



## wwPDB EM Validation Summary Report ⓘ

May 15, 2023 – 10:21 am BST

PDB ID : 7QXI  
EMDB ID : EMD-14200  
Title : Cryo-EM structure of RNA polymerase-sigma54 holo enzyme with promoter DNA closed complex  
Authors : Ye, F.Z.; Zhang, X.D.  
Deposited on : 2022-01-26  
Resolution : 3.40 Å (reported)  
Based on initial model : 5NSR

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.dev50  
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) : 2.32.2

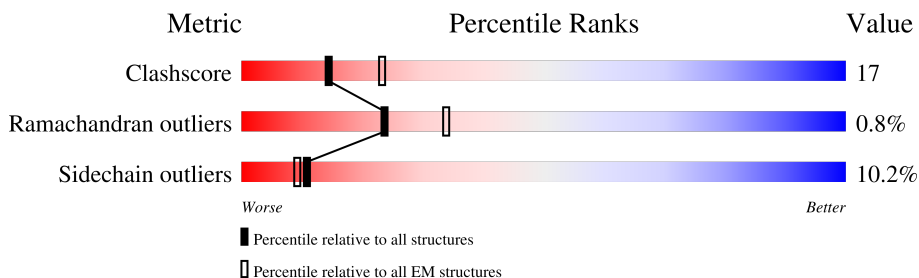
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 3.40 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  $\geq 3$ , 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	329	
1	B	329	
2	C	1342	
3	D	1407	
4	E	91	
5	M	497	
6	N	63	
7	T	63	

## 2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 28928 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 DNA-directed RNA polymerase subunit alpha.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	309	Total	C	N	O	S	0	0
			2316	1450	404	455	7		
1	B	222	Total	C	N	O	S	0	0
			1671	1042	293	331	5		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	C	1341	Total	C	N	O	S	0	0
			10125	6360	1761	1964	40		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	D	1334	Total	C	N	O	S	0	0
			9632	6049	1730	1814	39		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	E	74	Total	C	N	O	S	0	0
			546	335	109	101	1		

- Molecule 5 is a protein called RNA polymerase sigma-54 factor.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
5	M	403	Total	C	N	O	S	0	0
			3162	1974	551	626	11		

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
M	-19	MET	-	initiating methionine	UNP A0A0N9UTC1

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
M	-18	GLY	-	expression tag	UNP A0A0N9UTC1
M	-17	SER	-	expression tag	UNP A0A0N9UTC1
M	-16	SER	-	expression tag	UNP A0A0N9UTC1
M	-15	HIS	-	expression tag	UNP A0A0N9UTC1
M	-14	HIS	-	expression tag	UNP A0A0N9UTC1
M	-13	HIS	-	expression tag	UNP A0A0N9UTC1
M	-12	HIS	-	expression tag	UNP A0A0N9UTC1
M	-11	HIS	-	expression tag	UNP A0A0N9UTC1
M	-10	HIS	-	expression tag	UNP A0A0N9UTC1
M	-9	SER	-	expression tag	UNP A0A0N9UTC1
M	-8	SER	-	expression tag	UNP A0A0N9UTC1
M	-7	GLY	-	expression tag	UNP A0A0N9UTC1
M	-6	LEU	-	expression tag	UNP A0A0N9UTC1
M	-5	VAL	-	expression tag	UNP A0A0N9UTC1
M	-4	PRO	-	expression tag	UNP A0A0N9UTC1
M	-3	ARG	-	expression tag	UNP A0A0N9UTC1
M	-2	GLY	-	expression tag	UNP A0A0N9UTC1
M	-1	SER	-	expression tag	UNP A0A0N9UTC1
M	0	HIS	-	expression tag	UNP A0A0N9UTC1
M	49	GLU	GLN	conflict	UNP A0A0N9UTC1
M	80	GLU	ASP	conflict	UNP A0A0N9UTC1

- Molecule 6 is a DNA chain called Non-Template promoter DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
6	N	36	738	349	137	216	36	0	0

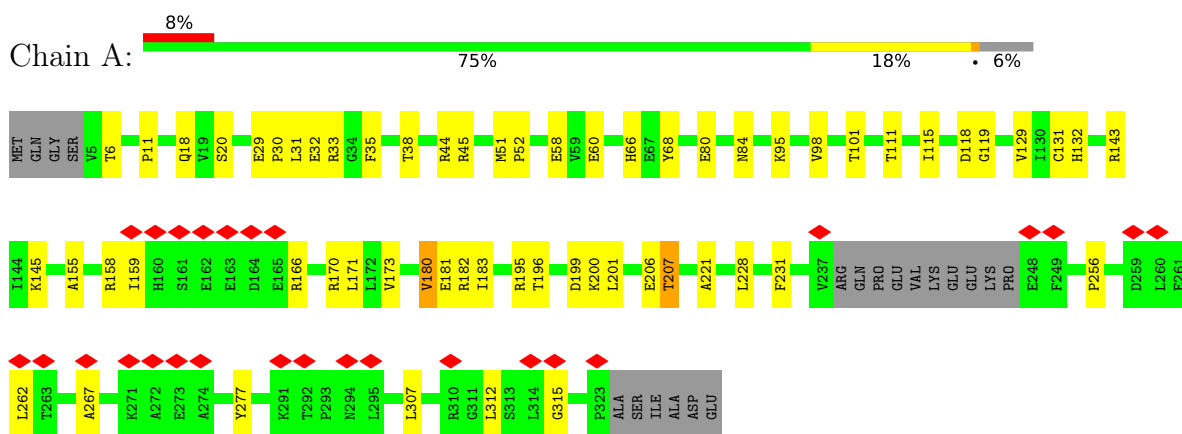
- Molecule 7 is a DNA chain called Template DNA promoter.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
7	T	36	738	349	137	216	36	0	0

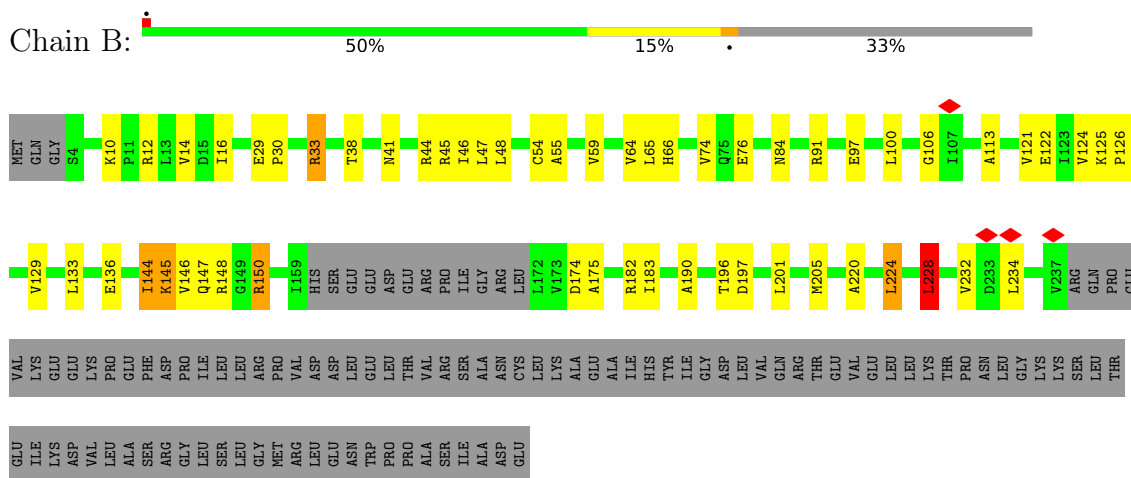
### 3 Residue-property plots [i](#)

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.

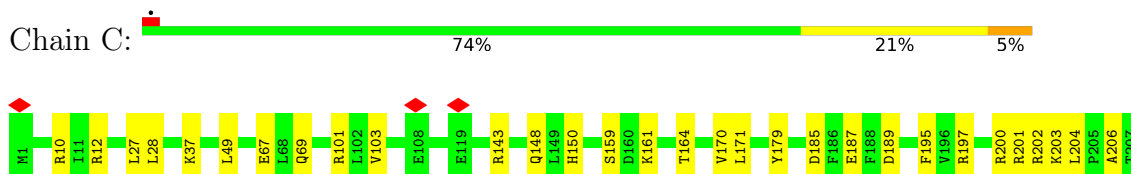
- Molecule 1: DNA-directed RNA polymerase subunit alpha

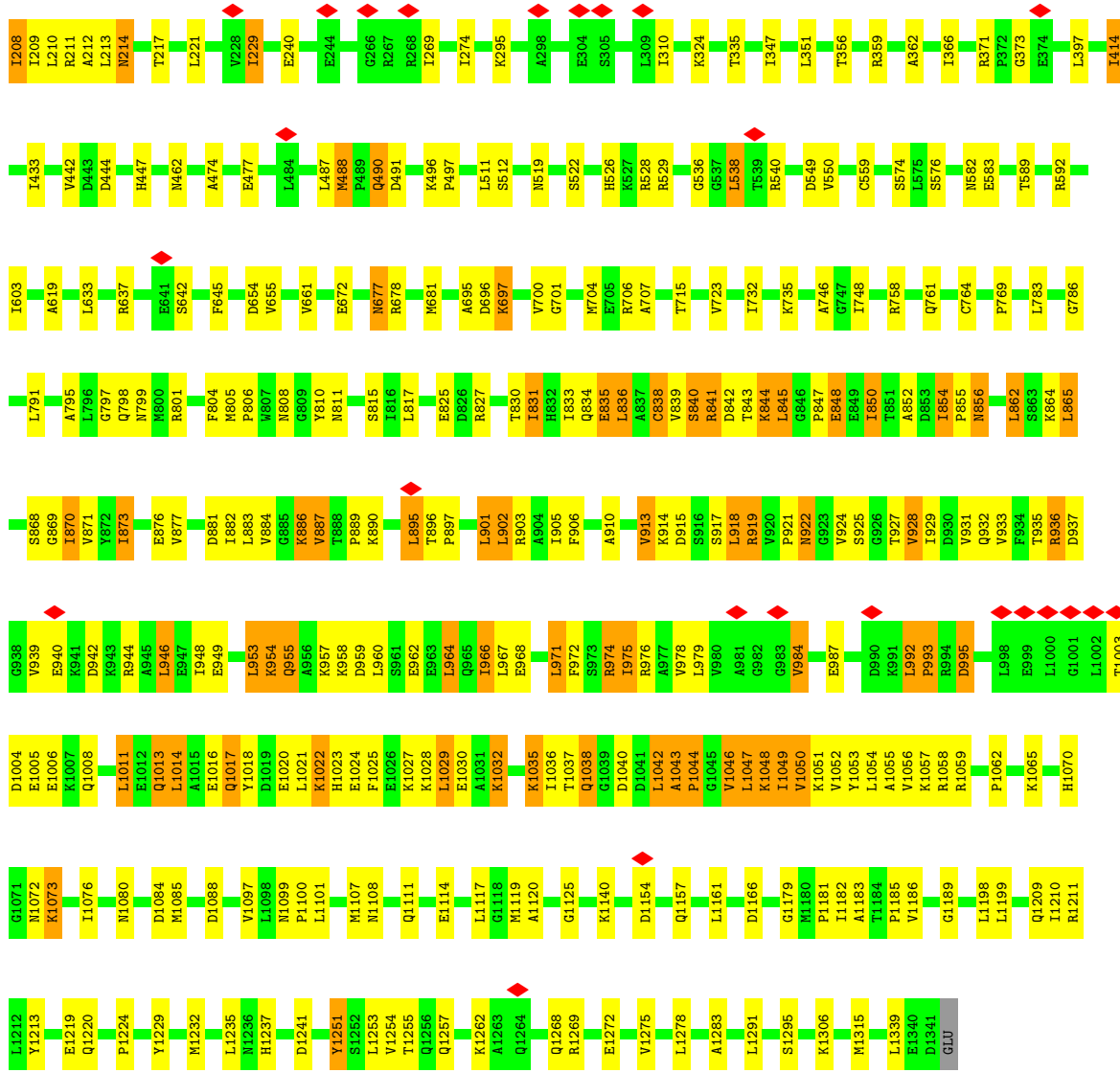


- Molecule 1: DNA-directed RNA polymerase subunit alpha

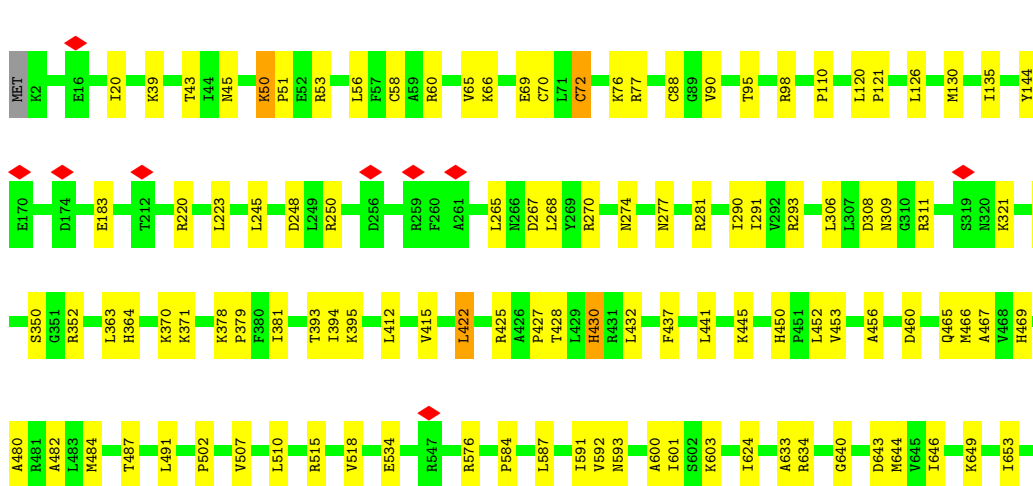
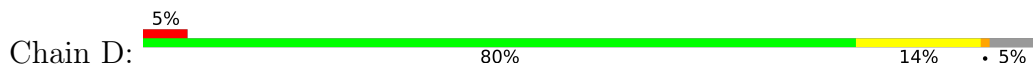


- Molecule 2: DNA-directed RNA polymerase subunit beta

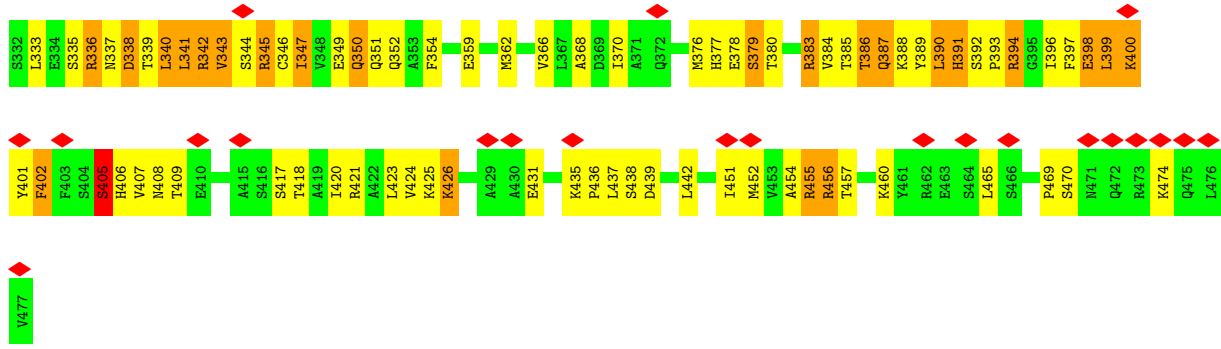




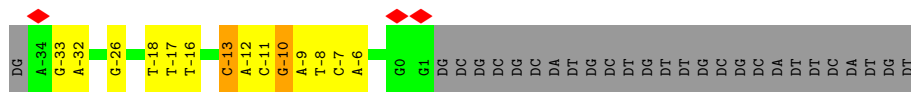
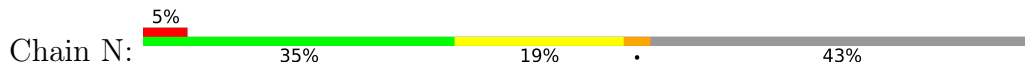
• Molecule 3: DNA-directed RNA polymerase subunit beta'



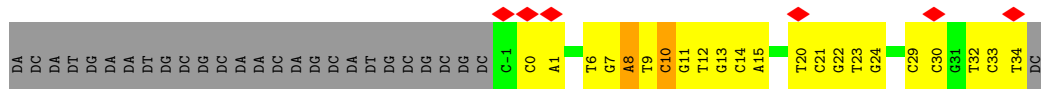
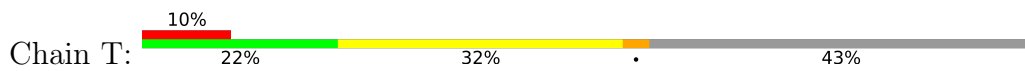




• Molecule 6: Non-Template promoter DNA



• Molecule 7: Template DNA promoter





## 4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	29321	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 ( $e^-/\text{\AA}^2$ )	50	Depositor
Minimum defocus (nm)	1400	Depositor
Maximum defocus (nm)	2600	Depositor
Magnification	81000	Depositor
Image detector	GATAN K3 (6k x 4k)	Depositor
Maximum map value	0.216	Depositor
Minimum map value	-0.119	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.006	Depositor
Recommended contour level	0.02	Depositor
Map size ( $\text{\AA}$ )	281.6, 281.6, 281.6	wwPDB
Map dimensions	256, 256, 256	wwPDB
Map angles ( $^\circ$ )	90.0, 90.0, 90.0	wwPDB
Pixel spacing ( $\text{\AA}$ )	1.1, 1.1, 1.1	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	0.23	0/2345	0.42	0/3195
1	B	0.23	0/1688	0.45	1/2293 (0.0%)
2	C	0.23	0/10283	0.40	0/13940
3	D	0.23	0/9766	0.41	1/13267 (0.0%)
4	E	0.22	0/547	0.37	0/740
5	M	0.72	3/3204 (0.1%)	0.88	7/4339 (0.2%)
6	N	0.73	2/827 (0.2%)	0.92	5/1274 (0.4%)
7	T	0.52	1/827 (0.1%)	0.95	3/1274 (0.2%)
All	All	0.35	6/29487 (0.0%)	0.53	17/40322 (0.0%)

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	N	-10	DG	C4'-O4'	15.90	1.60	1.45
6	N	-13	DC	C4'-O4'	-9.66	1.35	1.45
7	T	10	DC	C4'-O4'	9.52	1.54	1.45
5	M	402	PHE	CG-CD1	5.91	1.47	1.38
5	M	285	TRP	CG-CD1	-5.87	1.28	1.36

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	T	10	DC	O4'-C4'-C3'	-16.10	96.34	106.00
6	N	-10	DG	O3'-P-O5'	-11.20	82.72	104.00
6	N	-10	DG	C1'-O4'-C4'	-8.37	101.73	110.10
6	N	-10	DG	C5'-C4'-O4'	-6.74	96.49	109.30
5	M	287	VAL	CG1-CB-CG2	-5.72	101.75	110.90

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	2316	0	2297	55	0
1	B	1671	0	1674	32	0
2	C	10125	0	9829	279	0
3	D	9632	0	9238	112	0
4	E	546	0	548	5	0
5	M	3162	0	3169	510	0
6	N	738	0	404	39	0
7	T	738	0	404	61	0
All	All	28928	0	27563	965	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 17.

The worst 5 of 965 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)
5:M:226:LEU:CD1	5:M:235:LEU:HB2	1.31	1.57
5:M:144:ILE:CD1	5:M:161:ILE:HG21	1.42	1.45
5:M:226:LEU:HD12	5:M:235:LEU:CB	1.50	1.39
5:M:17:PRO:HB3	6:N:-12:DA:C2	1.57	1.36
5:M:23:ILE:HG21	7:T:12:DT:C6	1.66	1.29

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	305/329 (93%)	283 (93%)	22 (7%)	0	100	100
1	B	218/329 (66%)	207 (95%)	11 (5%)	0	100	100
2	C	1339/1342 (100%)	1239 (92%)	89 (7%)	11 (1%)	19	51
3	D	1322/1407 (94%)	1228 (93%)	94 (7%)	0	100	100
4	E	72/91 (79%)	71 (99%)	1 (1%)	0	100	100
5	M	399/497 (80%)	330 (83%)	52 (13%)	17 (4%)	2	17
All	All	3655/3995 (92%)	3358 (92%)	269 (7%)	28 (1%)	24	51

5 of 28 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	1043	ALA
2	C	1044	PRO
5	M	18	GLN
5	M	120	LEU
5	M	195	LEU

### 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	248/286 (87%)	242 (98%)	6 (2%)	49	74
1	B	180/286 (63%)	166 (92%)	14 (8%)	12	39
2	C	1047/1157 (90%)	920 (88%)	127 (12%)	5	18
3	D	913/1168 (78%)	882 (97%)	31 (3%)	37	65
4	E	53/75 (71%)	52 (98%)	1 (2%)	57	78
5	M	350/440 (80%)	243 (69%)	107 (31%)	0	1
All	All	2791/3412 (82%)	2505 (90%)	286 (10%)	11	26

5 of 286 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	M	259	ARG

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
5	M	273	ILE
5	M	342	ARG
2	C	957	LYS
2	C	953	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
5	M	322	ASN
5	M	205	GLN
3	D	450	HIS
5	M	127	GLN
3	D	45	ASN

### 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.

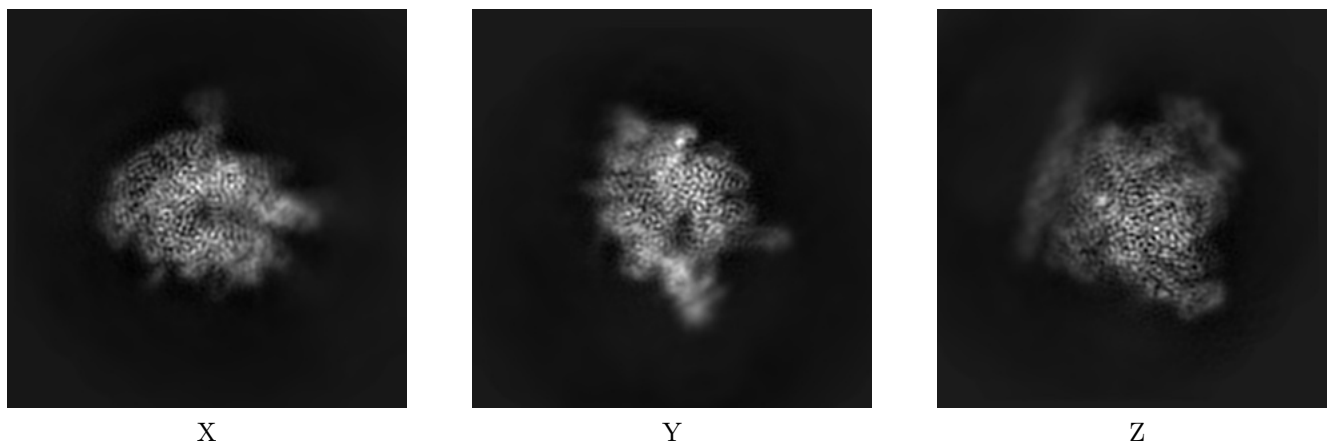
## 6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-14200. 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](#)

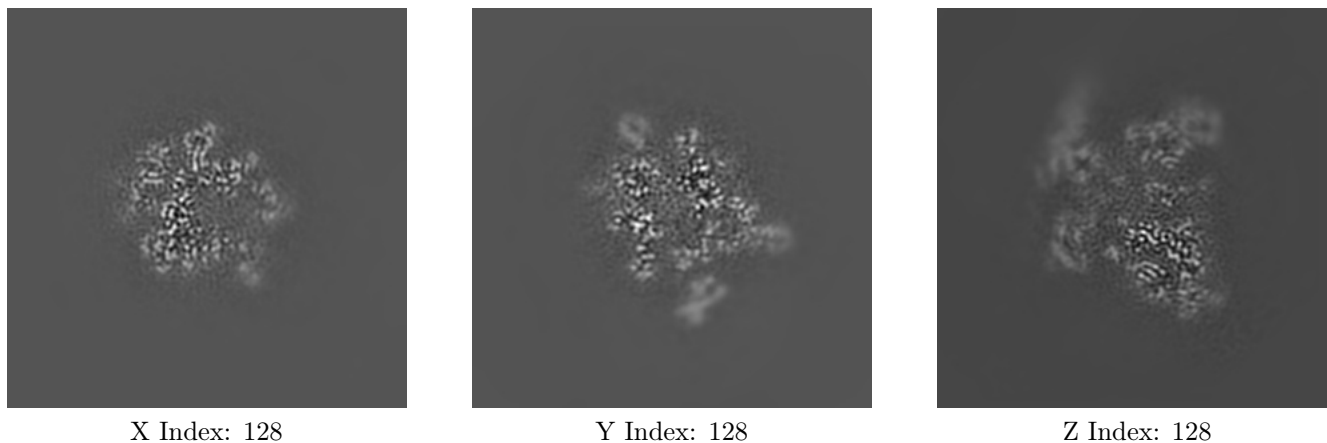
#### 6.1.1 Primary map



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

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

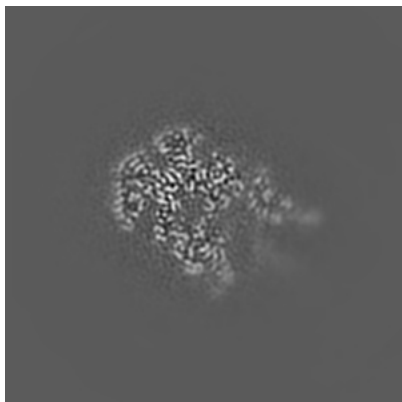
#### 6.2.1 Primary map



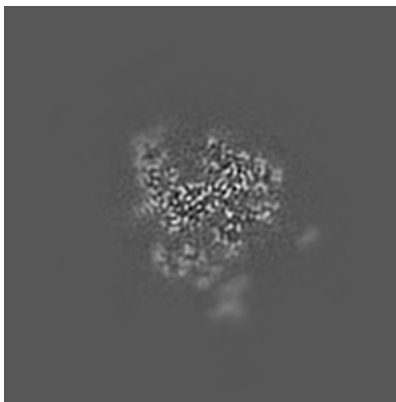
The images above show central slices of the map in three orthogonal directions.

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

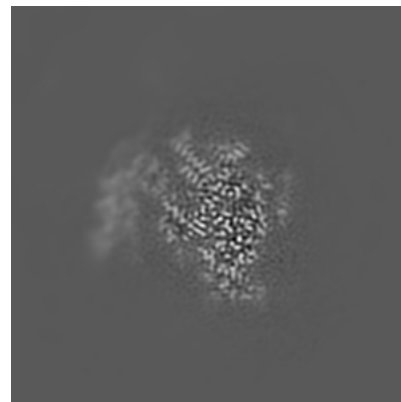
### 6.3.1 Primary map



X Index: 145



Y Index: 115

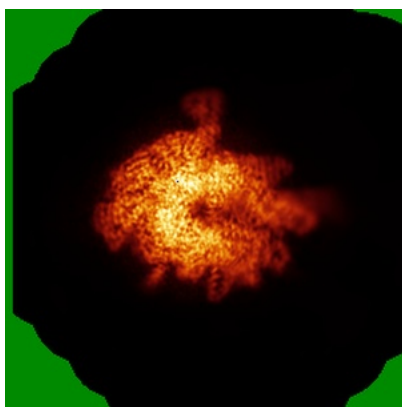


Z Index: 144

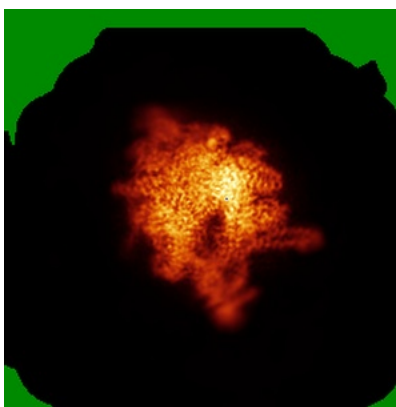
The images above show the largest variance slices of the map in three orthogonal directions.

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

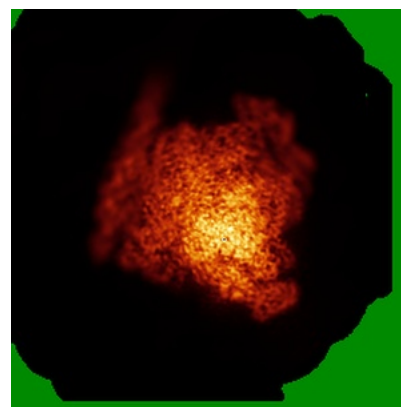
### 6.4.1 Primary map



X



Y

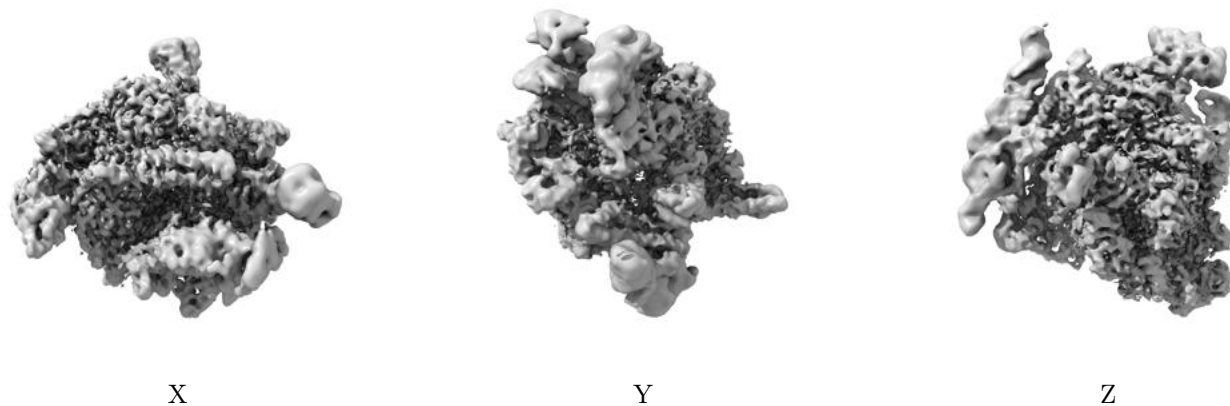


Z

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 0.02. 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](#)

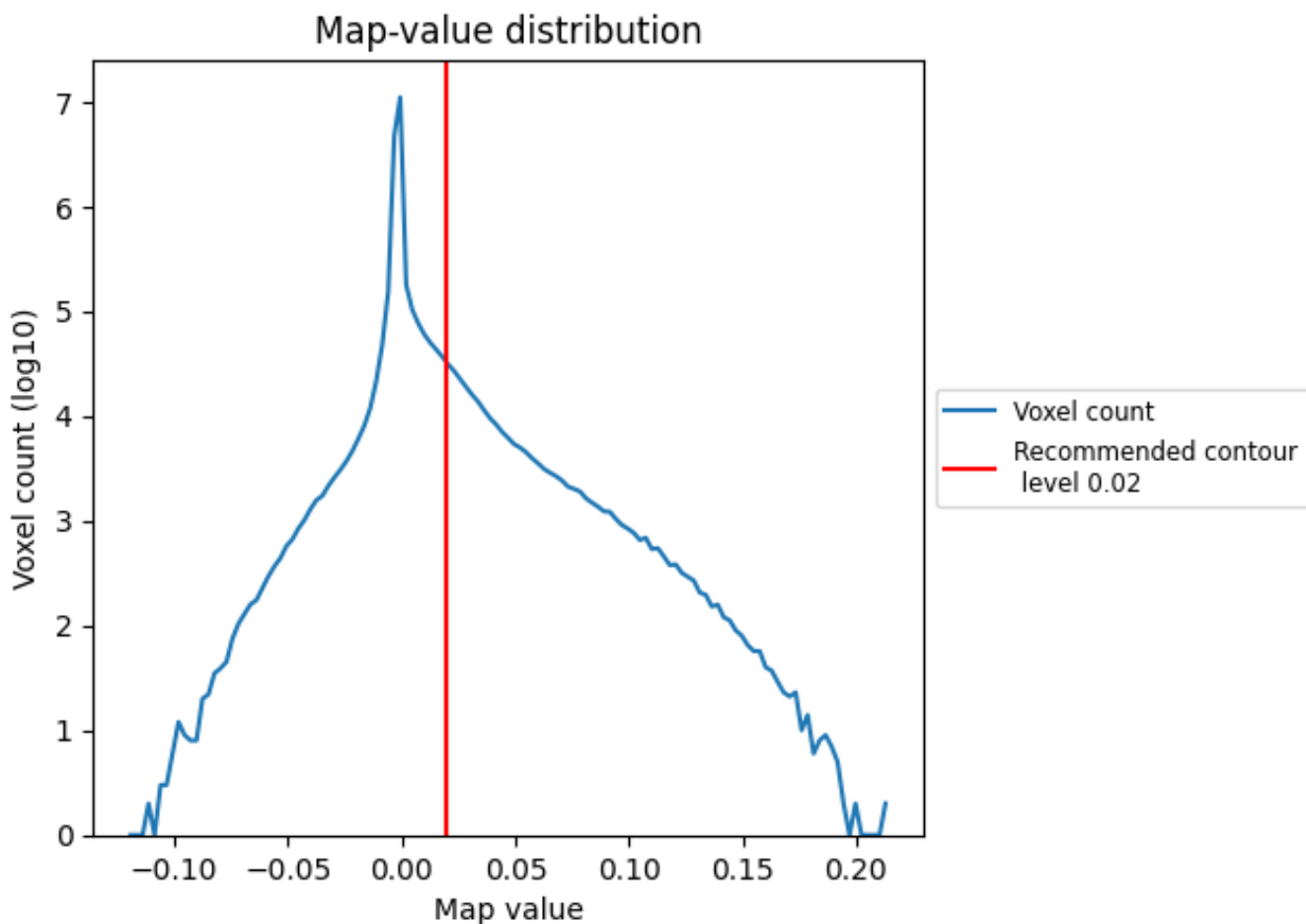
This section was not generated. No masks/segmentation were deposited.



## 7 Map analysis [i](#)

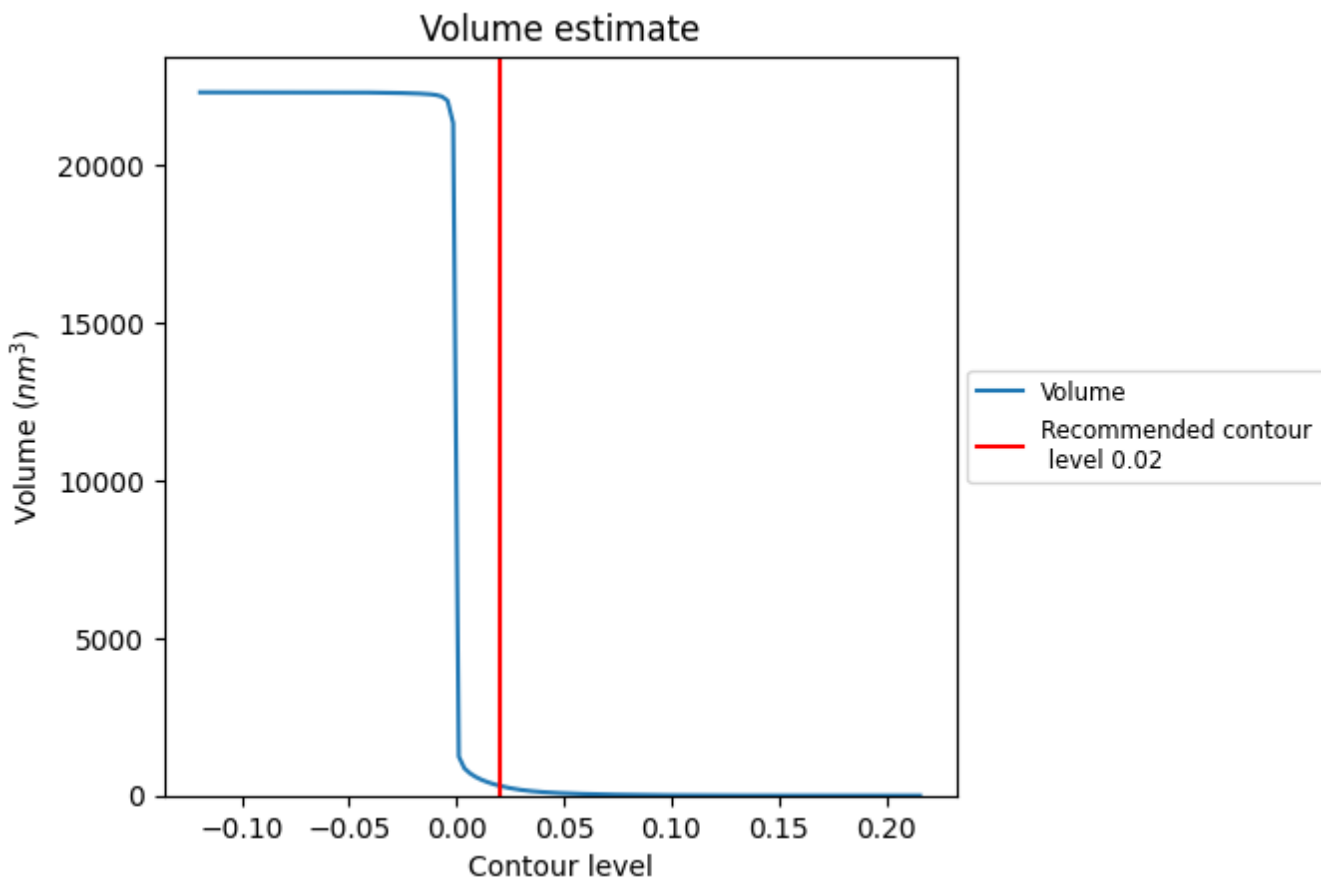
This section contains the results of statistical analysis of the map.

### 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.

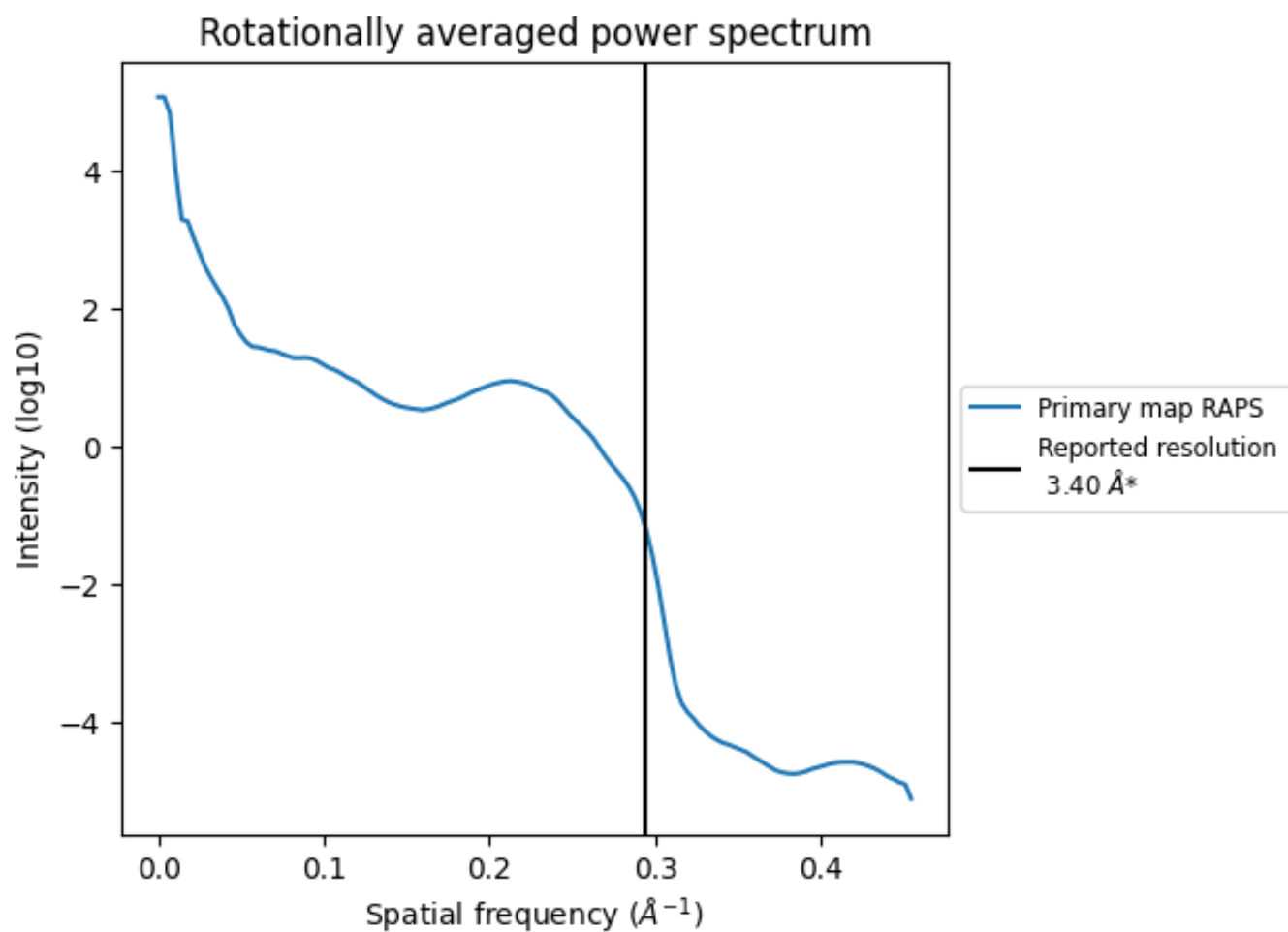
## 7.2 Volume estimate [i](#)



The volume at the recommended contour level is 312  $\text{nm}^3$ ; this corresponds to an approximate mass of 281 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  $0.294 \text{\AA}^{-1}$

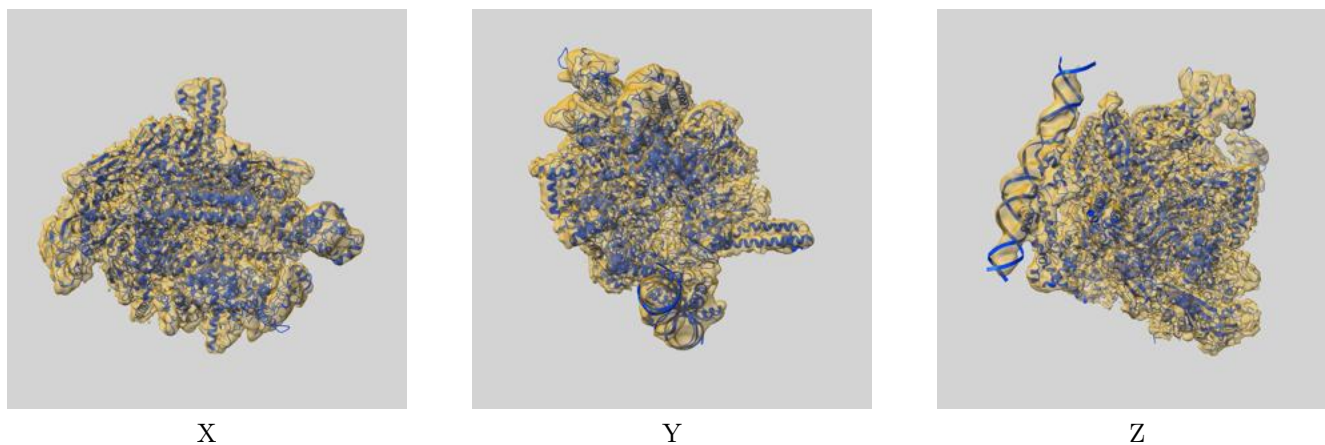
## 8 Fourier-Shell correlation

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-14200 and PDB model 7QXI. 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 0.02 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

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



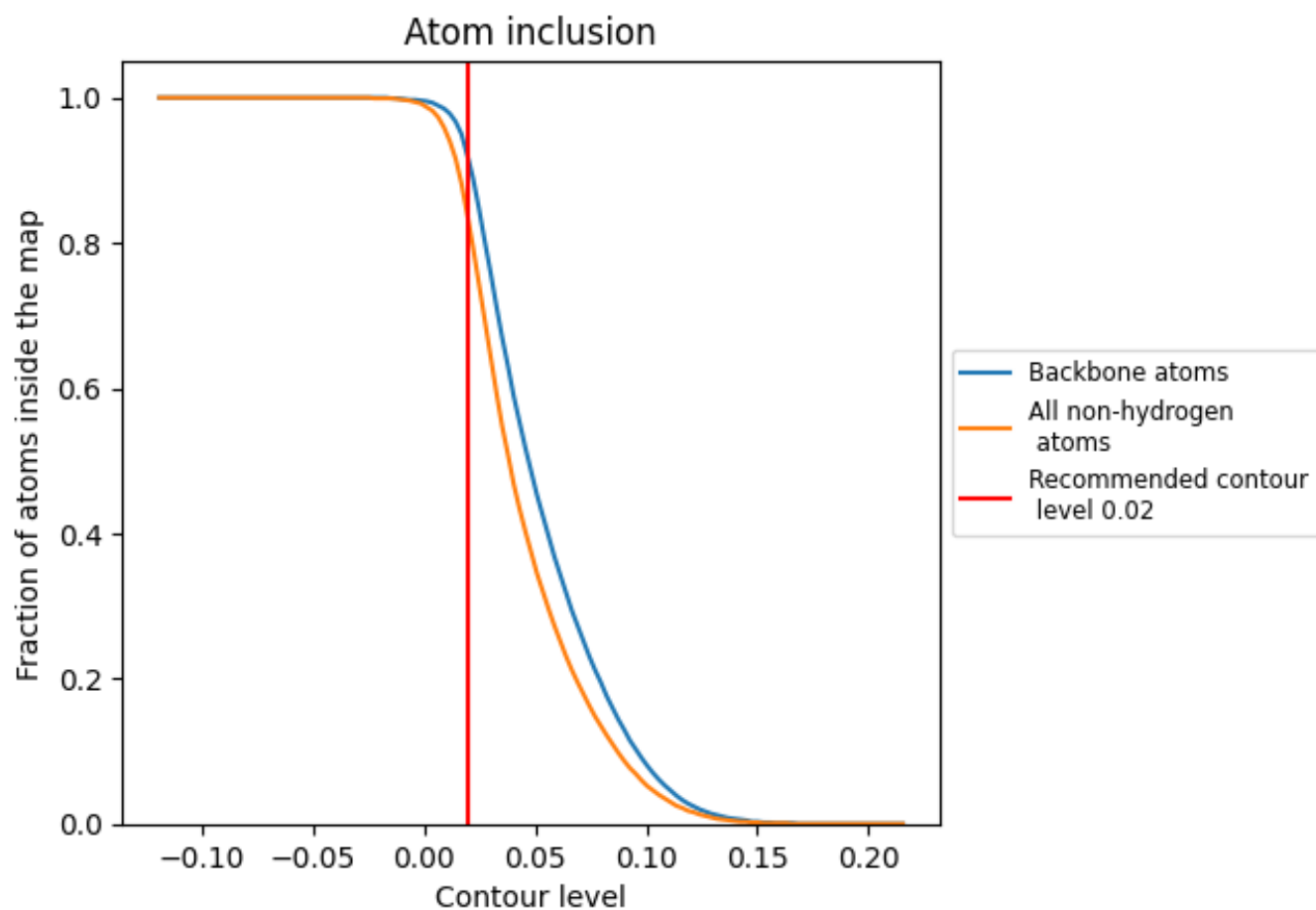
The images above show the model with each residue coloured according to its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions 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 (0.02).



















## 9.4 Atom inclusion [i](#)



At the recommended contour level, 91% of all backbone atoms, 83% of all non-hydrogen atoms, are inside the map.

## 9.5 Map-model fit summary

The table lists the average atom inclusion at the recommended contour level (0.02) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	 0.8290	 0.3640
A	 0.7870	 0.3820
B	 0.8450	 0.3680
C	 0.8610	 0.4040
D	 0.8310	 0.3630
E	 0.8770	 0.4130
M	 0.7920	 0.3010
N	 0.7470	 0.1710
T	 0.6800	 0.1780

