



# wwPDB EM Validation Summary Report ⓘ

Nov 25, 2024 – 06:22 PM JST

PDB ID : 8XA7  
EMDB ID : EMD-38196  
Title : Cryo-EM structure of Bacillus subtilis RNAP,sigA and SPO1 gp33 complex  
Authors : Chen, M.; Jin, Q.; Yuan, S.; Liu, B.  
Deposited on : 2023-12-02  
Resolution : 2.94 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:

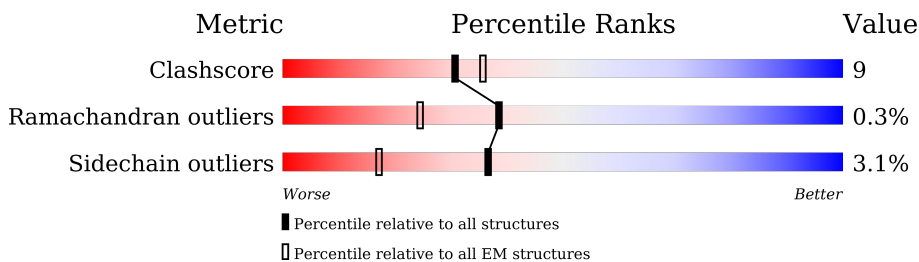
EMDB validation analysis : 0.0.1.dev113  
MolProbity : 4.02b-467  
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)  
MapQ : 1.9.13  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.40

# 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 2.94 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  $\geq 3$ , 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	A	314	
1	B	314	
2	C	1193	
3	D	1199	
4	F	67	
5	G	371	
6	H	101	
7	L	173	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
8	E	69	 83% 17%

## 2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 51724 atoms, of which 25990 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 DNA-directed RNA polymerase subunit alpha.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	H	N	O		
1	A	224	Total	C	H	N	O	0	0
			3456	1084	1736	290	346		
1	B	224	Total	C	H	N	O	0	0
			3456	1084	1736	290	346		

- Molecule 2 is a protein called DNA-directed RNA polymerase subunit beta.

Mol	Chain	Residues	Atoms					AltConf	Trace	
			Total	C	H	N	O			S
2	C	1139	Total	C	H	N	O	S	0	0
			17931	5643	8940	1572	1743	33		

- Molecule 3 is a protein called DNA-directed RNA polymerase subunit beta'.

Mol	Chain	Residues	Atoms					AltConf	Trace	
			Total	C	H	N	O			S
3	D	1176	Total	C	H	N	O	S	0	0
			18775	5841	9518	1650	1723	43		

- Molecule 4 is a protein called DNA-directed RNA polymerase subunit omega.

Mol	Chain	Residues	Atoms					AltConf	Trace	
			Total	C	H	N	O			S
4	F	61	Total	C	H	N	O	S	0	0
			990	305	505	83	93	4		

- Molecule 5 is a protein called RNA polymerase sigma factor SigA.

Mol	Chain	Residues	Atoms					AltConf	Trace	
			Total	C	H	N	O			S
5	G	178	Total	C	H	N	O	S	0	0
			2946	915	1488	267	273	3		

- Molecule 6 is a protein called Gene 33 protein.

Mol	Chain	Residues	Atoms						AltConf	Trace
			Total	C	H	N	O	S		
6	H	94	1554	495	778	129	148	4	0	0

- Molecule 7 is a protein called DNA-directed RNA polymerase subunit delta.

Mol	Chain	Residues	Atoms						AltConf	Trace
			Total	C	H	N	O	S		
7	L	92	1503	489	740	122	150	2	0	0

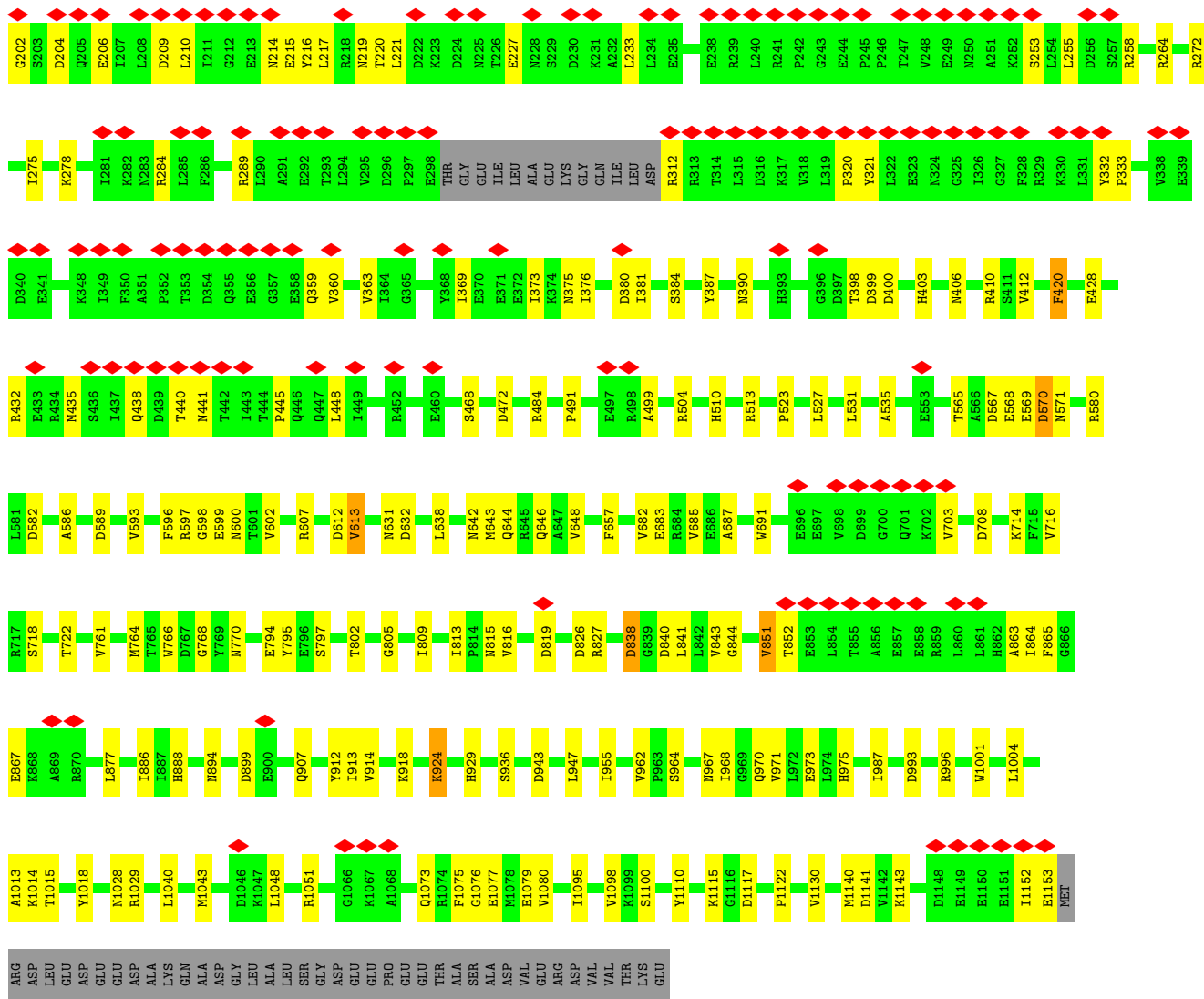
- Molecule 8 is a protein called DNA-directed RNA polymerase subunit epsilon.

Mol	Chain	Residues	Atoms						AltConf	Trace
			Total	C	H	N	O	S		
8	E	69	1113	365	549	89	109	1	0	0

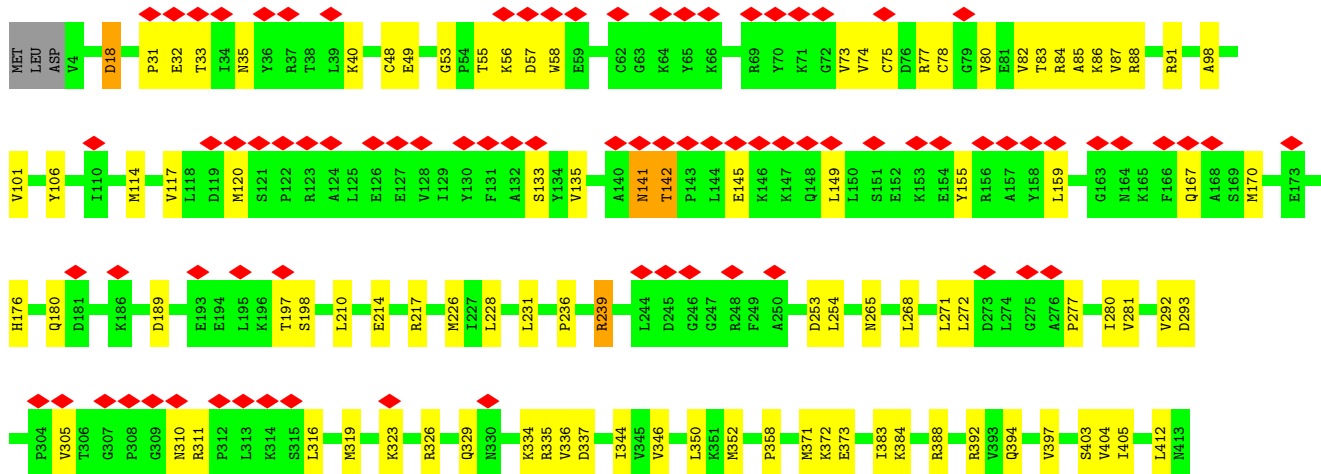
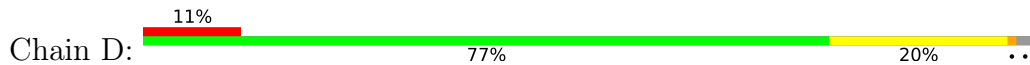
There is a discrepancy between the modelled and reference sequences:

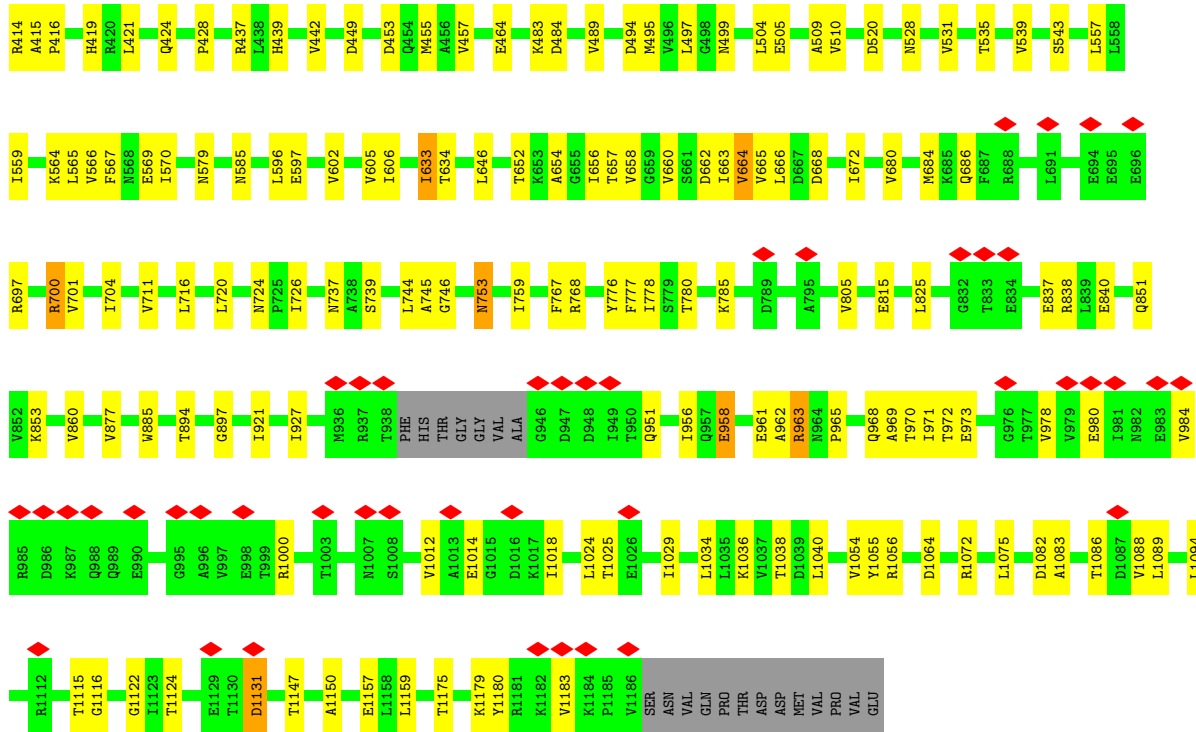
Chain	Residue	Modelled	Actual	Comment	Reference
E	33	ILE	VAL	engineered mutation	UNP O31718





● Molecule 3: DNA-directed RNA polymerase subunit beta'

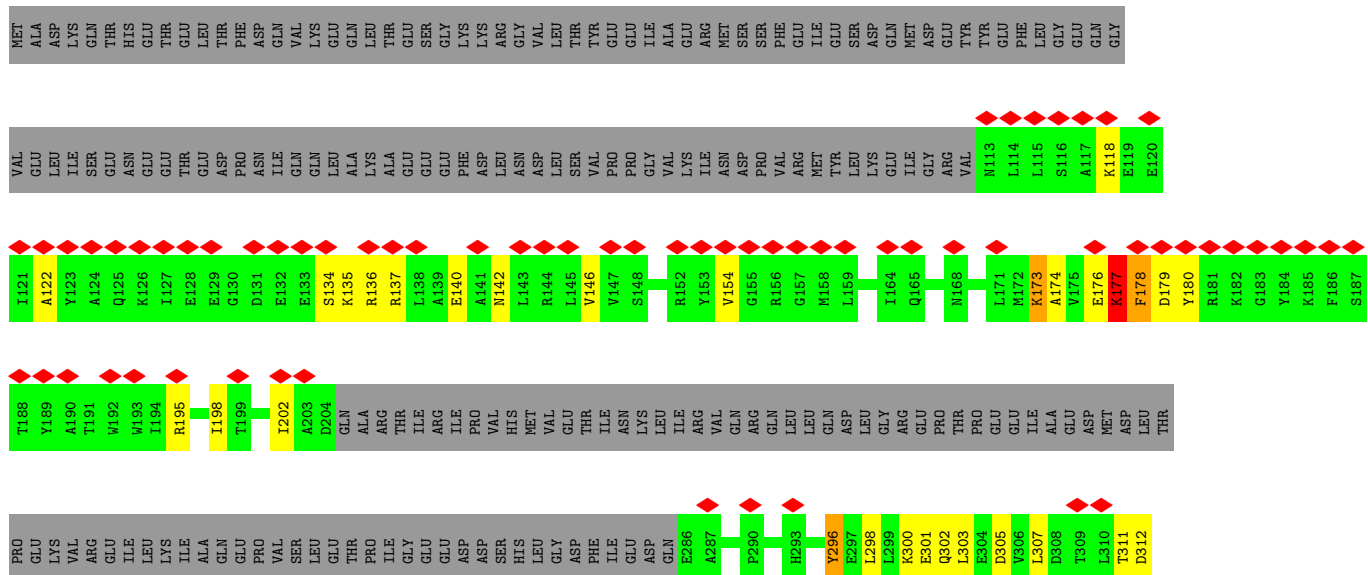
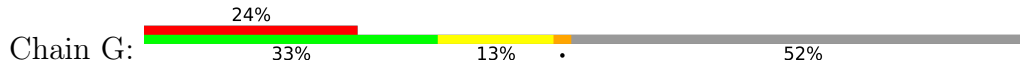




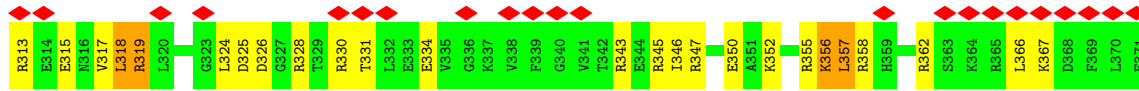
• Molecule 4: DNA-directed RNA polymerase subunit omega



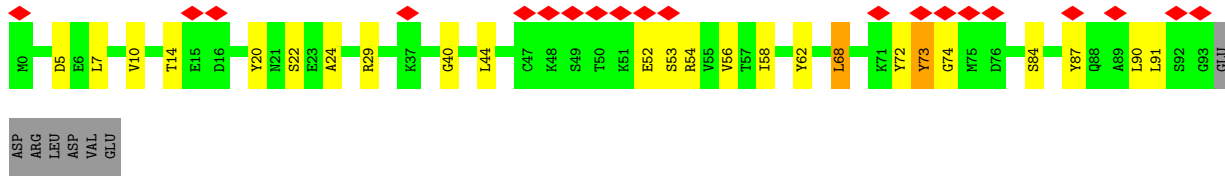
• Molecule 5: RNA polymerase sigma factor SigA



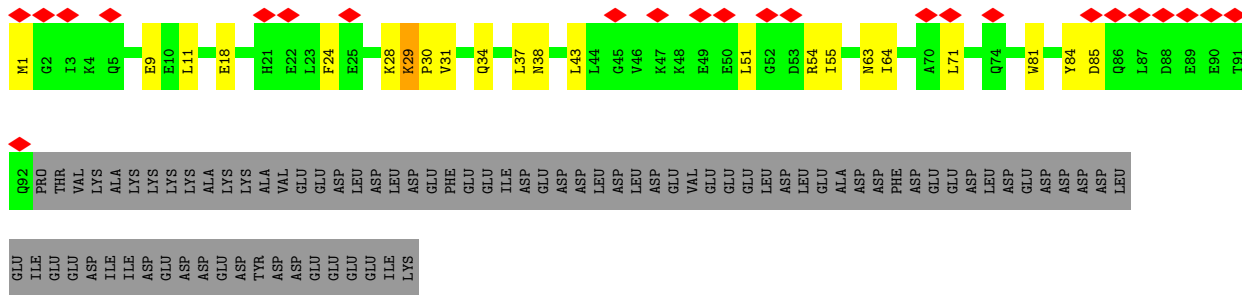
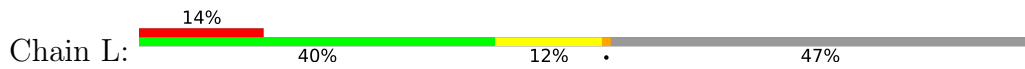




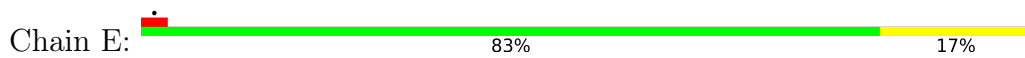
• Molecule 6: Gene 33 protein



• Molecule 7: DNA-directed RNA polymerase subunit delta



• Molecule 8: DNA-directed RNA polymerase subunit epsilon



## 4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	168575	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	NONE	Depositor
Microscope	JEOL CRYO ARM 300	Depositor
Voltage (kV)	300	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	40	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	3500	Depositor
Magnification	Not provided	
Image detector	GATAN K3 (6k x 4k)	Depositor
Maximum map value	1.827	Depositor
Minimum map value	-1.399	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.037	Depositor
Recommended contour level	0.195	Depositor
Map size ( $\text{\AA}$ )	380.0, 380.0, 380.0	wwPDB
Map dimensions	400, 400, 400	wwPDB
Map angles ( $^\circ$ )	90.0, 90.0, 90.0	wwPDB
Pixel spacing ( $\text{\AA}$ )	0.95, 0.95, 0.95	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	A	0.30	0/1746	0.52	0/2375
1	B	0.29	0/1746	0.51	0/2375
2	C	0.29	0/9149	0.52	0/12369
3	D	0.27	0/9405	0.52	0/12684
4	F	0.27	0/490	0.53	0/655
5	G	0.25	0/1476	0.55	0/1974
6	H	0.27	0/789	0.58	0/1056
7	L	0.27	0/779	0.49	0/1051
8	E	0.32	0/574	0.50	0/772
All	All	0.28	0/26154	0.52	0/35311

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts [i](#)

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

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1720	1736	1735	27	0
1	B	1720	1736	1735	30	0
2	C	8991	8940	8938	161	0
3	D	9257	9518	9516	195	0
4	F	485	505	505	13	0
5	G	1458	1488	1486	37	0
6	H	776	778	778	17	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	L	763	740	740	13	0
8	E	564	549	549	8	0
All	All	25734	25990	25982	444	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 9.

The worst 5 of 444 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)
2:C:766:TRP:O	3:D:652:THR:OG1	1.85	0.94
2:C:504:ARG:NH2	2:C:523:PRO:O	2.06	0.88
3:D:1036:LYS:O	7:L:84:TYR:OH	1.96	0.84
4:F:53:GLU:O	4:F:57:GLY:N	2.11	0.82
3:D:961:GLU:OE1	3:D:1124:THR:OG1	1.96	0.82

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	222/314 (71%)	215 (97%)	7 (3%)	0	100	100
1	B	222/314 (71%)	216 (97%)	6 (3%)	0	100	100
2	C	1135/1193 (95%)	1042 (92%)	91 (8%)	2 (0%)	44	66
3	D	1172/1199 (98%)	1093 (93%)	75 (6%)	4 (0%)	37	60
4	F	59/67 (88%)	53 (90%)	6 (10%)	0	100	100
5	G	174/371 (47%)	149 (86%)	21 (12%)	4 (2%)	5	15
6	H	92/101 (91%)	81 (88%)	11 (12%)	0	100	100
7	L	90/173 (52%)	80 (89%)	9 (10%)	1 (1%)	12	30

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
8	E	67/69 (97%)	63 (94%)	4 (6%)	0	100	100
All	All	3233/3801 (85%)	2992 (92%)	230 (7%)	11 (0%)	38	60

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	D	142	THR
5	G	177	LYS
7	L	29	LYS
3	D	32	GLU
3	D	633	ILE

### 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	192/273 (70%)	186 (97%)	6 (3%)	35	59
1	B	192/273 (70%)	188 (98%)	4 (2%)	48	71
2	C	982/1026 (96%)	957 (98%)	25 (2%)	42	66
3	D	1007/1027 (98%)	981 (97%)	26 (3%)	41	65
4	F	55/61 (90%)	55 (100%)	0	100	100
5	G	152/329 (46%)	138 (91%)	14 (9%)	7	19
6	H	83/90 (92%)	77 (93%)	6 (7%)	12	29
7	L	82/159 (52%)	78 (95%)	4 (5%)	21	43
8	E	60/65 (92%)	58 (97%)	2 (3%)	33	57
All	All	2805/3303 (85%)	2718 (97%)	87 (3%)	37	59

5 of 87 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	958	GLU
5	G	355	ARG

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	D	1064	ASP
5	G	296	TYR
6	H	22	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
2	C	791	HIS
2	C	862	HIS
3	D	349	HIS
5	G	196	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](#)

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.

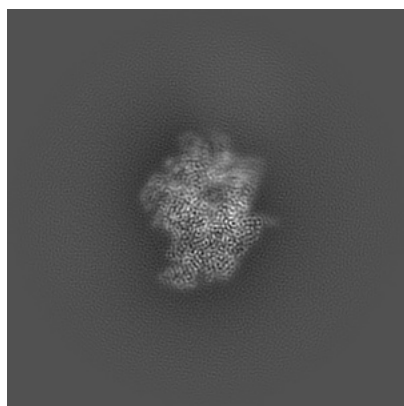
## 6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-38196. These allow visual inspection of the internal detail of the map and identification of artifacts.

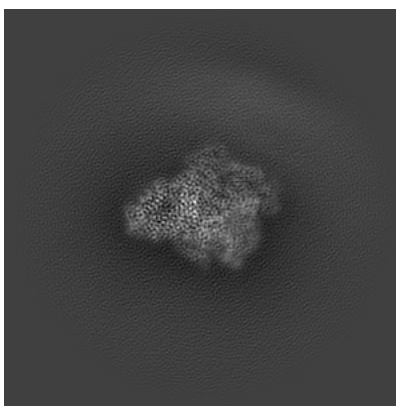
Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

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

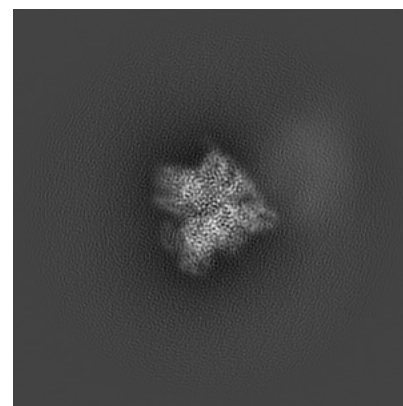
#### 6.1.1 Primary map



X

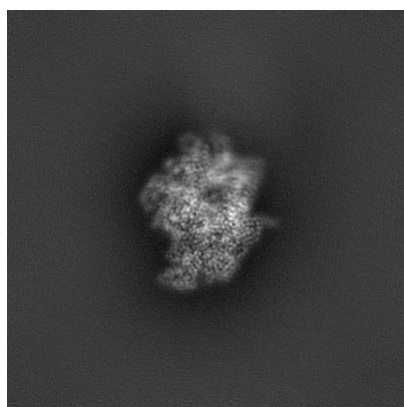


Y

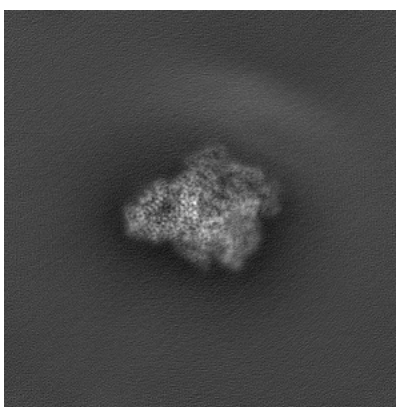


Z

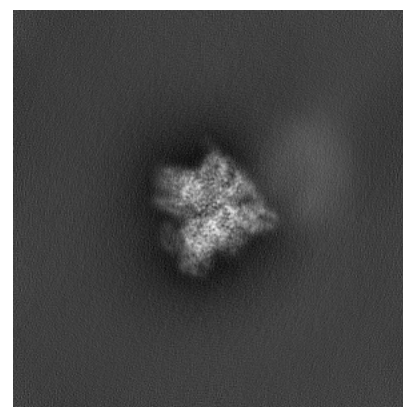
#### 6.1.2 Raw map



X



Y

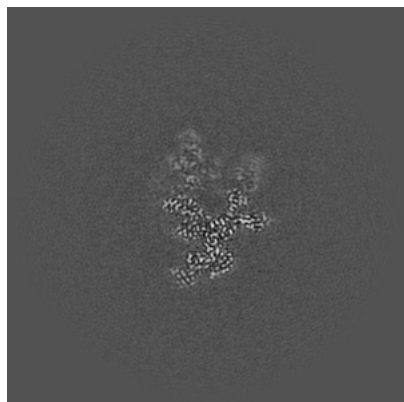


Z

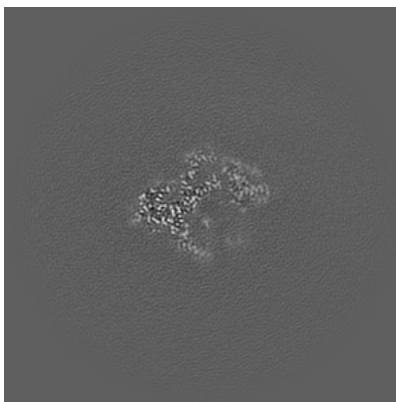
The images above show the map projected in three orthogonal directions.

## 6.2 Central slices [i](#)

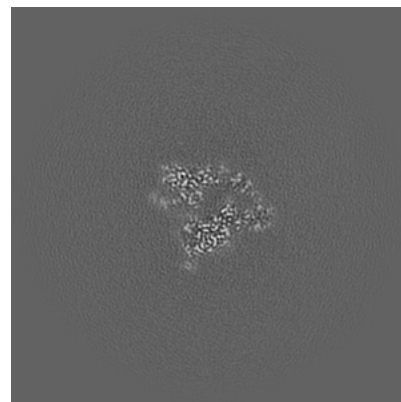
### 6.2.1 Primary map



X Index: 200



Y Index: 200

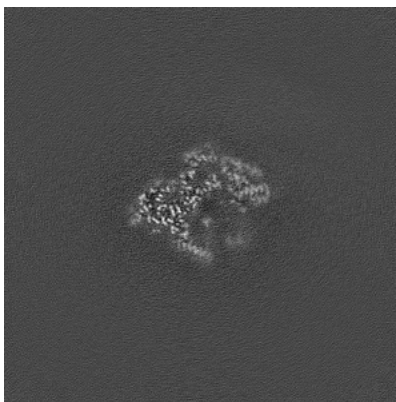


Z Index: 200

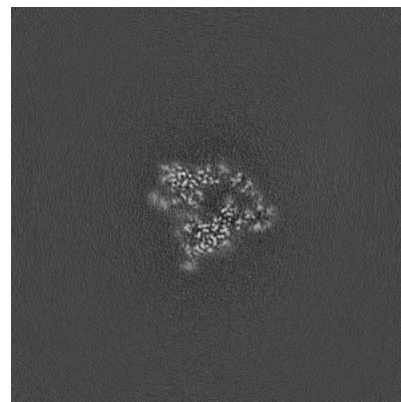
### 6.2.2 Raw map



X Index: 200



Y Index: 200



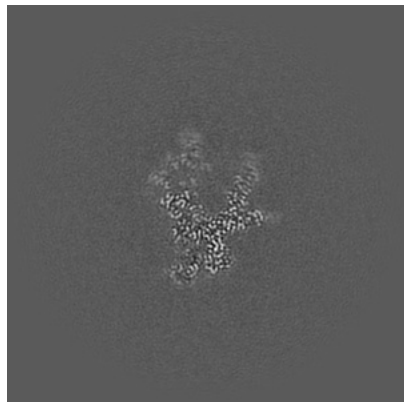
Z Index: 200

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

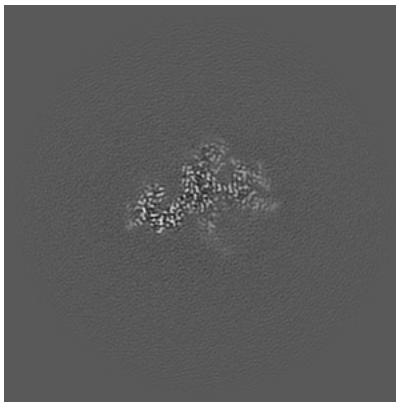


## 6.3 Largest variance slices [i](#)

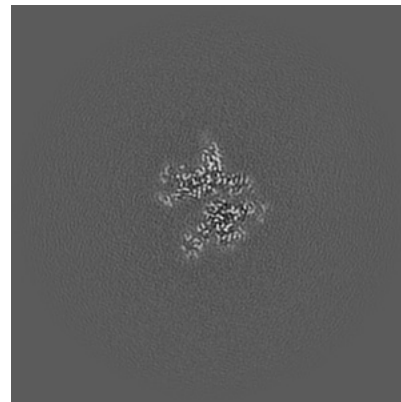
### 6.3.1 Primary map



X Index: 196



Y Index: 188

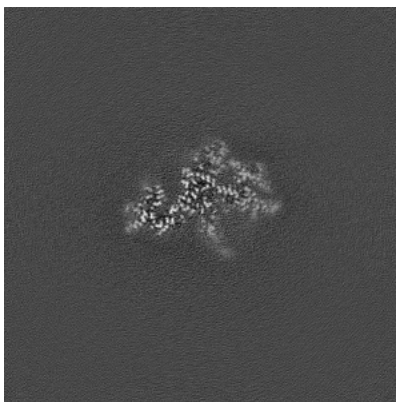


Z Index: 190

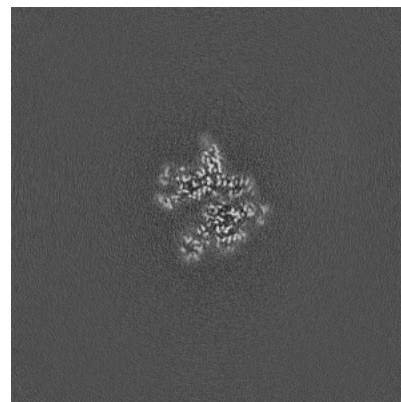
### 6.3.2 Raw map



X Index: 196



Y Index: 187

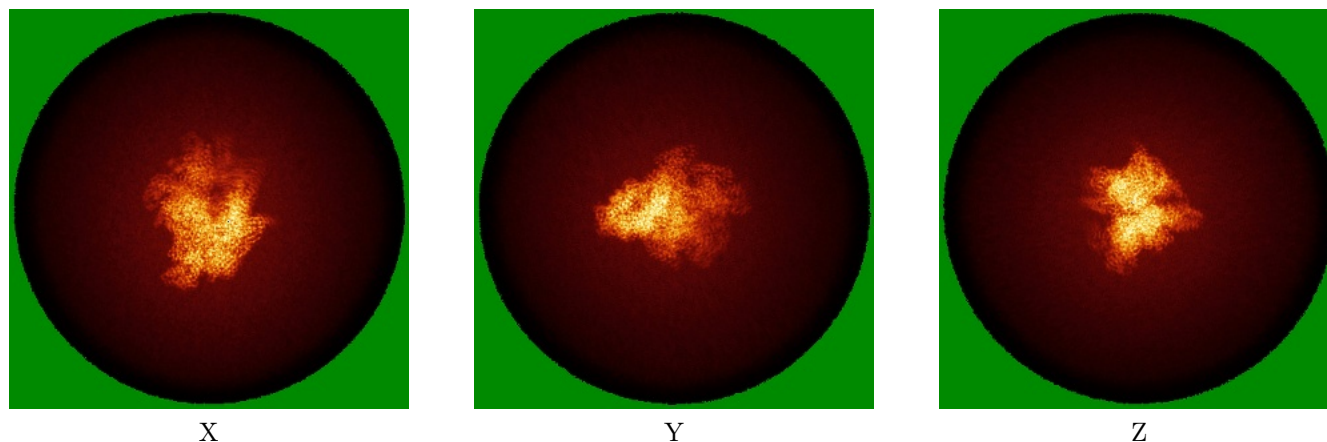


Z Index: 190

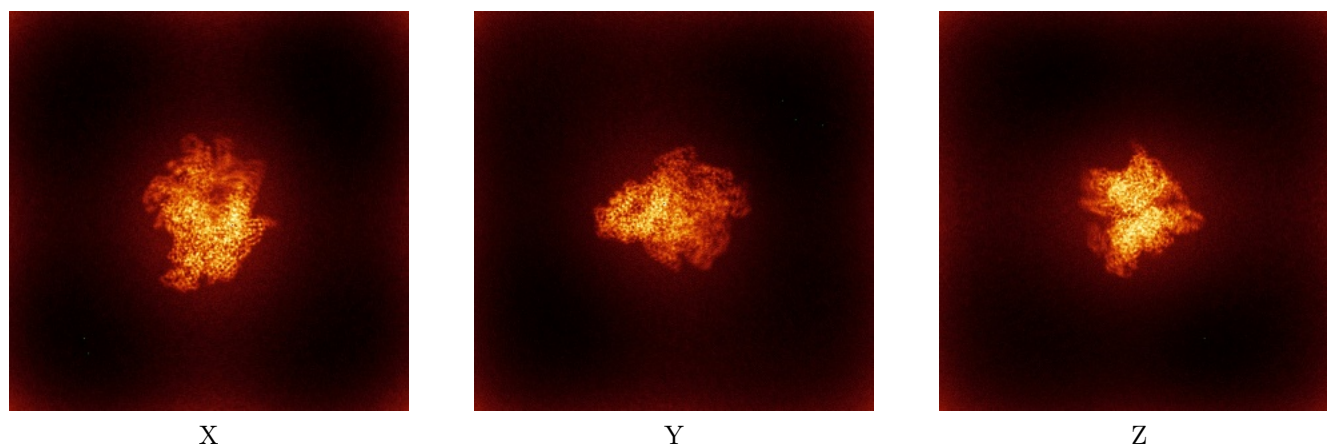
The images above show the largest variance slices of the map in three orthogonal directions.

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

### 6.4.1 Primary map



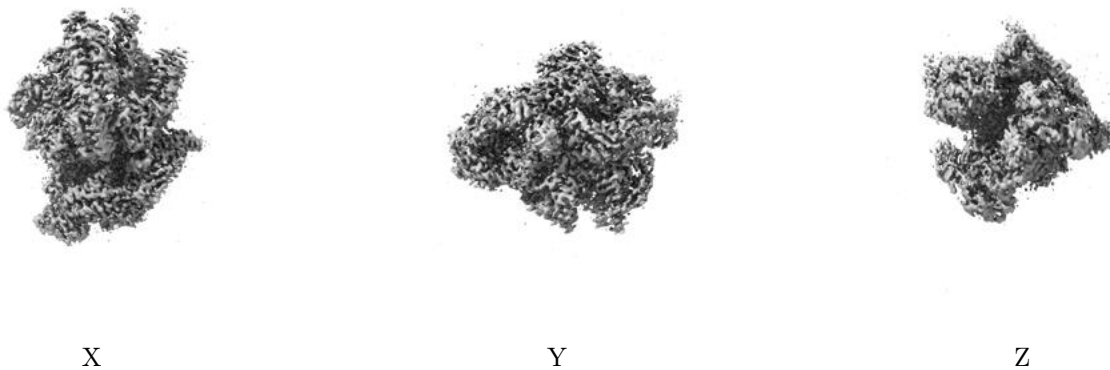
### 6.4.2 Raw map



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.

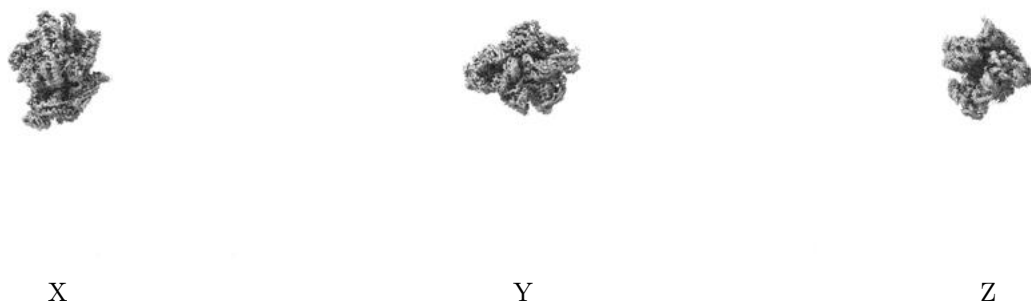
## 6.5 Orthogonal surface views [i](#)

### 6.5.1 Primary map



The images above show the 3D surface view of the map at the recommended contour level 0.195. 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



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.

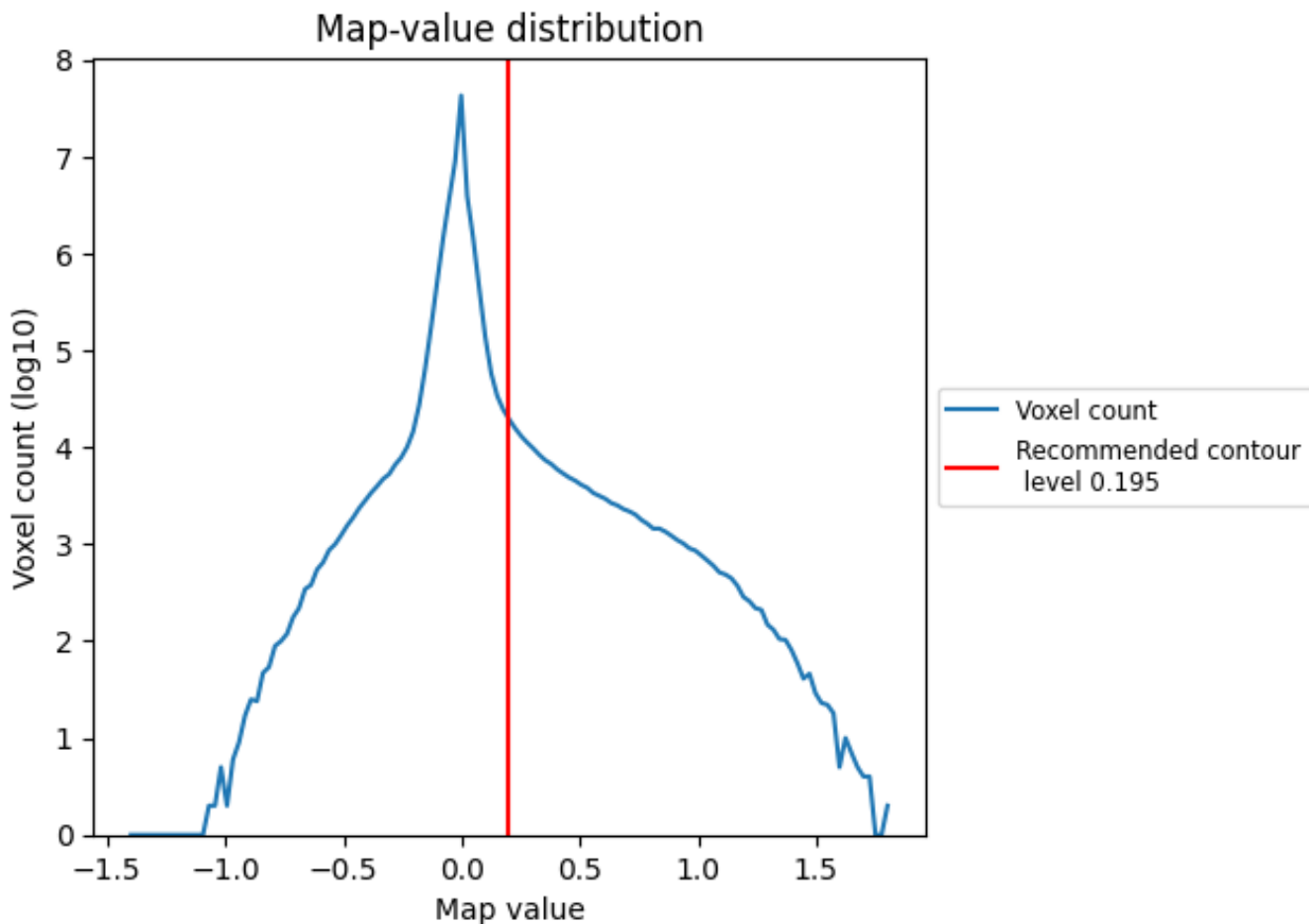
## 6.6 Mask visualisation [i](#)

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

## 7 Map analysis [i](#)

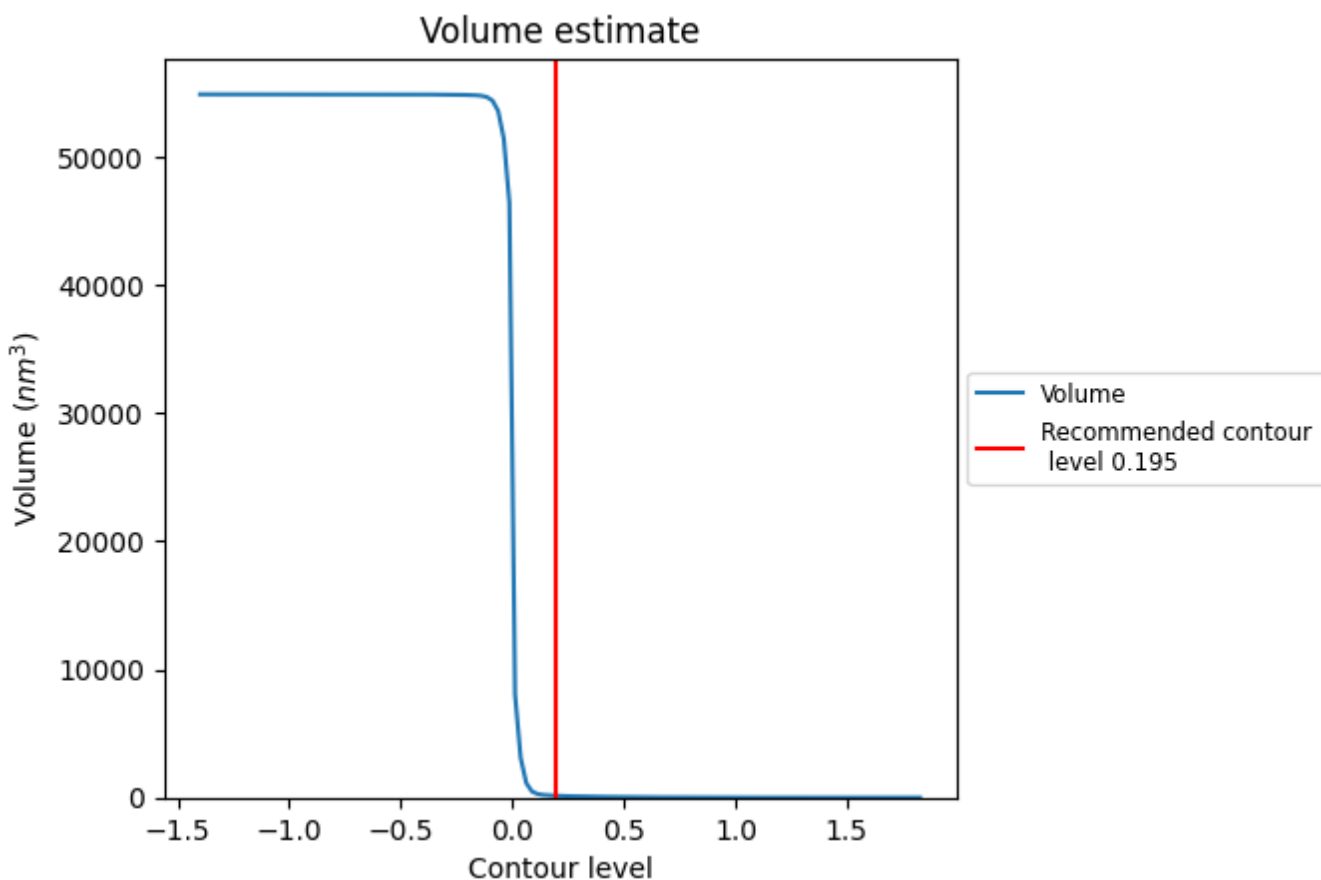
This section contains the results of statistical analysis of the map.

### 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.

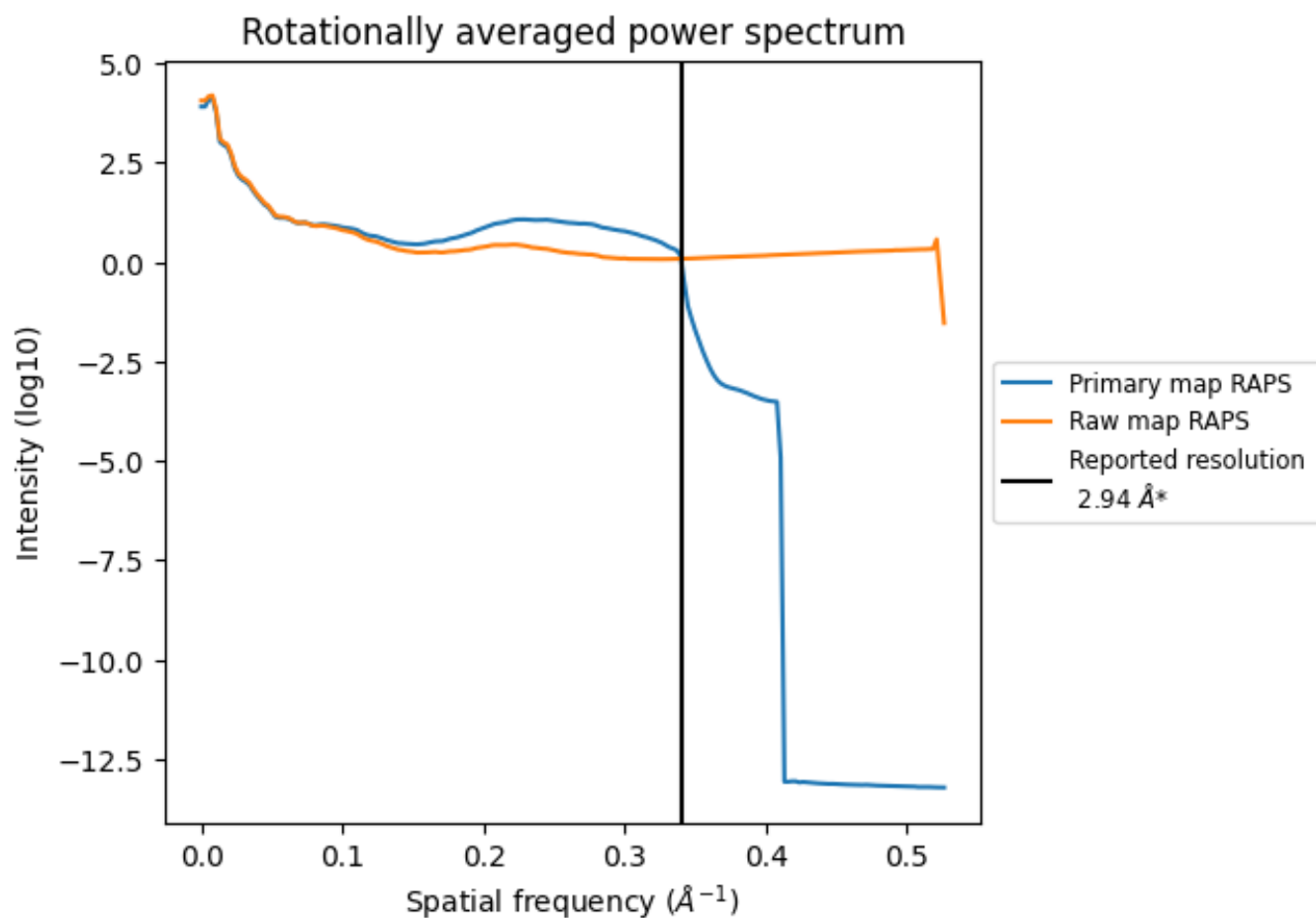
## 7.2 Volume estimate [i](#)



The volume at the recommended contour level is 142 nm<sup>3</sup>; this corresponds to an approximate mass of 128 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](#)

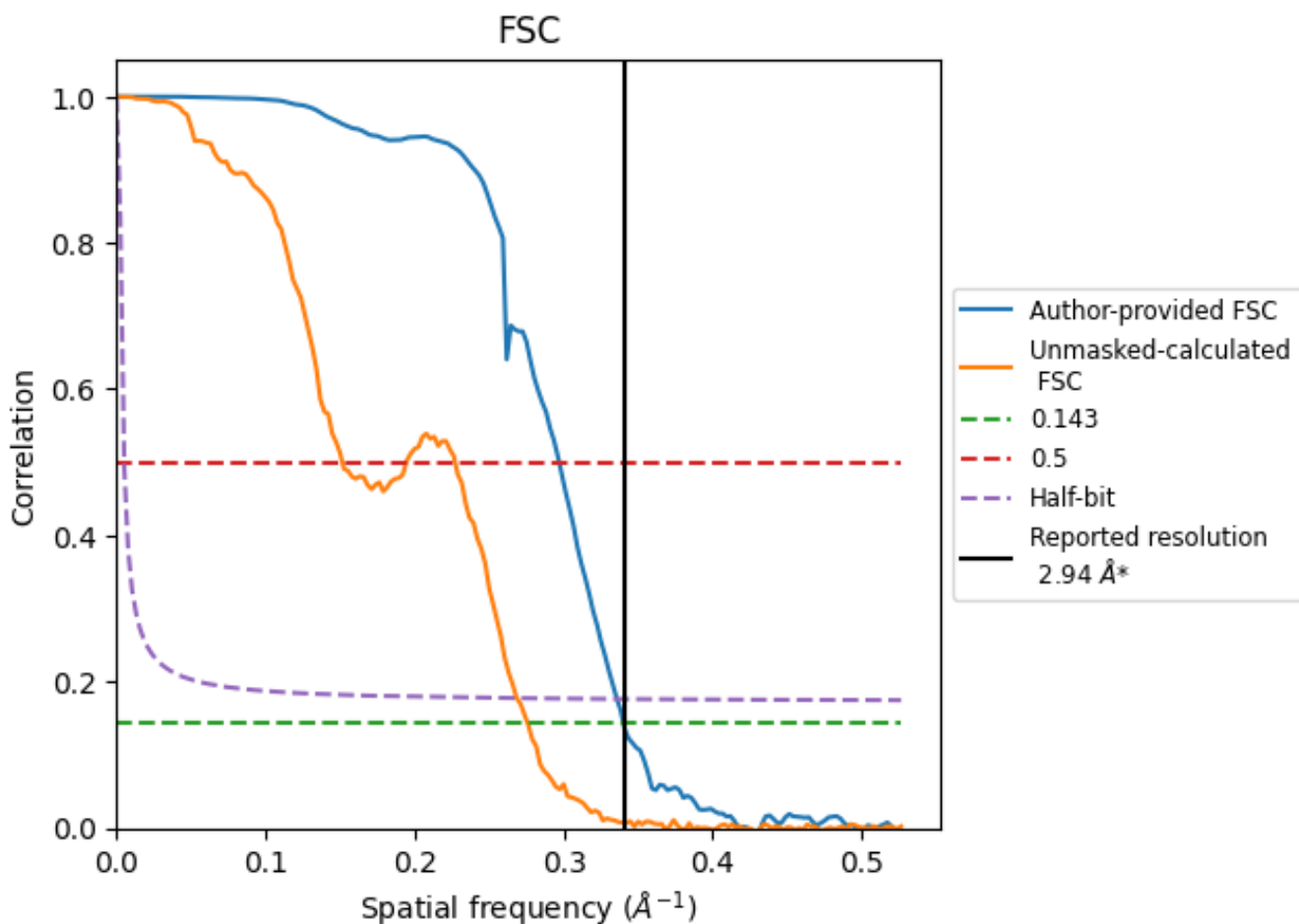


\*Reported resolution corresponds to spatial frequency of 0.340 Å<sup>-1</sup>

## 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 0.340 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	2.94	-	-
Author-provided FSC curve	2.94	3.37	2.98
Unmasked-calculated*	3.63	6.60	3.72

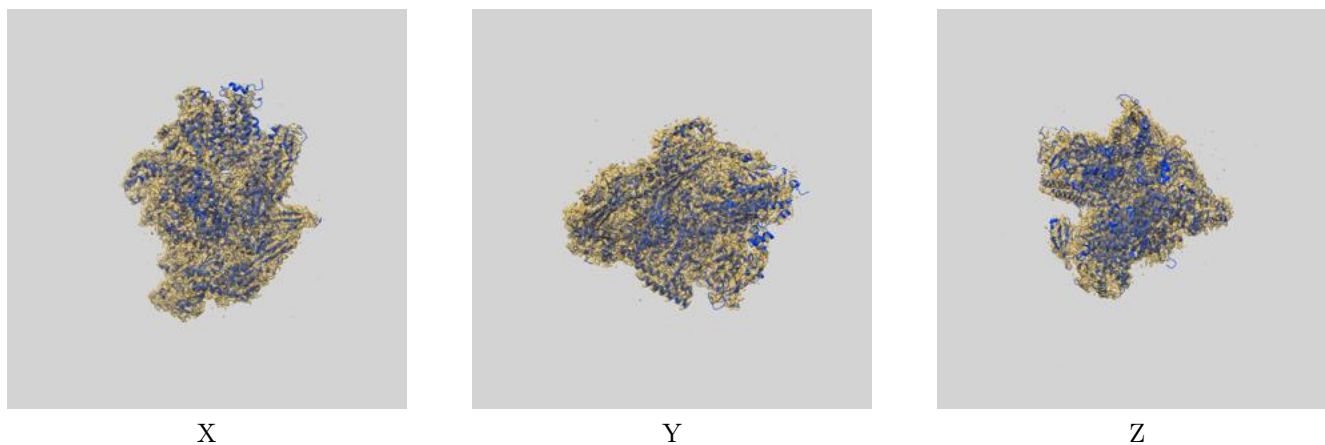
\*Resolution estimate based on FSC curve calculated by comparison of deposited half-maps. The value from deposited half-maps intersecting FSC 0.143 CUT-OFF 3.63 differs from the reported value 2.94 by more than 10 %



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

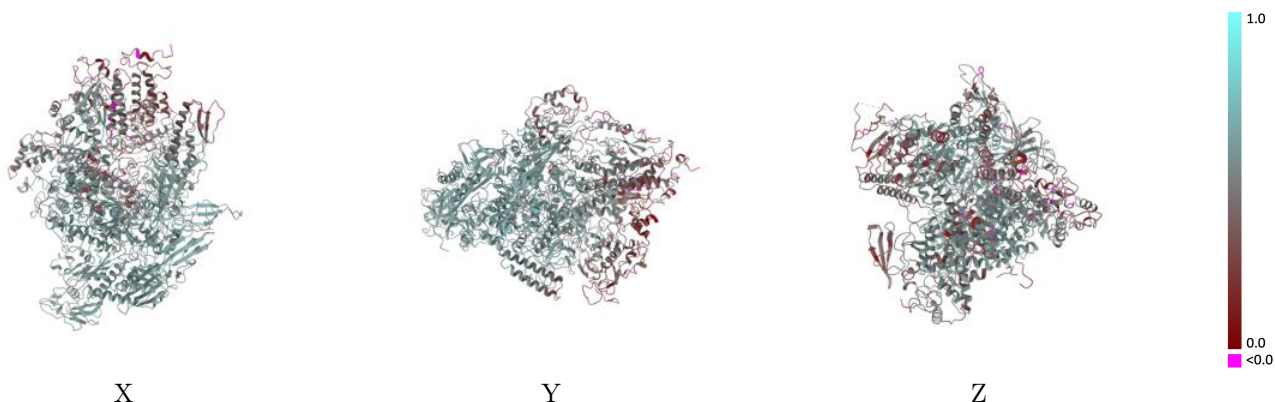
This section contains information regarding the fit between EMDB map EMD-38196 and PDB model 8XA7. Per-residue inclusion information can be found in section 3 on page 6.

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



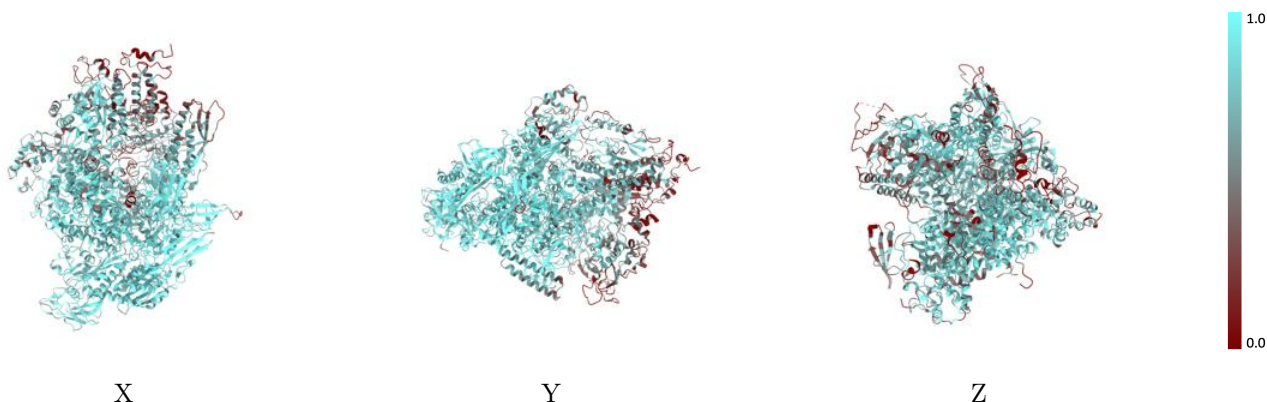
The images above show the 3D surface view of the map at the recommended contour level 0.195 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

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



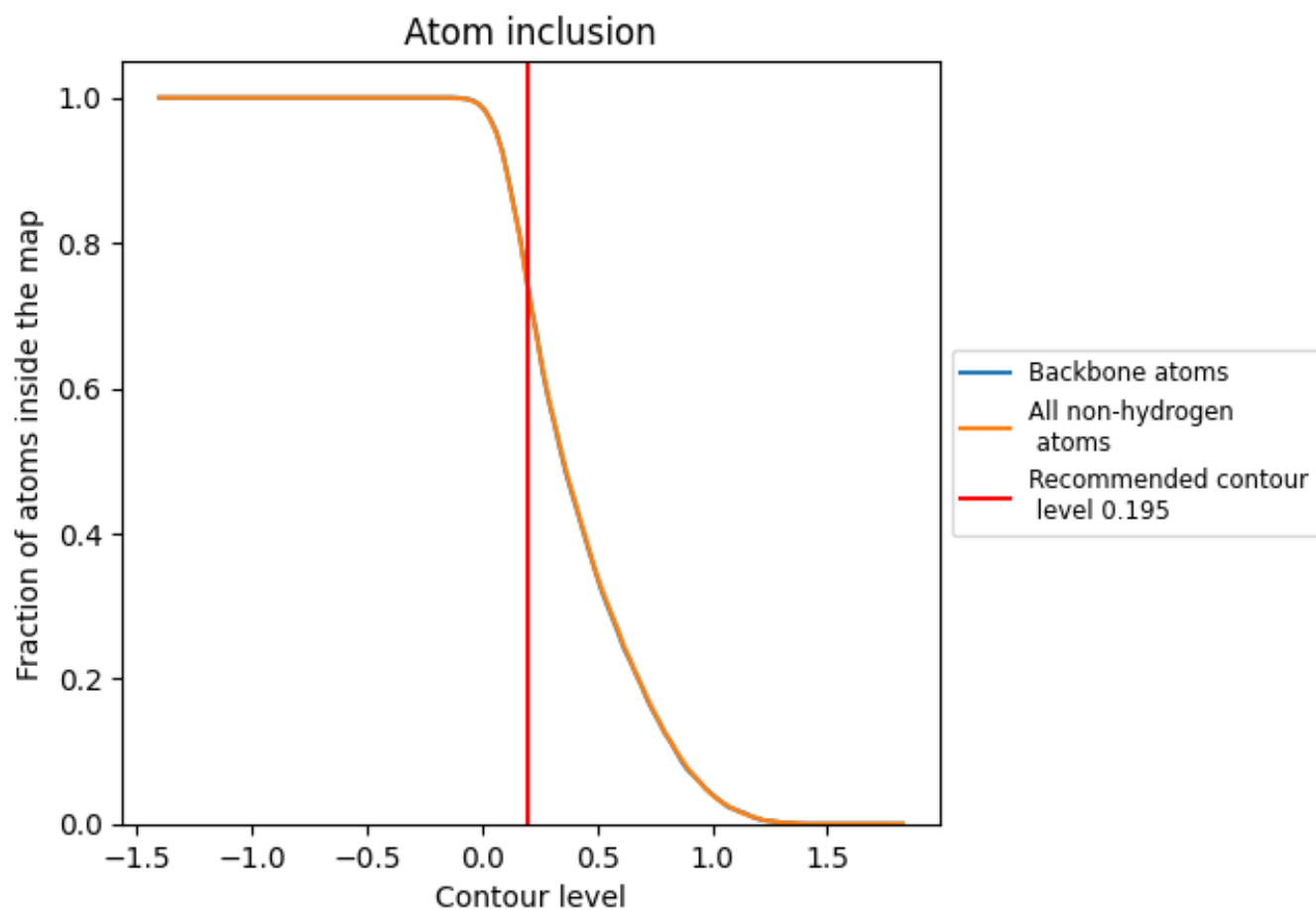
The images above show the model with each residue coloured according to its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions 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 (0.195).





















## 9.4 Atom inclusion [i](#)



At the recommended contour level, 75% of all backbone atoms, 75% 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 (0.195) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	 0.7470	 0.5110
A	 0.8960	 0.5830
B	 0.8460	 0.5600
C	 0.7470	 0.5110
D	 0.7690	 0.5200
E	 0.8680	 0.5940
F	 0.8470	 0.5560
G	 0.4100	 0.3400
H	 0.6510	 0.4430
L	 0.6130	 0.4350

