



## wwPDB EM Validation Summary Report ⓘ

Nov 8, 2022 – 05:26 PM EST

PDB ID : 6OEG  
EMDB ID : EMD-20020  
Title : Structure of CagX 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 : 2019-03-27  
Resolution : 3.80 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](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.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) : 2.31.2

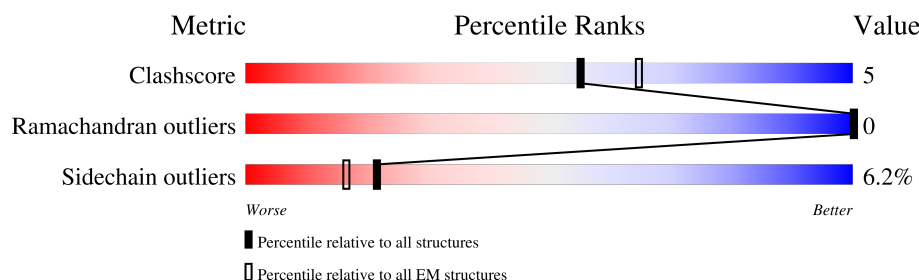
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*







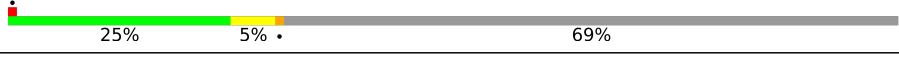

The reported resolution of this entry is 3.80 Å.

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	N	522	
1	O	522	
1	P	522	
1	Q	522	
1	R	522	
1	S	522	
1	T	522	
1	U	522	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	V	522	
1	W	522	
1	X	522	
1	Y	522	
1	Z	522	
1	a	522	

## 2 Entry composition [i](#)

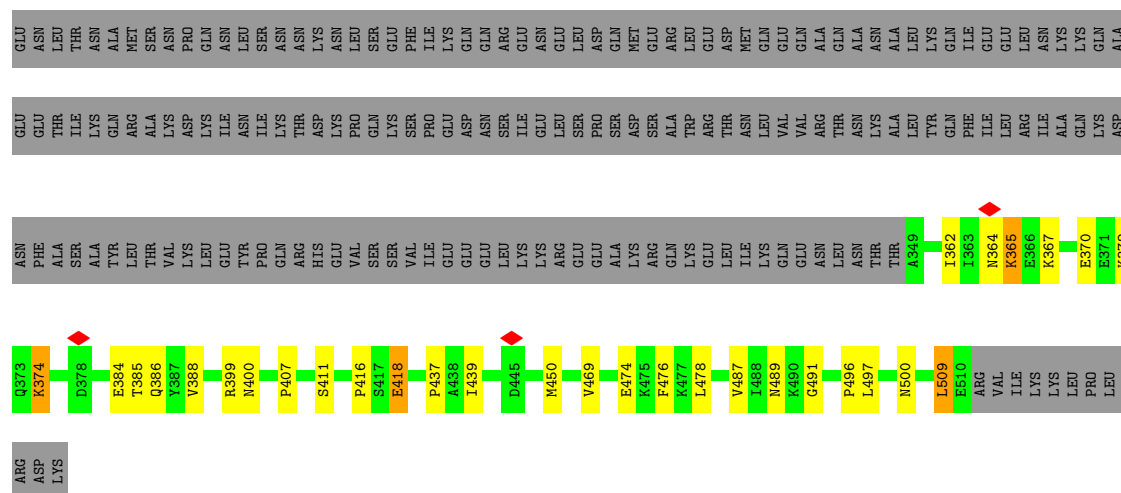
There is only 1 type of molecule in this entry. The entry contains 18564 atoms, of which 0 are hydrogens and 0 are deuteriums.

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.

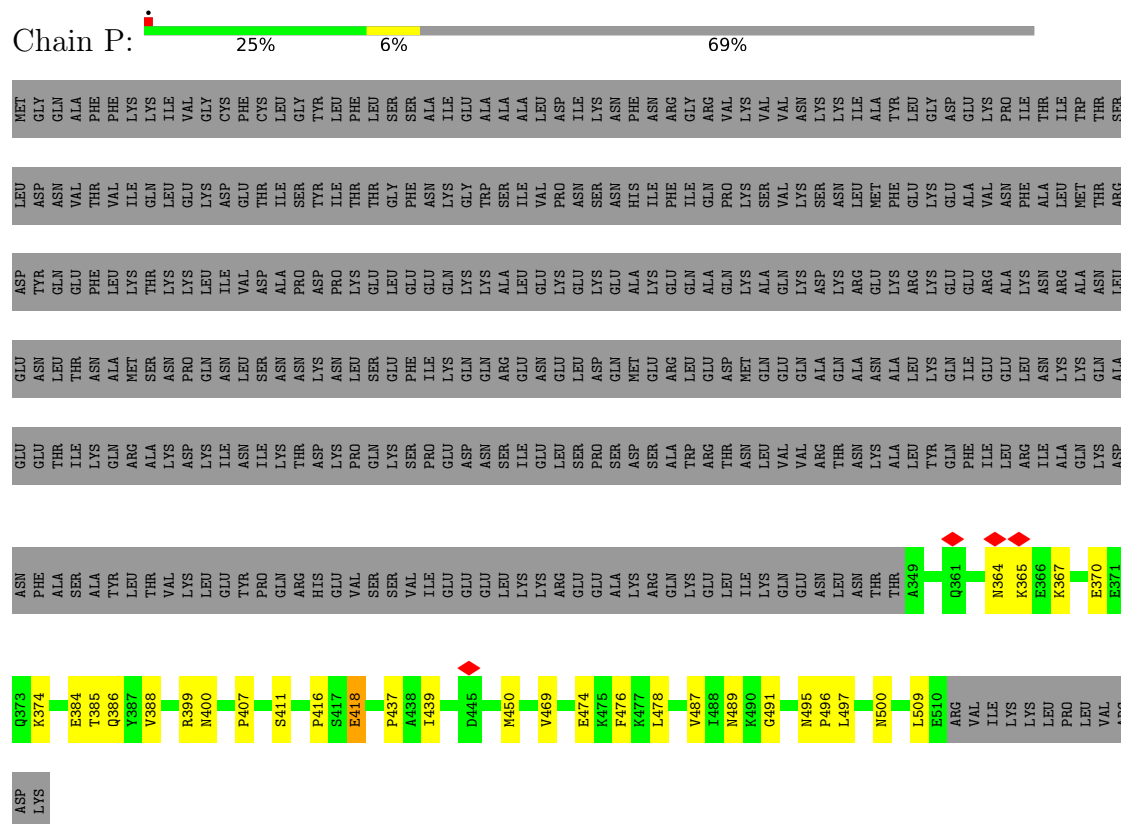
- Molecule 1 is a protein called Type IV secretion system apparatus protein CagX.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	N	162	Total	C	N	O	S	0	0
			1326	848	229	244	5		
1	O	162	Total	C	N	O	S	0	0
			1326	848	229	244	5		
1	P	162	Total	C	N	O	S	0	0
			1326	848	229	244	5		
1	Q	162	Total	C	N	O	S	0	0
			1326	848	229	244	5		
1	R	162	Total	C	N	O	S	0	0
			1326	848	229	244	5		
1	S	162	Total	C	N	O	S	0	0
			1326	848	229	244	5		
1	T	162	Total	C	N	O	S	0	0
			1326	848	229	244	5		
1	U	162	Total	C	N	O	S	0	0
			1326	848	229	244	5		
1	V	162	Total	C	N	O	S	0	0
			1326	848	229	244	5		
1	W	162	Total	C	N	O	S	0	0
			1326	848	229	244	5		
1	X	162	Total	C	N	O	S	0	0
			1326	848	229	244	5		
1	Y	162	Total	C	N	O	S	0	0
			1326	848	229	244	5		
1	Z	162	Total	C	N	O	S	0	0
			1326	848	229	244	5		
1	a	162	Total	C	N	O	S	0	0
			1326	848	229	244	5		

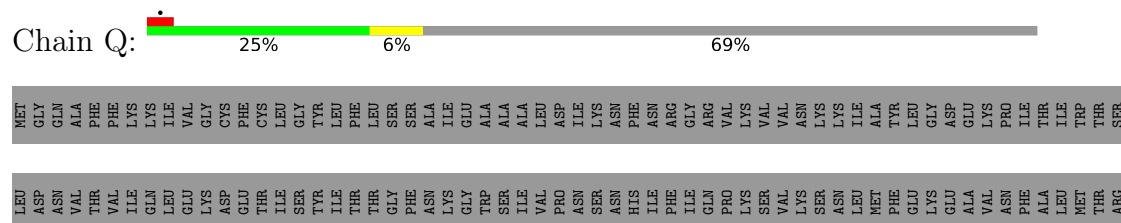




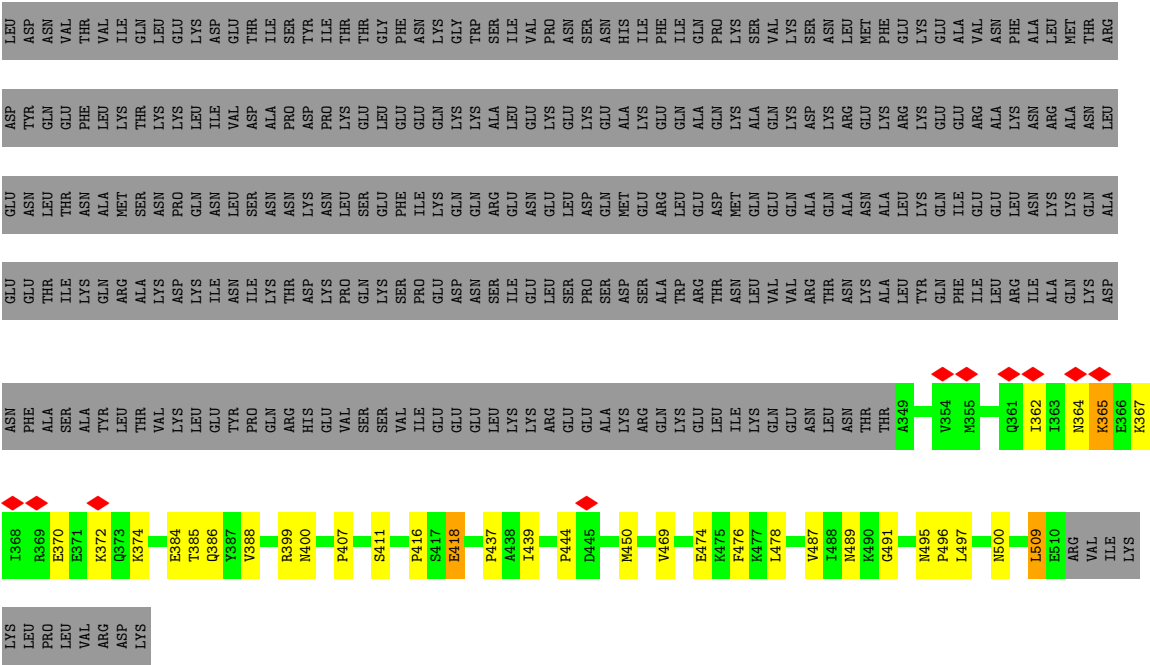
- Molecule 1: Type IV secretion system apparatus protein CagX



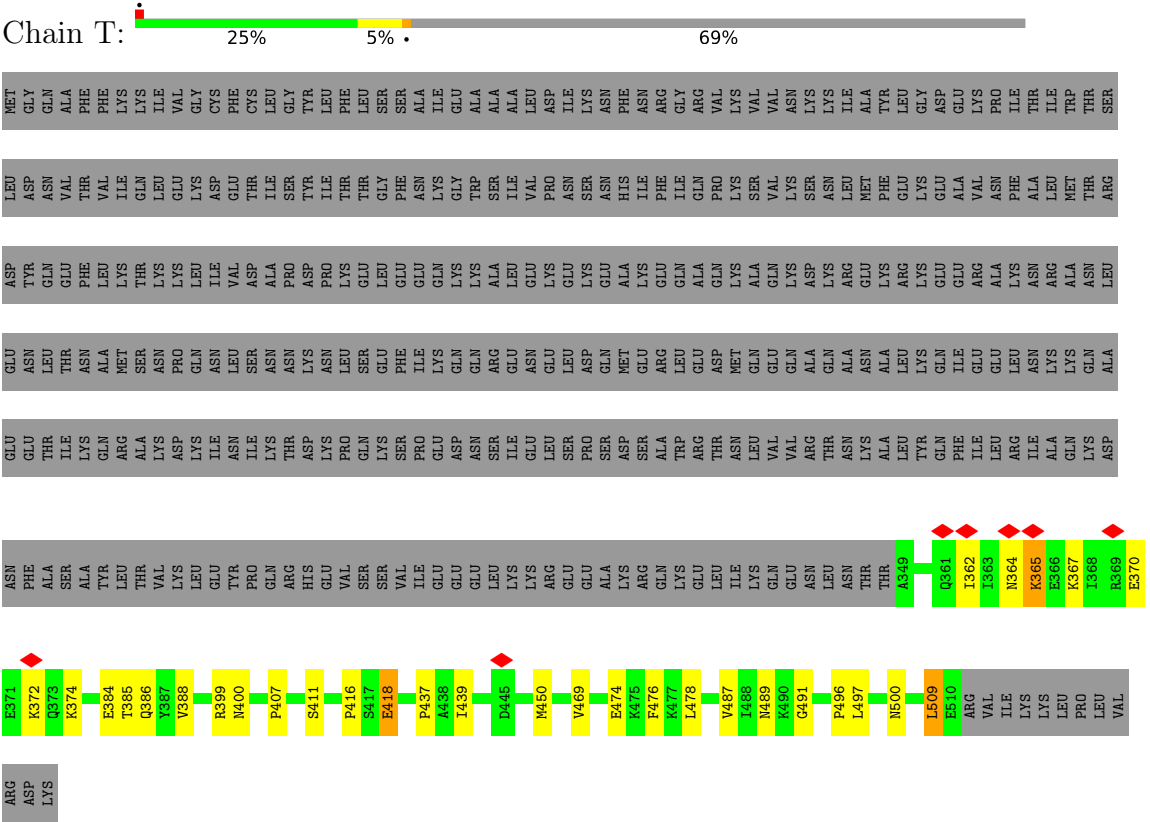
- Molecule 1: Type IV secretion system apparatus protein CagX







• Molecule 1: Type IV secretion system apparatus protein CagX



• Molecule 1: Type IV secretion system apparatus protein CagX







LEU VAL ARG ASP LYS	K372	ASN	GLU	GLU	GLU	ASP	GLU	ASP	LEU	MET
	Q373	PHE	THR	THR	ASN	LYN	ASN	GLN	GLY	
	K374	ALA	ILE	ILE	THR	GLU	THR	VAL	ALA	
	E384	SER	LYS	LYS	ASN	PHE	ASN	THR	PHE	
	T385	LYR	GLN	GLN	ALA	LEU	ALA	VAL	LYS	
	Q386	LEU	ARG	ARG	MET	LYS	MET	ILE	LYS	
	Y387	THR	ALA	ALA	SER	THR	ASN	GLN	LYS	
	V388	VAL	LYS	LYS	ASP	PRO	ASN	GLU	VAL	
	R399	LEU	LYS	LYS	GLN	LEU	ASN	LYS	GLY	
	M400	GLU	ILE	ILE	ASN	VAL	ASN	ASP	CYS	
	P407	TYR	ASN	ASN	LEU	VAL	LEU	GLU	PHE	
	R410	PRO	ILE	ILE	SER	ASP	ASN	THR	CYS	
	S411	ARG	LYS	THR	ASN	ALA	ASN	GLY	LYS	
	H412	HTS	GLU	ASP	LYS	ASP	LYS	TYR	TYR	
	H413	VAL	GLU	PRO	LEU	PRO	ASN	ILE	LEU	
	E418	SER	SER	GLN	SER	PHE	GLU	THR	PHE	
	P416	VAL	VAL	SER	GLU	GLU	GLU	GLY	SER	
	S417	ILE	GLU	PRO	ILE	GLN	GLN	LYS	ALA	
	E418	ILE	GLU	GLU	GLU	LYS	LYS	ILE	ILE	
	P437	GLU	GLU	ASP	ASN	LYS	GLY	GLY	GLU	
	A438	LEU	GLU	ASN	GLN	LYS	LYS	THR	ALA	
	T439	LEU	SER	SER	ARG	ALA	ALA	SER	ALA	
	M450	LYS	ILE	ILE	GLU	LEU	LEU	ILE	ALA	
	E469	LYS	GLU	GLU	ASN	GLU	GLU	VAL	LEU	
	E474	ARG	LEU	SER	GLU	LYS	LYS	PRO	ASP	
	K475	GLU	PRO	PRO	ASP	GLU	GLU	ASN	ILE	
	F476	LYS	ASP	ASP	MET	ALA	GLU	ASN	LYS	
	K477	ARG	SER	SER	GLU	LYS	LYS	ILE	ARG	
	L478	GLN	ALA	TRP	ARG	GLU	GLN	PHE	ARG	
	V487	LYS	LYS	LEU	LEU	GLN	GLN	ILE	GLY	
	T488	LEU	GLU	THR	GLU	ALA	GLN	ARG	VAL	
	M489	ILE	ILE	ASN	ASP	LYS	LYS	PRO	VAL	
	K490	LYS	LYS	LEU	MET	ALA	ALA	SER	LYS	
	G491	GLN	GLN	VAL	GLU	GLN	GLN	VAL	VAL	
	K494	ASN	ASN	THR	GLN	ASP	LYS	SER	ASN	
	M495	THR	ASN	LYS	ASN	ARG	GLU	MET	ALA	
	L497	THR	ALA	ALA	LEU	LYS	LYS	PHE	LEU	
	E500	ARG	TYR	TYR	LEU	ARG	ARG	GLU	GLY	
	L509	ILE	GLN	GLN	LYS	LYS	LYS	GLU	ASP	
	E510	ARG	PHE	ILE	ILE	GLU	GLU	ALA	GLY	
	V516	VAL	LEU	LEU	GLU	ARG	VAL	VAL	LYS	
	ILE	ARG	ARG	ARG	LEU	ALA	ASN	ASN	PRO	
	LYS	ILE	ILE	ILE	ASN	LYS	ASN	PHE	THR	
	LYS	ALA	ALA	ALA	LYS	ARG	ARG	ALA	ILE	
	LYS	LYS	GLN	GLN	LYS	ALA	LYS	MET	TRP	
	LEU	LEU	ASN	ASN	GLN	LEU	ASN	THR	SER	
	E370	ALA	ASN	ASN	ALA	LEU	LEU	THR	THR	
	E370	ALA	ASN	ASN	ALA	LEU	LEU	THR	THR	

- Chain X:  25% 6% 69%

[illegible]





## 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

### 5.1 Standard geometry

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	N	0.44	0/1353	0.62	1/1824 (0.1%)
1	O	0.44	0/1353	0.62	1/1824 (0.1%)
1	P	0.44	0/1353	0.62	1/1824 (0.1%)
1	Q	0.44	0/1353	0.62	1/1824 (0.1%)
1	R	0.44	0/1353	0.62	1/1824 (0.1%)
1	S	0.44	0/1353	0.62	1/1824 (0.1%)
1	T	0.44	0/1353	0.62	1/1824 (0.1%)
1	U	0.44	0/1353	0.62	1/1824 (0.1%)
1	V	0.44	0/1353	0.62	1/1824 (0.1%)
1	W	0.44	0/1353	0.62	1/1824 (0.1%)
1	X	0.44	0/1353	0.62	1/1824 (0.1%)
1	Y	0.44	0/1353	0.62	1/1824 (0.1%)
1	Z	0.44	0/1353	0.62	1/1824 (0.1%)
1	a	0.44	0/1353	0.62	1/1824 (0.1%)
All	All	0.44	0/18942	0.62	14/25536 (0.1%)

There are no bond length outliers.

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	U	418	GLU	CA-CB-CG	5.82	126.20	113.40
1	V	418	GLU	CA-CB-CG	5.82	126.19	113.40
1	Z	418	GLU	CA-CB-CG	5.81	126.18	113.40
1	R	418	GLU	CA-CB-CG	5.81	126.17	113.40
1	N	418	GLU	CA-CB-CG	5.80	126.17	113.40

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	N	1326	0	1343	14	0
1	O	1326	0	1343	15	0
1	P	1326	0	1343	12	0
1	Q	1326	0	1343	14	0
1	R	1326	0	1343	19	0
1	S	1326	0	1343	17	0
1	T	1326	0	1343	13	0
1	U	1326	0	1343	13	0
1	V	1326	0	1343	16	0
1	W	1326	0	1343	18	0
1	X	1326	0	1343	14	0
1	Y	1326	0	1343	14	0
1	Z	1326	0	1343	13	0
1	a	1326	0	1343	0	0
All	All	18564	0	18802	184	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.

The worst 5 of 184 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:R:432:ASN:HD21	1:S:497:LEU:HB2	1.49	0.77
1:R:432:ASN:ND2	1:S:497:LEU:HB2	2.18	0.59
1:S:400:ASN:ND2	1:S:491:GLY:O	2.38	0.57
1:T:400:ASN:ND2	1:T:491:GLY:O	2.38	0.57
1:Y:400:ASN:ND2	1:Y:491:GLY:O	2.38	0.57

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	N	160/522 (31%)	147 (92%)	13 (8%)	0	100	100
1	O	160/522 (31%)	147 (92%)	13 (8%)	0	100	100
1	P	160/522 (31%)	147 (92%)	13 (8%)	0	100	100
1	Q	160/522 (31%)	147 (92%)	13 (8%)	0	100	100
1	R	160/522 (31%)	147 (92%)	13 (8%)	0	100	100
1	S	160/522 (31%)	147 (92%)	13 (8%)	0	100	100
1	T	160/522 (31%)	147 (92%)	13 (8%)	0	100	100
1	U	160/522 (31%)	147 (92%)	13 (8%)	0	100	100
1	V	160/522 (31%)	147 (92%)	13 (8%)	0	100	100
1	W	160/522 (31%)	147 (92%)	13 (8%)	0	100	100
1	X	160/522 (31%)	147 (92%)	13 (8%)	0	100	100
1	Y	160/522 (31%)	147 (92%)	13 (8%)	0	100	100
1	Z	160/522 (31%)	147 (92%)	13 (8%)	0	100	100
1	a	160/522 (31%)	147 (92%)	13 (8%)	0	100	100
All	All	2240/7308 (31%)	2058 (92%)	182 (8%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains

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	N	144/470 (31%)	135 (94%)	9 (6%)	18	49
1	O	144/470 (31%)	135 (94%)	9 (6%)	18	49
1	P	144/470 (31%)	135 (94%)	9 (6%)	18	49
1	Q	144/470 (31%)	135 (94%)	9 (6%)	18	49
1	R	144/470 (31%)	135 (94%)	9 (6%)	18	49
1	S	144/470 (31%)	135 (94%)	9 (6%)	18	49
1	T	144/470 (31%)	135 (94%)	9 (6%)	18	49
1	U	144/470 (31%)	135 (94%)	9 (6%)	18	49
1	V	144/470 (31%)	135 (94%)	9 (6%)	18	49
1	W	144/470 (31%)	135 (94%)	9 (6%)	18	49
1	X	144/470 (31%)	135 (94%)	9 (6%)	18	49
1	Y	144/470 (31%)	135 (94%)	9 (6%)	18	49
1	Z	144/470 (31%)	135 (94%)	9 (6%)	18	49
1	a	144/470 (31%)	135 (94%)	9 (6%)	18	49
All	All	2016/6580 (31%)	1890 (94%)	126 (6%)	21	49

5 of 126 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	T	374	LYS
1	Z	372	LYS
1	U	509	LEU
1	Z	365	LYS
1	a	365	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 28 such sidechains are listed below:

Mol	Chain	Res	Type
1	U	500	ASN
1	a	505	ASN
1	V	505	ASN
1	Z	500	ASN
1	V	500	ASN

### 5.3.3 RNA ⓘ

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.

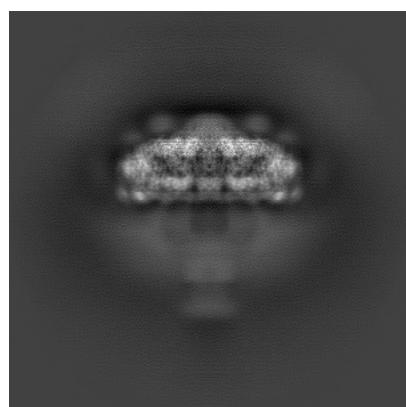
## 6 Map visualisation [i](#)

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

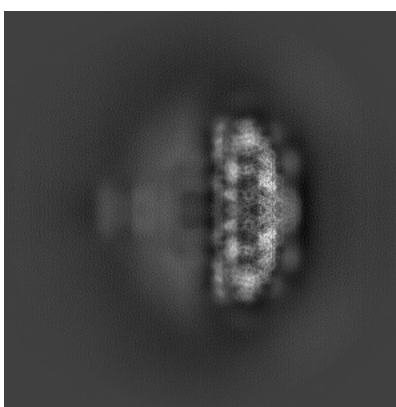
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

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

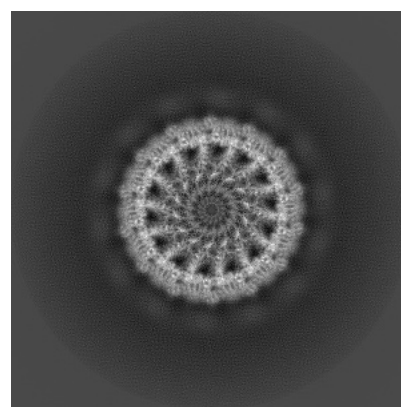
#### 6.1.1 Primary map



X



Y

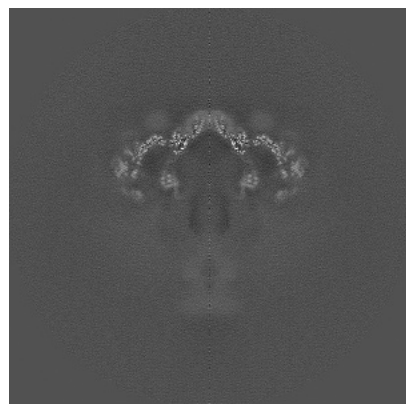


Z

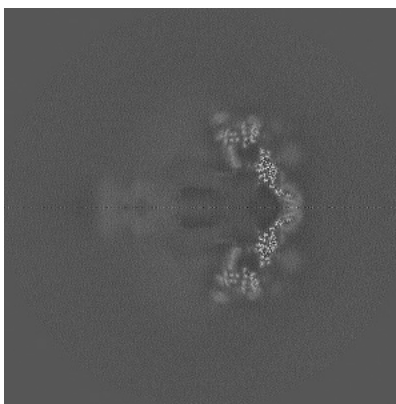
The images above show the map projected in three orthogonal directions.

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

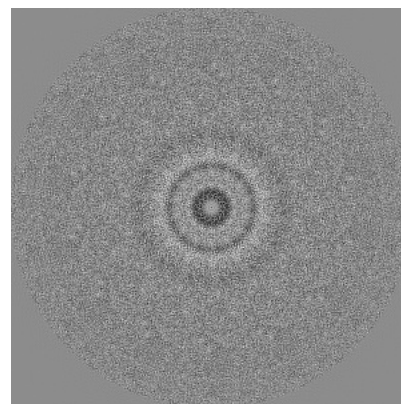
#### 6.2.1 Primary map



X Index: 255



Y Index: 255

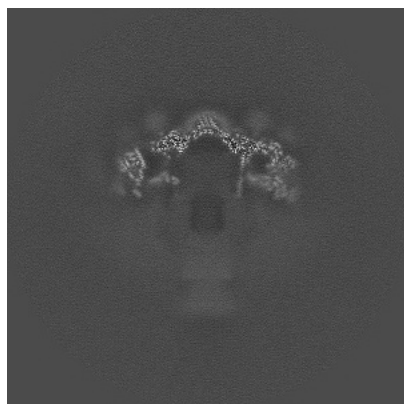


Z Index: 255

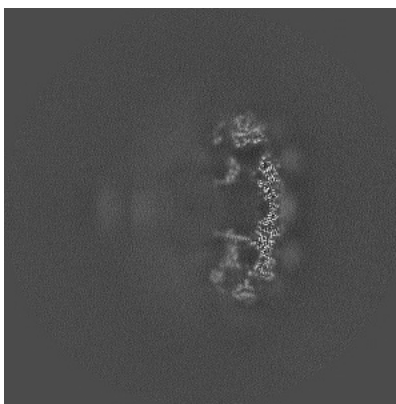
The images above show central slices of the map in three orthogonal directions.

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

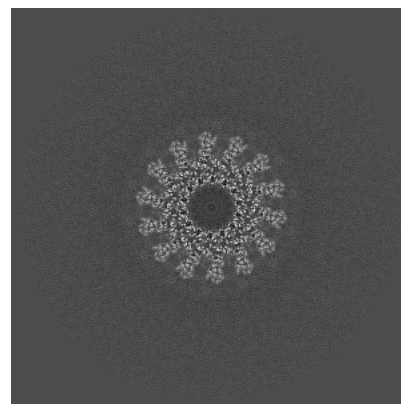
### 6.3.1 Primary map



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 surface views [i](#)

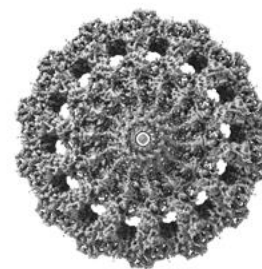
### 6.4.1 Primary map



X



Y



Z

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

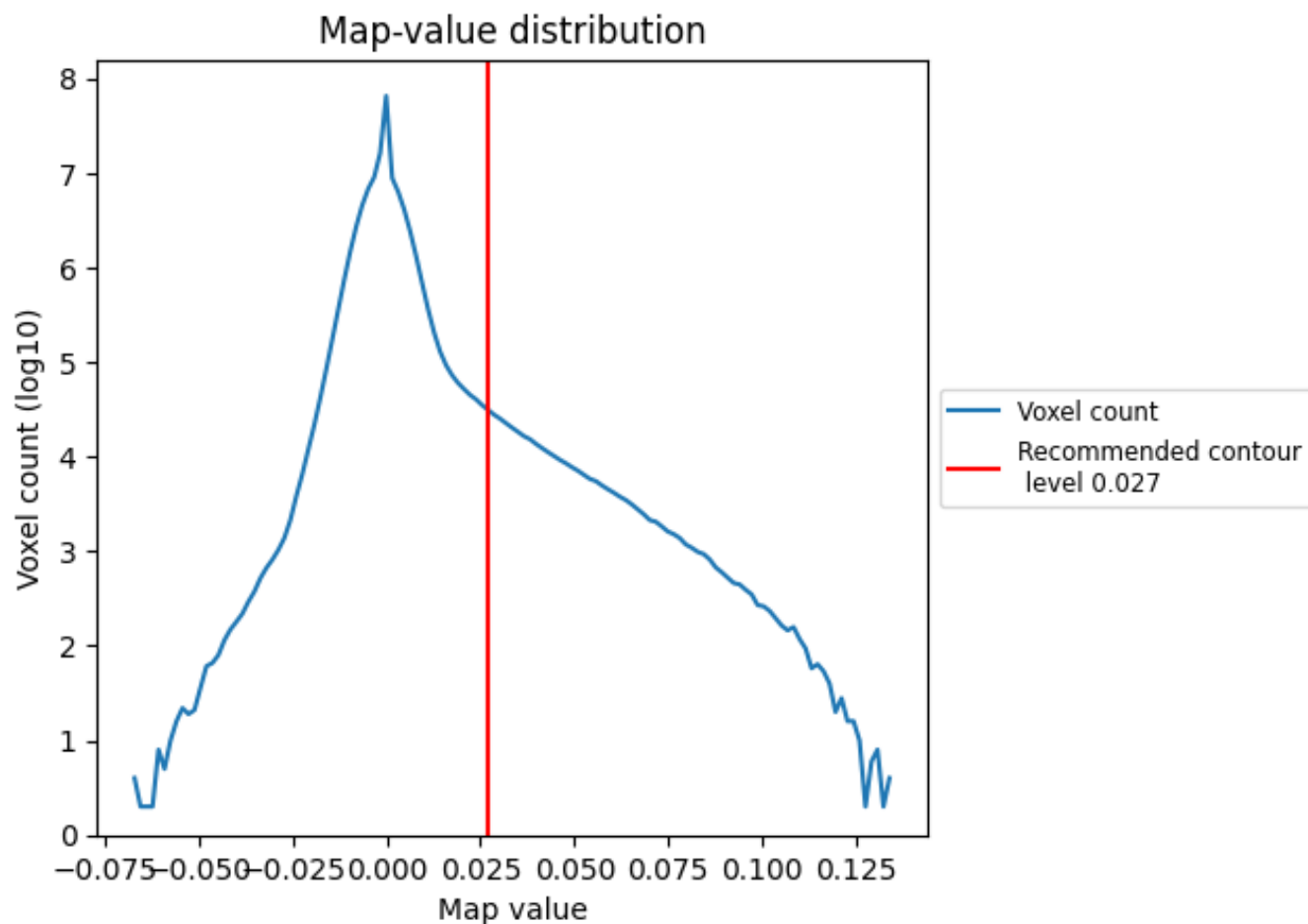
## 6.5 Mask visualisation

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

## 7 Map analysis [i](#)

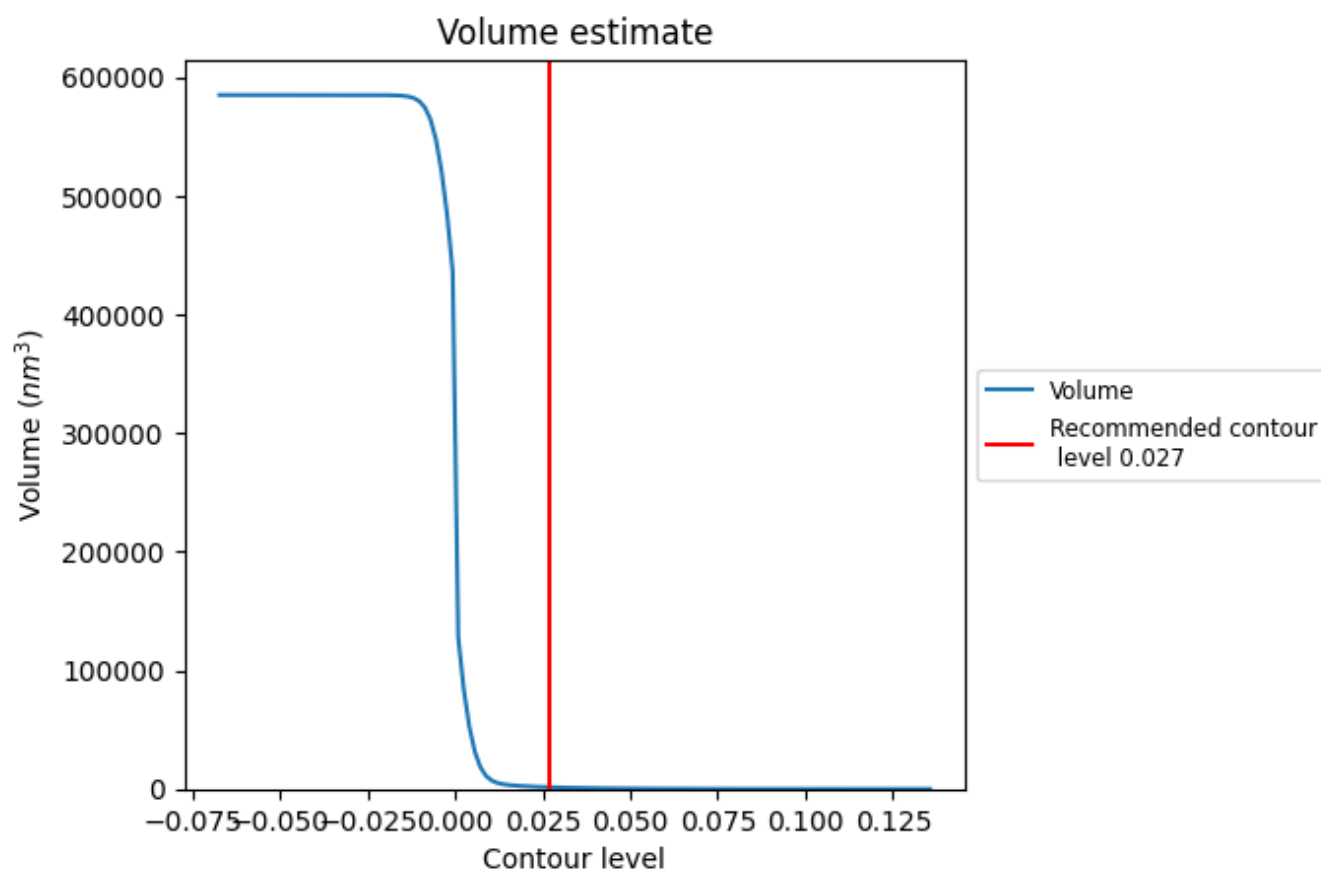
This section contains the results of statistical analysis of the map.

### 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.

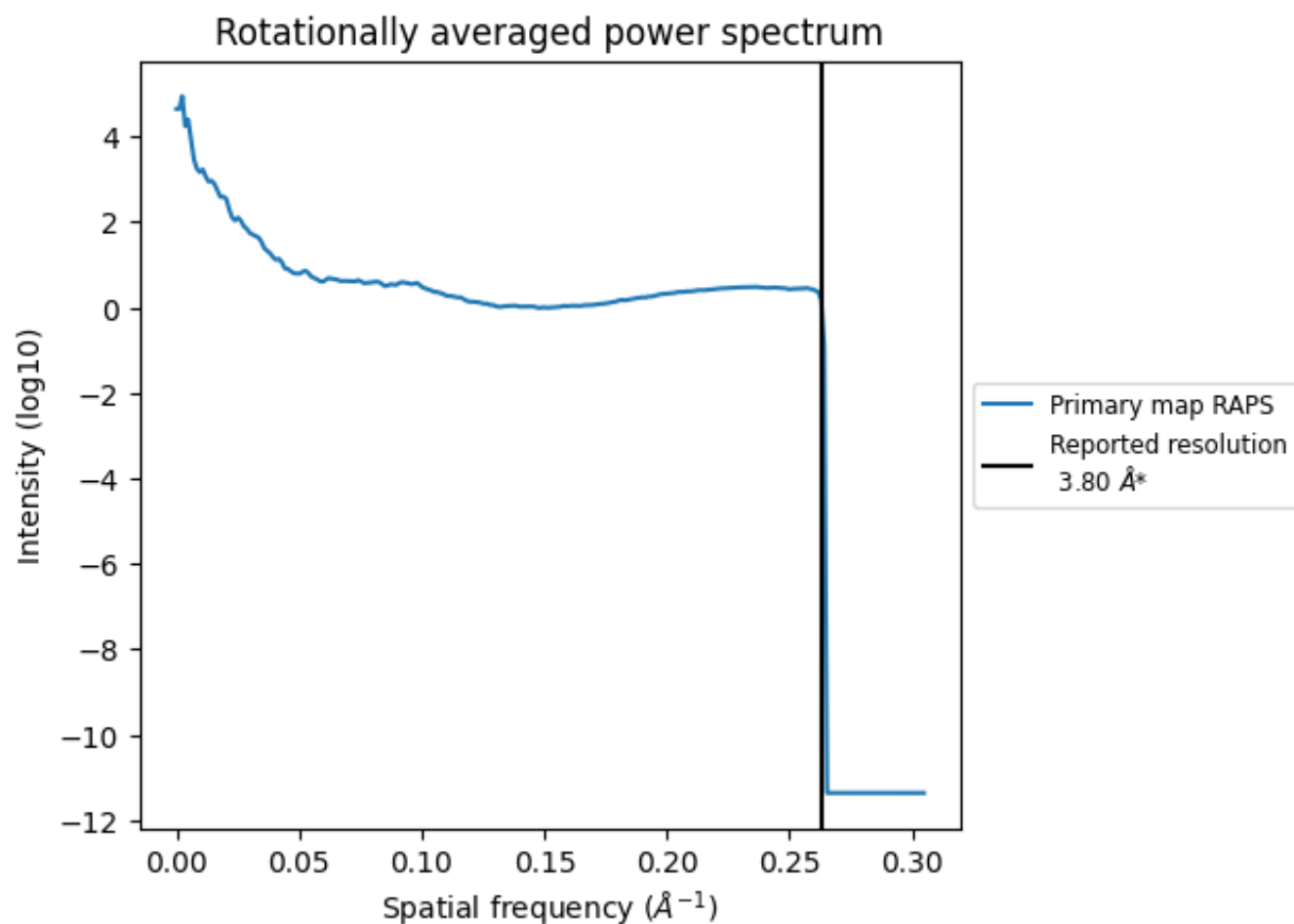
## 7.2 Volume estimate [i](#)



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 ⓘ



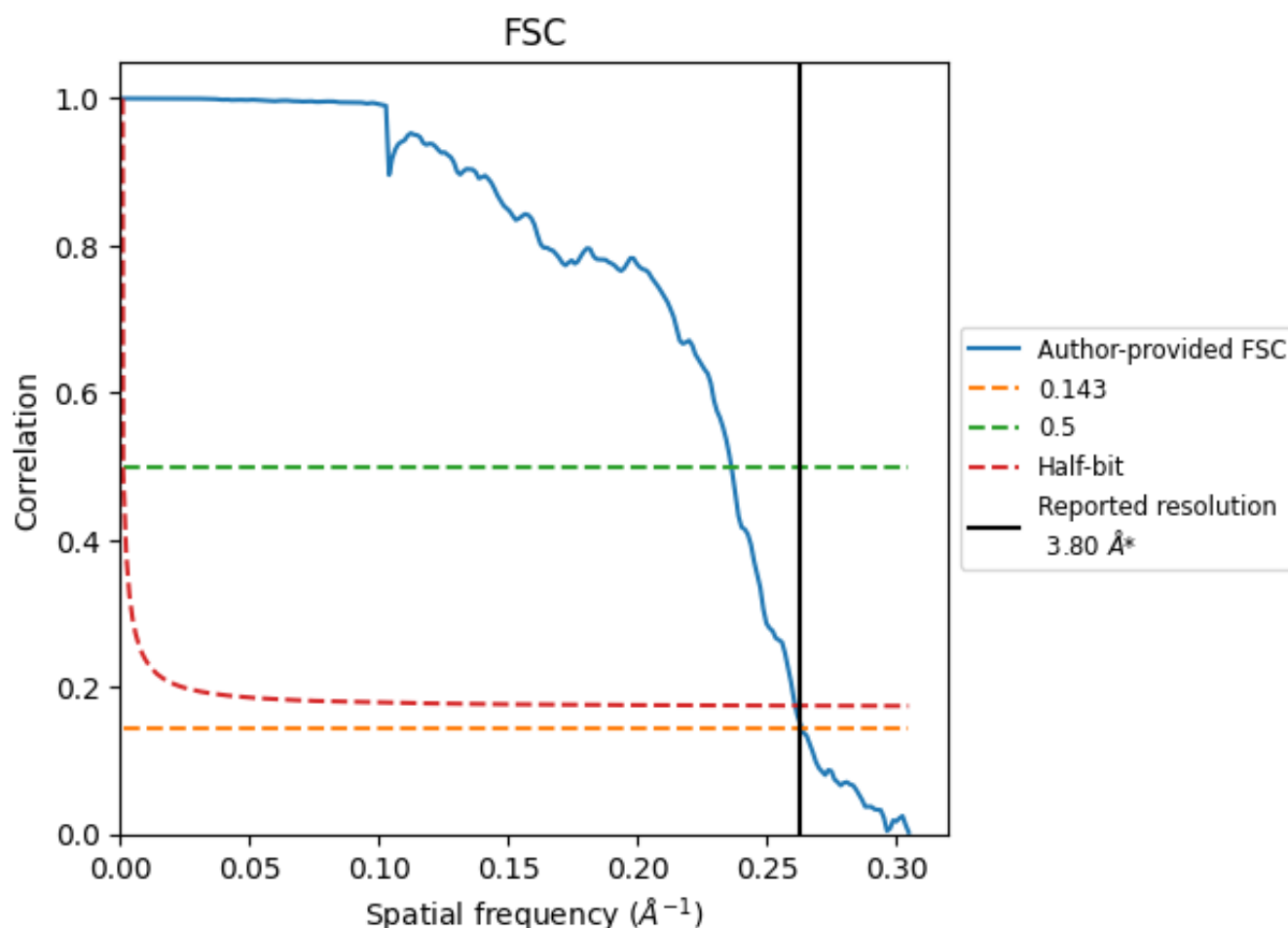
\*Reported resolution corresponds to spatial frequency of 0.263 Å<sup>-1</sup>



## 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](#)

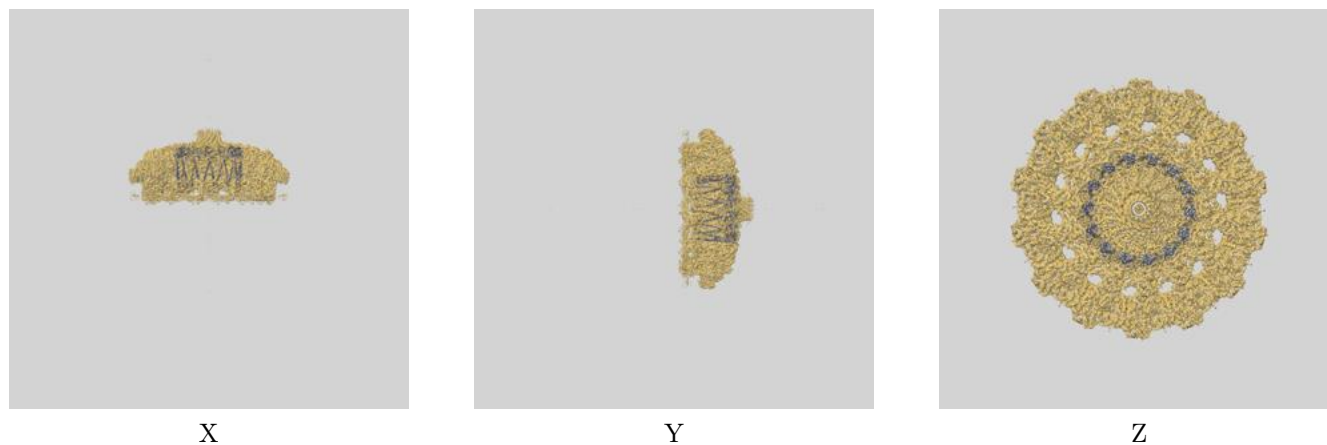
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.80	-	-
Author-provided FSC curve	3.79	4.23	3.83
Unmasked-calculated*	-	-	-

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

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

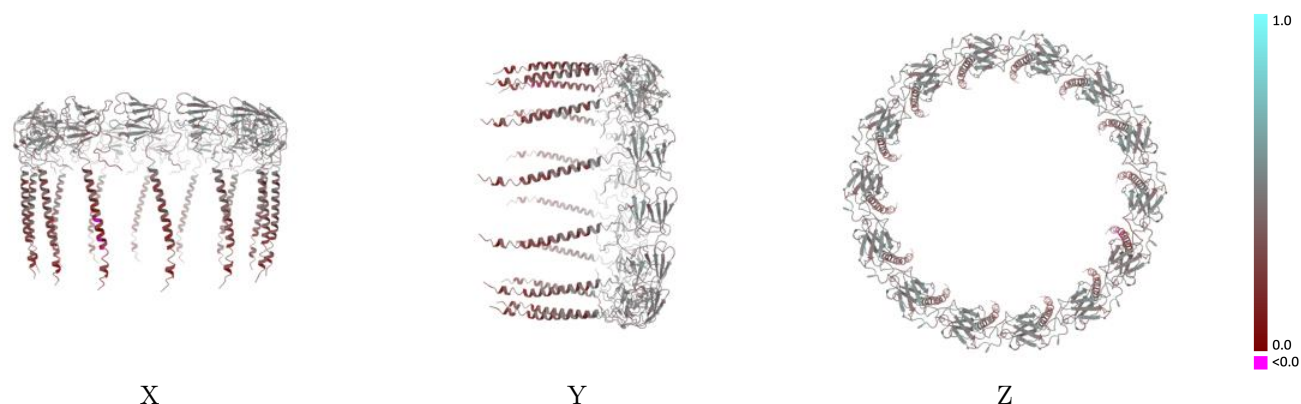
This section contains information regarding the fit between EMDB map EMD-20020 and PDB model 6OEG. Per-residue inclusion information can be found in [section 3](#) on [page 5](#).

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



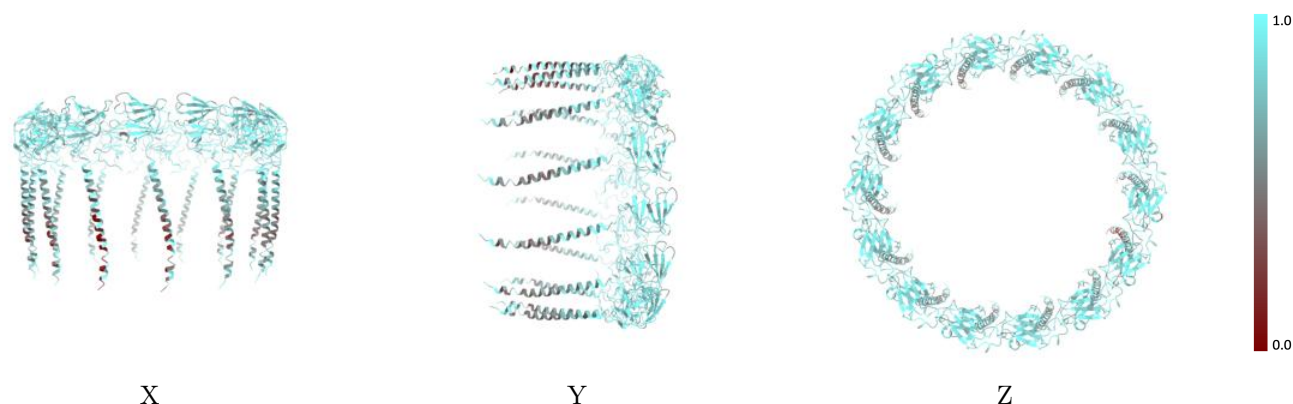
The images above show the 3D surface view of the map at the recommended contour level 0.027 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

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



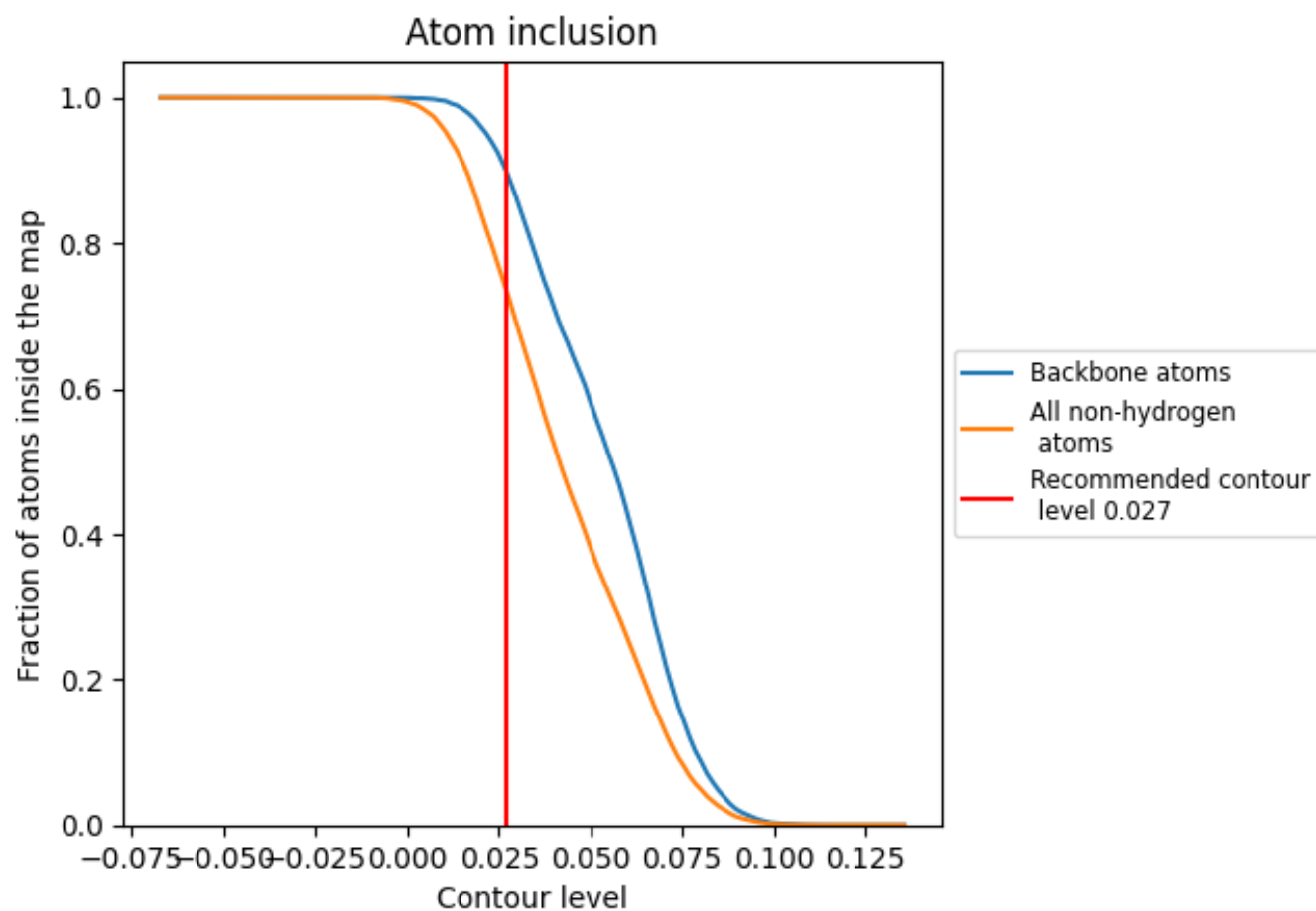
The images above show the model with each residue coloured according to its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions 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, 90% of all backbone atoms, 74% 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	 0.7369	 0.4090
N	 0.7433	 0.4200
O	 0.7479	 0.4200
P	 0.7502	 0.4180
Q	 0.7348	 0.4020
R	 0.6733	 0.3420
S	 0.7394	 0.4020
T	 0.7394	 0.4130
U	 0.7510	 0.4220
V	 0.7279	 0.4110
W	 0.7448	 0.4130
X	 0.7294	 0.4070
Y	 0.7517	 0.4230
Z	 0.7502	 0.4220
a	 0.7325	 0.4140

