wwPDB EM Validation Summary Report (i)

Dec 10, 2022 – 09:25 pm GMT

PDB ID	:	6QYJ
EMDB ID	:	EMD-4678
Title	:	The cryo-EM structure of the connector of the mature bacteriophage phi29
Authors	:	Xu, J.; Wang, D.; Gui, M.; Xiang, Y.
Deposited on	:	2019-03-09
Resolution	:	3.40 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:

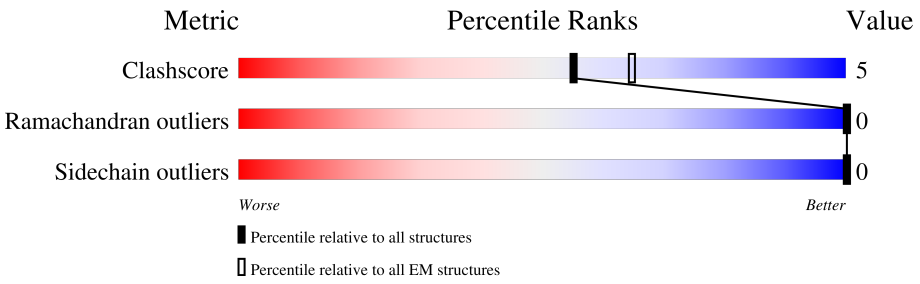
EMDB validation analysis	:	0.0.1.dev43
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.9
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.3

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 3.40 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	0a	309	85% 15%
1	0b	309	85% 15%
1	0c	309	85% 15%
1	0d	309	85% 15%
1	0e	309	85% 15%
1	0f	309	85% 15%
1	0g	309	85% 15%
1	0h	309	85% 15%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	0i	309	85% 15%
1	0j	309	85% 15%
1	0k	309	85% 15%
1	0l	309	85% 15%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 25956 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.

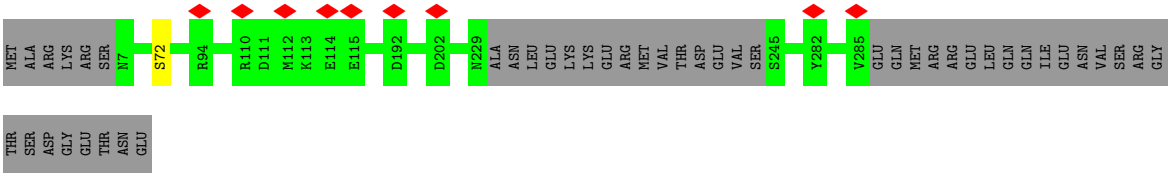
Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	0a	264	2163	1384	360	412	7	0	0
1	0b	264	2163	1384	360	412	7	0	0
1	0c	264	2163	1384	360	412	7	0	0
1	0d	264	2163	1384	360	412	7	0	0
1	0e	264	2163	1384	360	412	7	0	0
1	0f	264	2163	1384	360	412	7	0	0
1	0g	264	2163	1384	360	412	7	0	0
1	0h	264	2163	1384	360	412	7	0	0
1	0i	264	2163	1384	360	412	7	0	0
1	0j	264	2163	1384	360	412	7	0	0
1	0k	264	2163	1384	360	412	7	0	0
1	0l	264	2163	1384	360	412	7	0	0

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.

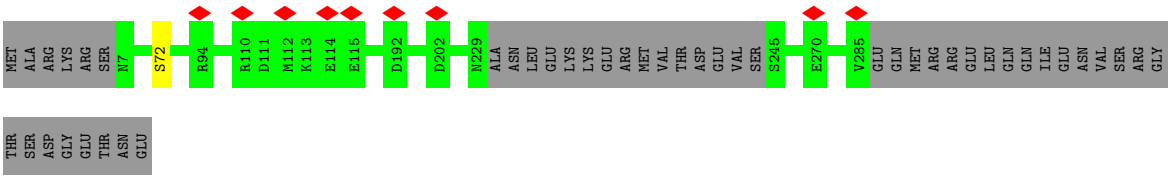
- Molecule 1: Portal protein

Chain 0a:



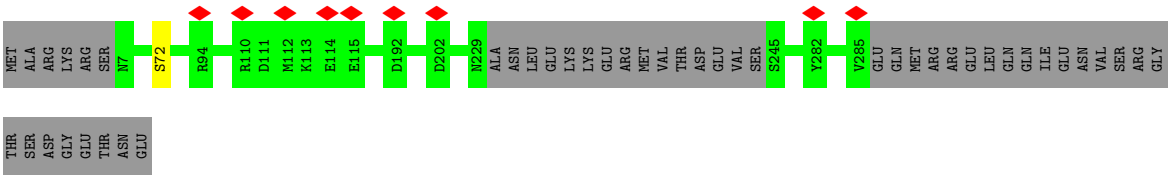
- Molecule 1: Portal protein

Chain 0b:



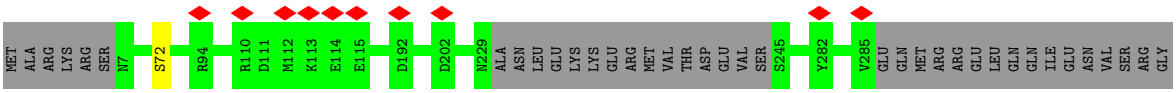
- Molecule 1: Portal protein

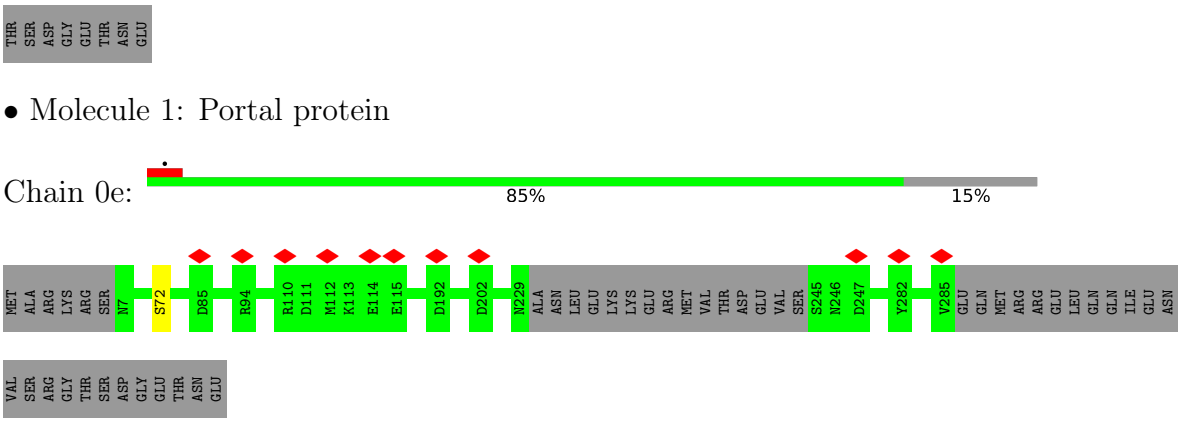
Chain 0c:

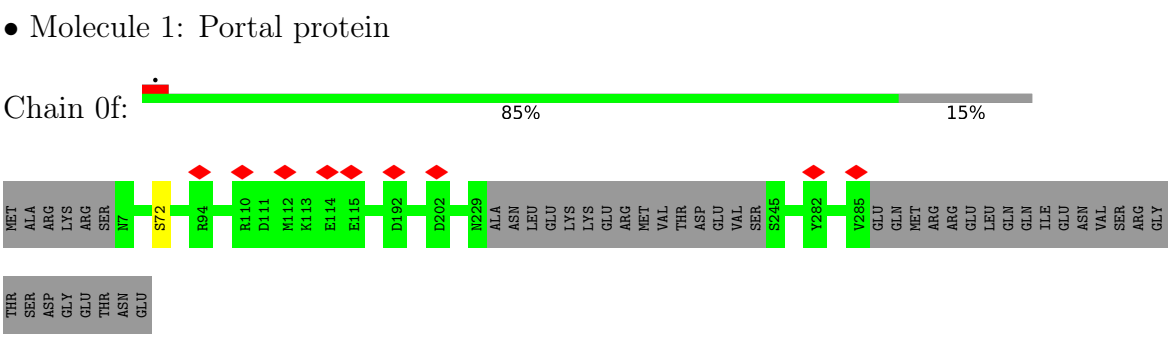


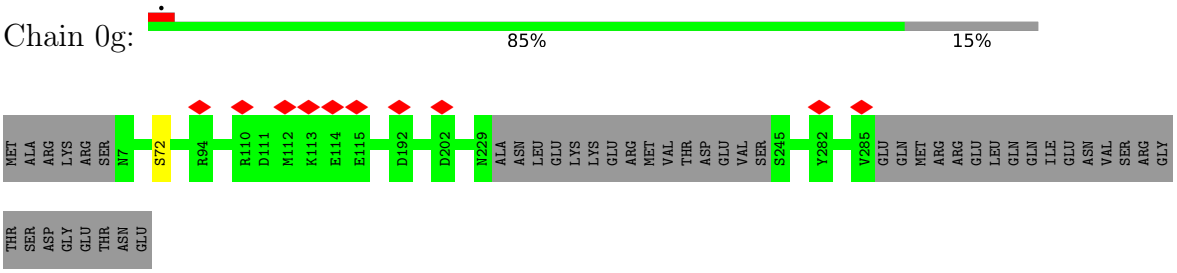
- Molecule 1: Portal protein

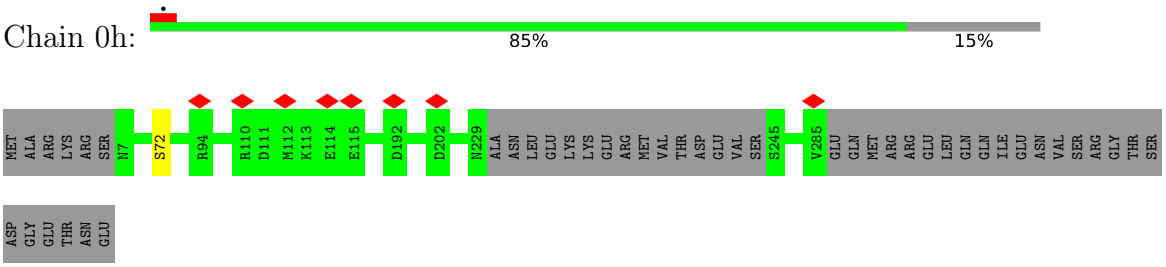
Chain 0d:

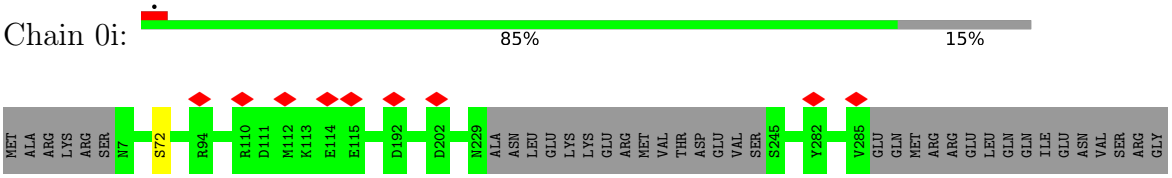


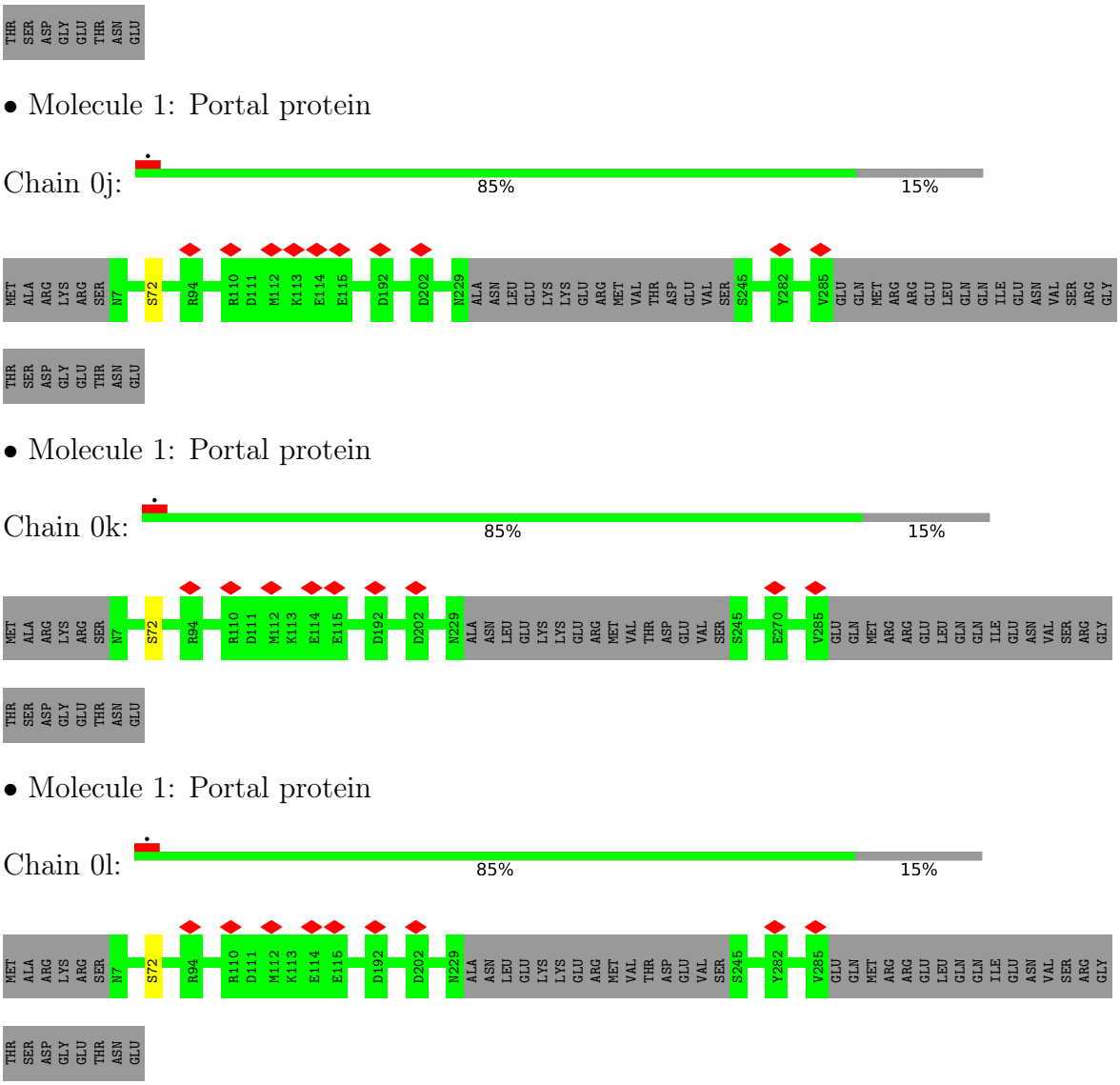

THR SER ASP GLY GLU THR ASN GLU
• Molecule 1: Portal protein
Chain 0e:
85%
15%
MET ALA ARG LYS ARG SER N7 S72 D85 R94 R110 D111 M112 K113 E114 E115 D192 D202 N229 ALA ASN LEU GLU GLU LYS LYS GLU ARG MET VAL THR ASP GLU VAL SER S245 N246 D247 Y282 V285 GLU GLN MET ARG ARG GLU LEU GLN GLN ILE GLU ASN
VAL SER ARG GLY THR SER ASP GLY GLU THR ASN GLU


• Molecule 1: Portal protein
Chain 0f:
85%
15%
MET ALA ARG LYS ARG SER N7 S72 R94 R110 D111 M112 K113 E114 E115 D192 D202 N229 ALA ASN LEU GLU GLU LYS LYS GLU ARG MET VAL THR ASP GLU VAL SER S245 Y282 V285 GLU GLN MET ARG ARG GLU LEU GLN GLN ILE GLU ASN VAL SER ARG GLY
THR SER ASP GLY GLU THR ASN GLU


• Molecule 1: Portal protein
Chain 0g:
85%
15%
MET ALA ARG LYS ARG SER N7 S72 R94 R110 D111 M112 K113 E114 E115 D192 D202 N229 ALA ASN LEU GLU GLU LYS LYS GLU ARG MET VAL THR ASP GLU VAL SER S245 Y282 V285 GLU GLN MET ARG ARG GLU LEU GLN GLN ILE GLU ASN VAL SER ARG GLY
THR SER ASP GLY GLU THR ASN GLU


• Molecule 1: Portal protein
Chain 0h:
85%
15%
MET ALA ARG LYS ARG SER N7 S72 R94 R110 D111 M112 K113 E114 E115 D192 D202 N229 ALA ASN LEU GLU GLU LYS LYS GLU ARG MET VAL THR ASP GLU VAL SER S245 Y285 GLU GLN MET ARG ARG GLU LEU GLN GLN ILE GLU ASN VAL SER ARG GLY THR SER
ASP GLY GLU THR ASN GLU


• Molecule 1: Portal protein
Chain 0i:
85%
15%
MET ALA ARG LYS ARG SER N7 S72 R94 R110 D111 M112 K113 E114 E115 D192 D202 N229 ALA ASN LEU GLU GLU LYS LYS GLU ARG MET VAL THR ASP GLU VAL SER S245 Y282 V285 GLU GLN MET ARG ARG GLU LEU GLN GLN ILE GLU ASN VAL SER ARG GLY


THR SER ASP GLY GLU THR ASN GLU
• Molecule 1: Portal protein
Chain 0j:
85%
15%
MET ALA ARG LYS ARG SER N7 S72 R94 R110 D111 M112 K113 E114 E115 D192 D202 N229 ALA ASN LEU GLU GLU LYS LYS GLU ARG MET VAL THR ASP GLU VAL SER S245 Y282 V285 GLU GLN MET ARG ARG GLU LEU GLN GLN ILE GLU ASN VAL SER ARG GLY
THR SER ASP GLY GLU THR ASN GLU
• Molecule 1: Portal protein
Chain 0k:
85%
15%
MET ALA ARG LYS ARG SER N7 S72 R94 R110 D111 M112 K113 E114 E115 D192 D202 N229 ALA ASN LEU GLU GLU LYS LYS GLU ARG MET VAL THR ASP GLU VAL SER S245 E270 V285 GLU GLN MET ARG ARG GLU LEU GLN GLN ILE GLU ASN VAL SER ARG GLY
THR SER ASP GLY GLU THR ASN GLU
• Molecule 1: Portal protein
Chain 0l:
85%
15%
MET ALA ARG LYS ARG SER N7 S72 R94 R110 D111 M112 K113 E114 E115 D192 D202 N229 ALA ASN LEU GLU GLU LYS LYS GLU ARG MET VAL THR ASP GLU VAL SER S245 Y282 V285 GLU GLN MET ARG ARG GLU LEU GLN GLN ILE GLU ASN VAL SER ARG GLY
THR SER ASP GLY GLU THR ASN GLU

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	36730	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 ($e^-/Å^2$)	40	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	42.791	Depositor
Minimum map value	-0.417	Depositor
Average map value	0.005	Depositor
Map value standard deviation	0.271	Depositor
Recommended contour level	3.33	Depositor
Map size (Å)	743.6, 743.6, 743.6	wwPDB
Map dimensions	572, 572, 572	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.3, 1.3, 1.3	Depositor

5 Model quality (i)

5.1 Standard geometry (i)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\|>5$	RMSZ	$\#\|Z\|>5$
1	0a	0.30	0/2211	0.48	0/2996
1	0b	0.30	0/2211	0.48	0/2996
1	0c	0.30	0/2211	0.48	0/2996
1	0d	0.30	0/2211	0.48	0/2996
1	0e	0.30	0/2211	0.48	0/2996
1	0f	0.30	0/2211	0.48	0/2996
1	0g	0.30	0/2211	0.48	0/2996
1	0h	0.30	0/2211	0.48	0/2996
1	0i	0.30	0/2211	0.48	0/2996
1	0j	0.30	0/2211	0.48	0/2996
1	0k	0.30	0/2211	0.49	0/2996
1	0l	0.30	0/2211	0.48	0/2996
All	All	0.30	0/26532	0.48	0/35952

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	0a	0	1
1	0b	0	1
1	0c	0	1
1	0d	0	1
1	0e	0	1
1	0f	0	1
1	0g	0	1
1	0h	0	1
1	0i	0	1
1	0j	0	1
1	0k	0	1
1	0l	0	1
All	All	0	12

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	0a	72	SER	Peptide
1	0b	72	SER	Peptide
1	0c	72	SER	Peptide
1	0d	72	SER	Peptide
1	0e	72	SER	Peptide

5.2 Too-close contacts (i)

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

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	0a	2163	0	2104	0	0
1	0b	2163	0	2104	0	0
1	0c	2163	0	2104	0	0
1	0d	2163	0	2104	0	0
1	0e	2163	0	2104	0	0
1	0f	2163	0	2104	0	0
1	0g	2163	0	2104	0	0
1	0h	2163	0	2104	0	0
1	0i	2163	0	2104	0	0
1	0j	2163	0	2104	0	0
1	0k	2163	0	2104	0	0
1	0l	2163	0	2104	0	0
All	All	25956	0	25248	0	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 5.

There are no clashes within the asymmetric unit.

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	0a	260/309 (84%)	243 (94%)	17 (6%)	0	100	100
1	0b	260/309 (84%)	243 (94%)	17 (6%)	0	100	100
1	0c	260/309 (84%)	243 (94%)	17 (6%)	0	100	100
1	0d	260/309 (84%)	243 (94%)	17 (6%)	0	100	100
1	0e	260/309 (84%)	243 (94%)	17 (6%)	0	100	100
1	0f	260/309 (84%)	243 (94%)	17 (6%)	0	100	100
1	0g	260/309 (84%)	243 (94%)	17 (6%)	0	100	100
1	0h	260/309 (84%)	243 (94%)	17 (6%)	0	100	100
1	0i	260/309 (84%)	243 (94%)	17 (6%)	0	100	100
1	0j	260/309 (84%)	243 (94%)	17 (6%)	0	100	100
1	0k	260/309 (84%)	243 (94%)	17 (6%)	0	100	100
1	0l	260/309 (84%)	243 (94%)	17 (6%)	0	100	100
All	All	3120/3708 (84%)	2916 (94%)	204 (6%)	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	0a	236/277 (85%)	236 (100%)	0	100	100
1	0b	236/277 (85%)	236 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	0c	236/277 (85%)	236 (100%)	0	100	100
1	0d	236/277 (85%)	236 (100%)	0	100	100
1	0e	236/277 (85%)	236 (100%)	0	100	100
1	0f	236/277 (85%)	236 (100%)	0	100	100
1	0g	236/277 (85%)	236 (100%)	0	100	100
1	0h	236/277 (85%)	236 (100%)	0	100	100
1	0i	236/277 (85%)	236 (100%)	0	100	100
1	0j	236/277 (85%)	236 (100%)	0	100	100
1	0k	236/277 (85%)	236 (100%)	0	100	100
1	0l	236/277 (85%)	236 (100%)	0	100	100
All	All	2832/3324 (85%)	2832 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

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

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

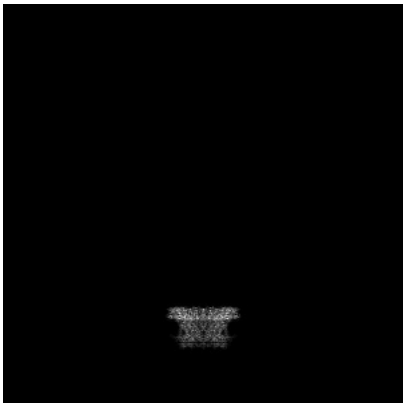
6 Map visualisation (i)

This section contains visualisations of the EMDB entry EMD-4678. These allow visual inspection of the internal detail of the map and identification of artifacts.

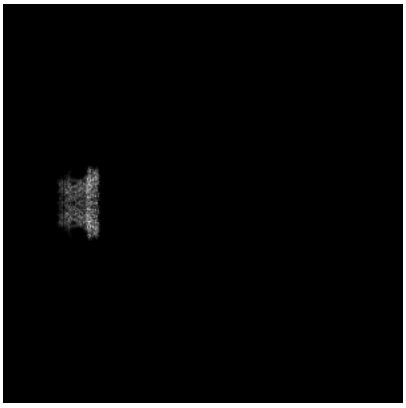
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections (i)

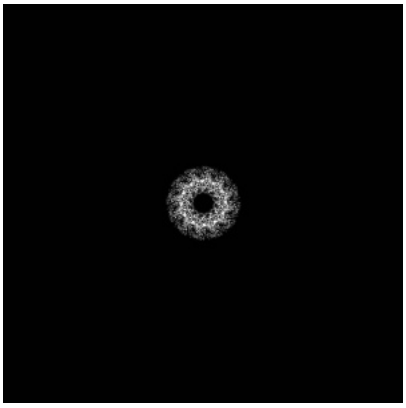
6.1.1 Primary map



X



Y

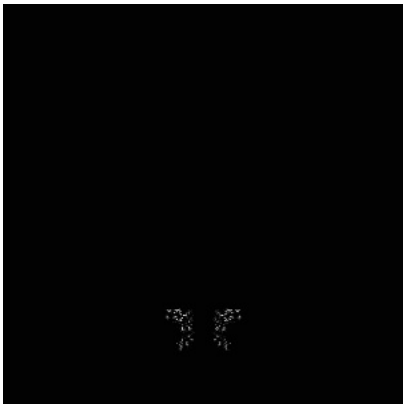


Z

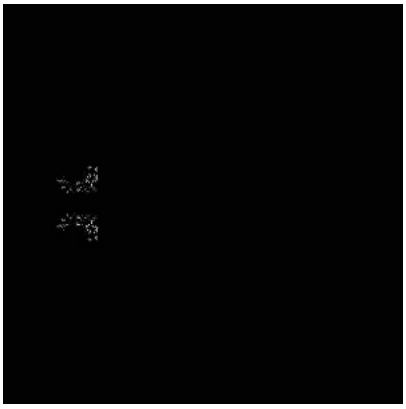
The images above show the map projected in three orthogonal directions.

6.2 Central slices (i)

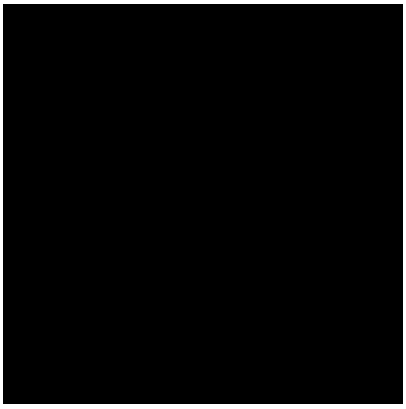
6.2.1 Primary map



X Index: 286



Y Index: 286

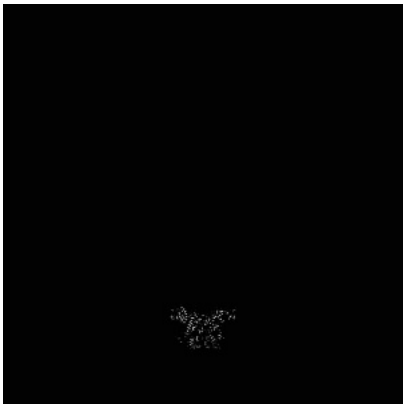


Z Index: 286

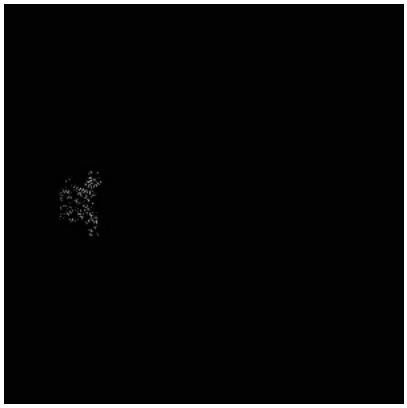
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

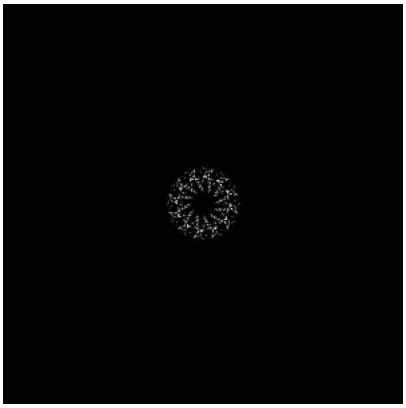
6.3.1 Primary map



X Index: 260



Y Index: 260

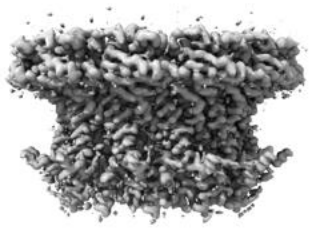


Z Index: 121

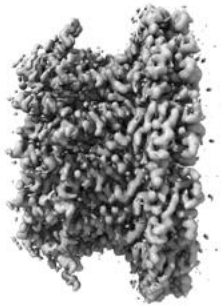
The images above show the largest variance slices of the map in three orthogonal directions.

6.4 Orthogonal surface views (i)

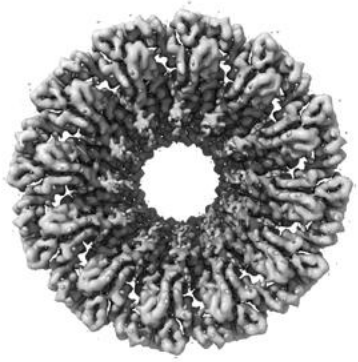
6.4.1 Primary map



X



Y



Z

The images above show the 3D surface view of the map at the recommended contour level 3.33. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

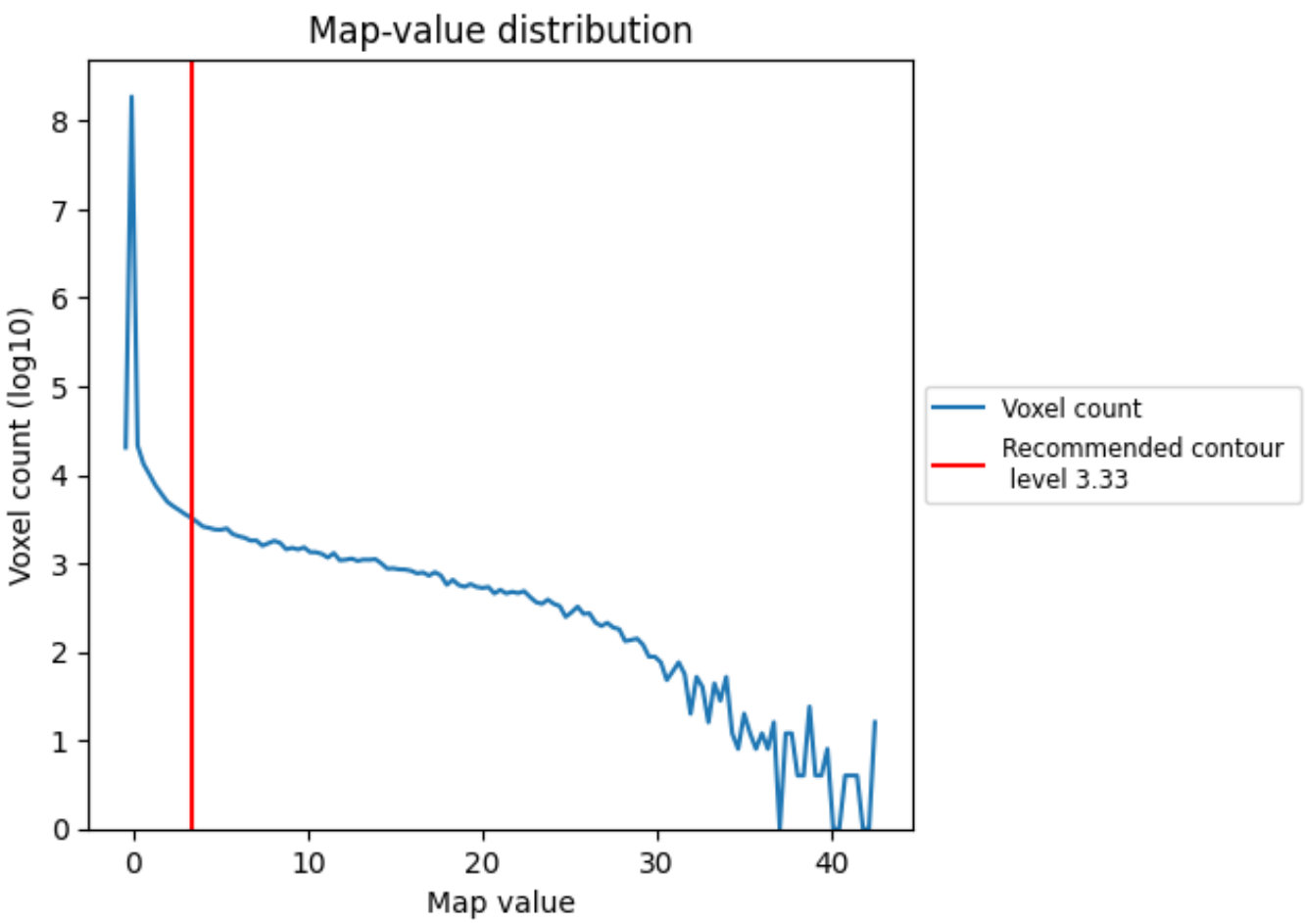
6.5 Mask visualisation (i)

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

7 Map analysis (i)

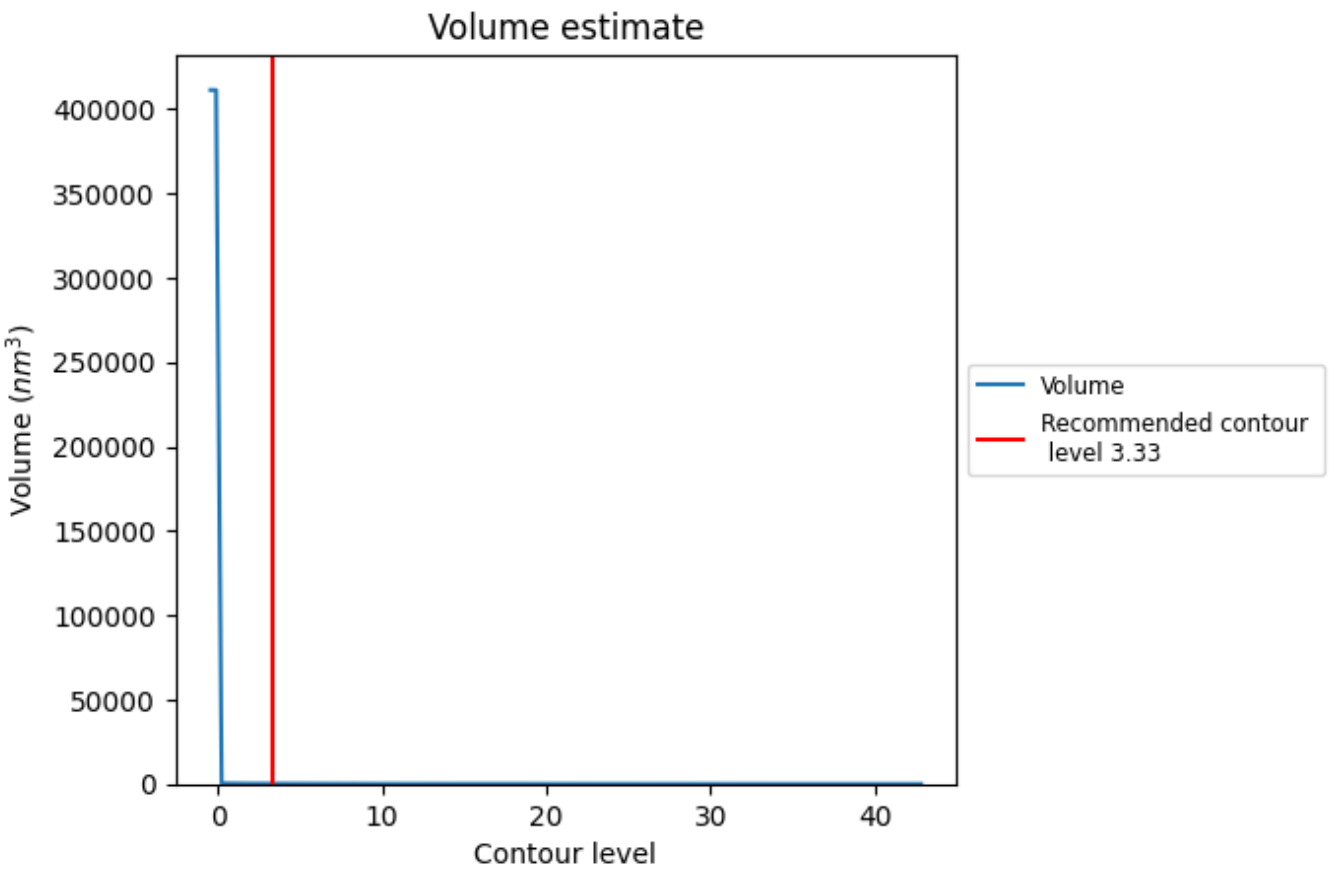
This section contains the results of statistical analysis of the map.

7.1 Map-value distribution (i)



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

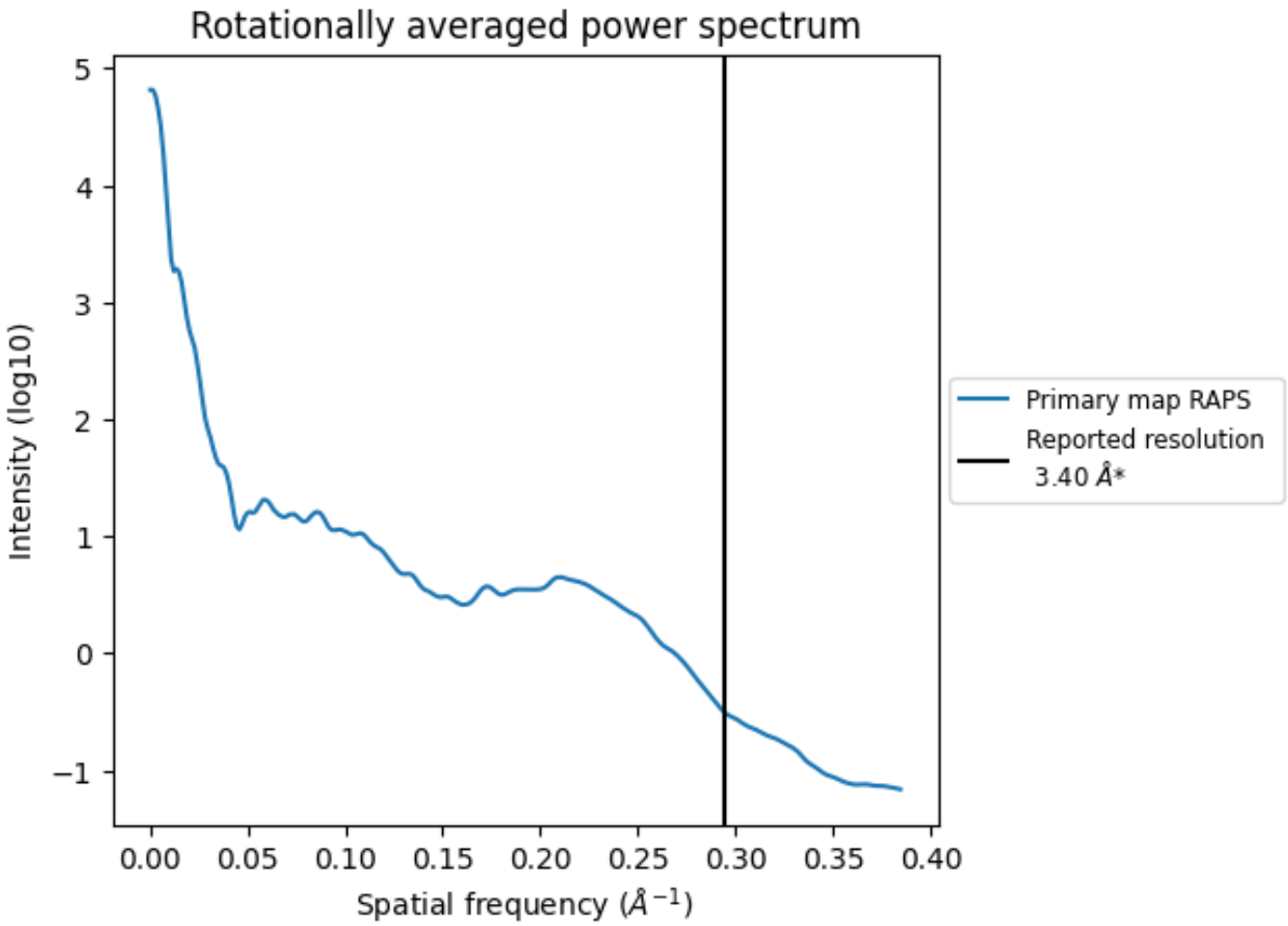
7.2 Volume estimate (i)



The volume at the recommended contour level is 170 nm^3; this corresponds to an approximate mass of 154 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 0.294 Å^{-1}

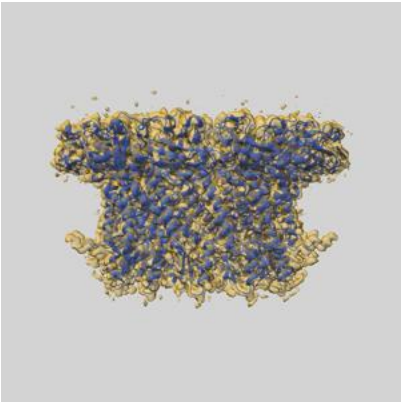
8 Fourier-Shell correlation (i)

This section was not generated. No FSC curve or half-maps provided.

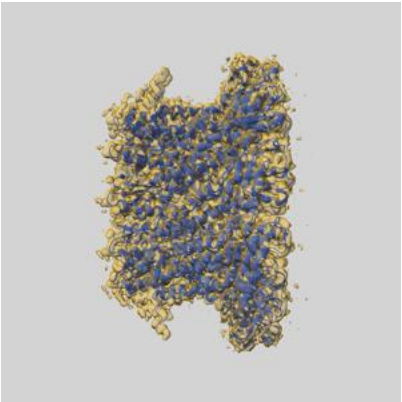
9 Map-model fit (i)

This section contains information regarding the fit between EMDB map EMD-4678 and PDB model 6QYJ. Per-residue inclusion information can be found in section 3 on page 5.

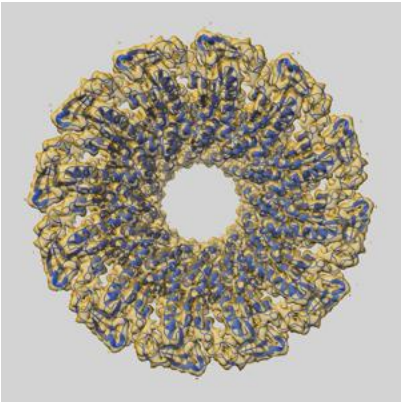
9.1 Map-model overlay (i)



X



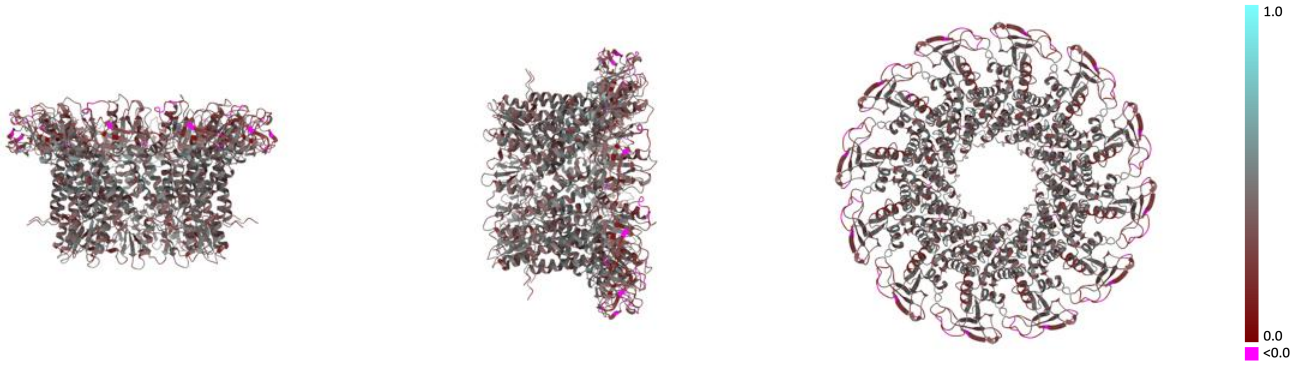
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 3.33 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.

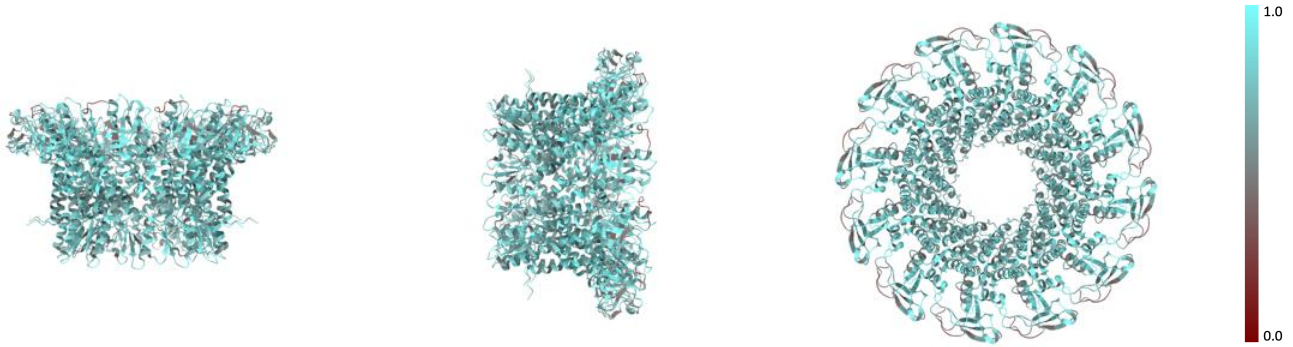
9.2 Q-score mapped to coordinate model (i)



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.

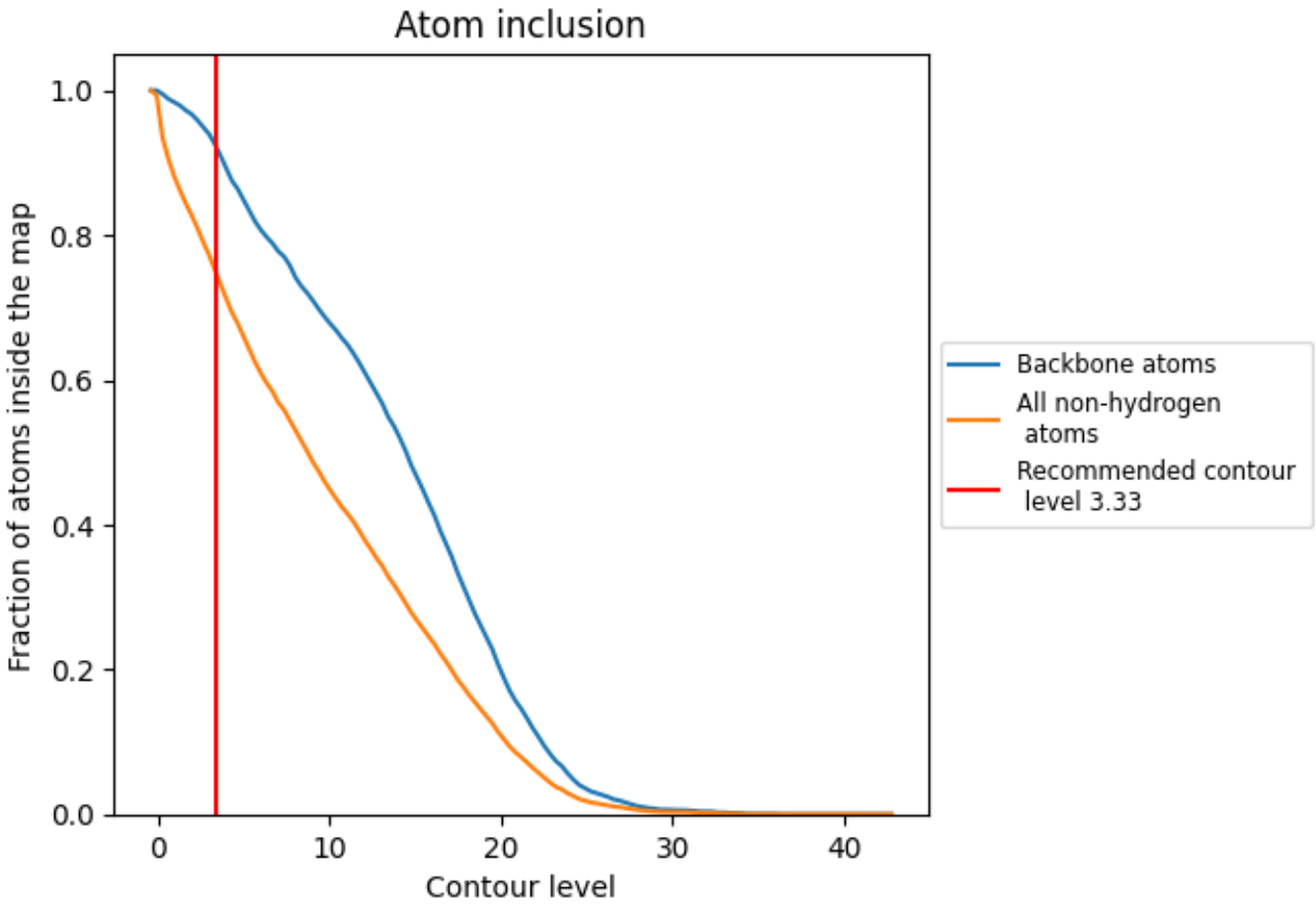
9.3 Atom inclusion mapped to coordinate model (i)



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 (3.33).

9.4 Atom inclusion (i)



At the recommended contour level, 93% of all backbone atoms, 75% of all non-hydrogen atoms, are inside the map.

9.5 Map-model fit summary (i)

The table lists the average atom inclusion at the recommended contour level (3.33) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.7509	0.3620
0a	0.7522	0.3620
0b	0.7508	0.3620
0c	0.7499	0.3610
0d	0.7518	0.3620
0e	0.7513	0.3620
0f	0.7494	0.3630
0g	0.7522	0.3620
0h	0.7522	0.3630
0i	0.7494	0.3640
0j	0.7522	0.3610
0k	0.7504	0.3630
0l	0.7494	0.3630