wwPDB EM Validation Summary Report

Mar 9, 2026 – 01:05 PM UTC

PDB ID	:	6P1K / pdb_00006p1k
EMDB ID	:	EMD-20230
Title	:	Cryo-EM structure of Escherichia coli sigma70 bound RNAP polymerase holoenzyme
Authors	:	Chen, J.; Chiu, C.E.; Campbell, E.A.; Darst, S.A.
Deposited on	:	2019-05-20
Resolution	:	4.05 Å(reported)

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

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the ⓘ 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:

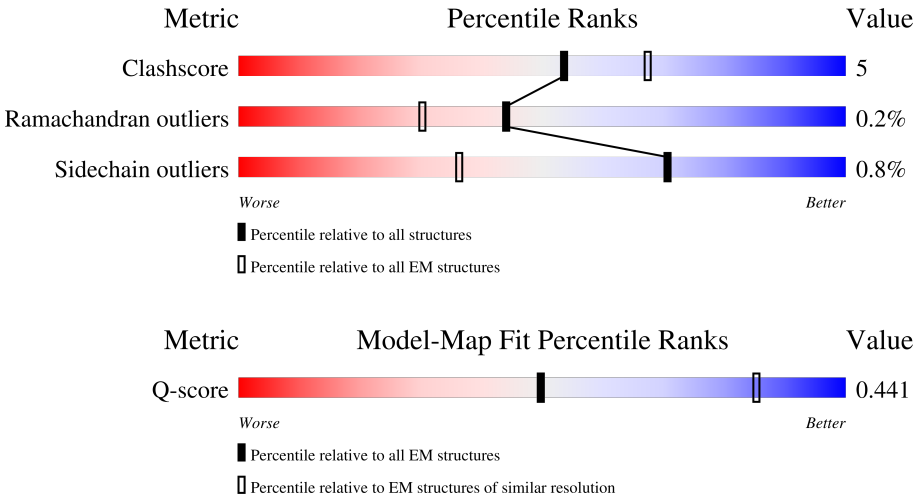
EMDB validation analysis	:	0.0.1.dev132
MolProbity	:	4-5-2 with Phenix2.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
EM percentile statistics	:	202505.v01 (Using data in the EMDB archive up until May 2025)
MapQ	:	1.9.13
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

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 4.05 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.



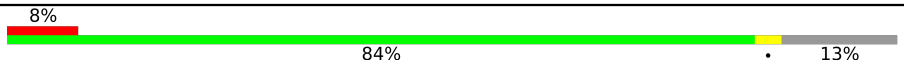
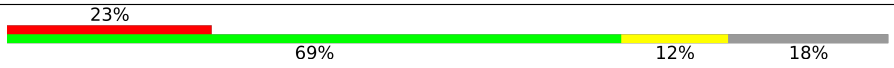
Metric	Whole archive (#Entries)	EM structures (#Entries)	Similar EM resolution (#Entries, resolution range(Å))
Clashscore	229148	23984	-
Ramachandran outliers	224038	23583	-
Sidechain outliers	223484	23102	-
Q-score	-	25397	6569 (3.55 - 4.55)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the EM map (all-atom inclusion < 40%). The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	G	329	62% 7% 31%
1	H	329	58% 8% 34%
2	I	1342	85% 15%
3	J	1430	6% 79% 13% 7%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	K	91	
5	L	616	

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 29011 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
1	G	228	Total	C	N	O	S	0	0
			1756	1094	312	344	6		
1	H	217	Total	C	N	O	S	0	0
			1667	1041	293	327	6		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
2	I	1340	Total	C	N	O	S	0	0
			10570	6631	1841	2055	43		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
3	J	1325	Total	C	N	O	S	0	0
			10295	6470	1831	1945	49		

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
J	1	VAL	-	expression tag	UNP P0A8T7
J	1408	LEU	-	expression tag	UNP P0A8T7
J	1409	GLU	-	expression tag	UNP P0A8T7
J	1410	LEU	-	expression tag	UNP P0A8T7
J	1411	GLU	-	expression tag	UNP P0A8T7
J	1412	VAL	-	expression tag	UNP P0A8T7
J	1413	LEU	-	expression tag	UNP P0A8T7
J	1414	PHE	-	expression tag	UNP P0A8T7
J	1415	GLN	-	expression tag	UNP P0A8T7
J	1416	GLY	-	expression tag	UNP P0A8T7
J	1417	PRO	-	expression tag	UNP P0A8T7
J	1418	SER	-	expression tag	UNP P0A8T7
J	1419	SER	-	expression tag	UNP P0A8T7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
J	1420	GLY	-	expression tag	UNP P0A8T7
J	1421	HIS	-	expression tag	UNP P0A8T7
J	1422	HIS	-	expression tag	UNP P0A8T7
J	1423	HIS	-	expression tag	UNP P0A8T7
J	1424	HIS	-	expression tag	UNP P0A8T7
J	1425	HIS	-	expression tag	UNP P0A8T7
J	1426	HIS	-	expression tag	UNP P0A8T7
J	1427	HIS	-	expression tag	UNP P0A8T7
J	1428	HIS	-	expression tag	UNP P0A8T7
J	1429	HIS	-	expression tag	UNP P0A8T7
J	1430	HIS	-	expression tag	UNP P0A8T7

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

Mol	Chain	Residues	Atoms					AltConf	Trace
4	K	79	Total	C	N	O	S	0	0
			627	382	118	126	1		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
5	L	504	Total	C	N	O	S	0	0
			4093	2563	726	780	24		

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
L	-2	SER	-	expression tag	UNP Q0P6L9
L	-1	GLU	-	expression tag	UNP Q0P6L9
L	0	PHE	-	expression tag	UNP Q0P6L9

- Molecule 6 is MAGNESIUM ION (CCD ID: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		AltConf
6	J	1	Total	Mg	0
			1	1	

- Molecule 7 is ZINC ION (CCD ID: ZN) (formula: Zn).

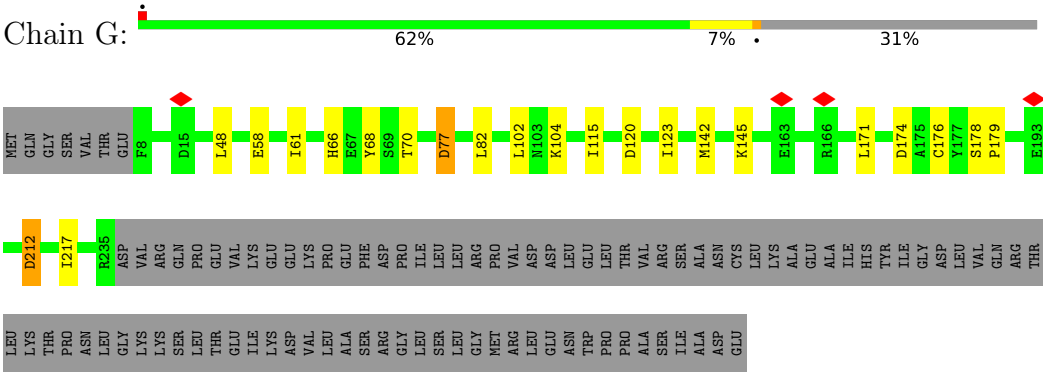
Mol	Chain	Residues	Atoms		AltConf
7	J	2	Total 2	Zn 2	0

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.

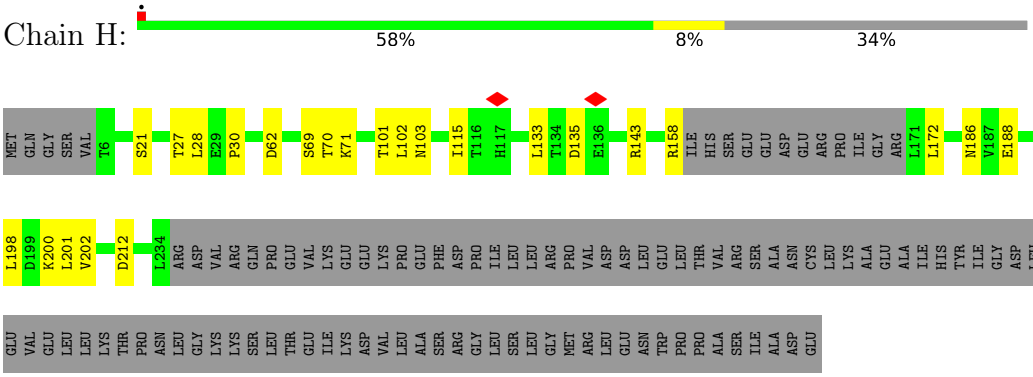
• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain G:



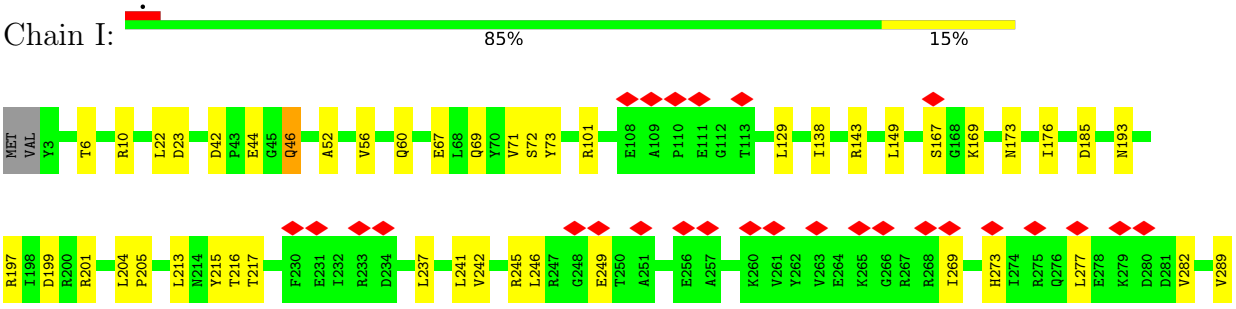
• Molecule 1: DNA-directed RNA polymerase subunit alpha

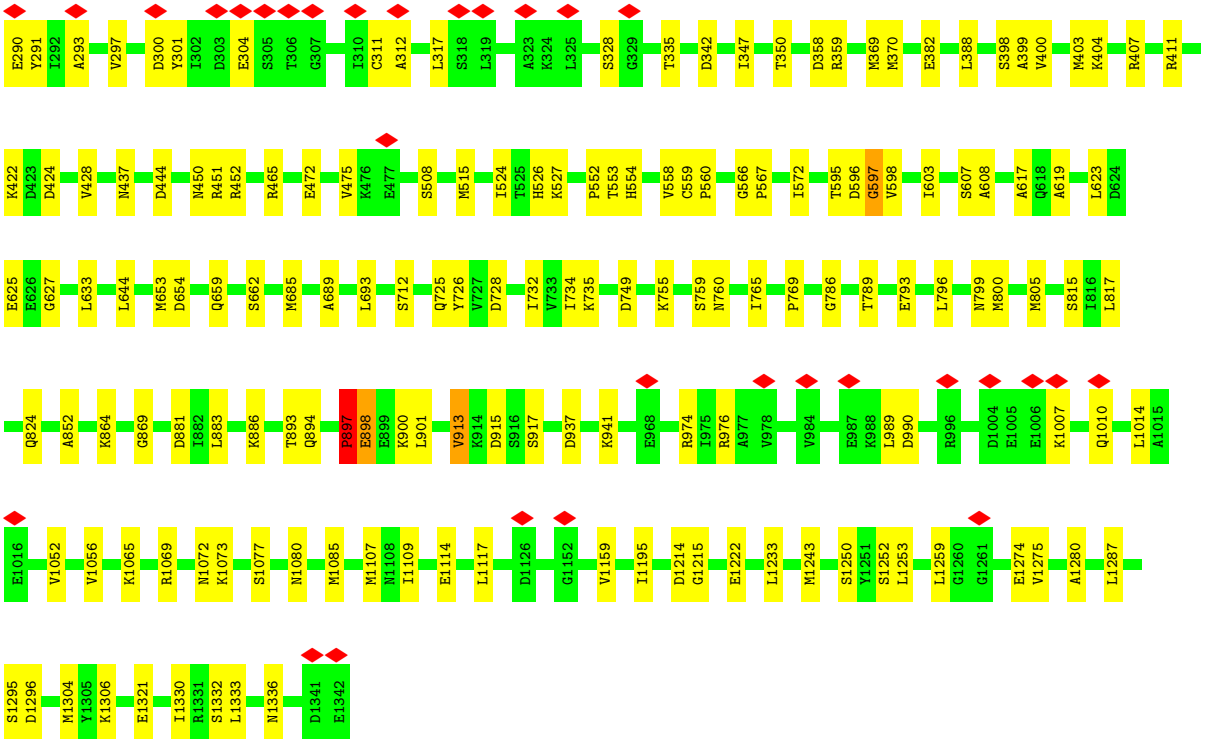
Chain H:



• Molecule 2: DNA-directed RNA polymerase subunit beta

Chain I:

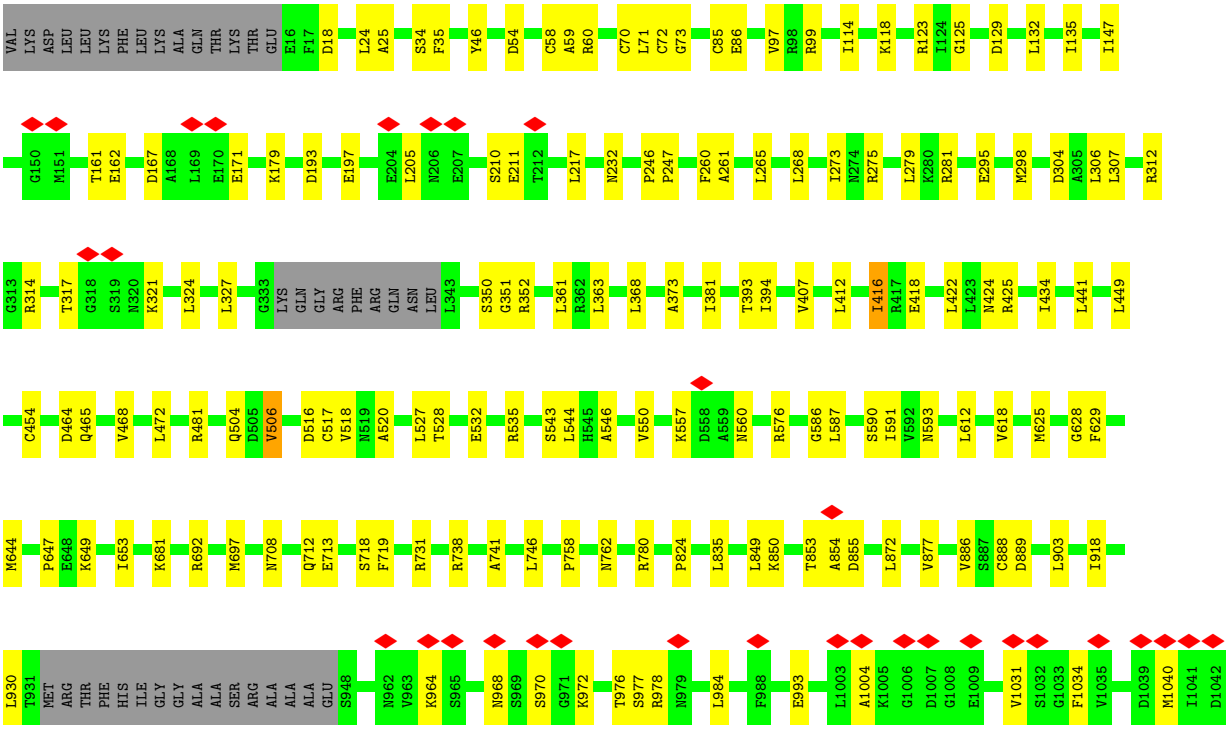


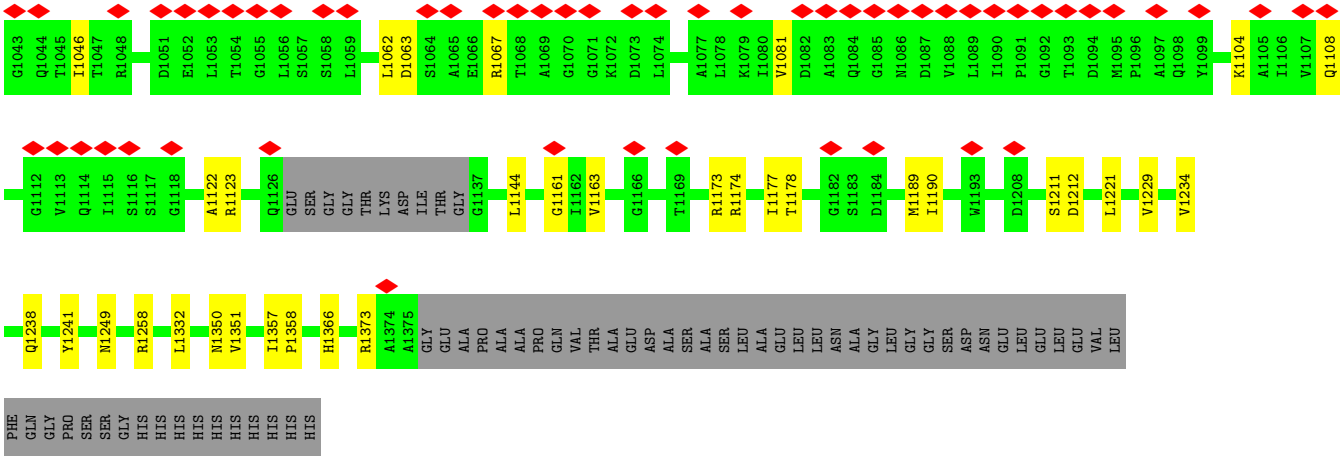


• Molecule 3: DNA-directed RNA polymerase subunit beta'

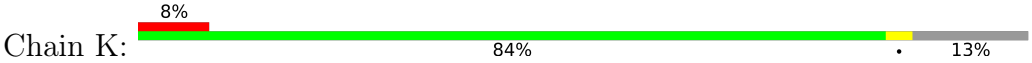
Chain J:

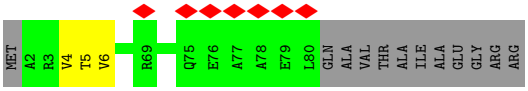






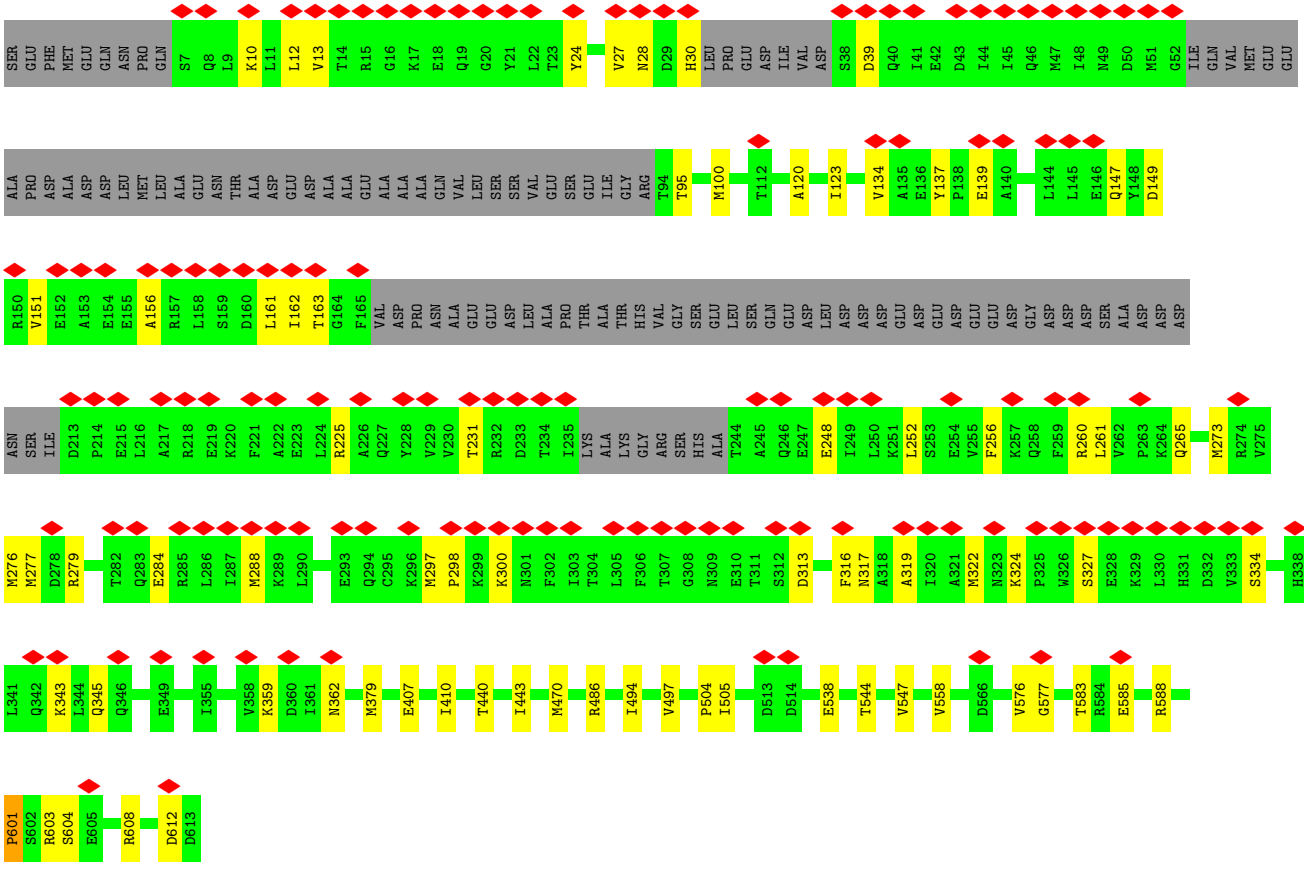
• Molecule 4: DNA-directed RNA polymerase subunit omega





• Molecule 5: RNA polymerase sigma factor RpoD





4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	358725	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TALOS ARCTICA	Depositor
Voltage (kV)	200	Depositor
Electron dose ($e^-/Å^2$)	0.71	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	0.374	Depositor
Minimum map value	-0.304	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.010	Depositor
Recommended contour level	0.025	Depositor
Map size (Å)	300.0, 300.0, 300.0	wwPDB
Map dimensions	200, 200, 200	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.5, 1.5, 1.5	Depositor

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, MG

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	G	0.35	0/1778	0.73	0/2410
1	H	0.30	0/1686	0.63	0/2285
2	I	0.37	0/10739	0.68	6/14489 (0.0%)
3	J	0.35	0/10450	0.64	1/14112 (0.0%)
4	K	0.29	0/629	0.68	1/847 (0.1%)
5	L	0.25	0/4145	0.64	0/5571
All	All	0.34	0/29427	0.66	8/39714 (0.0%)

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	H	0	1
2	I	0	6
3	J	0	1
All	All	0	8

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	K	6	VAL	N-CA-C	-6.37	105.13	111.88
2	I	56	VAL	N-CA-C	-6.17	106.75	111.62
2	I	595	THR	CA-C-N	6.03	133.05	121.54
2	I	595	THR	C-N-CA	6.03	133.05	121.54
3	J	97	VAL	N-CA-C	-5.27	107.22	111.91

There are no chirality outliers.

5 of 8 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	H	135	ASP	Peptide
2	I	205	PRO	Peptide
2	I	552	PRO	Peptide
2	I	596	ASP	Peptide
2	I	597	GLY	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	G	1756	0	1771	17	0
1	H	1667	0	1689	17	0
2	I	10570	0	10582	120	0
3	J	10295	0	10510	114	0
4	K	627	0	634	1	0
5	L	4093	0	4141	50	0
6	J	1	0	0	0	0
7	J	2	0	0	0	0
All	All	29011	0	29327	302	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 302 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:I:301:TYR:HB2	2:I:311:CYS:O	1.74	0.88
3:J:1034:PHE:HB2	3:J:1081:VAL:O	1.88	0.73
3:J:58:CYS:SG	3:J:59:ALA:N	2.67	0.67
2:I:1306:LYS:HD3	5:L:538:GLU:HG2	1.78	0.65
2:I:167:SER:HB3	2:I:169:LYS:HE2	1.78	0.65

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	G	226/329 (69%)	191 (84%)	35 (16%)	0	100	100
1	H	213/329 (65%)	183 (86%)	30 (14%)	0	100	100
2	I	1338/1342 (100%)	1176 (88%)	159 (12%)	3 (0%)	43	75
3	J	1317/1430 (92%)	1174 (89%)	141 (11%)	2 (0%)	43	75
4	K	77/91 (85%)	69 (90%)	8 (10%)	0	100	100
5	L	494/616 (80%)	458 (93%)	35 (7%)	1 (0%)	43	75
All	All	3665/4137 (89%)	3251 (89%)	408 (11%)	6 (0%)	44	75

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	I	897	PRO
3	J	854	ALA
2	I	898	GLU
5	L	601	PRO
3	J	46	TYR

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	G	192/286 (67%)	188 (98%)	4 (2%)	47	65
1	H	183/286 (64%)	183 (100%)	0	100	100
2	I	1155/1157 (100%)	1144 (99%)	11 (1%)	68	76

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
3	J	1110/1189 (93%)	1101 (99%)	9 (1%)	73 78
4	K	67/75 (89%)	67 (100%)	0	100 100
5	L	446/543 (82%)	446 (100%)	0	100 100
All	All	3153/3536 (89%)	3129 (99%)	24 (1%)	70 78

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	I	1159	VAL
3	J	407	VAL
3	J	361	LEU
3	J	422	LEU
2	I	215	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 36 such sidechains are listed below:

Mol	Chain	Res	Type
5	L	301	ASN
5	L	589	GLN
5	L	317	ASN
5	L	400	GLN
2	I	725	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

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

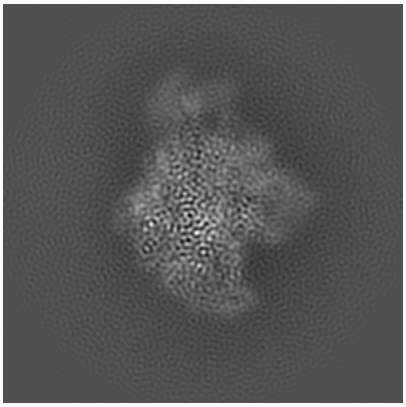
6 Map visualisation

This section contains visualisations of the EMDB entry EMD-20230. These allow visual inspection of the internal detail of the map and identification of artifacts.

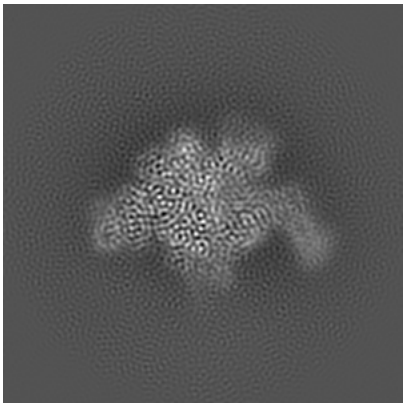
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections

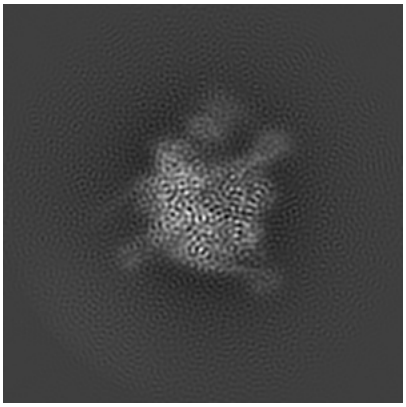
6.1.1 Primary map



X



Y

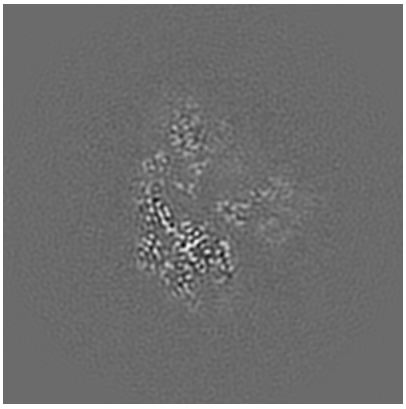


Z

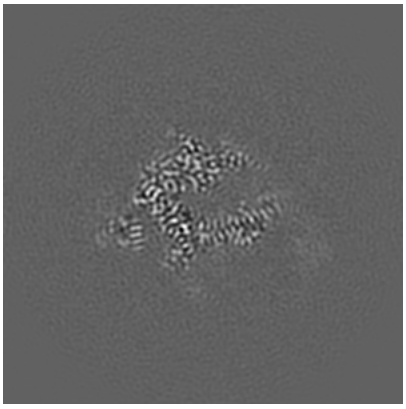
The images above show the map projected in three orthogonal directions.

6.2 Central slices

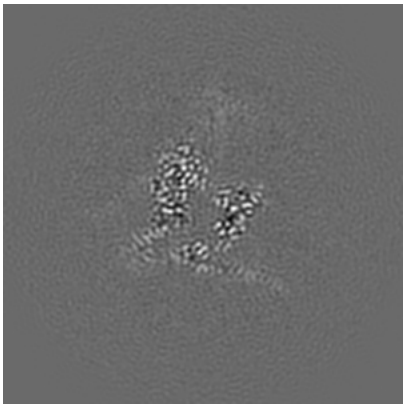
6.2.1 Primary map



X Index: 100



Y Index: 100

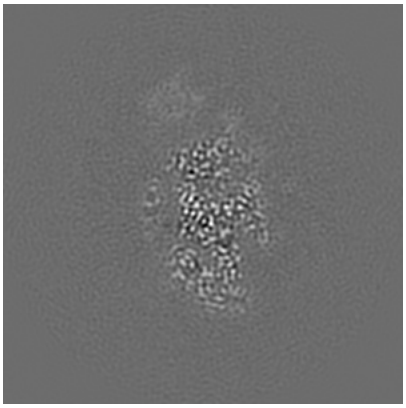


Z Index: 100

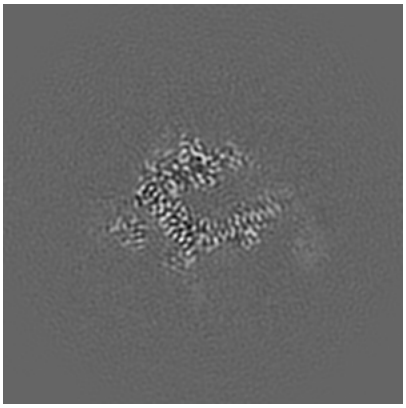
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

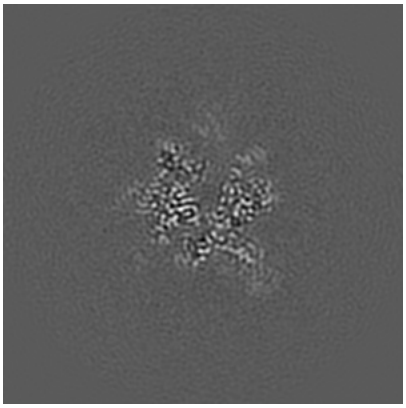
6.3.1 Primary map



X Index: 87



Y Index: 99

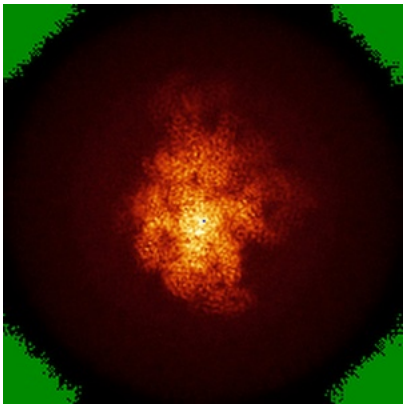


Z Index: 92

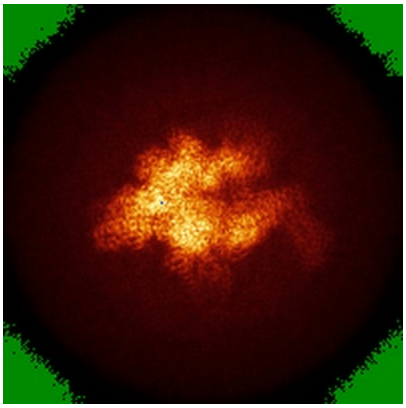
The images above show the largest variance slices of the map in three orthogonal directions.

6.4 Orthogonal standard-deviation projections (False-color) (i)

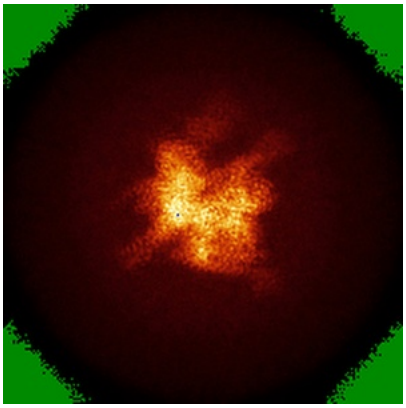
6.4.1 Primary map



X



Y

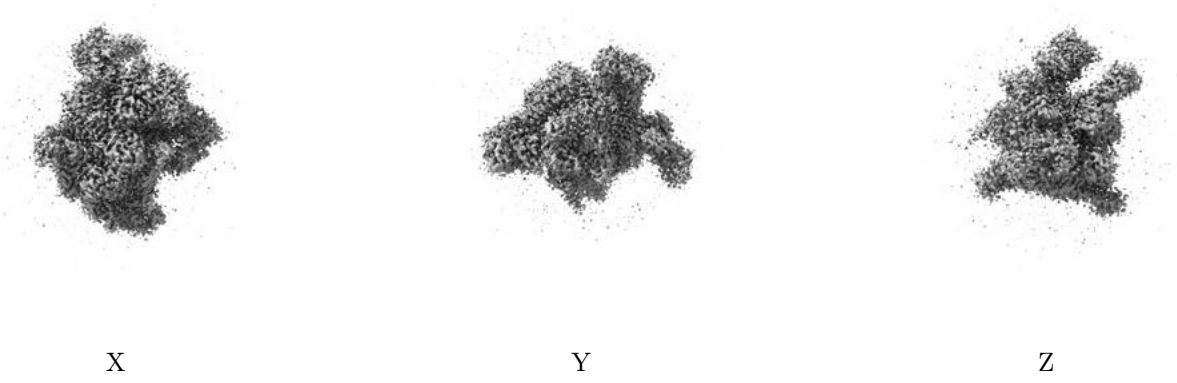


Z

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



X Y Z

The images above show the 3D surface view of the map at the recommended contour level 0.025. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

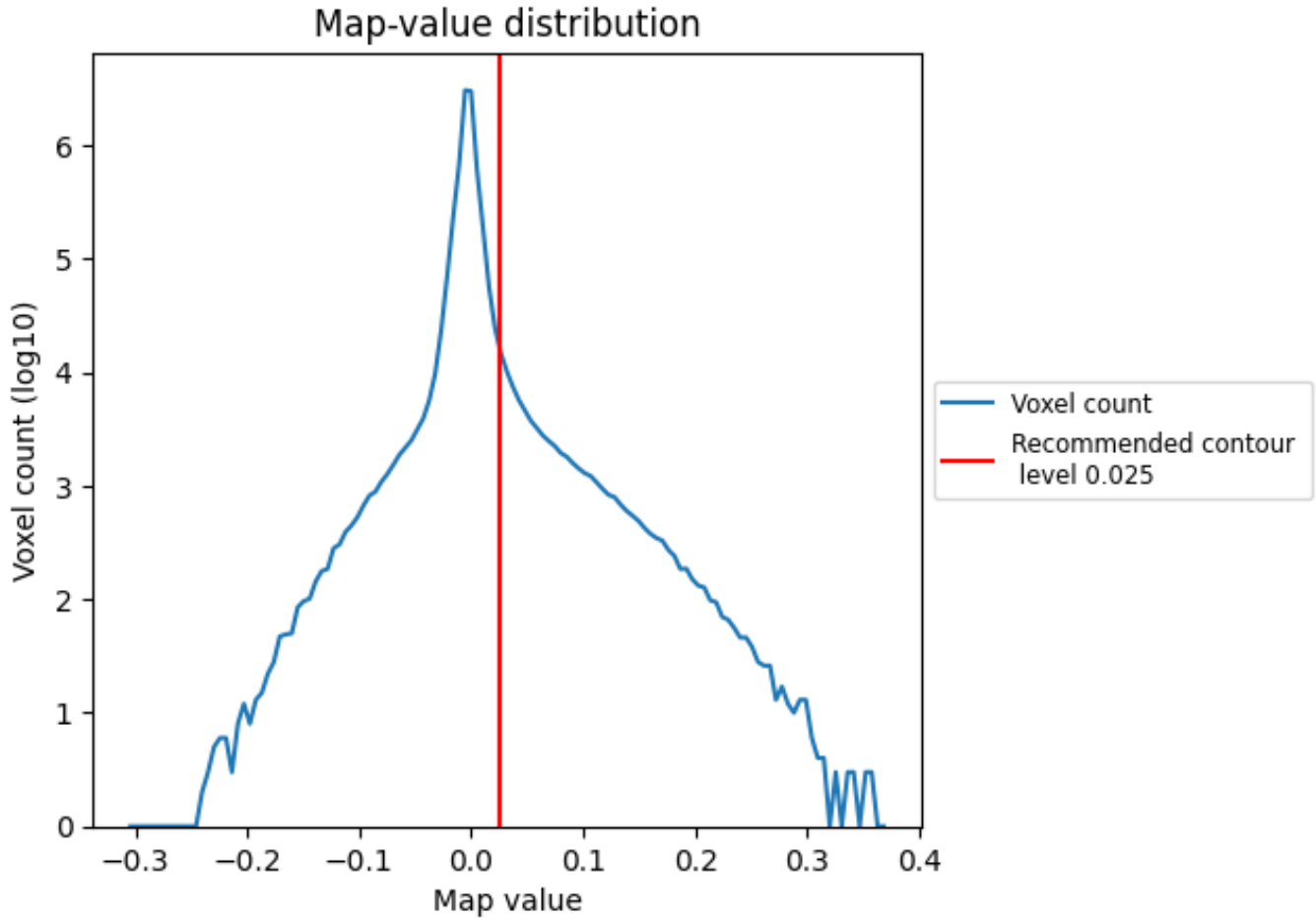
6.6 Mask visualisation (i)

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

7 Map analysis

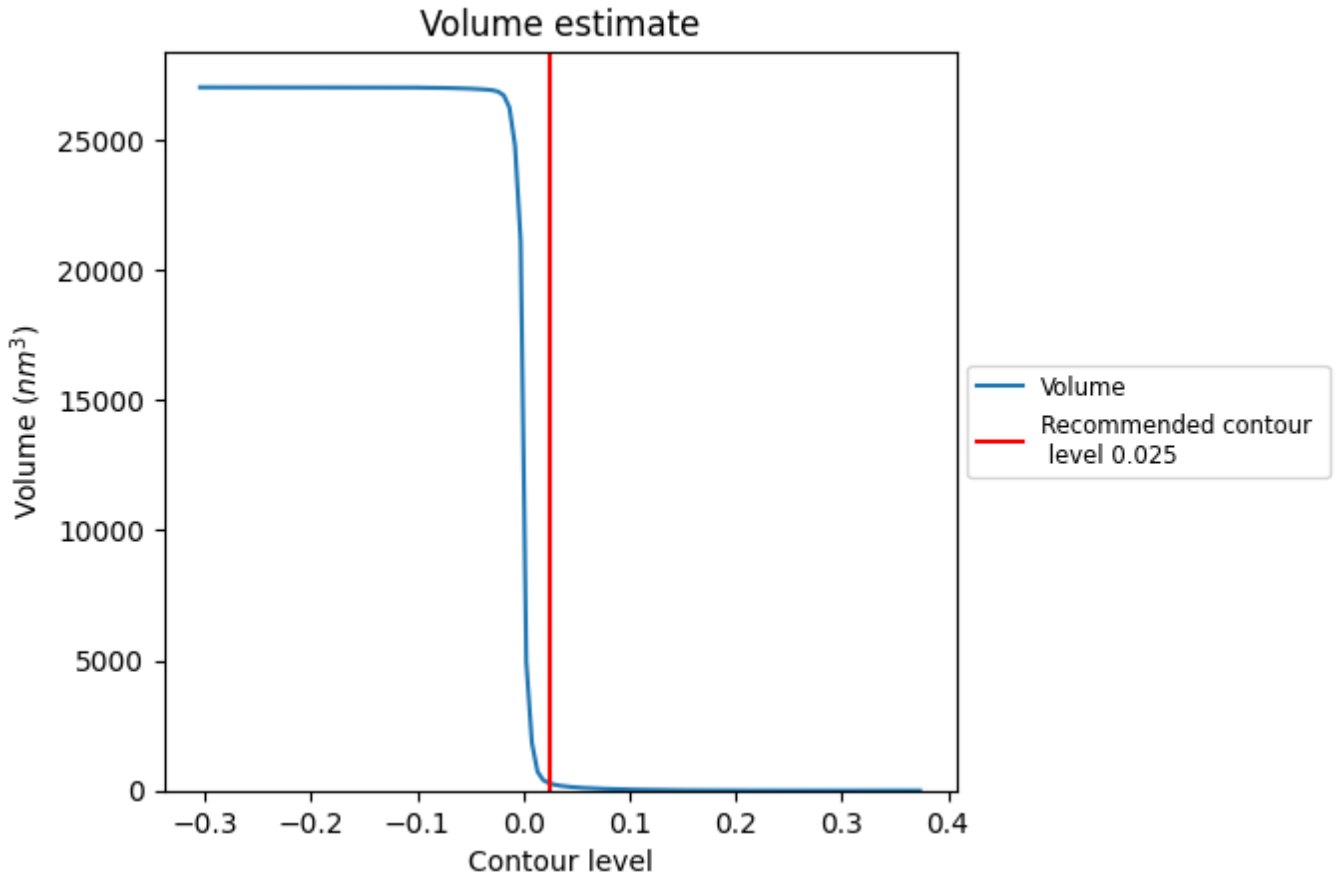
This section contains the results of statistical analysis of the map.

7.1 Map-value distribution



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.

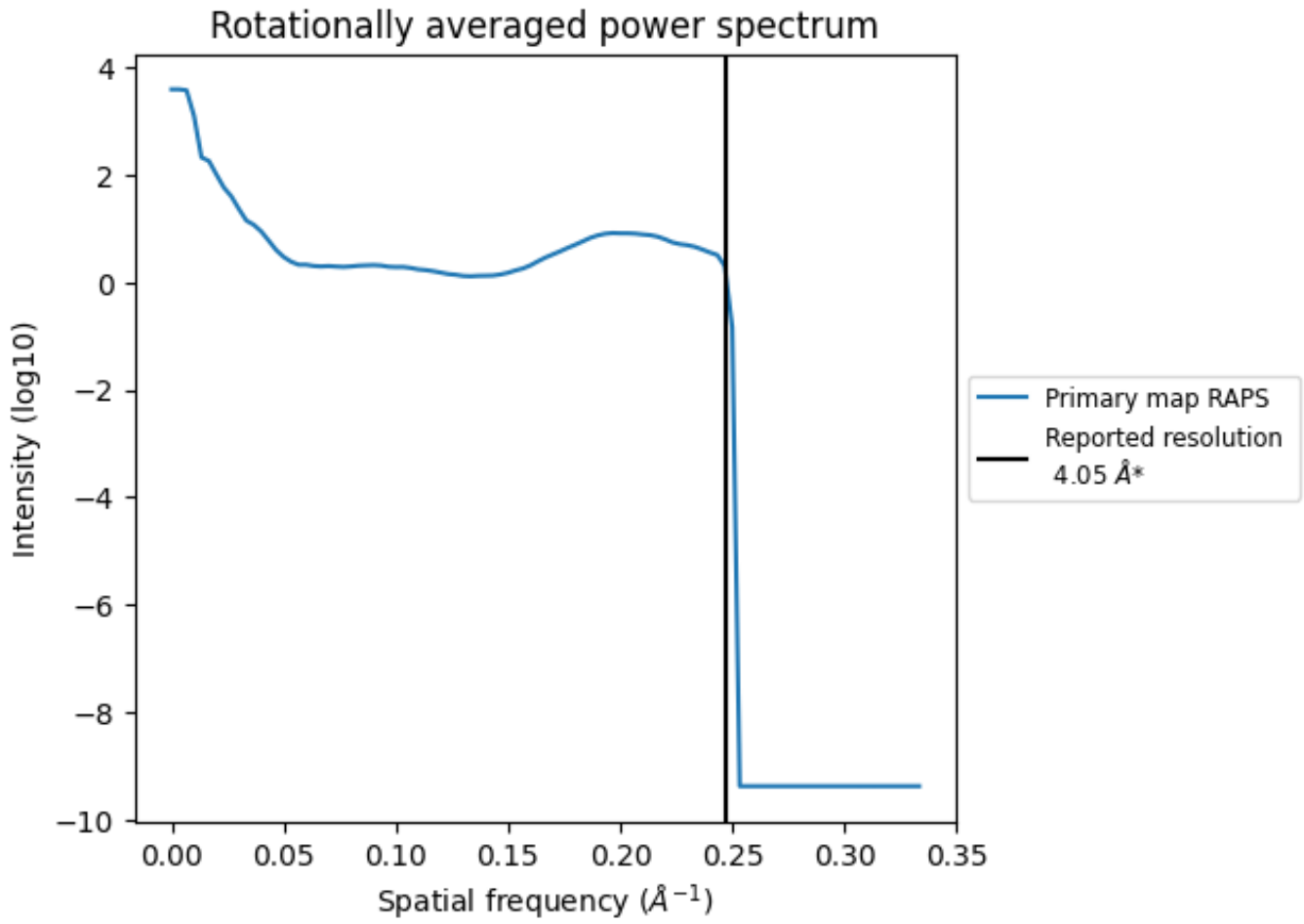
7.2 Volume estimate (i)



The volume at the recommended contour level is 276 nm^3; this corresponds to an approximate mass of 250 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)

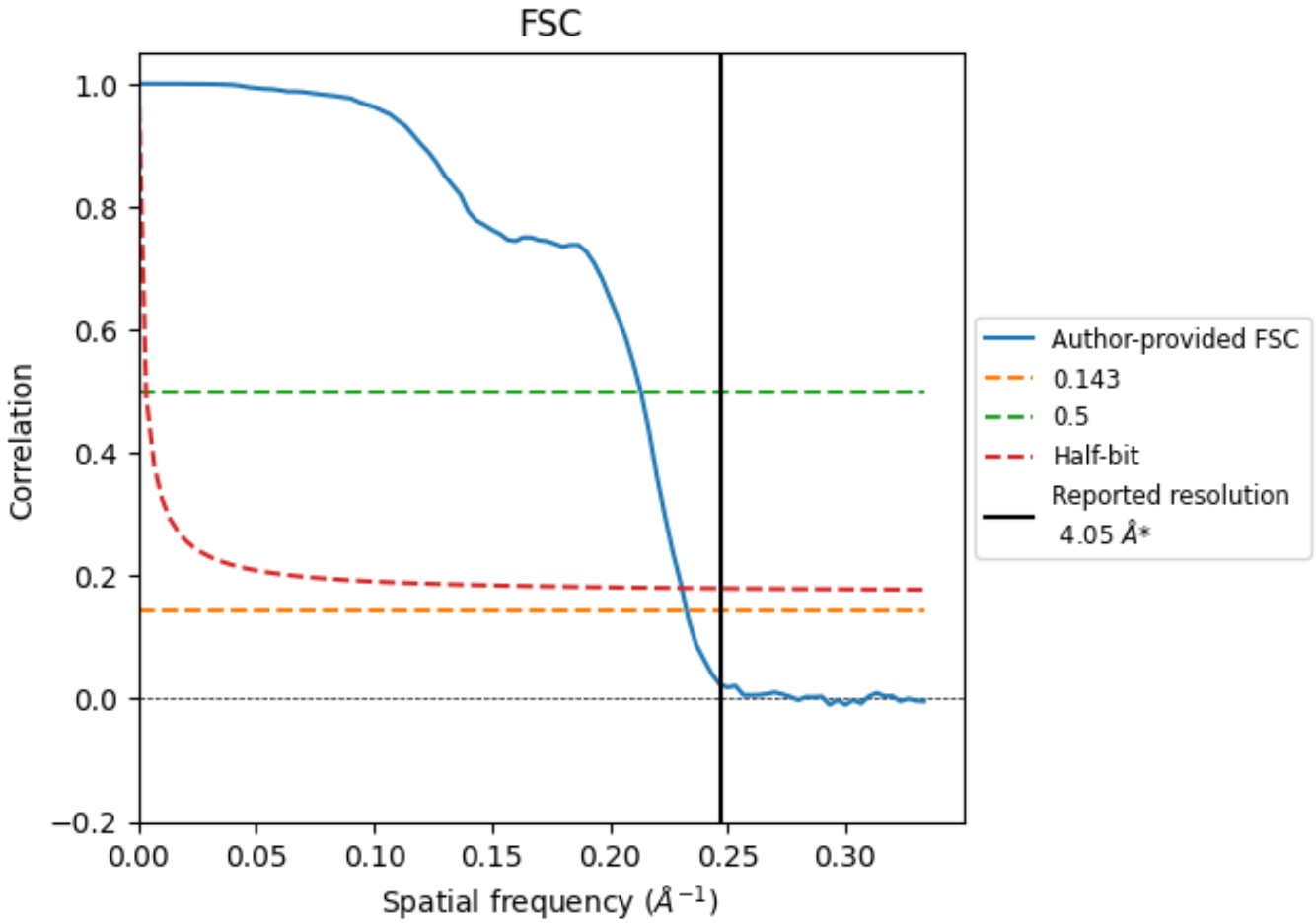


*Reported resolution corresponds to spatial frequency of 0.247 Å^{-1}

8 Fourier-Shell correlation

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



*Reported resolution corresponds to spatial frequency of 0.247 Å^{-1}

8.2 Resolution estimates (i)

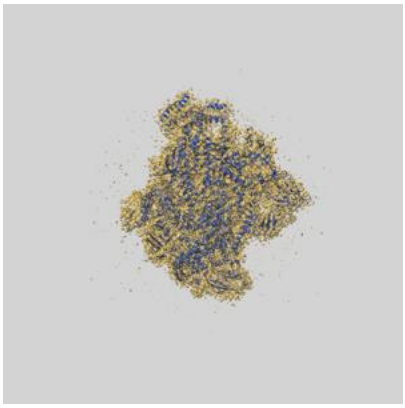
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	4.05	-	-
Author-provided FSC curve	4.30	4.69	4.34
Unmasked-calculated*	-	-	-

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

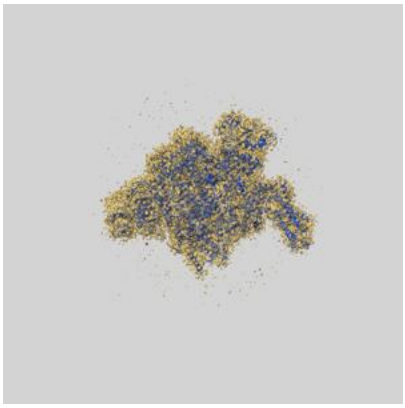
9 Map-model fit

This section contains information regarding the fit between EMDB map EMD-20230 and PDB model 6P1K. Per-residue inclusion information can be found in section 3 on page 7.

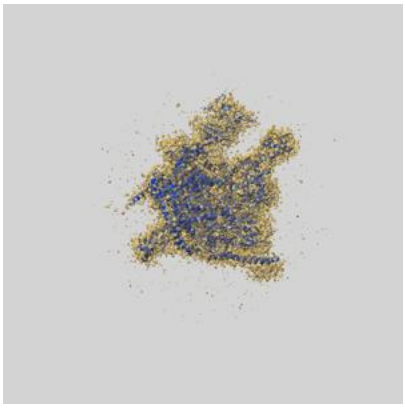
9.1 Map-model overlay



X



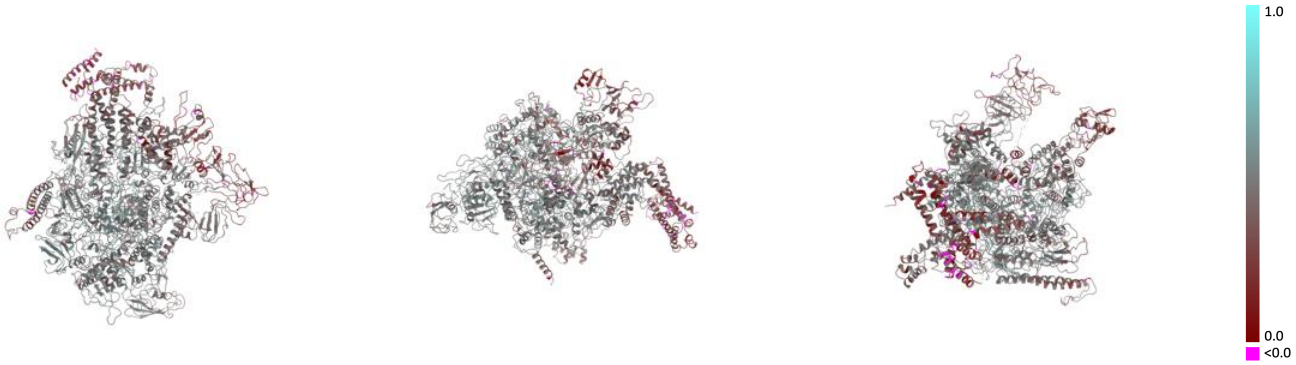
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.025 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

9.2 Q-score mapped to coordinate model (i)



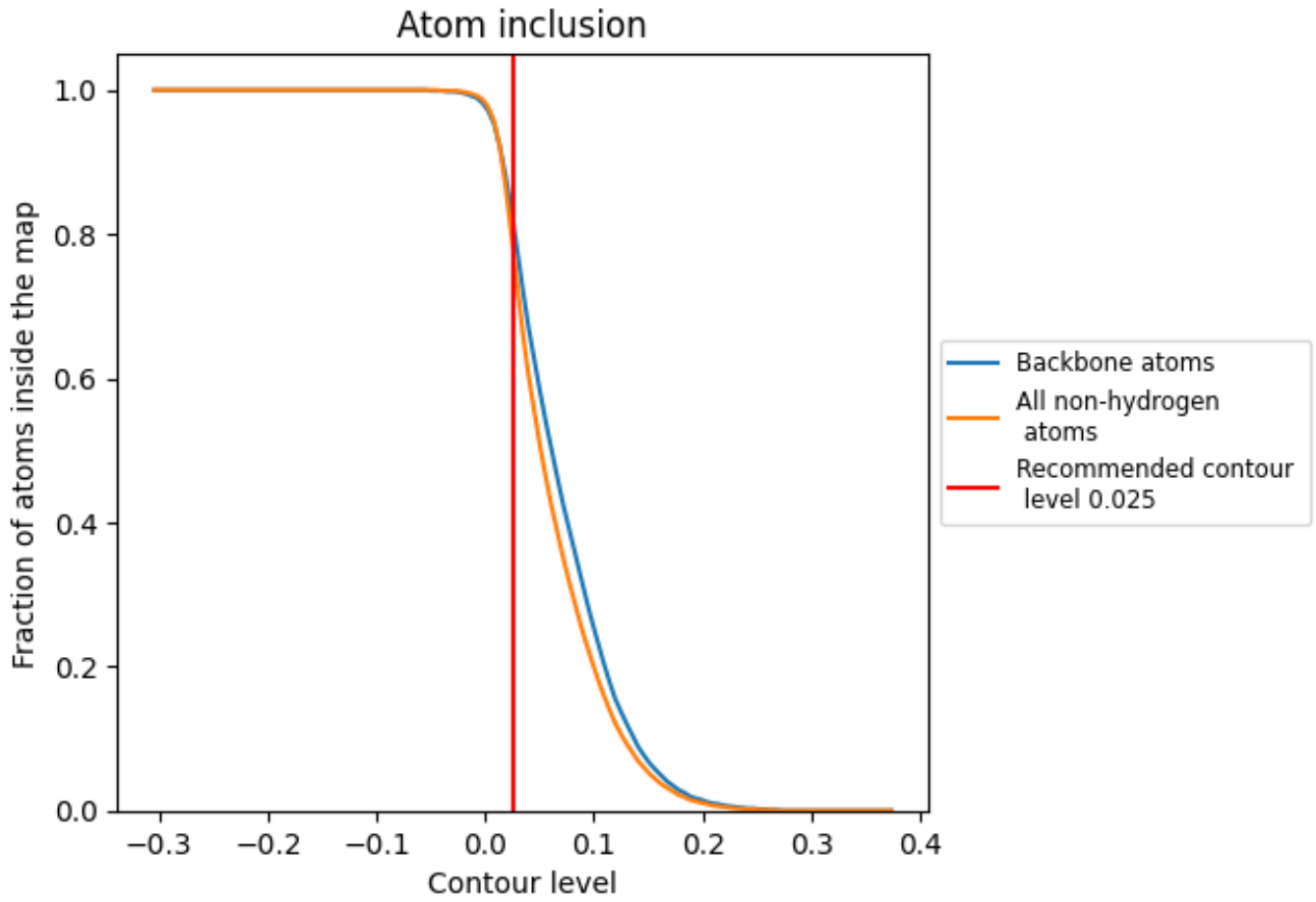
X Y Z

The images above show the model with each residue coloured according its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions at which atoms can be resolved. Low Q-score values may therefore be expected for many entries.

9.3 Atom inclusion mapped to coordinate model (i)

This section was not generated.

9.4 Atom inclusion (i)



At the recommended contour level, 82% of all backbone atoms, 79% 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 (0.025) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.7870	0.4410
G	0.8680	0.4910
H	0.8320	0.4670
I	0.8250	0.4610
J	0.8060	0.4570
K	0.7410	0.4280
L	0.5930	0.3190

