



## Full wwPDB EM Validation Report ⓘ

Mar 16, 2024 – 08:15 AM EDT

PDB ID : 6ODI  
EMDB ID : EMD-20020  
Title : Structure of CagY from a cryo-EM reconstruction of a T4SS  
Authors : Chung, J.M.; Sheedlo, M.J.; Campbell, A.; Sawhney, N.; Frick-Cheng, A.E.;  
Lacy, D.B.; Cover, T.L.; Ohi, M.D.  
Deposited on : 2019-03-26  
Resolution : 3.80 Å(reported)

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

We welcome your comments at [validation@mail.wwpdb.org](mailto: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

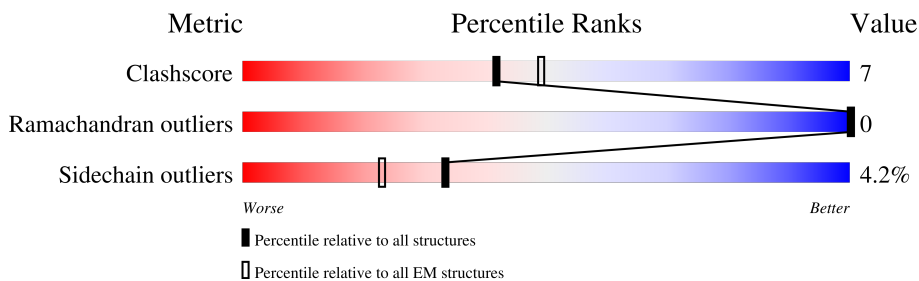
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 3.80 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  $\geq 3$ , 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  $\leq 5\%$ . 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	b	1927	10% 90%
1	c	1927	10% 90%
1	d	1927	10% 90%
1	e	1927	10% 90%
1	f	1927	10% 90%
1	g	1927	10% 90%
1	h	1927	10% 90%
1	i	1927	10% 90%

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain	
1	j	1927	 10%	90%
1	k	1927	 10%	90%
1	l	1927	 10%	90%
1	m	1927	 10%	90%
1	n	1927	 10%	90%
1	o	1927	 10%	90%

## 2 Entry composition [i](#)

There is only 1 type of molecule in this entry. The entry contains 20902 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.

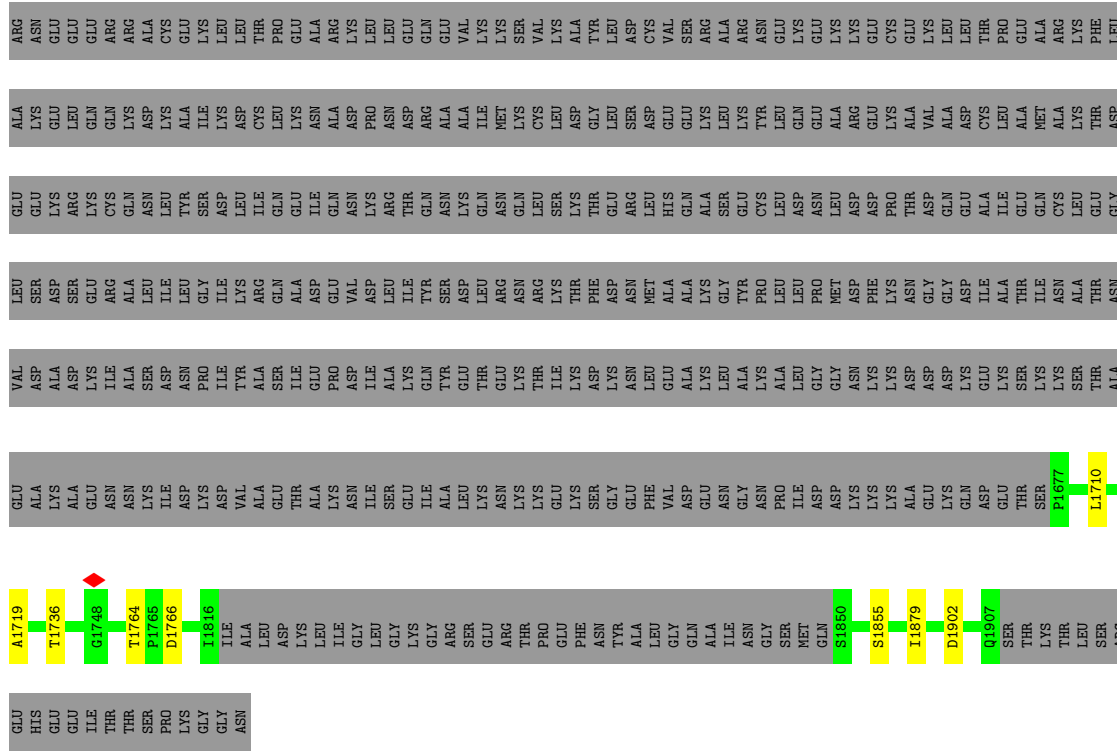
- Molecule 1 is a protein called Type IV secretion system apparatus protein CagY.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	b	198	1493	959	241	284	9	0	0
1	c	198	1493	959	241	284	9	0	0
1	d	198	1493	959	241	284	9	0	0
1	e	198	1493	959	241	284	9	0	0
1	f	198	1493	959	241	284	9	0	0
1	g	198	1493	959	241	284	9	0	0
1	h	198	1493	959	241	284	9	0	0
1	i	198	1493	959	241	284	9	0	0
1	j	198	1493	959	241	284	9	0	0
1	k	198	1493	959	241	284	9	0	0
1	l	198	1493	959	241	284	9	0	0
1	m	198	1493	959	241	284	9	0	0
1	n	198	1493	959	241	284	9	0	0
1	o	198	1493	959	241	284	9	0	0



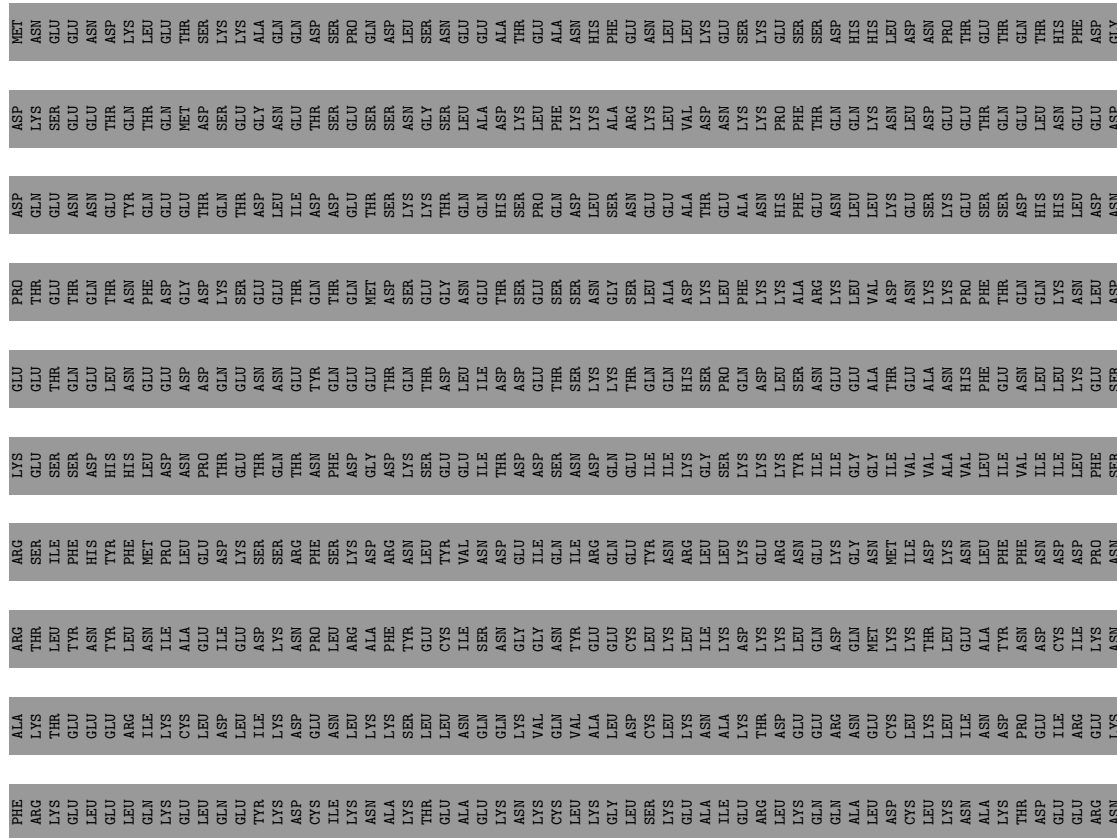






● Molecule 1: Type IV secretion system apparatus protein CagY

Chain d: 10% 90%



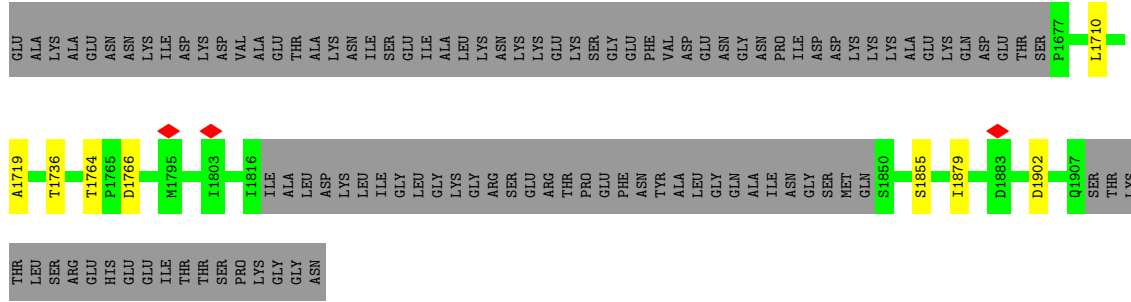






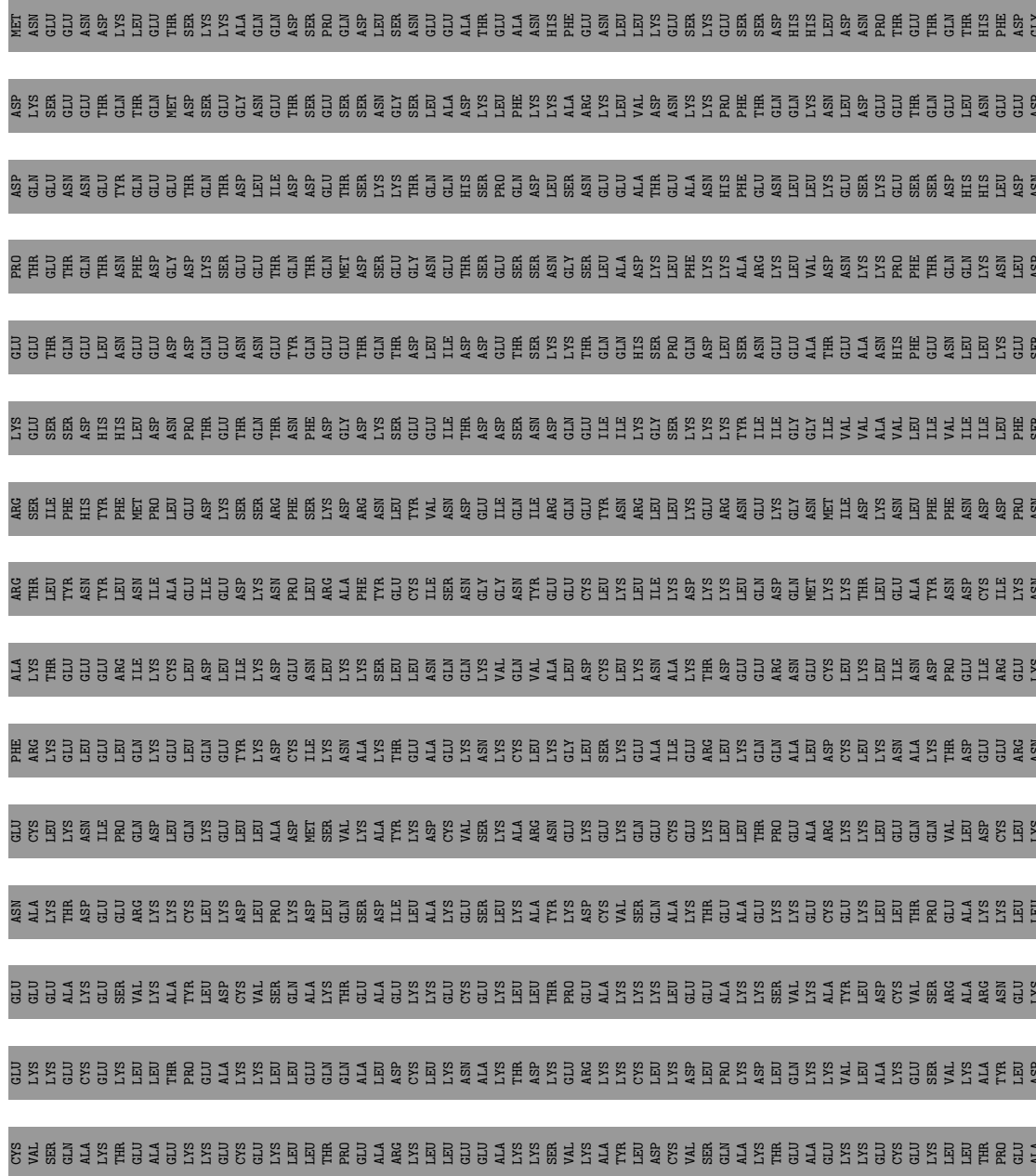






● Molecule 1: Type IV secretion system apparatus protein CagY

Chain g:  10%  90%





















● Molecule 1: Type IV secretion system apparatus protein CagY

Chain 1:  10%  90%

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















## 4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C14	Depositor
Number of particles used	17159	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	59.2	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	0.136	Depositor
Minimum map value	-0.067	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.004	Depositor
Recommended contour level	0.027	Depositor
Map size (Å)	836.39996, 836.39996, 836.39996	wwPDB
Map dimensions	510, 510, 510	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.64, 1.64, 1.64	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	b	0.41	0/1517	0.58	0/2056
1	c	0.41	0/1517	0.58	0/2056
1	d	0.41	0/1517	0.58	0/2056
1	e	0.41	0/1517	0.58	0/2056
1	f	0.41	0/1517	0.58	0/2056
1	g	0.41	0/1517	0.58	0/2056
1	h	0.41	0/1517	0.58	0/2056
1	i	0.41	0/1517	0.58	0/2056
1	j	0.41	0/1517	0.58	0/2056
1	k	0.41	0/1517	0.58	0/2056
1	l	0.41	0/1517	0.58	0/2056
1	m	0.41	0/1517	0.58	0/2056
1	n	0.41	0/1517	0.58	0/2056
1	o	0.41	0/1517	0.58	0/2056
All	All	0.41	0/21238	0.58	0/28784

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

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	#Chirality outliers	#Planarity outliers
1	m	0	1
1	n	0	1
1	o	0	1
All	All	0	14

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (14) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	b	1719	ALA	Peptide
1	c	1719	ALA	Peptide
1	d	1719	ALA	Peptide
1	e	1719	ALA	Peptide
1	f	1719	ALA	Peptide
1	g	1719	ALA	Peptide
1	h	1719	ALA	Peptide
1	i	1719	ALA	Peptide
1	j	1719	ALA	Peptide
1	k	1719	ALA	Peptide
1	l	1719	ALA	Peptide
1	m	1719	ALA	Peptide
1	n	1719	ALA	Peptide
1	o	1719	ALA	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	b	1493	0	1538	0	0
1	c	1493	0	1538	0	0
1	d	1493	0	1538	0	0
1	e	1493	0	1538	0	0
1	f	1493	0	1538	0	0
1	g	1493	0	1538	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	h	1493	0	1538	0	0
1	i	1493	0	1538	0	0
1	j	1493	0	1538	0	0
1	k	1493	0	1538	0	0
1	l	1493	0	1538	0	0
1	m	1493	0	1538	0	0
1	n	1493	0	1538	0	0
1	o	1493	0	1538	0	0
All	All	20902	0	21532	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 7.

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	b	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	c	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	d	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	e	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	f	194/1927 (10%)	176 (91%)	18 (9%)	0	100	100
1	g	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	h	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	i	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	j	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	k	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	l	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	m	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	n	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	o	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
All	All	2716/26978 (10%)	2451 (90%)	265 (10%)	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	b	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	c	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	d	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	e	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	f	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	g	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	h	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	i	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	j	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	k	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	l	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	m	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	n	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	o	167/1725 (10%)	160 (96%)	7 (4%)	30	58
All	All	2338/24150 (10%)	2240 (96%)	98 (4%)	33	58

All (98) residues with a non-rotameric sidechain are listed below:



Mol	Chain	Res	Type
1	b	1710	LEU
1	b	1736	THR
1	b	1764	THR
1	b	1766	ASP
1	b	1855	SER
1	b	1879	ILE
1	b	1902	ASP
1	c	1710	LEU
1	c	1736	THR
1	c	1764	THR
1	c	1766	ASP
1	c	1855	SER
1	c	1879	ILE
1	c	1902	ASP
1	d	1710	LEU
1	d	1736	THR
1	d	1764	THR
1	d	1766	ASP
1	d	1855	SER
1	d	1879	ILE
1	d	1902	ASP
1	e	1710	LEU
1	e	1736	THR
1	e	1764	THR
1	e	1766	ASP
1	e	1855	SER
1	e	1879	ILE
1	e	1902	ASP
1	f	1710	LEU
1	f	1736	THR
1	f	1764	THR
1	f	1766	ASP
1	f	1855	SER
1	f	1879	ILE
1	f	1902	ASP
1	g	1710	LEU
1	g	1736	THR
1	g	1764	THR
1	g	1766	ASP
1	g	1855	SER
1	g	1879	ILE
1	g	1902	ASP
1	h	1710	LEU

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>
1	h	1736	THR
1	h	1764	THR
1	h	1766	ASP
1	h	1855	SER
1	h	1879	ILE
1	h	1902	ASP
1	i	1710	LEU
1	i	1736	THR
1	i	1764	THR
1	i	1766	ASP
1	i	1855	SER
1	i	1879	ILE
1	i	1902	ASP
1	j	1710	LEU
1	j	1736	THR
1	j	1764	THR
1	j	1766	ASP
1	j	1855	SER
1	j	1879	ILE
1	j	1902	ASP
1	k	1710	LEU
1	k	1736	THR
1	k	1764	THR
1	k	1766	ASP
1	k	1855	SER
1	k	1879	ILE
1	k	1902	ASP
1	l	1710	LEU
1	l	1736	THR
1	l	1764	THR
1	l	1766	ASP
1	l	1855	SER
1	l	1879	ILE
1	l	1902	ASP
1	m	1710	LEU
1	m	1736	THR
1	m	1764	THR
1	m	1766	ASP
1	m	1855	SER
1	m	1879	ILE
1	m	1902	ASP
1	n	1710	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	n	1736	THR
1	n	1764	THR
1	n	1766	ASP
1	n	1855	SER
1	n	1879	ILE
1	n	1902	ASP
1	o	1710	LEU
1	o	1736	THR
1	o	1764	THR
1	o	1766	ASP
1	o	1855	SER
1	o	1879	ILE
1	o	1902	ASP

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

There are no chain breaks in this entry.

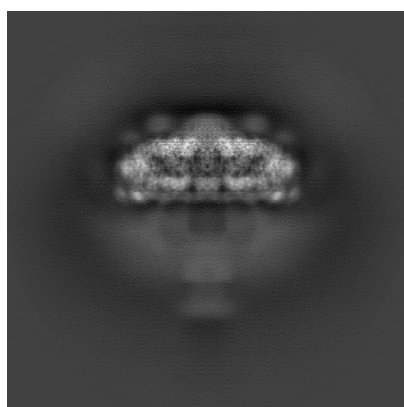
## 6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-20020. These allow visual inspection of the internal detail of the map and identification of artifacts.

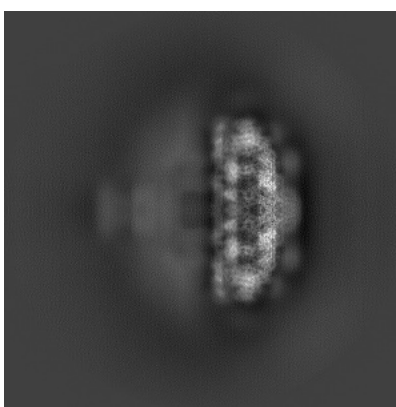
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

### 6.1 Orthogonal projections [i](#)

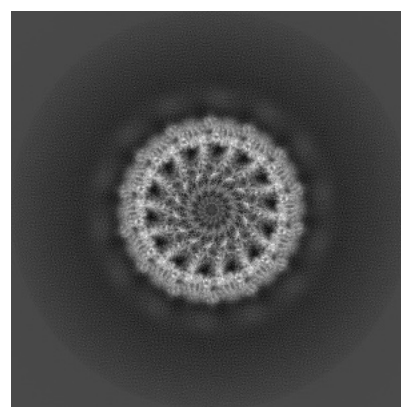
#### 6.1.1 Primary map



X



Y

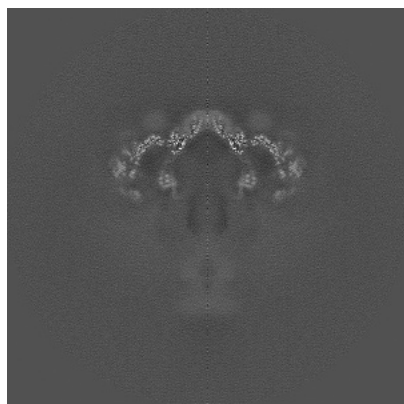


Z

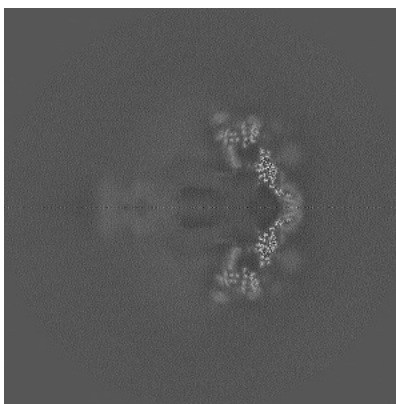
The images above show the map projected in three orthogonal directions.

### 6.2 Central slices [i](#)

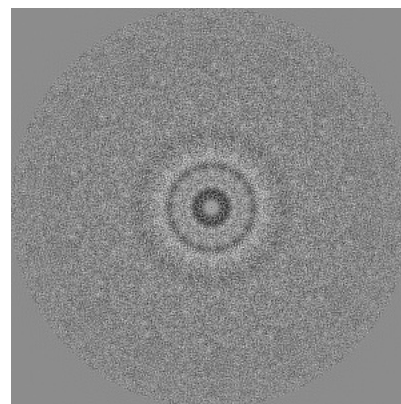
#### 6.2.1 Primary map



X Index: 255



Y Index: 255

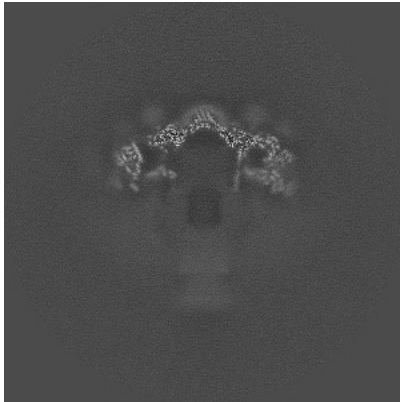


Z Index: 255

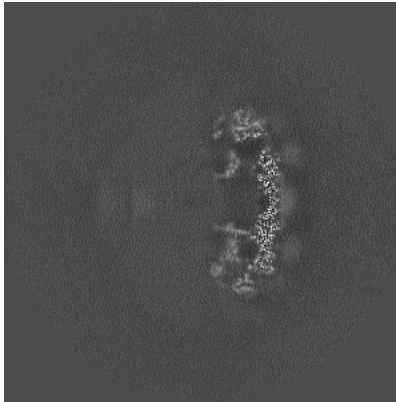
The images above show central slices of the map in three orthogonal directions.

## 6.3 Largest variance slices [i](#)

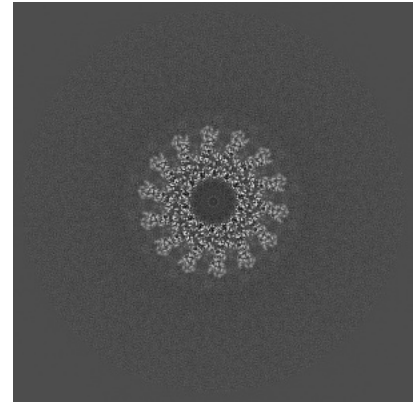
### 6.3.1 Primary map



X Index: 241



Y Index: 228

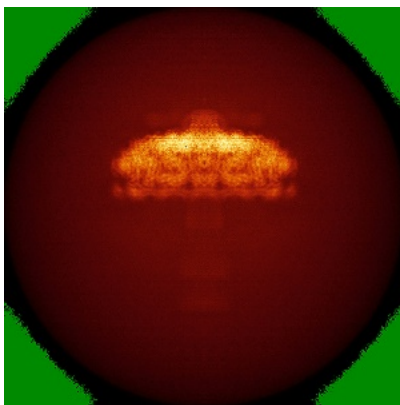


Z Index: 335

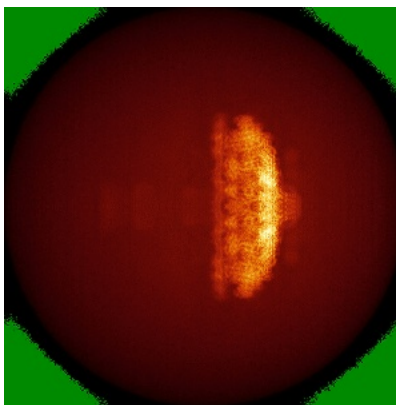
The images above show the largest variance slices of the map in three orthogonal directions.

## 6.4 Orthogonal standard-deviation projections (False-color) [i](#)

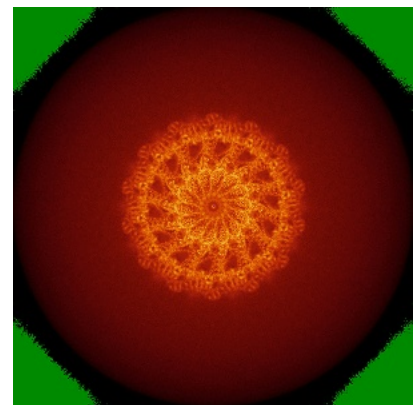
### 6.4.1 Primary map



X



Y

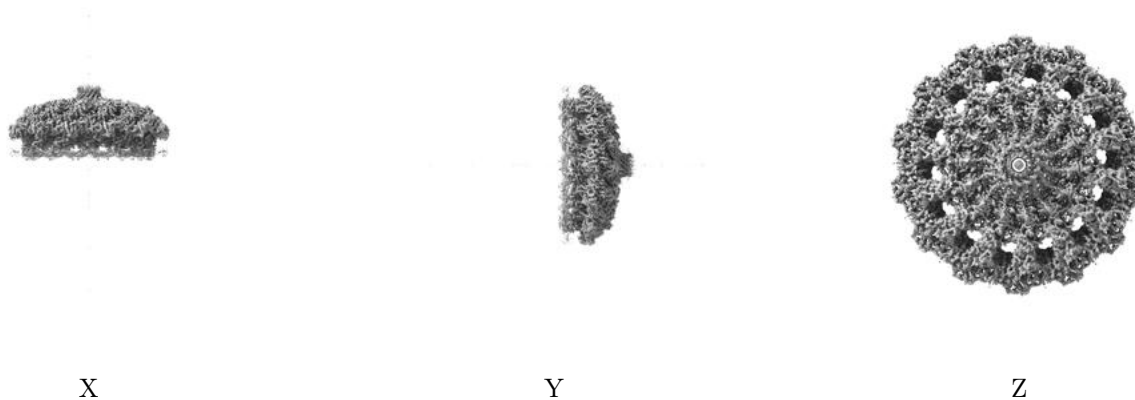


Z

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



The images above show the 3D surface view of the map at the recommended contour level 0.027. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

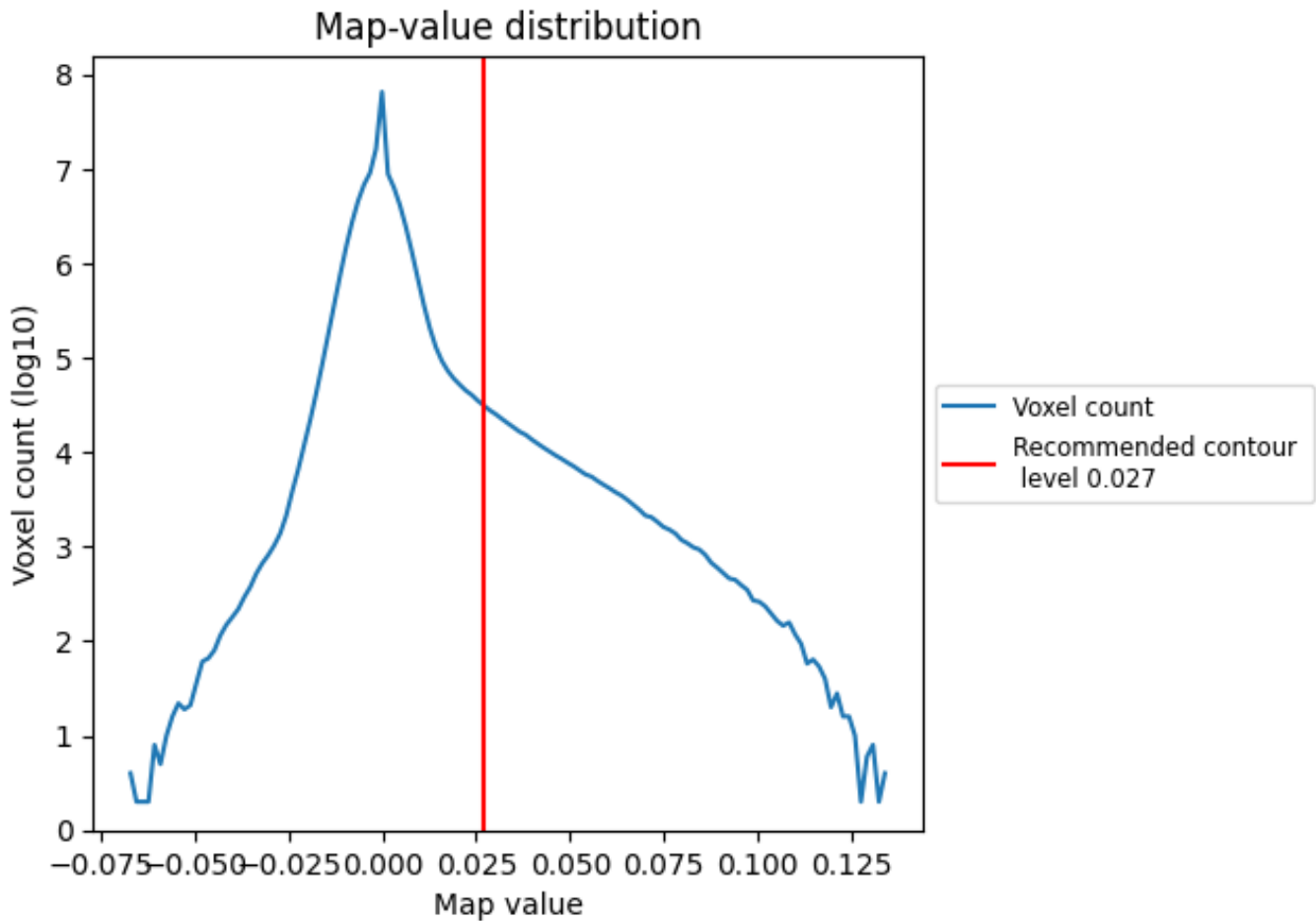
## 6.6 Mask visualisation [i](#)

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

## 7 Map analysis [i](#)

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

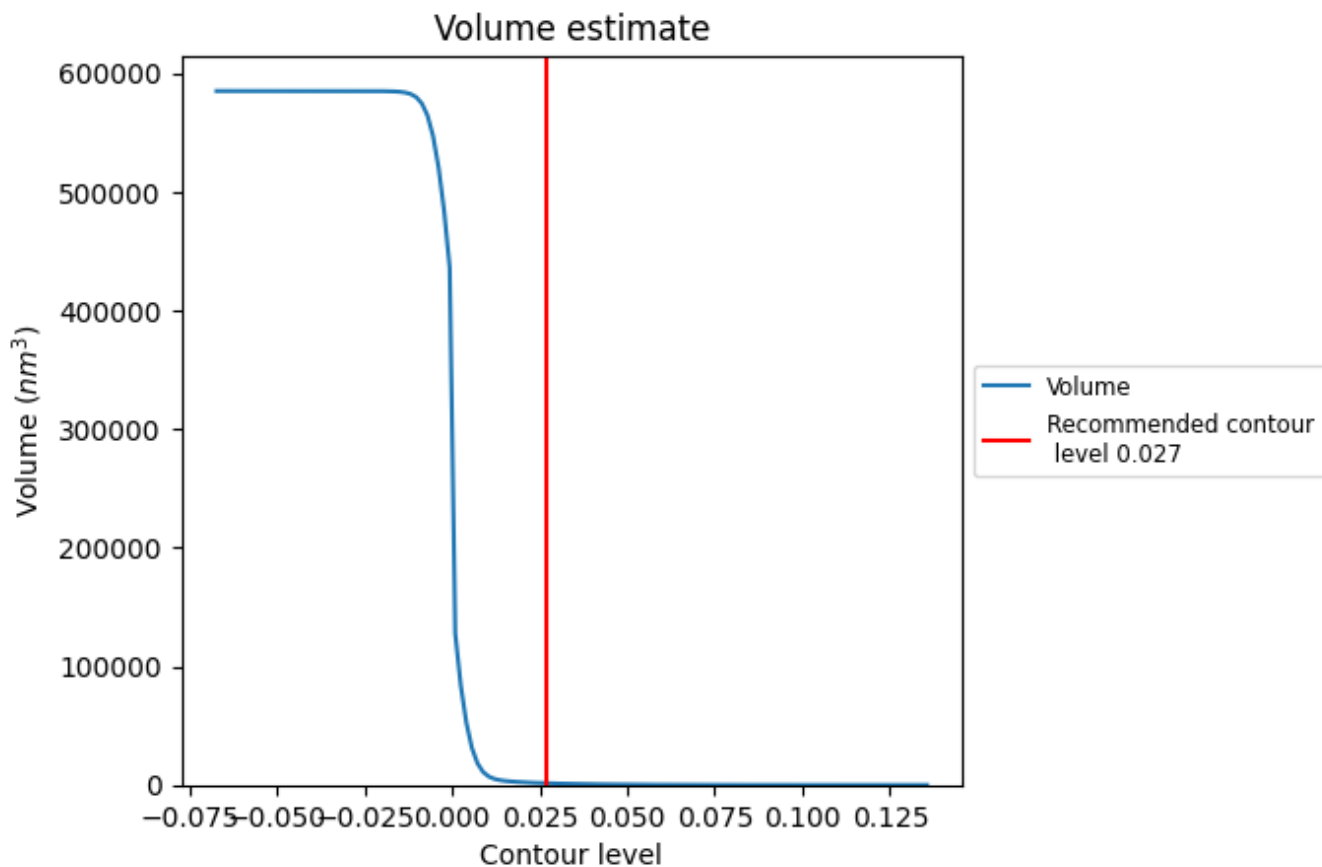
### 7.1 Map-value distribution [i](#)



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.



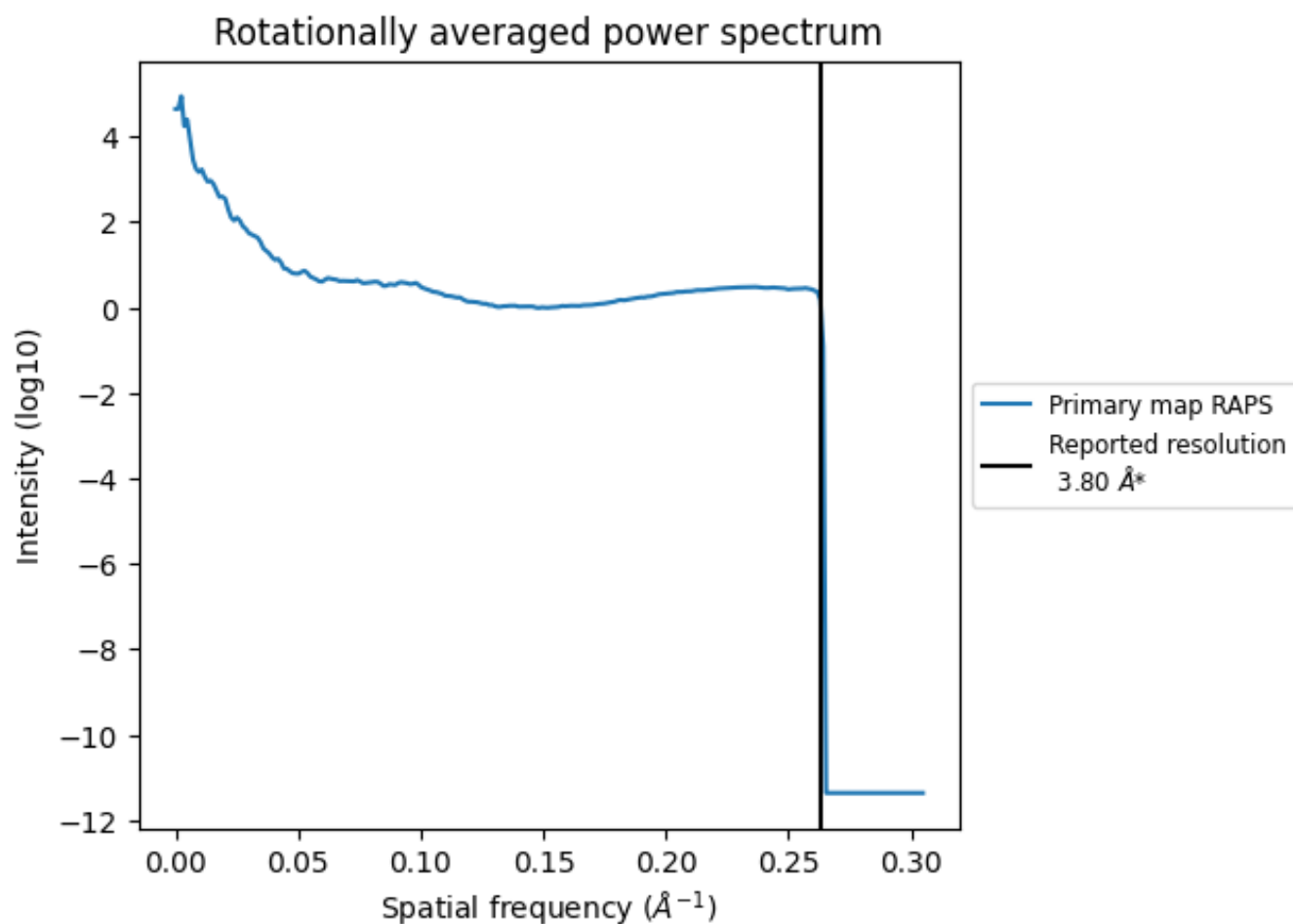
## 7.2 Volume estimate [i](#)



The volume at the recommended contour level is 1439 nm<sup>3</sup>; this corresponds to an approximate mass of 1300 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 [\(i\)](#)

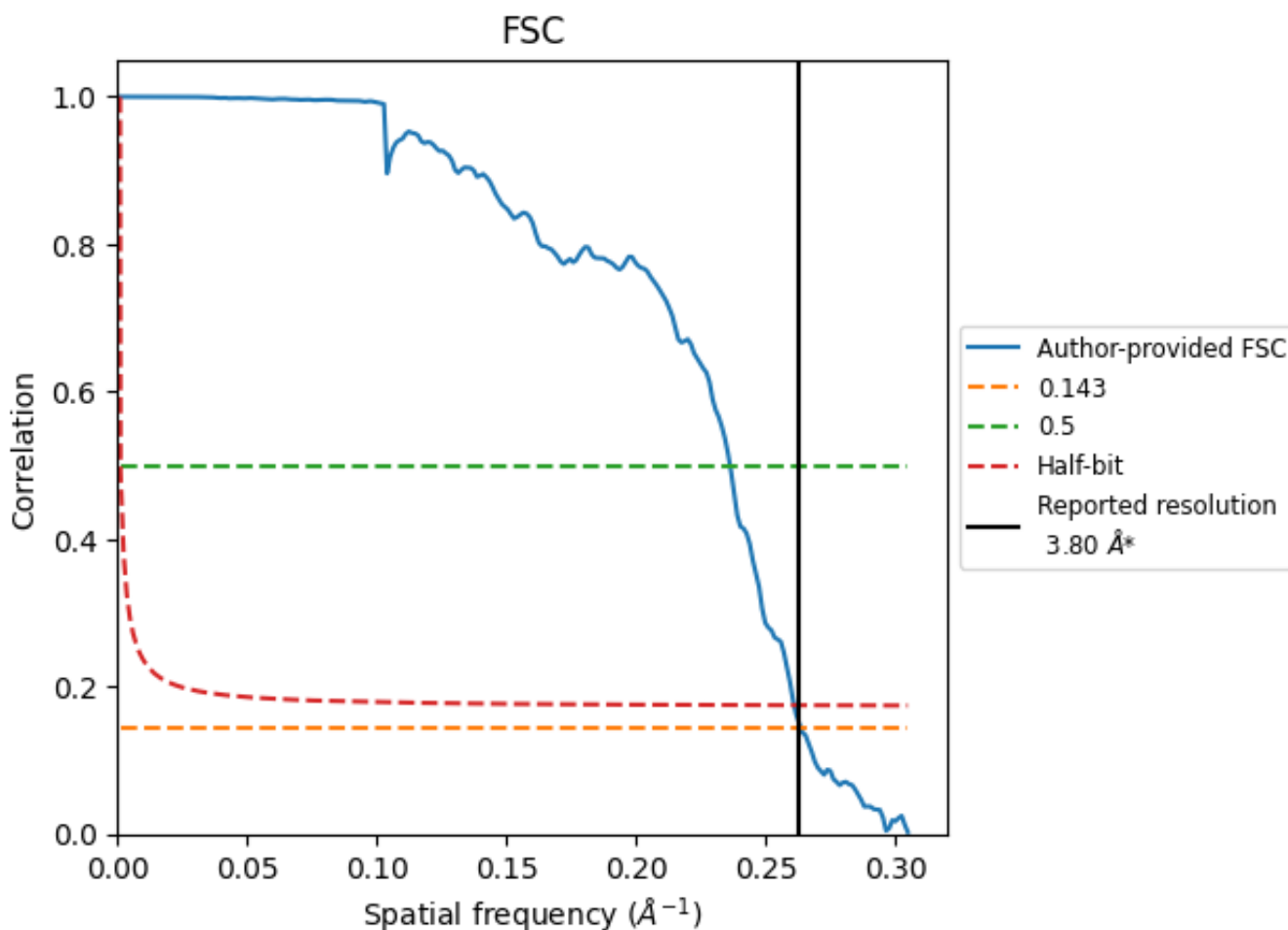


\*Reported resolution corresponds to spatial frequency of 0.263 Å<sup>-1</sup>

## 8 Fourier-Shell correlation [i](#)

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 [i](#)



\*Reported resolution corresponds to spatial frequency of 0.263 Å<sup>-1</sup>

## 8.2 Resolution estimates [i](#)

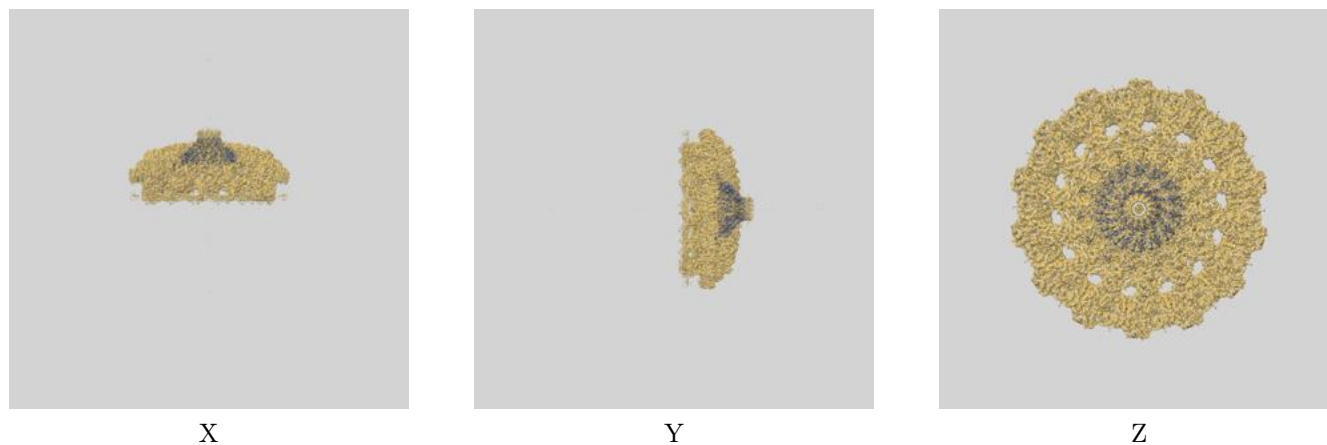
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.80	-	-
Author-provided FSC curve	3.79	4.23	3.83
Unmasked-calculated*	-	-	-

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

## 9 Map-model fit [i](#)

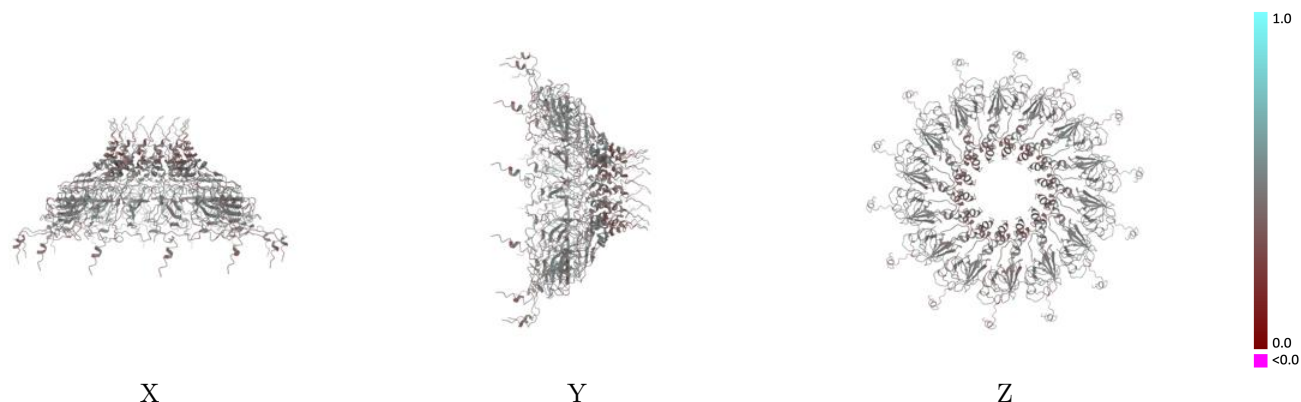
This section contains information regarding the fit between EMDB map EMD-20020 and PDB model 6ODI. Per-residue inclusion information can be found in section 3 on page 5.

### 9.1 Map-model overlay [i](#)



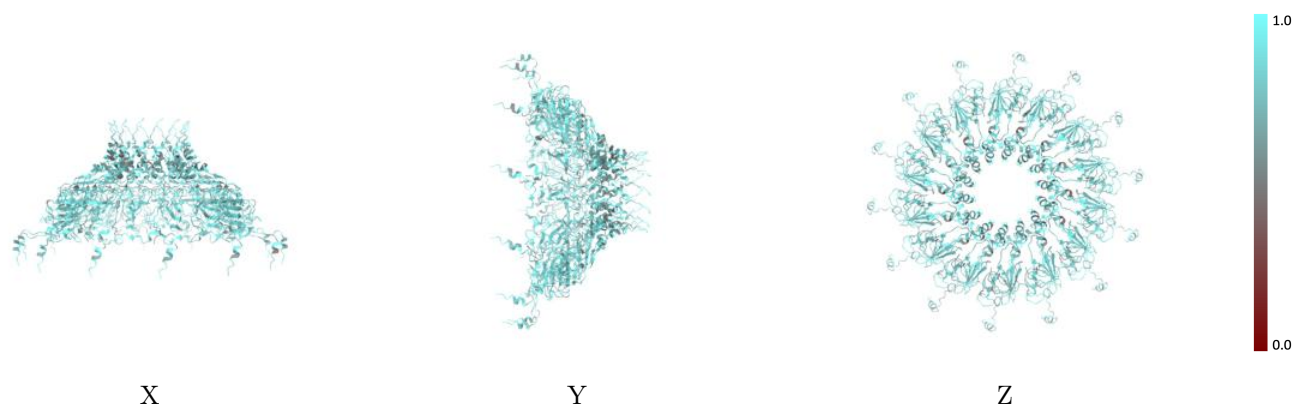
The images above show the 3D surface view of the map at the recommended contour level 0.027 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

## 9.2 Q-score mapped to coordinate model [i](#)



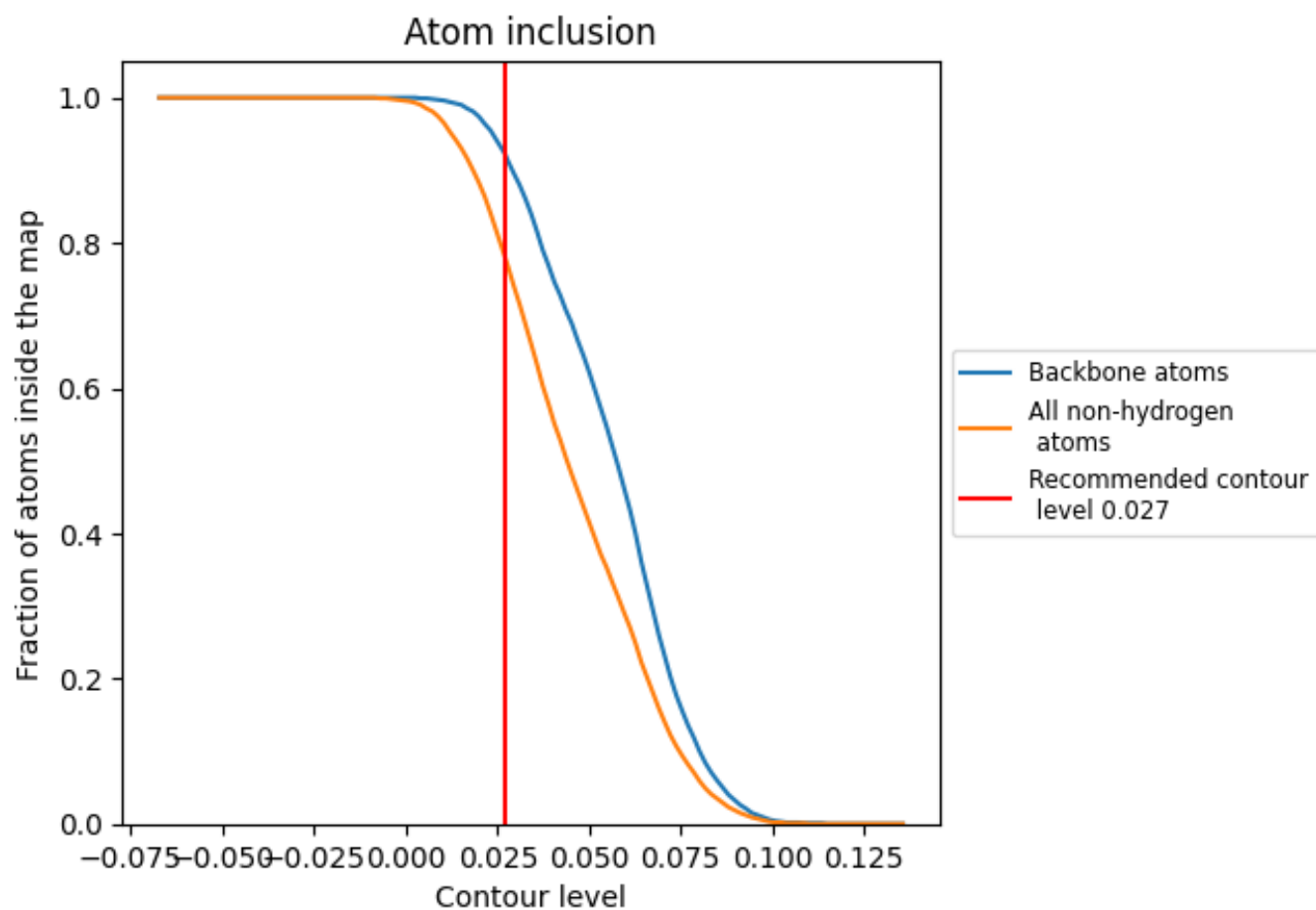
The images above show the model with each residue coloured according to its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions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



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 (0.027).































## 9.4 Atom inclusion [i](#)



At the recommended contour level, 92% 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 (0.027) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	 0.7820	 0.4500
b	 0.7710	 0.4370
c	 0.7800	 0.4500
d	 0.7930	 0.4580
e	 0.7890	 0.4600
f	 0.7890	 0.4580
g	 0.7780	 0.4510
h	 0.7530	 0.4180
i	 0.7980	 0.4550
j	 0.7970	 0.4580
k	 0.7930	 0.4590
l	 0.7700	 0.4460
m	 0.7700	 0.4450
n	 0.7830	 0.4450
o	 0.7900	 0.4570

