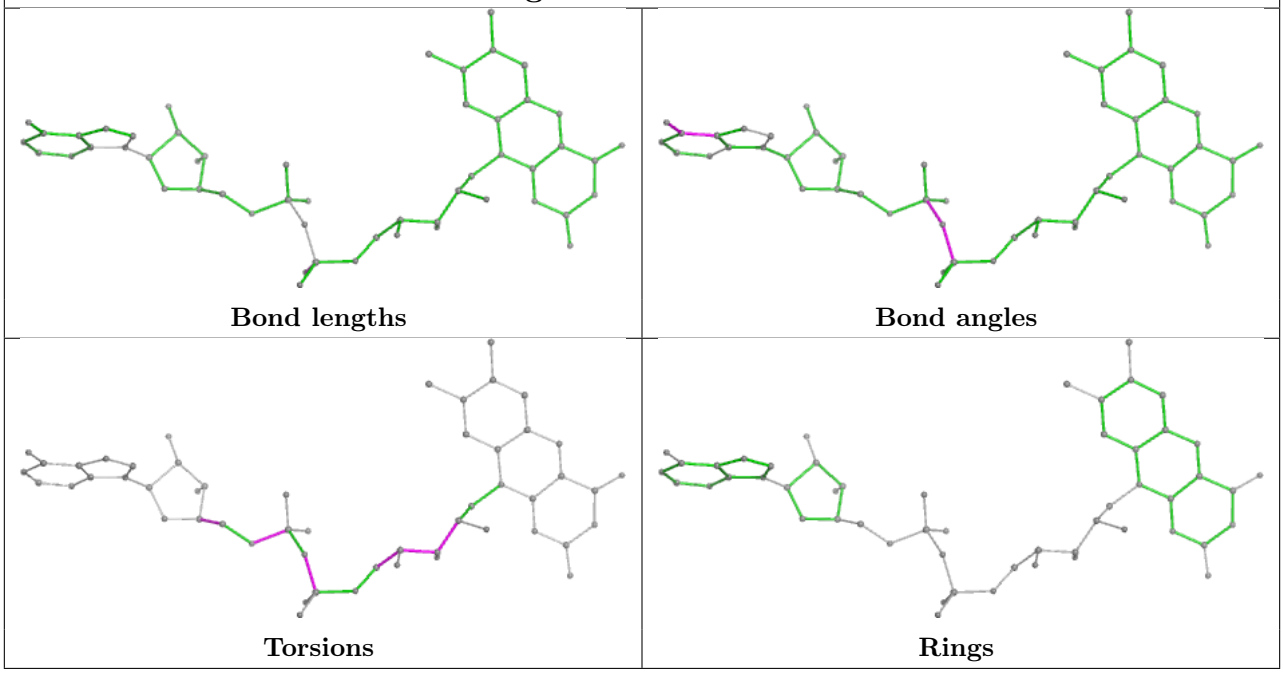
There are no ring outliers.

2 monomers are involved in 8 short contacts:

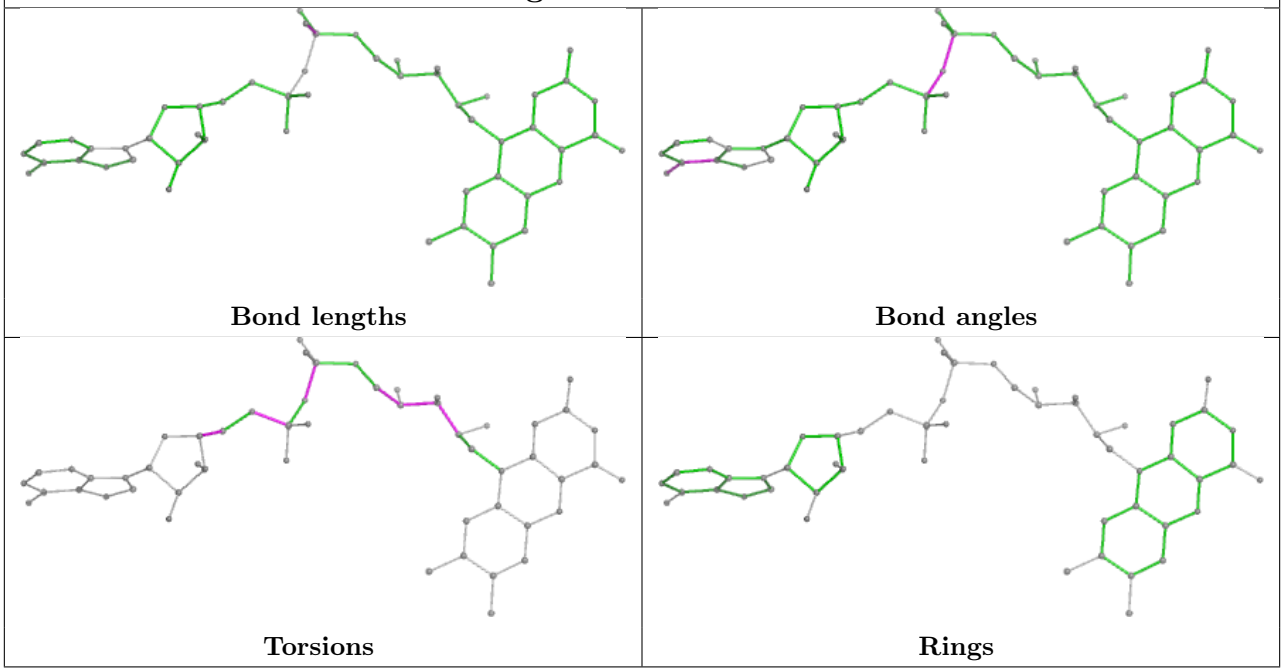
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	901	FAD	4	0
2	B	901	FAD	4	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand FAD A 901



Ligand FAD B 901



5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

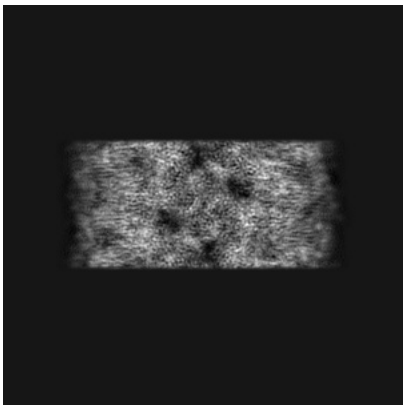
6 Map visualisation

This section contains visualisations of the EMDB entry EMD-14354. These allow visual inspection of the internal detail of the map and identification of artifacts.

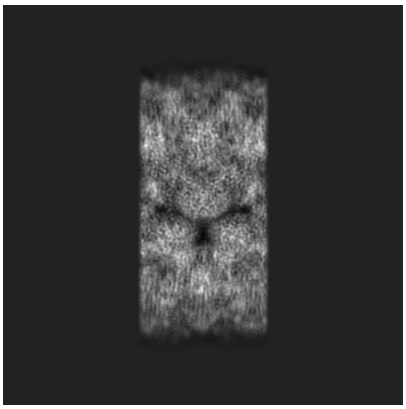
Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

6.1 Orthogonal projections

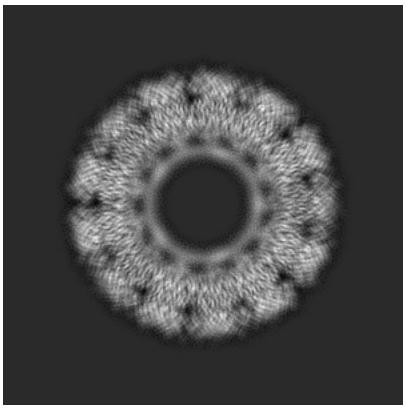
6.1.1 Primary map



X

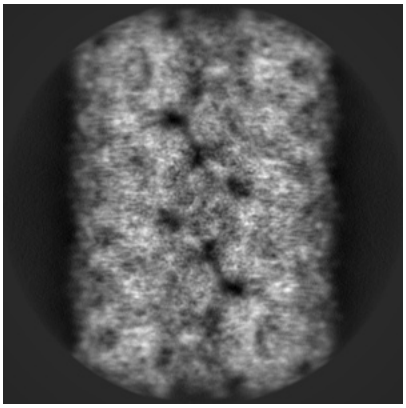


Y

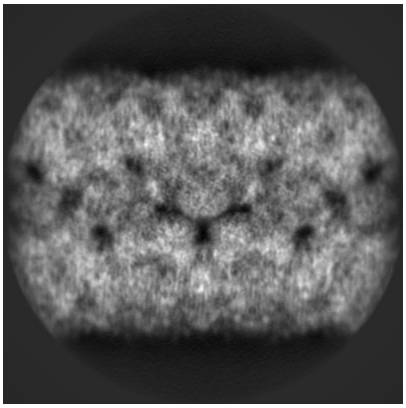


Z

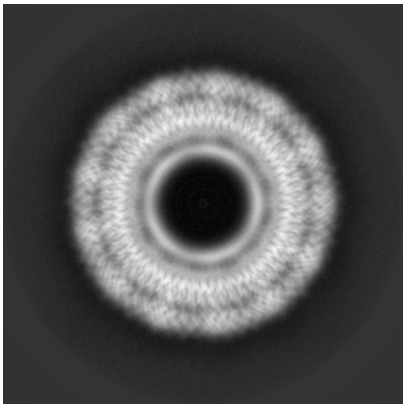
6.1.2 Raw map



X



Y

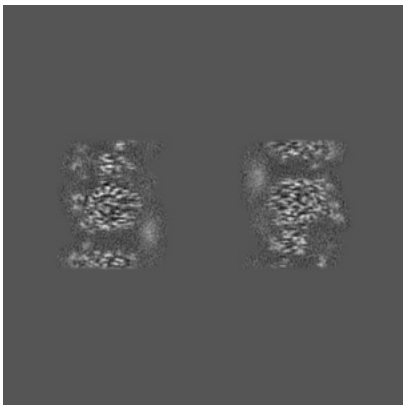


Z

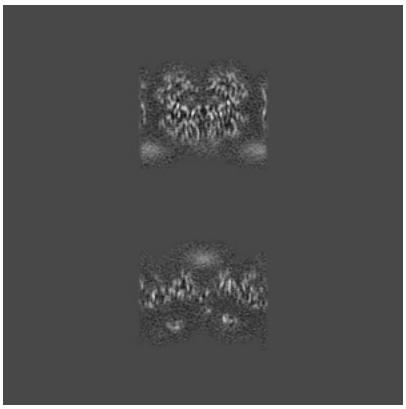
The images above show the map projected in three orthogonal directions.

6.2 Central slices (i)

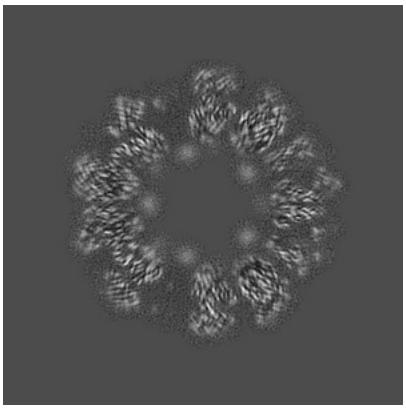
6.2.1 Primary map



X Index: 190

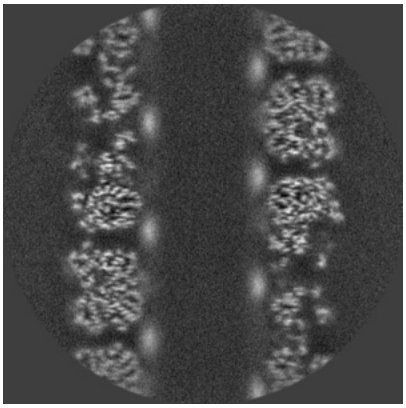


Y Index: 190

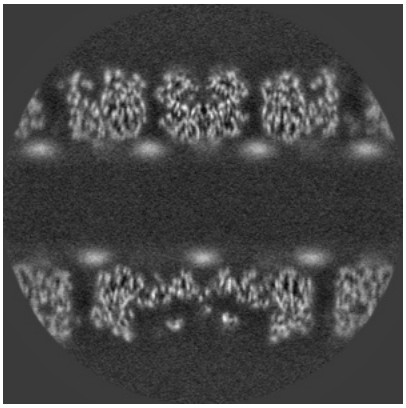


Z Index: 190

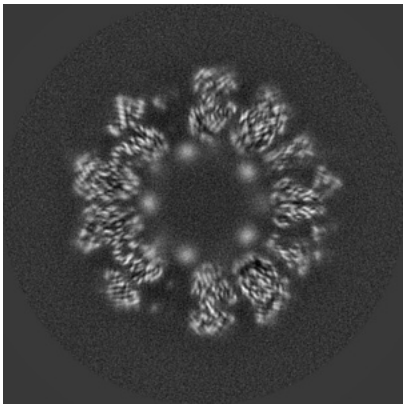
6.2.2 Raw map



X Index: 190



Y Index: 190

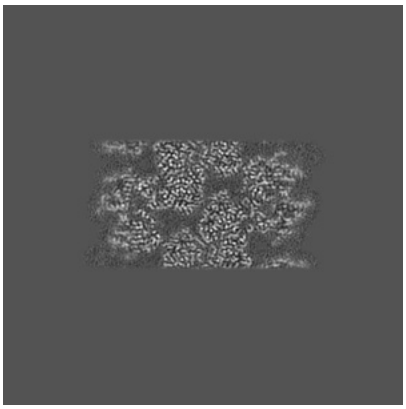


Z Index: 190

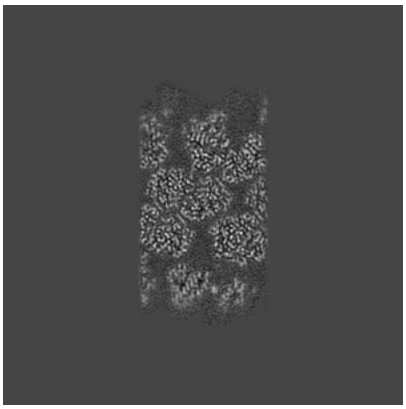
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

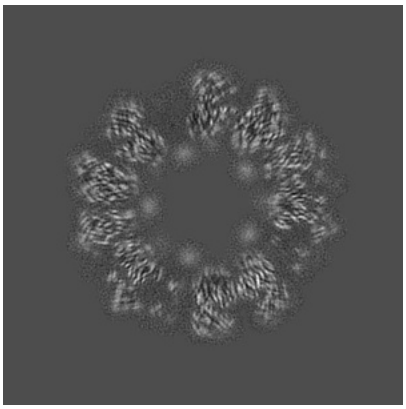
6.3.1 Primary map



X Index: 110

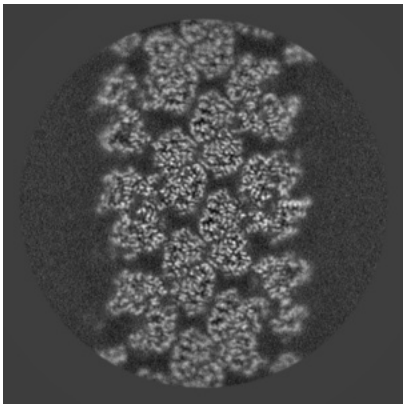


Y Index: 269

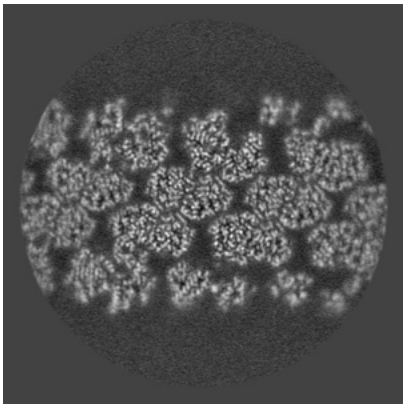


Z Index: 186

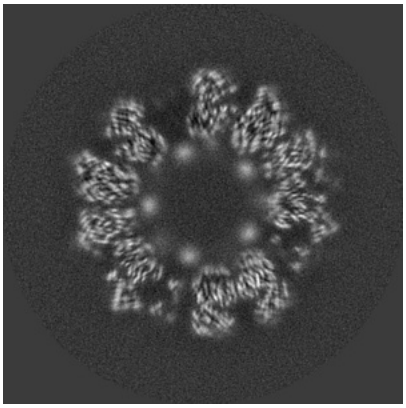
6.3.2 Raw map



X Index: 112



Y Index: 269

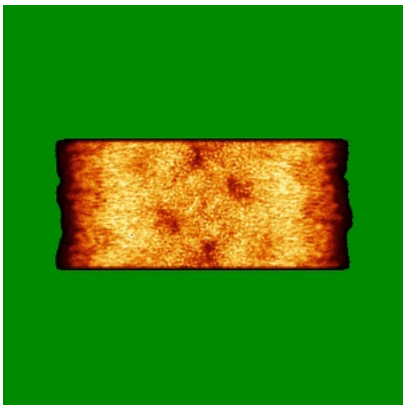


Z Index: 186

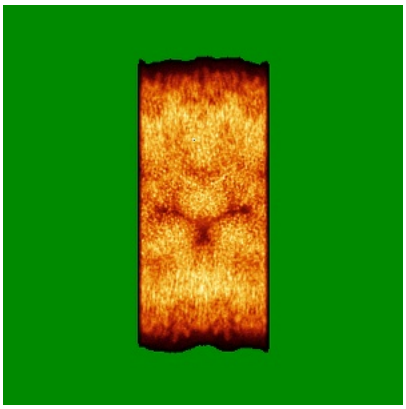
The images above show the largest variance slices of the map in three orthogonal directions.

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

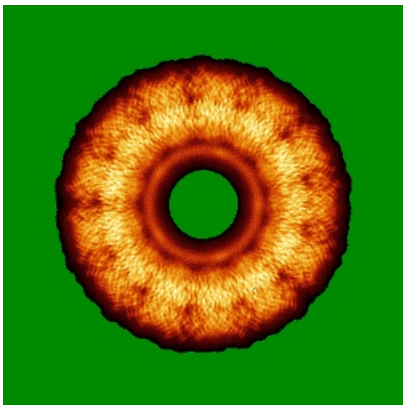
6.4.1 Primary map



X

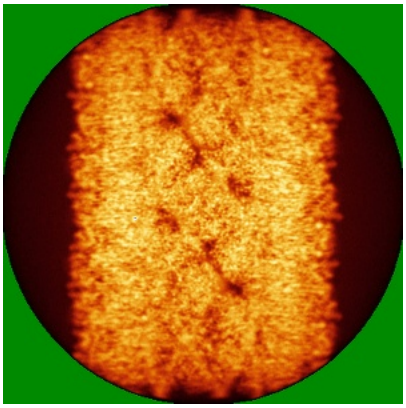


Y

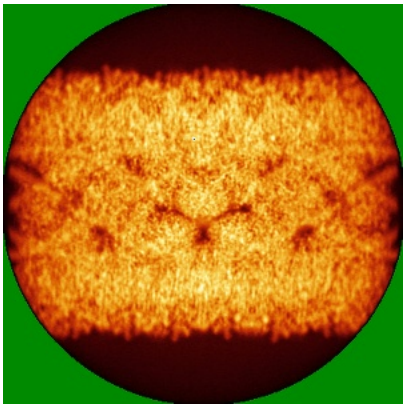


Z

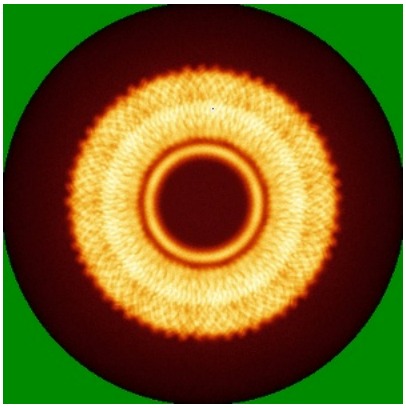
6.4.2 Raw map



X



Y

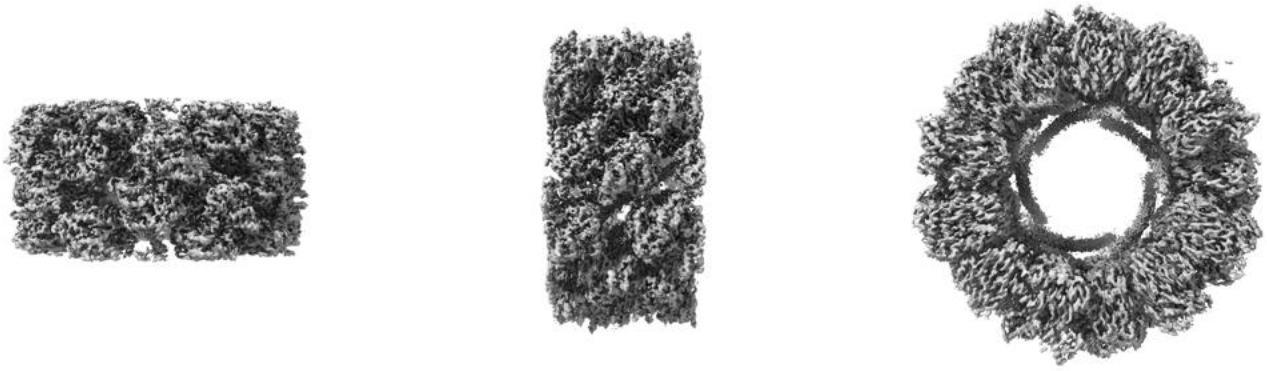


Z

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)

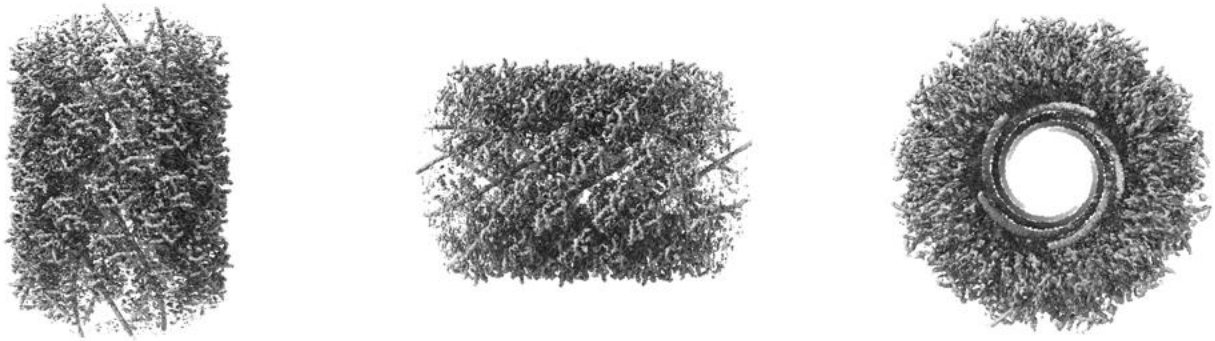
6.5.1 Primary map



X Y Z

The images above show the 3D surface view of the map at the recommended contour level 0.02. 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.

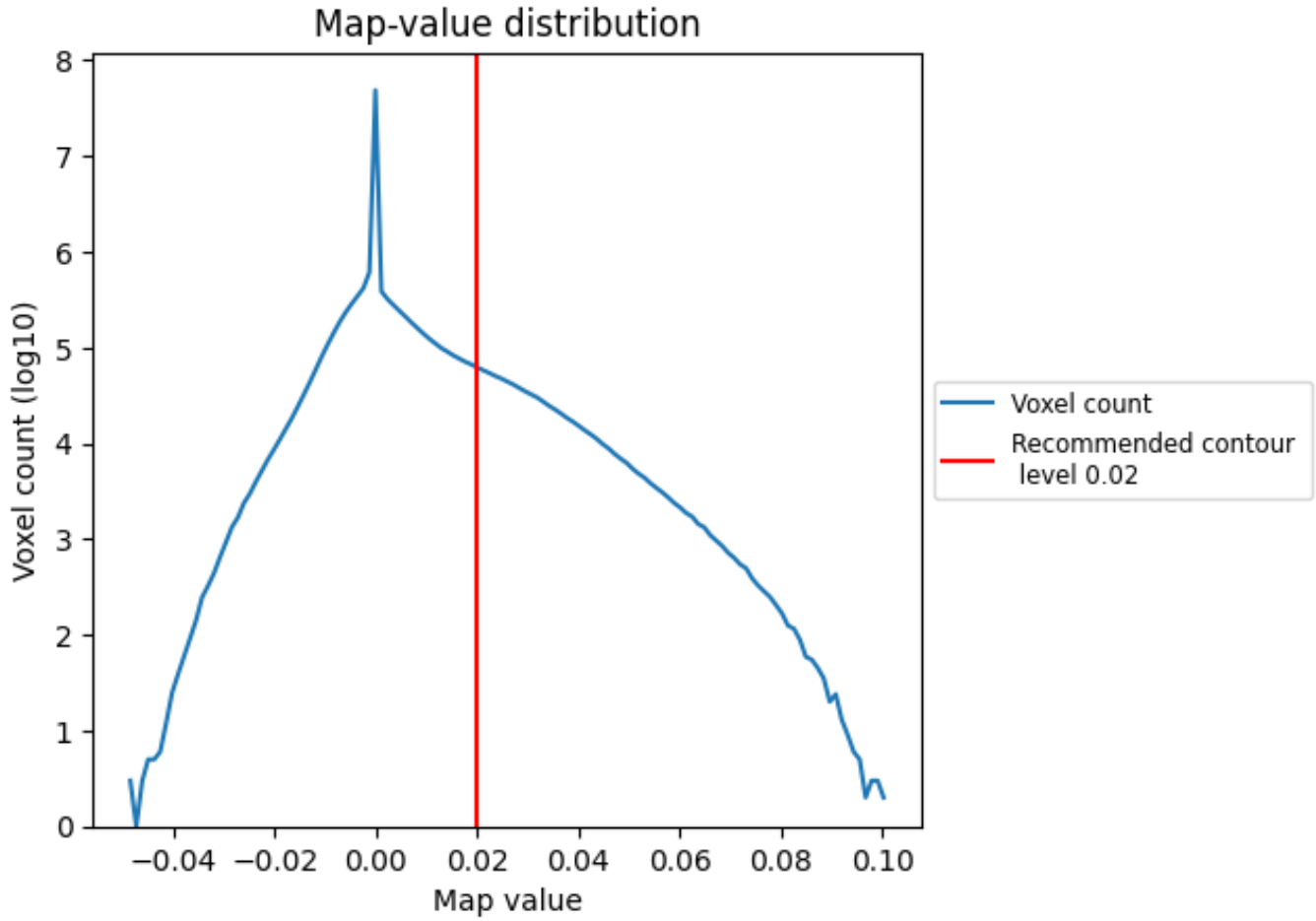
6.6 Mask visualisation (i)

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

7 Map analysis

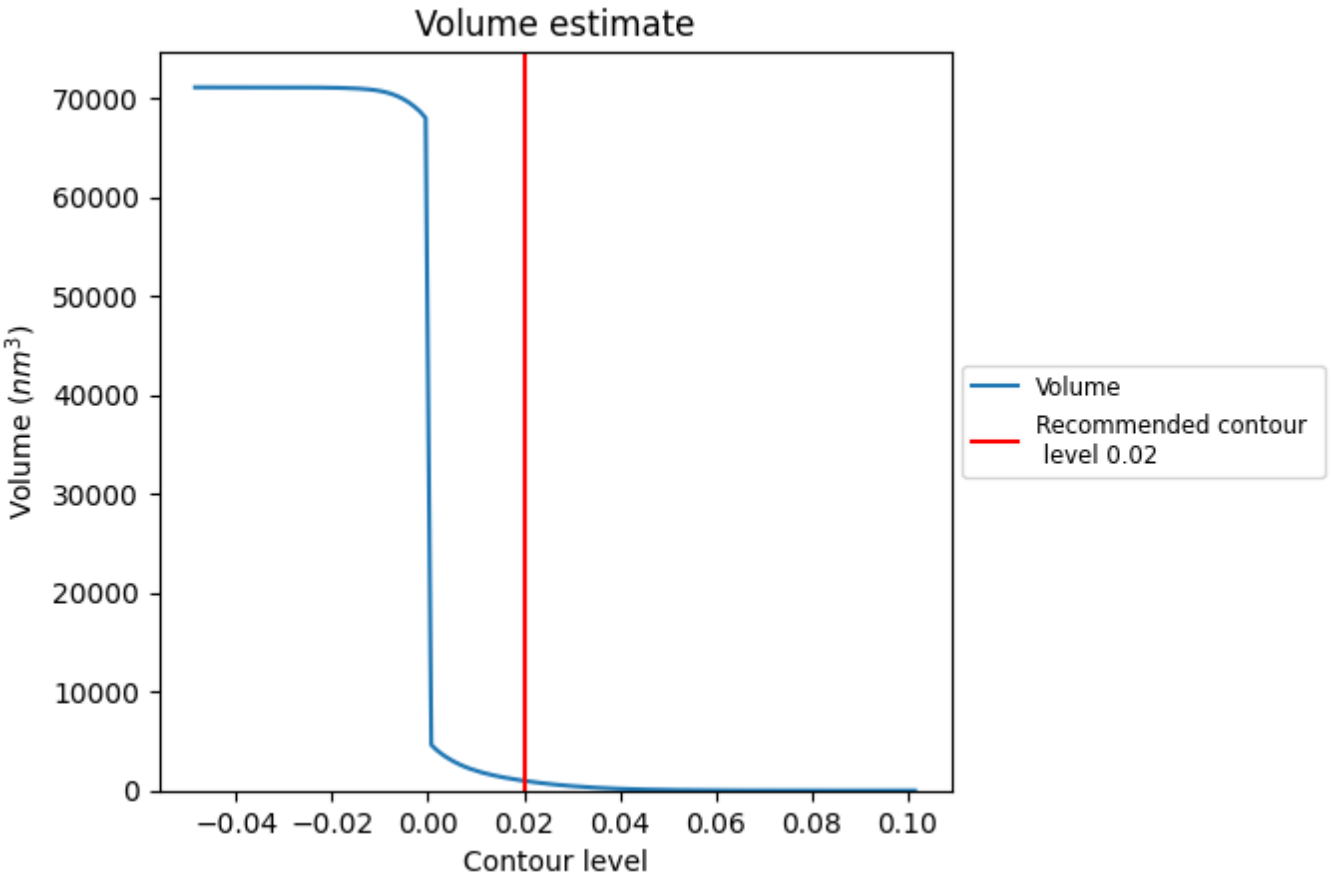
This section contains the results of statistical analysis of the map.

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.

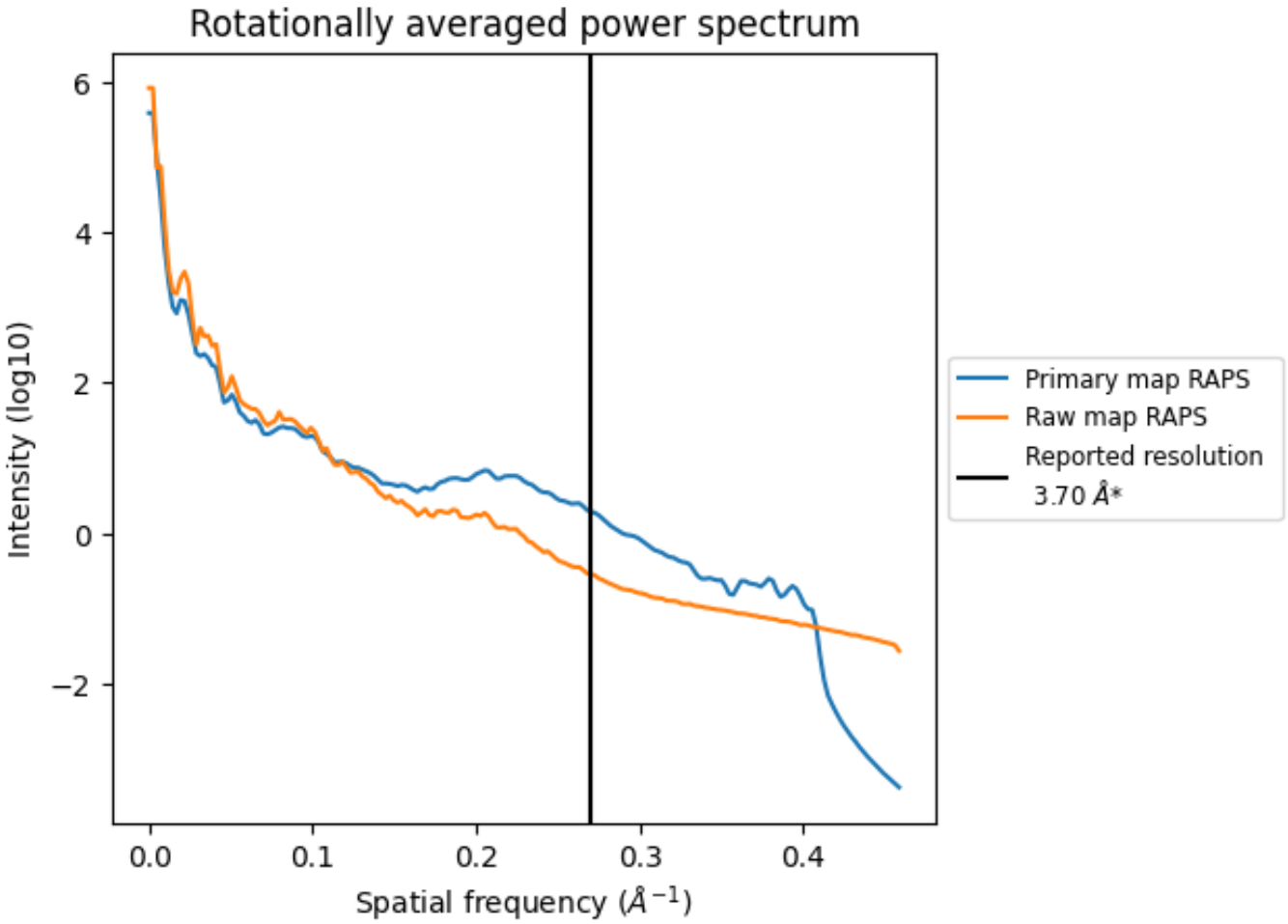
7.2 Volume estimate



The volume at the recommended contour level is 989 nm^3; this corresponds to an approximate mass of 894 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)

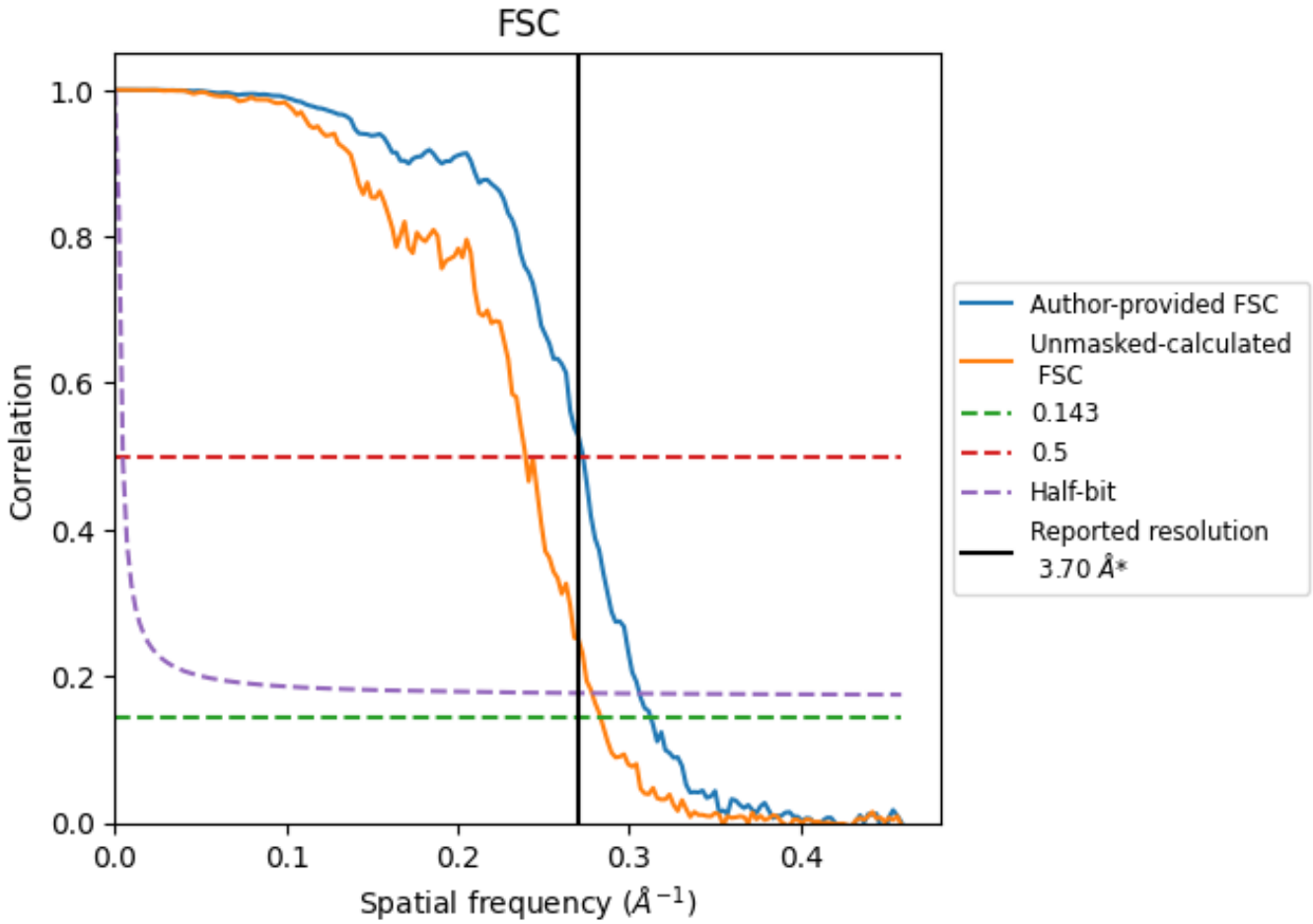


*Reported resolution corresponds to spatial frequency of 0.270 $Å^{-1}$

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 0.270 Å^{-1}

8.2 Resolution estimates (i)

Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	-	3.70	-
Author-provided FSC curve	3.20	3.66	3.27
Unmasked-calculated*	3.53	4.18	3.59

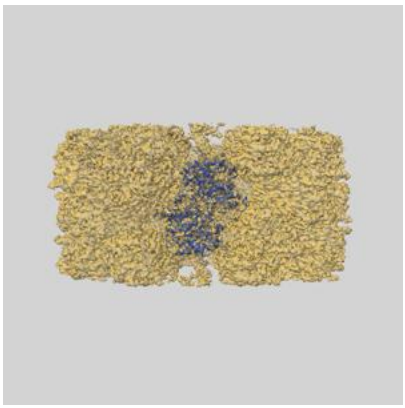
*Resolution estimate based on FSC curve calculated by comparison of deposited half-maps. The value from deposited half-maps intersecting FSC 0.5 CUT-OFF 4.18 differs from the reported value 3.7 by more than 10 %

9 Map-model fit

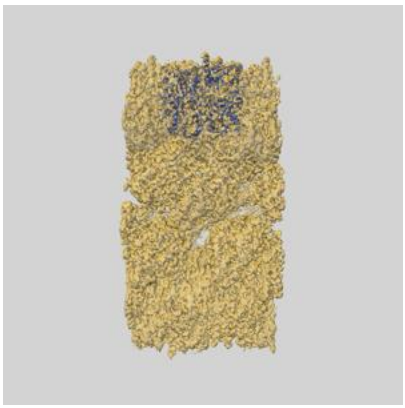
This section contains information regarding the fit between EMDB map EMD-14354 and PDB model 7YX4. Per-residue inclusion information can be found in section 3 on page 4.

9.1 Map-model overlays

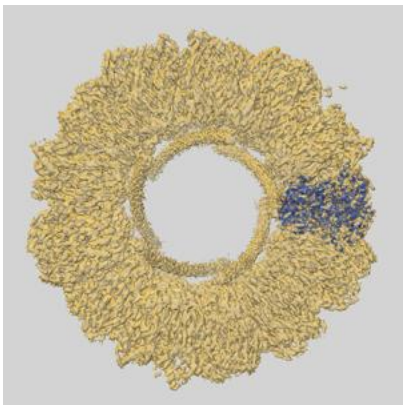
9.1.1 Map-model overlay



X

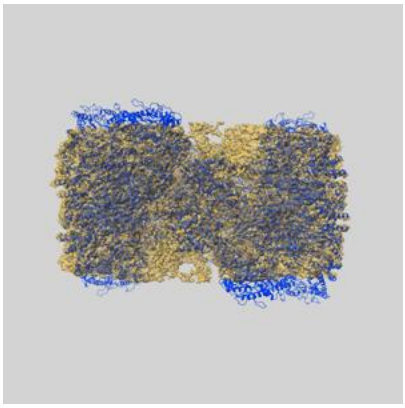


Y

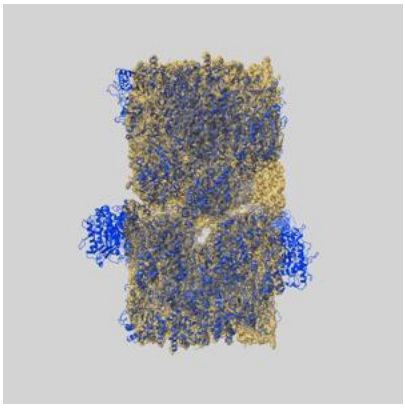


Z

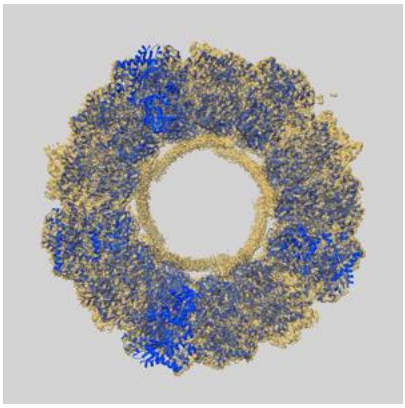
9.1.2 Map-model assembly overlay



X



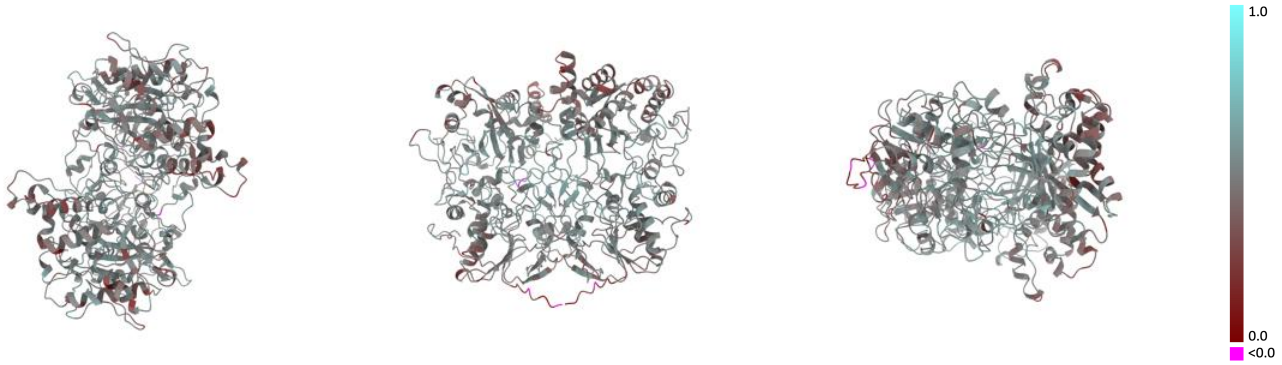
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.02 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.

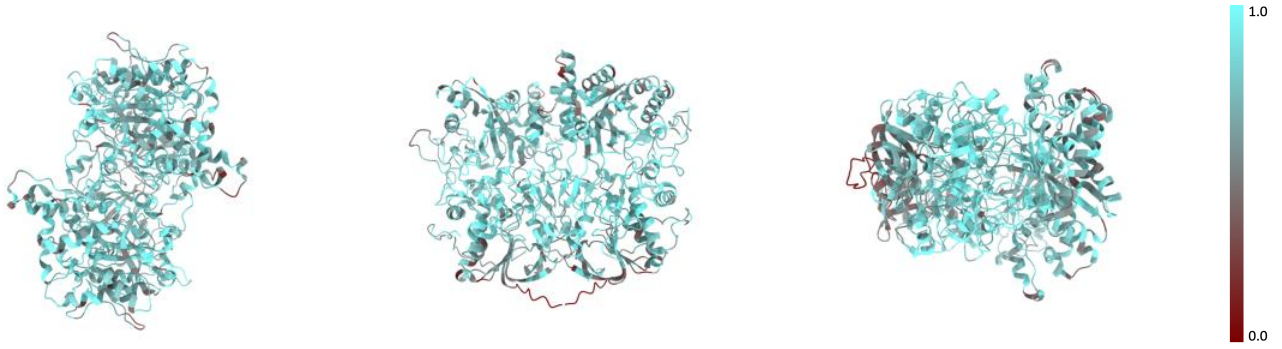
9.2 Q-score mapped to coordinate model



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.

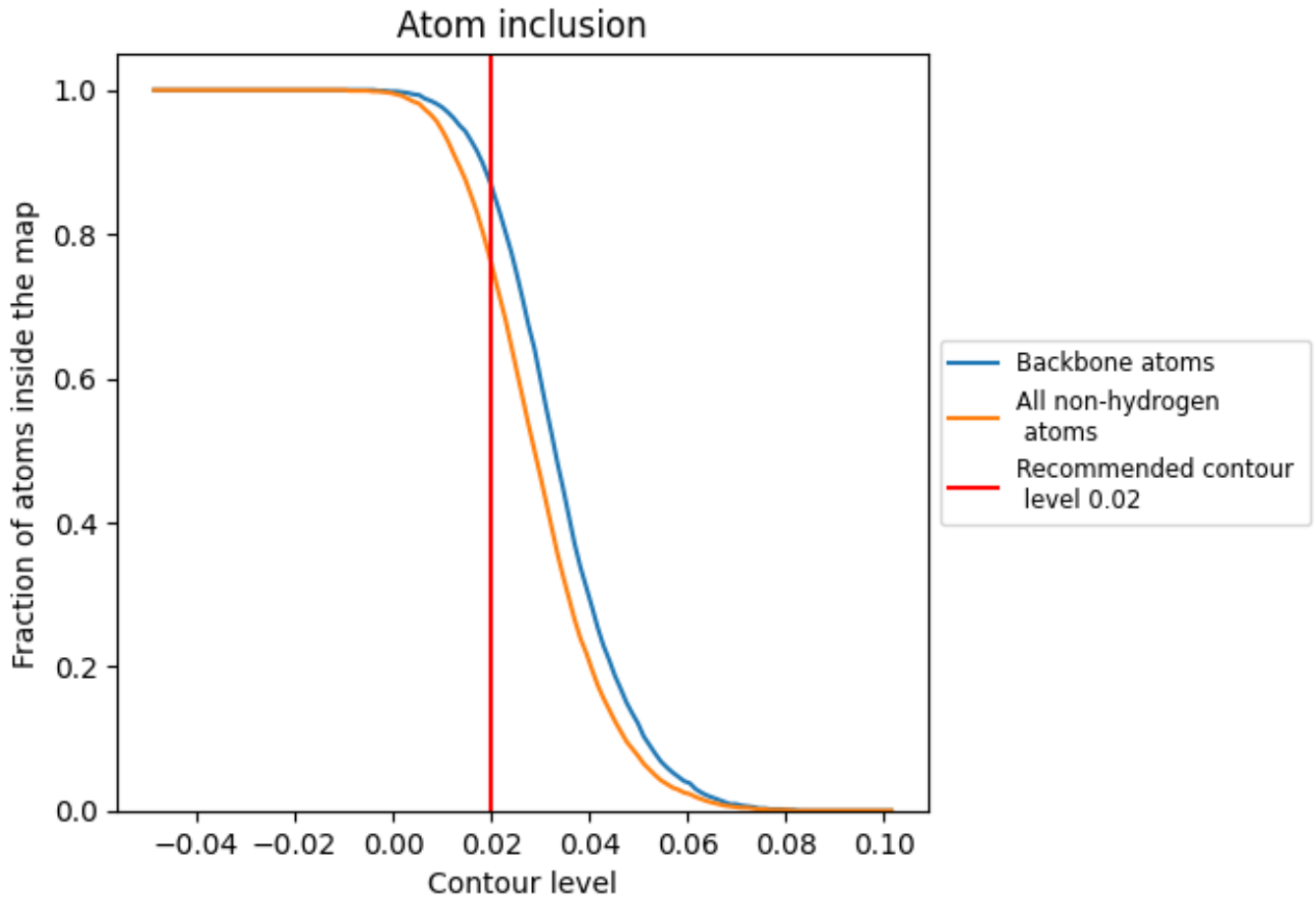
9.3 Atom inclusion mapped to coordinate model



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 (0.02).

9.4 Atom inclusion (i)



At the recommended contour level, 87% of all backbone atoms, 76% 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 (0.02) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.7610	0.4810
A	0.7620	0.4820
B	0.7590	0.4790