wwPDB EM Validation Summary Report (i)

Jul 3, 2024 – 11:19 am BST

PDB ID	:	6YYS
EMDB ID	:	EMD-11004
Title	:	Structure of Mycobacterium smegmatis HelD protein in complex with RNA polymerase core - State II, primary channel engaged and active site interfering
Authors	:	Kouba, T.; Koval, T.; Krasny, L.; Dohnalek, J.
Deposited on	:	2020-05-05
Resolution	:	3.08 Å(reported)
Based on initial models	:	6VSX, 6F6W

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 (i) 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 (i)) were used in the production of this report:

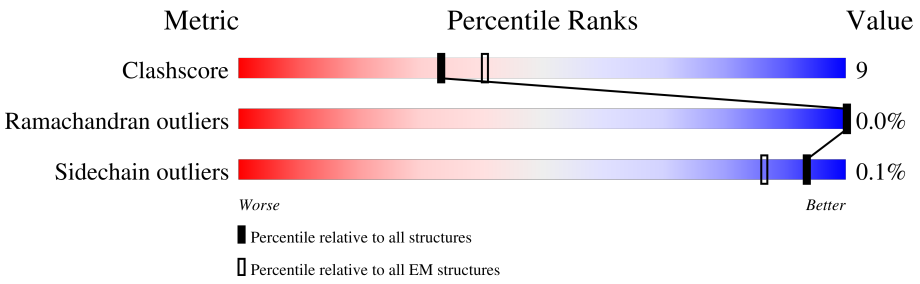
EMDB validation analysis	:	0.0.1.dev92
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.13
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 3.08 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 >=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	A	350	• 56% 8% 36%
1	B	350	• 53% 14% 33%
2	C	1169	• 75% 18% 7%
3	D	1317	19% 76% 21% •
4	E	107	9% 60% 18% 22%
5	H	736	20% 74% 20% 6%

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 27930 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	224	1700	1070	293	334	3	0	0
1	B	233	1771	1112	305	351	3	0	0

- 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	1089	8433	5276	1476	1646	35	0	0

- 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	1278	9983	6249	1809	1884	41	0	0

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

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
4	E	83	646	411	108	127	0	0

- Molecule 5 is a protein called RNA polymerase-associated transcription factor HelD.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
5	H	690	5394	3364	972	1042	16	0	0

- Molecule 6 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		AltConf
			Total	Zn	
6	D	2	2	2	0

- Molecule 7 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

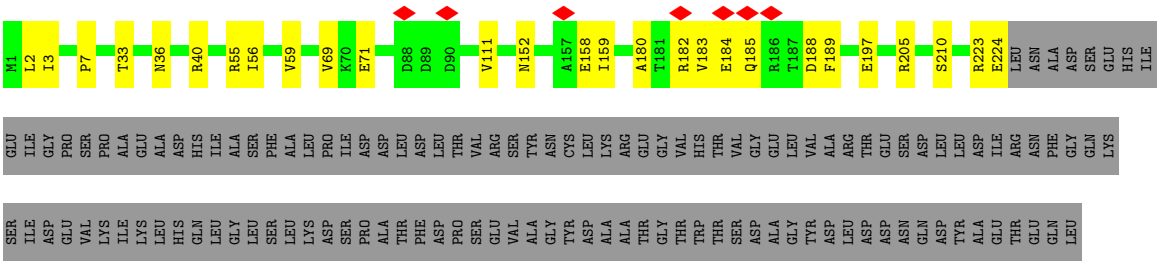
Mol	Chain	Residues	Atoms		AltConf
			Total	Mg	
7	D	1	1	1	0

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.

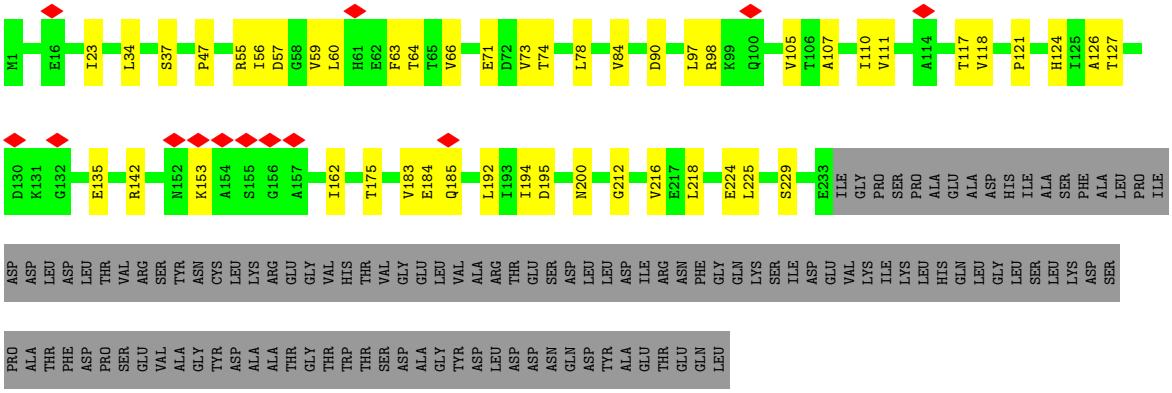
• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain A:



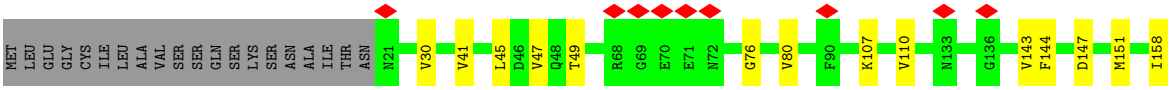
• Molecule 1: DNA-directed RNA polymerase subunit alpha

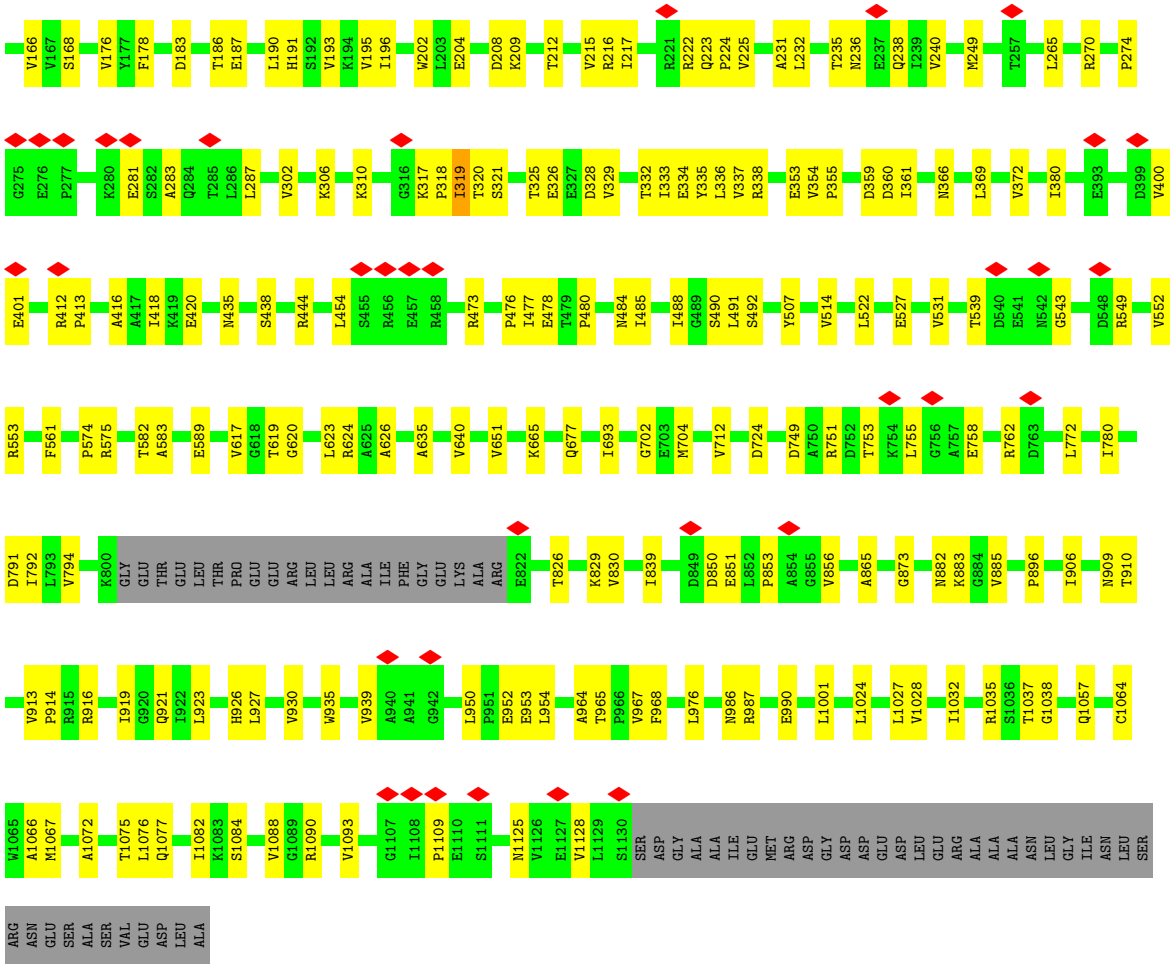
Chain B:



• Molecule 2: DNA-directed RNA polymerase subunit beta

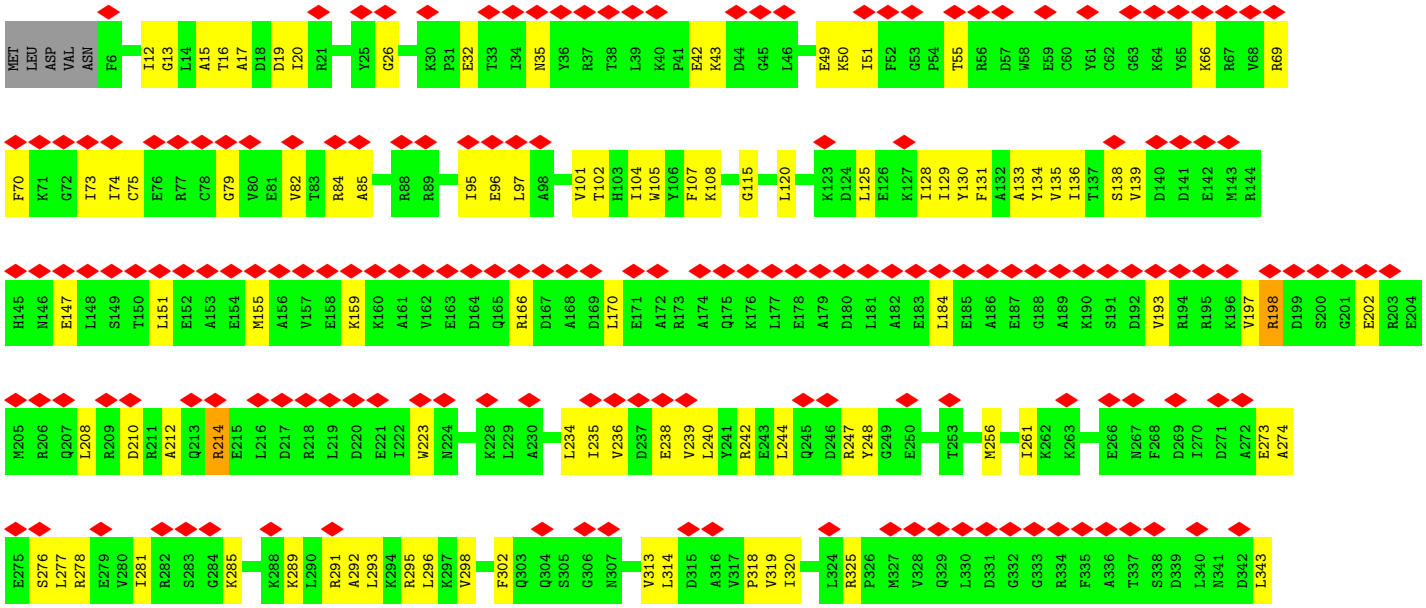
Chain C:

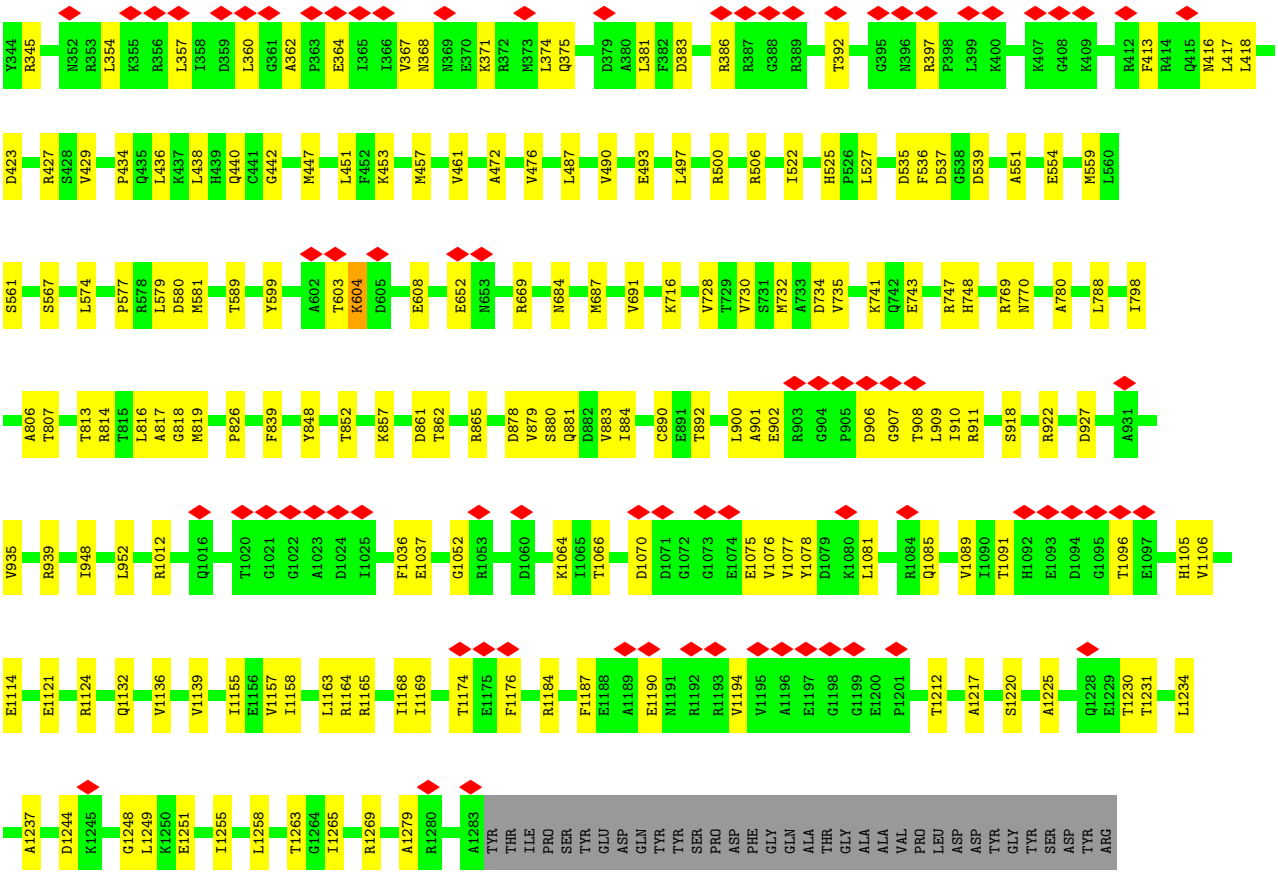




- Molecule 3: DNA-directed RNA polymerase subunit beta'

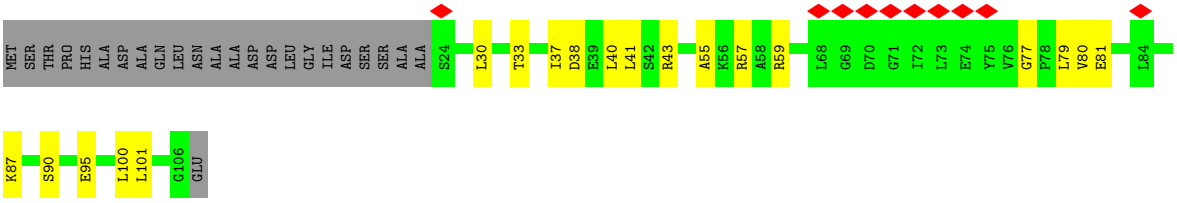
Chain D:





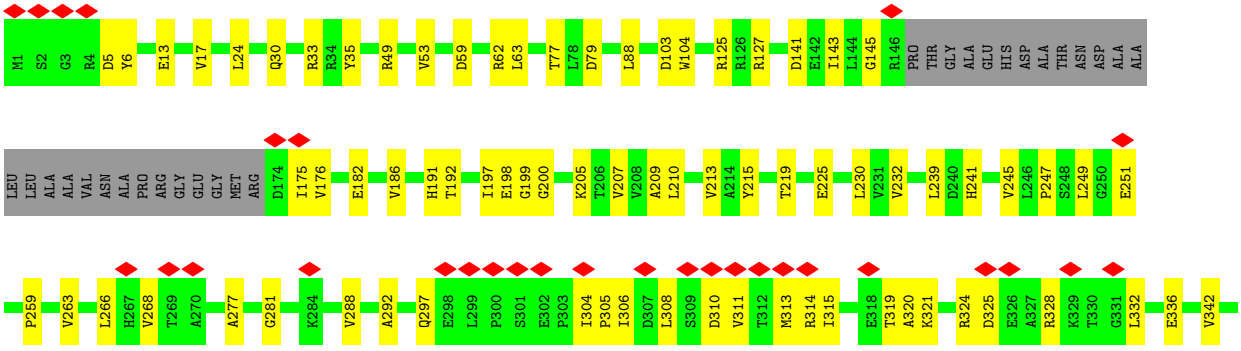
- Molecule 4: DNA-directed RNA polymerase subunit omega

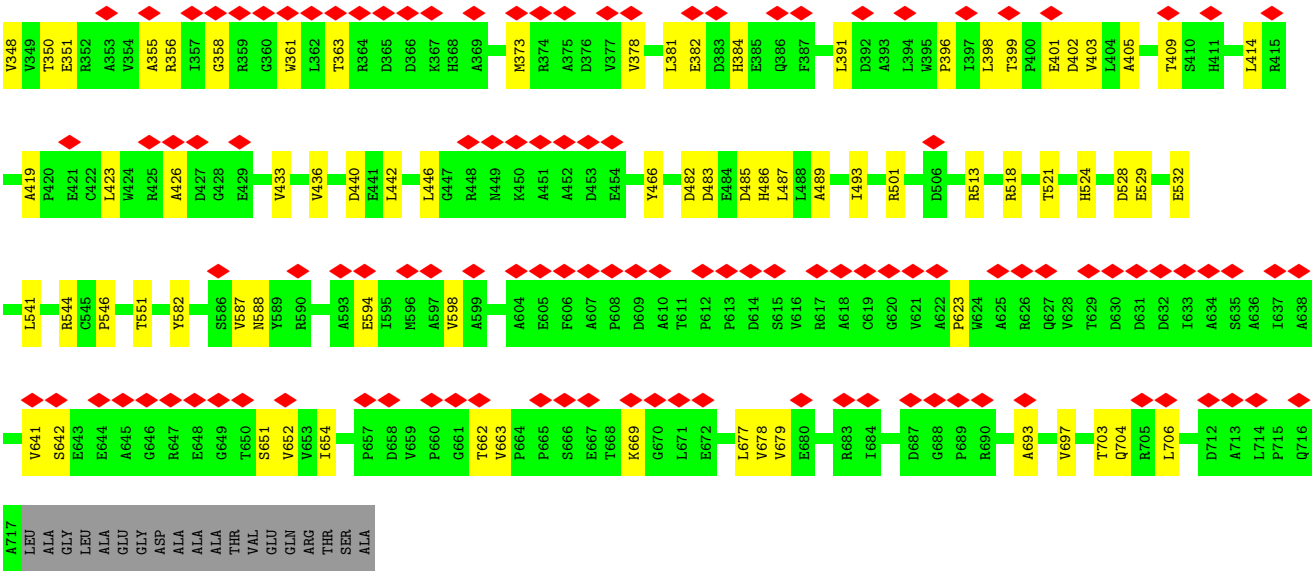
Chain E:



- Molecule 5: RNA polymerase-associated transcription factor HelD

Chain H:




V348 V349 T350 E351 R352 A353 V354 A355 R356 I357 G358 R359 G360 W361 L362 T363 R364 D365 D366 K367 H368 A369 M373 R374 A375 D376 V377 V378 L381 E382 D383 H384 E385 Q386 F387 L391 D392 A393 L394 W395 P396 I397 L398 T399 P400 E401 D402 V403 L404 A405 T409 S410 H411 L414 R415
A419 P420 E421 C422 L423 W424 R425 A426 D427 G428 E429 V433 V436 D440 E441 L442 L446 G447 R448 N449 K450 A451 A452 D453 E454 Y466 D482 D483 E484 D485 H486 L487 L488 A489 I493 R501 D506 R513 R518 T521 H524 D528 E529 E532
L541 R544 C545 P546 T551 Y582 S586 V587 N588 Y589 R590 A593 E594 I595 M596 A597 V598 A599 A604 E605 F606 A607 P608 D609 A610 T611 P612 P613 D614 S615 V616 R617 A618 C619 G620 V621 A622 P623 W624 A625 R626 Q627 V628 T629 D630 D631 D632 I633 A634 S635 A636 I637 A638
V641 S642 E643 E644 A645 G646 R647 E648 G649 T650 S651 V652 V653 I654 P657 D658 V659 P660 G661 T662 V663 P664 P665 S666 E667 T668 K669 G670 L671 E672 L677 V678 V679 E680 R683 I684 D687 G688 P689 R690 A693 V697 T703 Q704 R705 L706 D712 A713 L714 P715 Q716
A717 LEU ALA GLY LEU ALA GLU GLY ASP ALA ALA ALA THR VAL GLU GLN ARG THR SER ALA

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	173500	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 ($e^-/Å^2$)	50	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	0.728	Depositor
Minimum map value	-0.227	Depositor
Average map value	0.003	Depositor
Map value standard deviation	0.025	Depositor
Recommended contour level	0.17	Depositor
Map size (Å)	299.19598, 299.19598, 299.19598	wwPDB
Map dimensions	360, 360, 360	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	0.8310999, 0.8310999, 0.8310999	Depositor

5 Model quality (i)

5.1 Standard geometry (i)

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

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	0.32	0/1726	0.50	0/2348
1	B	0.30	0/1798	0.49	0/2446
2	C	0.32	0/8583	0.49	0/11632
3	D	0.29	0/10145	0.48	0/13709
4	E	0.29	0/659	0.48	0/897
5	H	0.28	0/5498	0.46	0/7476
All	All	0.30	0/28409	0.48	0/38508

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	1700	0	1742	21	0
1	B	1771	0	1803	36	0
2	C	8433	0	8377	154	0
3	D	9983	0	10103	210	0
4	E	646	0	642	15	0
5	H	5394	0	5269	108	0
6	D	2	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	D	1	0	0	0	0
All	All	27930	0	27936	510	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 510 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)
3:D:599:TYR:OH	3:D:608:GLU:OE1	1.93	0.84
3:D:125:LEU:HD11	3:D:261:ILE:HD11	1.56	0.84
2:C:910:THR:HG23	3:D:730:VAL:HG23	1.65	0.78
1:A:210:SER:OG	1:B:229:SER:O	2.00	0.76
2:C:623:LEU:HD12	2:C:702:GLY:O	1.86	0.75

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	222/350 (63%)	212 (96%)	10 (4%)	0	100	100
1	B	231/350 (66%)	214 (93%)	17 (7%)	0	100	100
2	C	1085/1169 (93%)	994 (92%)	90 (8%)	1 (0%)	51	82
3	D	1276/1317 (97%)	1172 (92%)	104 (8%)	0	100	100
4	E	81/107 (76%)	71 (88%)	10 (12%)	0	100	100
5	H	686/736 (93%)	601 (88%)	85 (12%)	0	100	100
All	All	3581/4029 (89%)	3264 (91%)	316 (9%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	319	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	191/297 (64%)	191 (100%)	0	100	100
1	B	199/297 (67%)	199 (100%)	0	100	100
2	C	921/984 (94%)	921 (100%)	0	100	100
3	D	1062/1095 (97%)	1058 (100%)	4 (0%)	91	95
4	E	69/86 (80%)	69 (100%)	0	100	100
5	H	553/583 (95%)	553 (100%)	0	100	100
All	All	2995/3342 (90%)	2991 (100%)	4 (0%)	93	97

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	69	ARG
3	D	198	ARG
3	D	214	ARG
3	D	604	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	505	HIS
5	H	406	GLN
3	D	552	GLN
5	H	505	GLN
3	D	1092	HIS

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)

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 (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

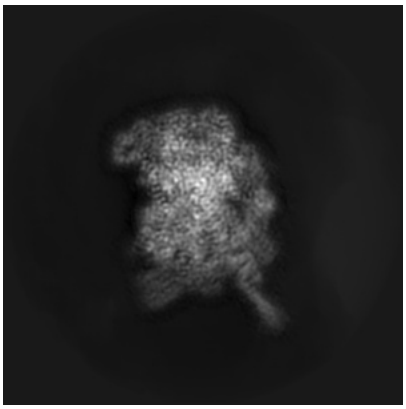
6 Map visualisation (i)

This section contains visualisations of the EMDB entry EMD-11004. These allow visual inspection of the internal detail of the map and identification of artifacts.

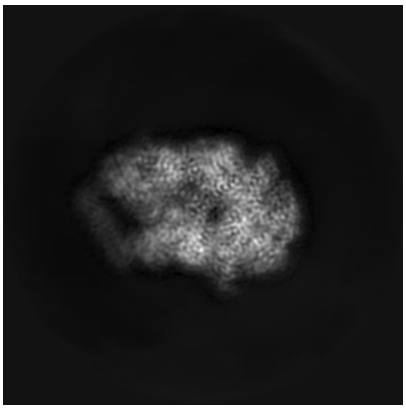
Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

6.1 Orthogonal projections (i)

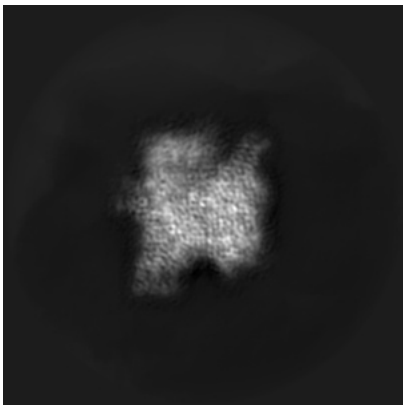
6.1.1 Primary map



X

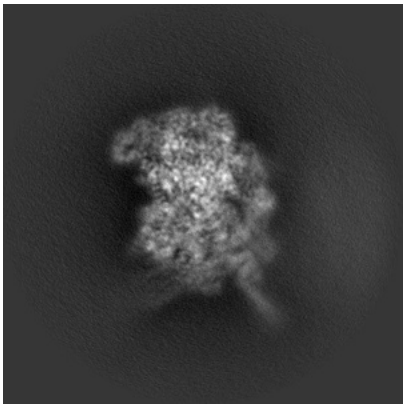


Y

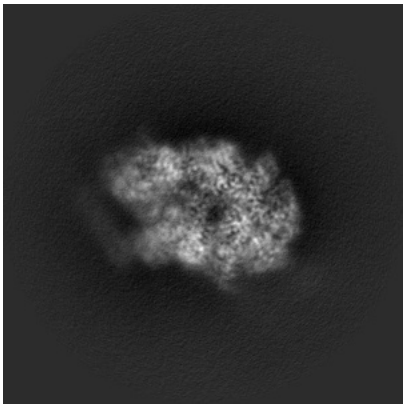


Z

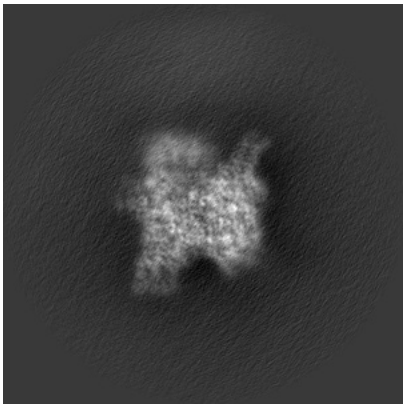
6.1.2 Raw map



X



Y

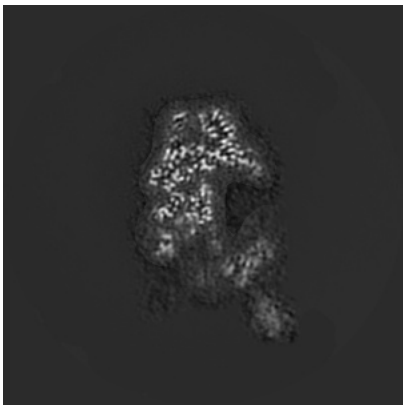


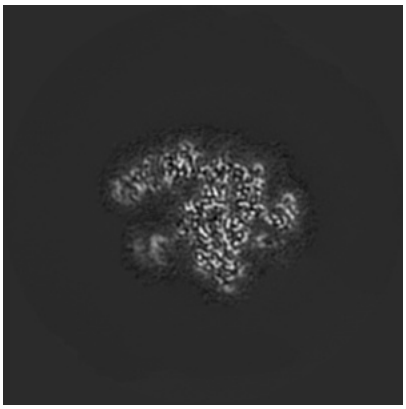
Z

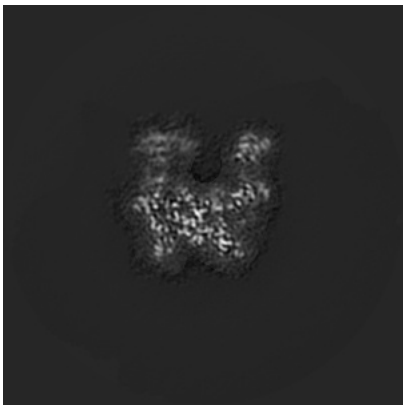
The images above show the map projected in three orthogonal directions.

6.2 Central slices (i)

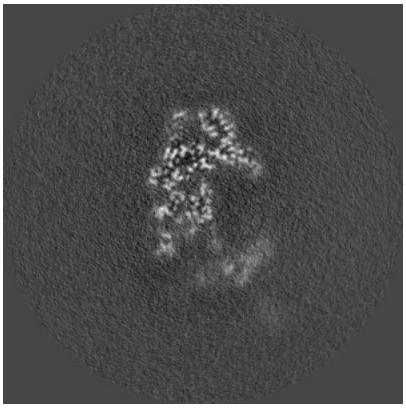
6.2.1 Primary map

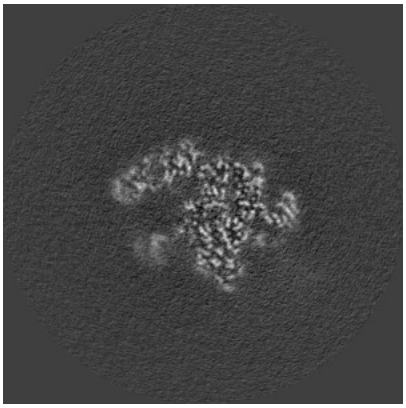

X Index: 180

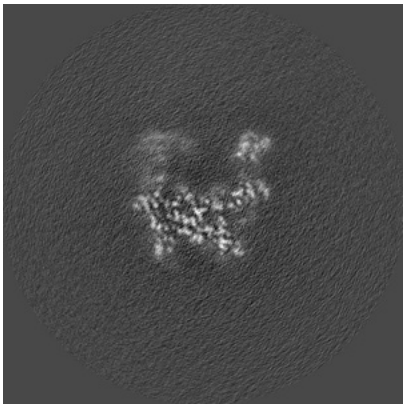

Y Index: 180


Z Index: 180

6.2.2 Raw map


X Index: 180


Y Index: 180

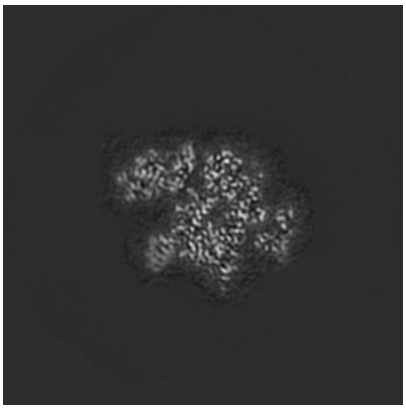

Z Index: 180

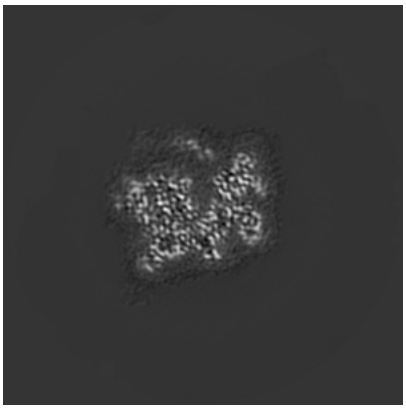
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

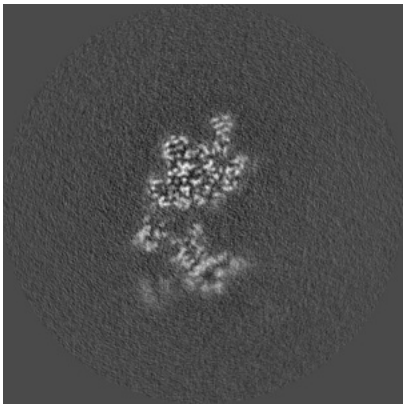
6.3.1 Primary map

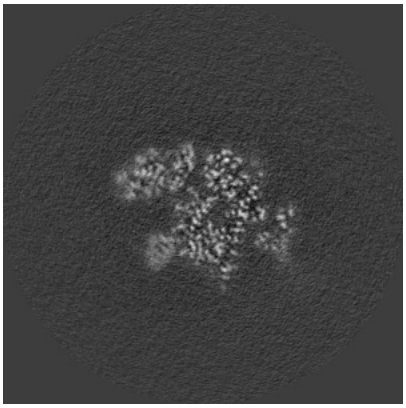
X Index: 146

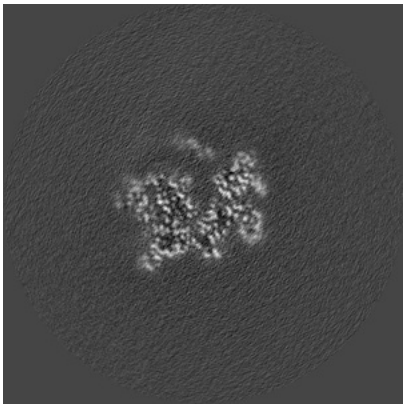

Y Index: 173


Z Index: 204

6.3.2 Raw map


X Index: 196

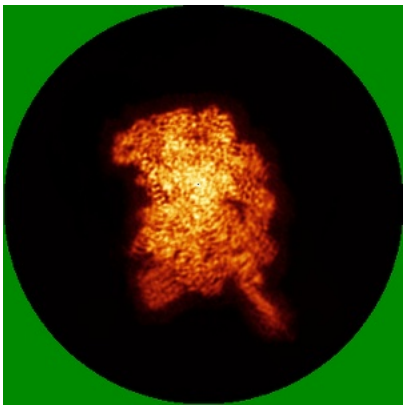

Y Index: 173

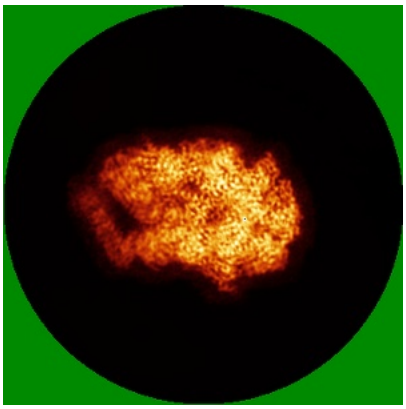

Z Index: 204

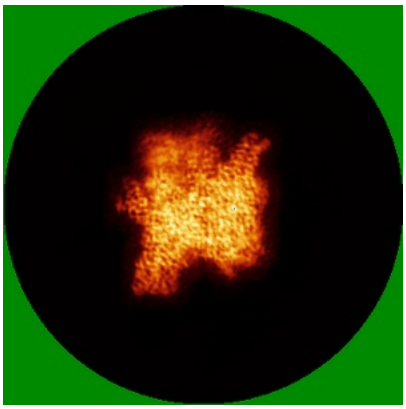
The images above show the largest variance slices of the map in three orthogonal directions.

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

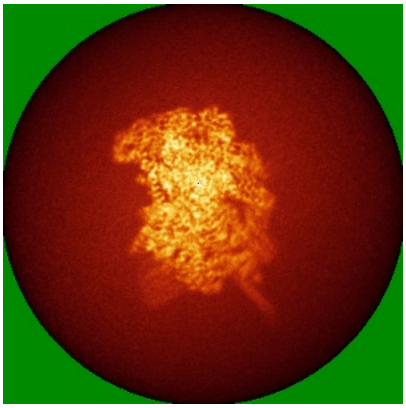
6.4.1 Primary map

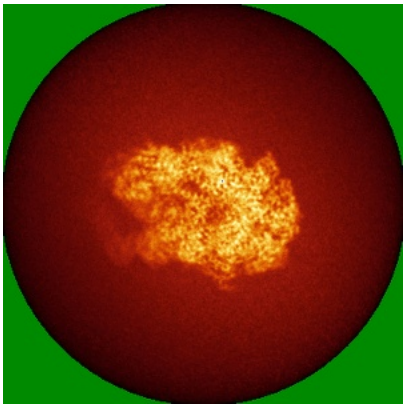

X

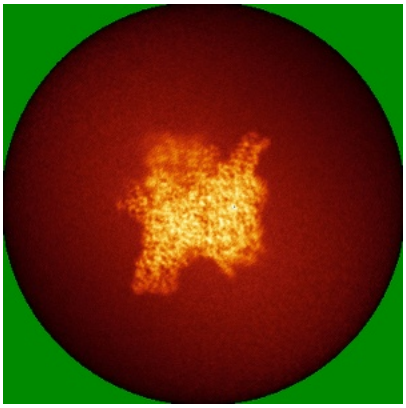

Y


Z

6.4.2 Raw map


X

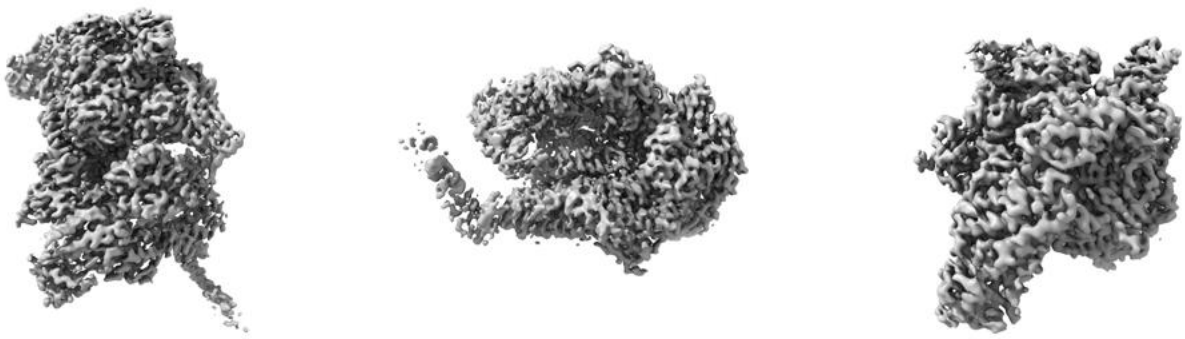

Y


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 0.17. 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

X Y Z

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.

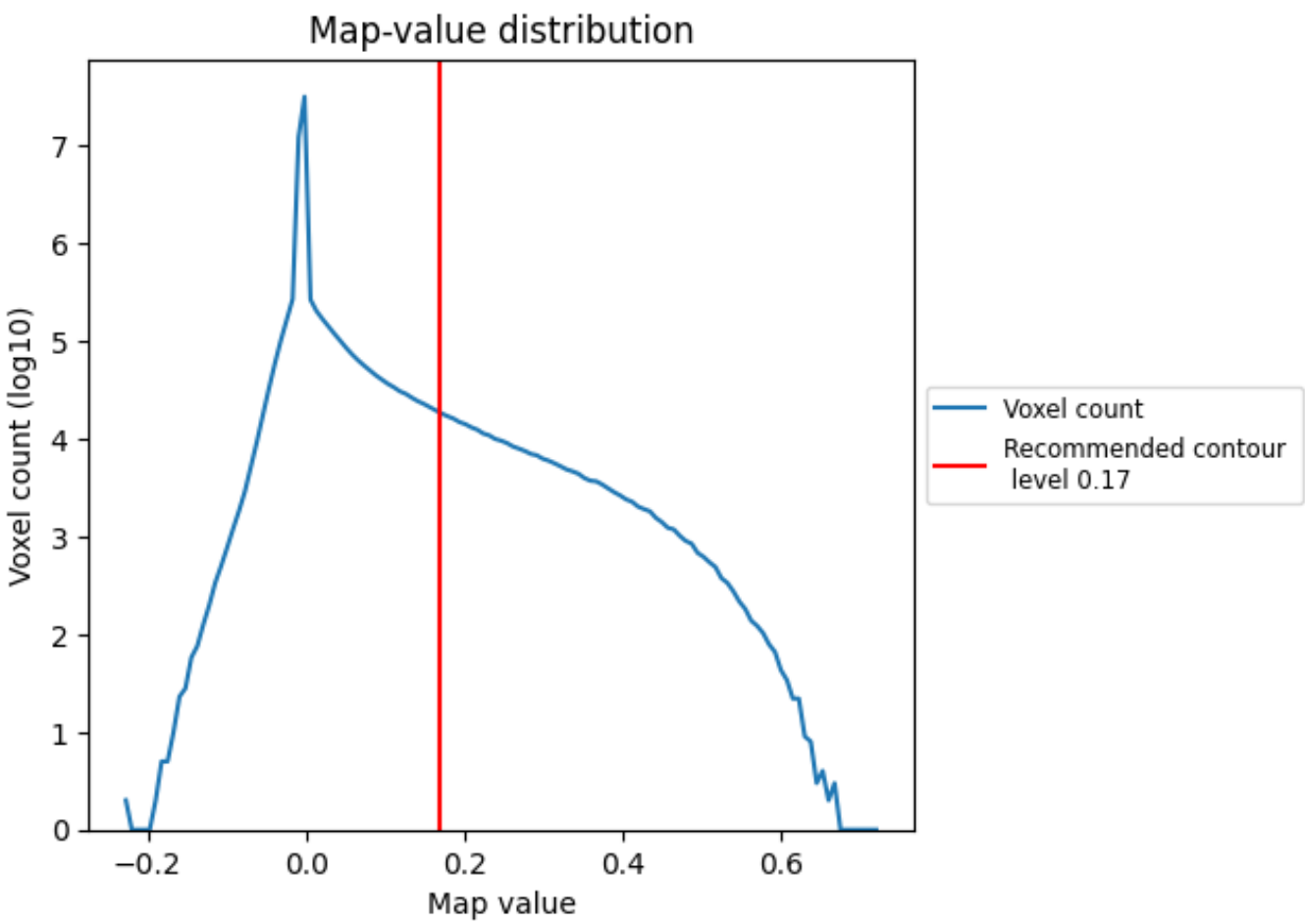
6.6 Mask visualisation (i)

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

7 Map analysis (i)

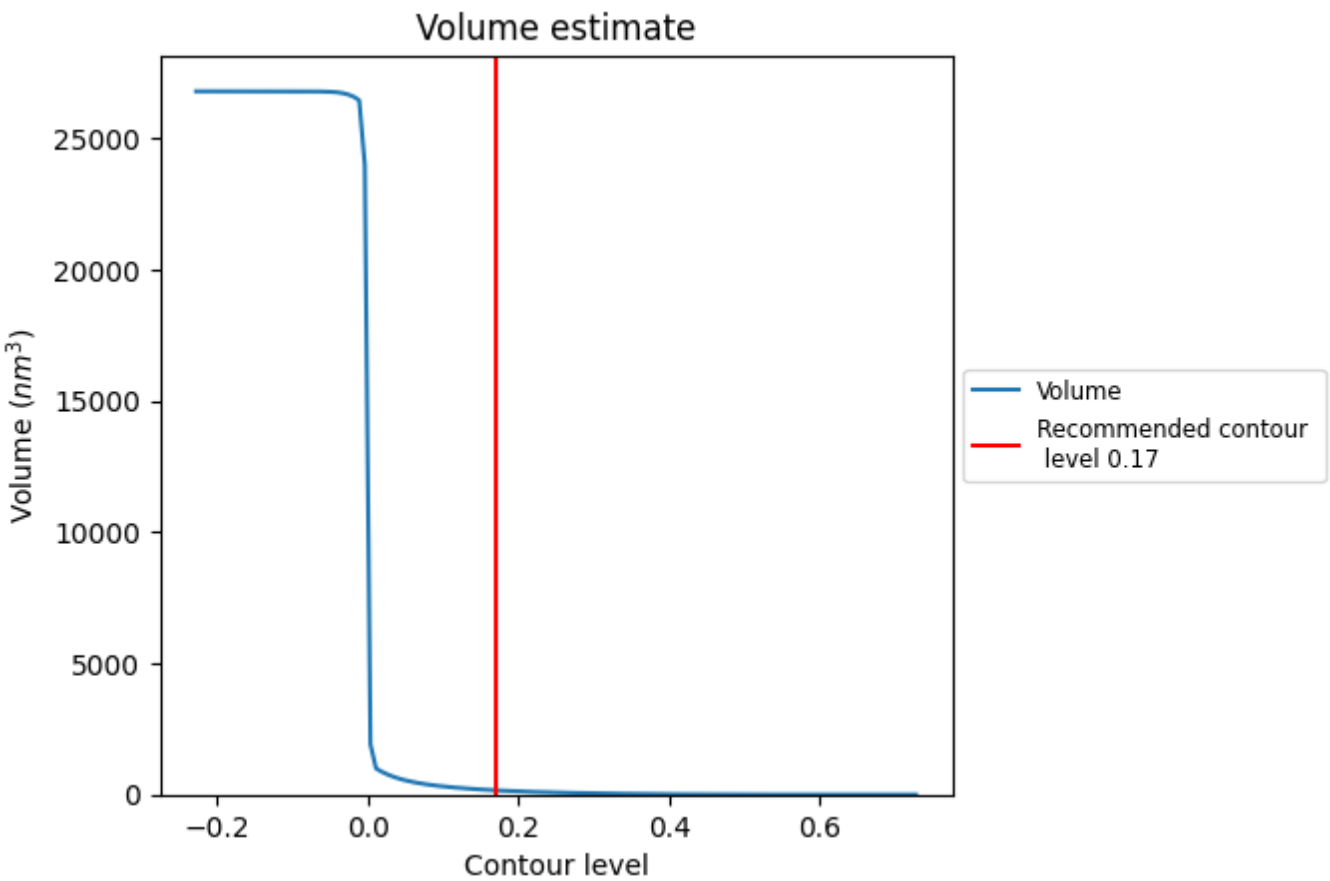
This section contains the results of statistical analysis of the map.

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.

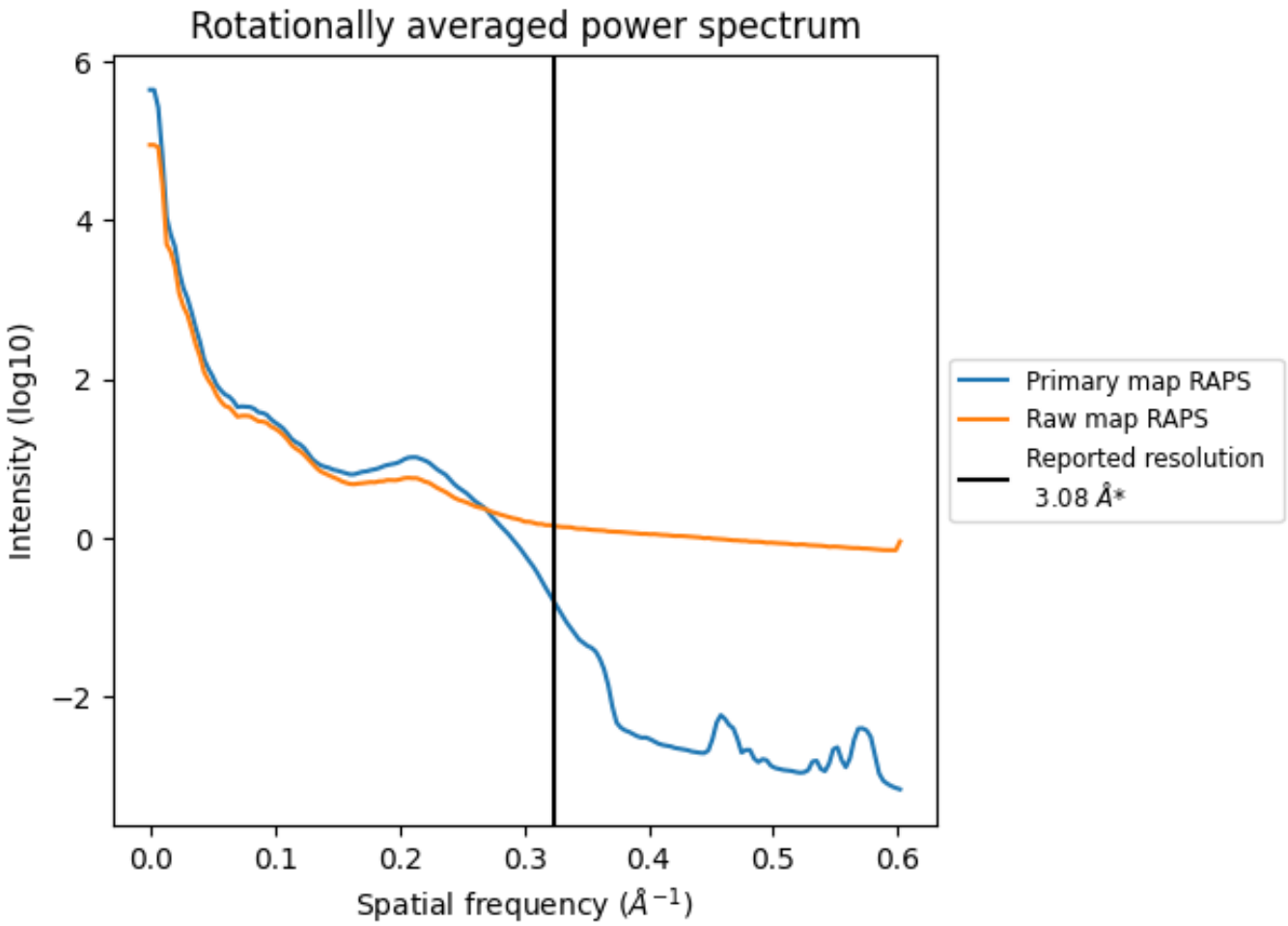
7.2 Volume estimate (i)



The volume at the recommended contour level is 161 nm^3; this corresponds to an approximate mass of 146 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)

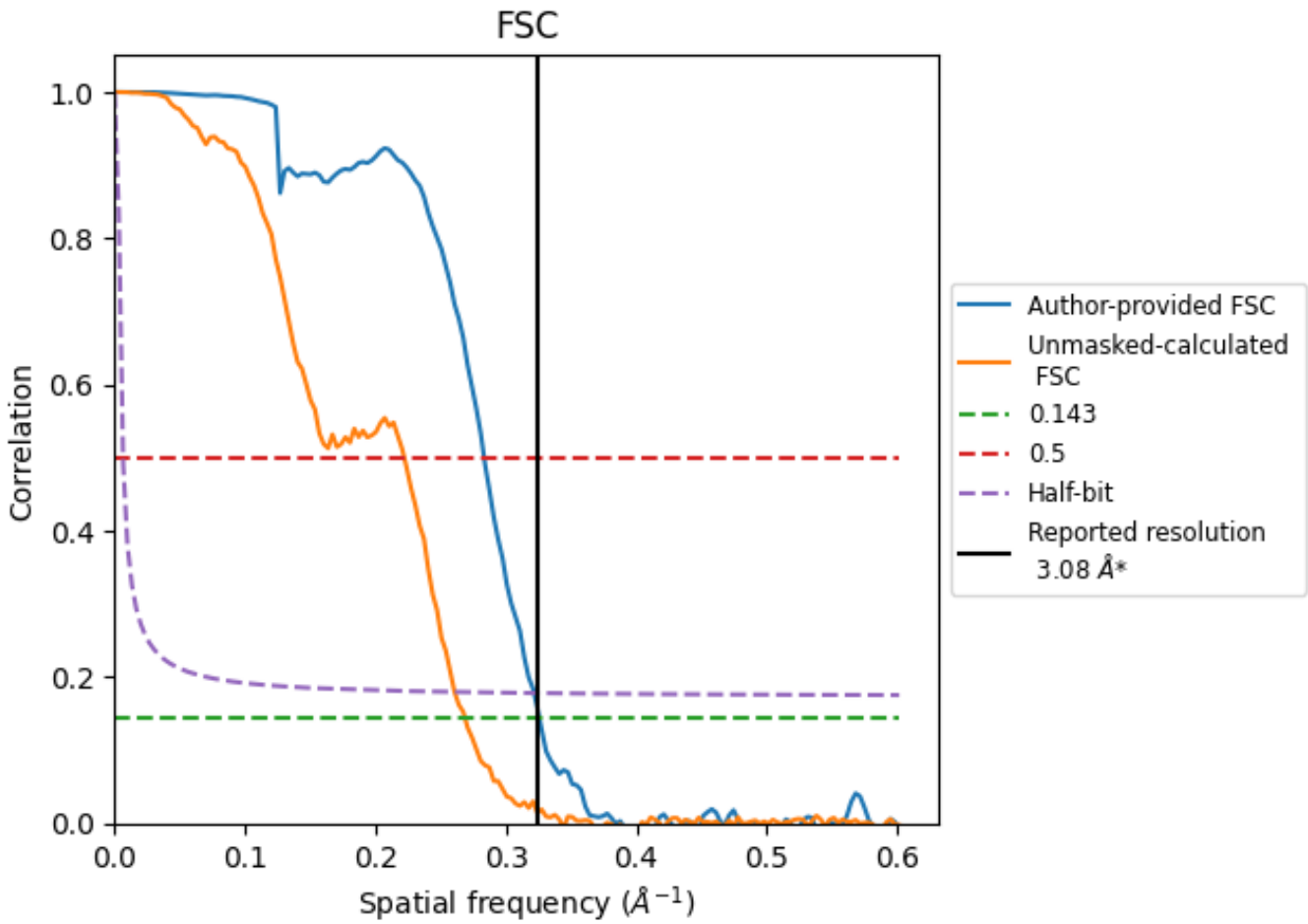


*Reported resolution corresponds to spatial frequency of 0.325 Å^{-1}

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 0.325 Å^{-1}

8.2 Resolution estimates (i)

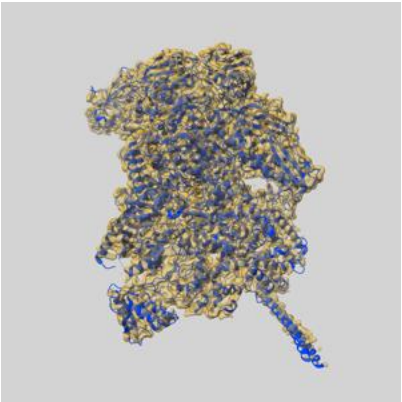
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.08	-	-
Author-provided FSC curve	3.07	3.53	3.11
Unmasked-calculated*	3.72	4.50	3.84

*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 3.72 differs from the reported value 3.08 by more than 10 %

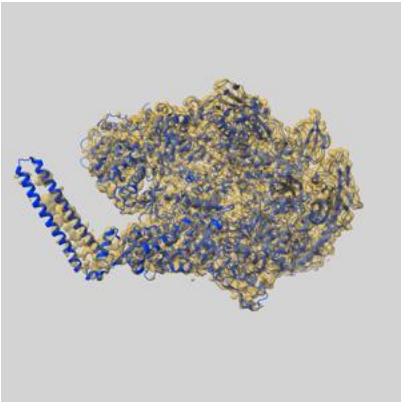
9 Map-model fit (i)

This section contains information regarding the fit between EMDB map EMD-11004 and PDB model 6YYS. Per-residue inclusion information can be found in section 3 on page 5.

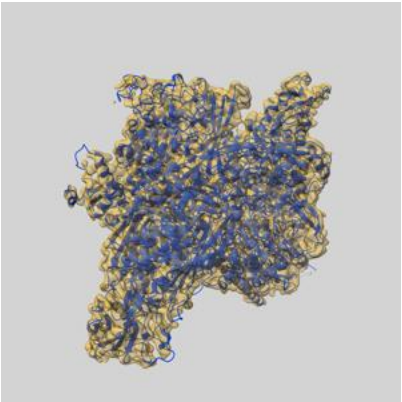
9.1 Map-model overlay (i)



X



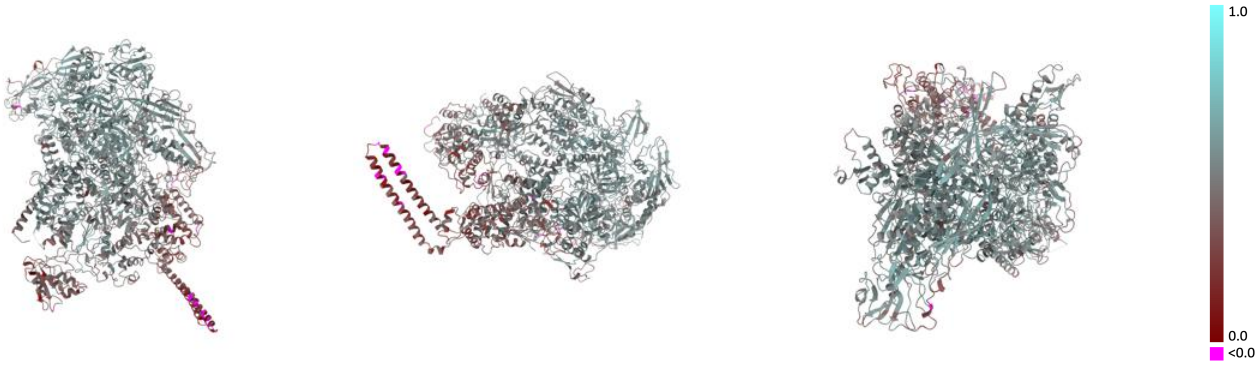
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.17 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.

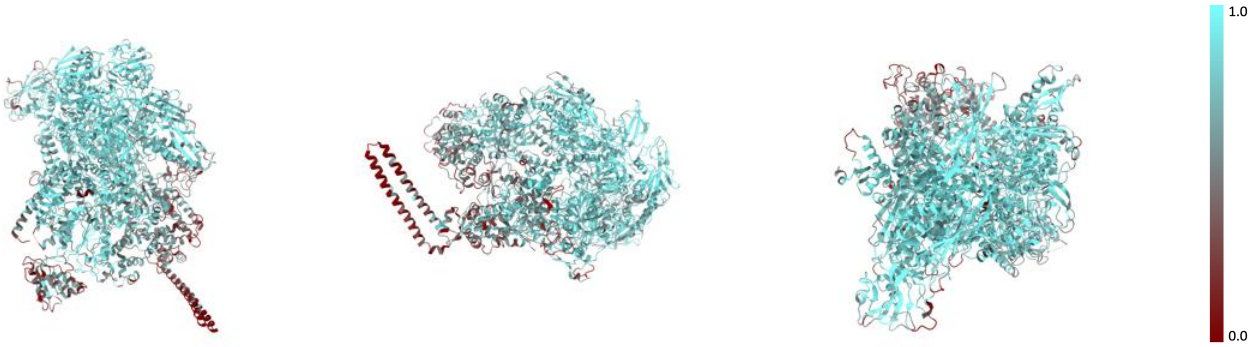
9.2 Q-score mapped to coordinate model (i)



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.

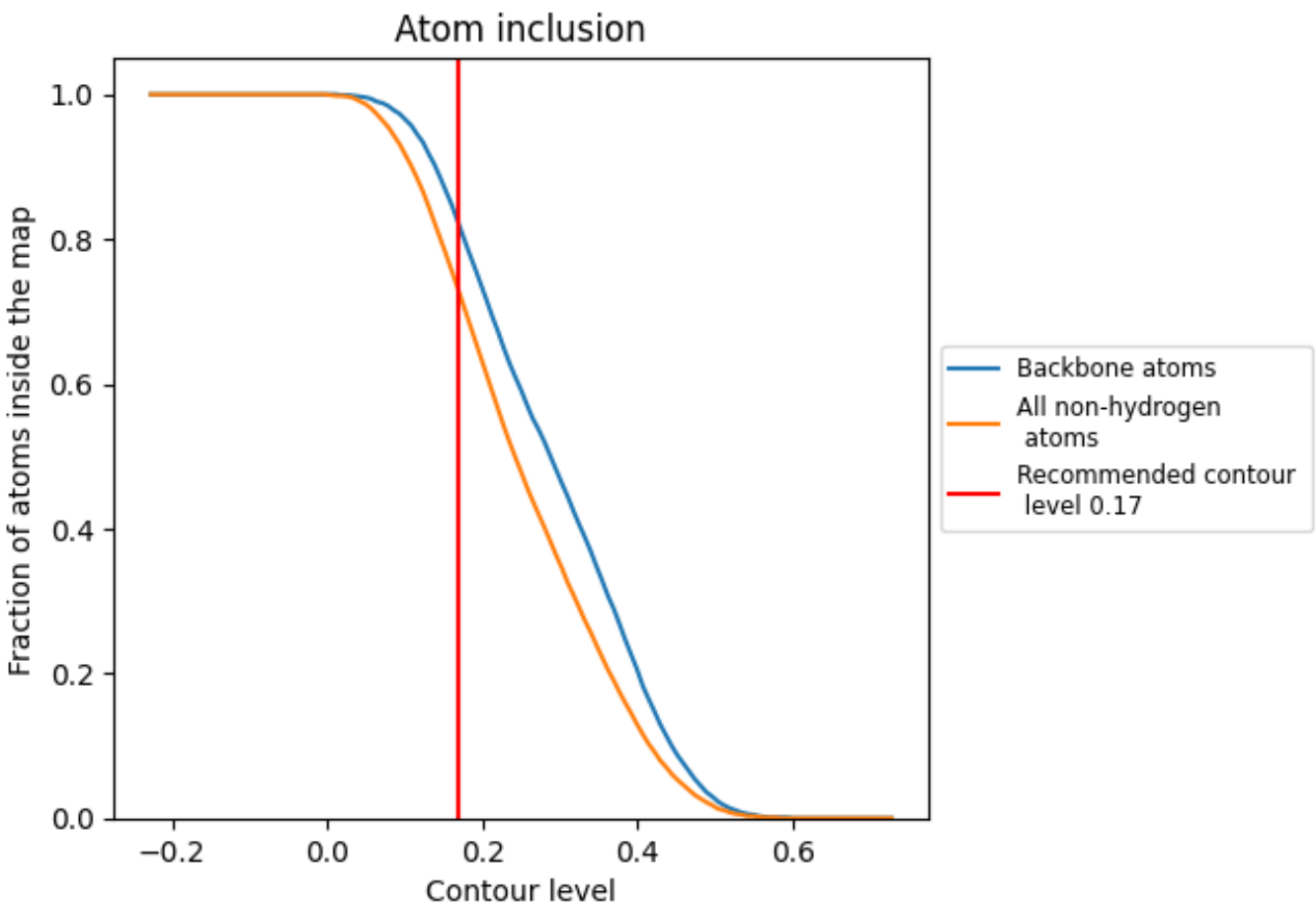
9.3 Atom inclusion mapped to coordinate model (i)



X Y Z

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 (0.17).

9.4 Atom inclusion (i)



At the recommended contour level, 82% of all backbone atoms, 73% of all non-hydrogen atoms, are inside the map.

9.5 Map-model fit summary (i)

The table lists the average atom inclusion at the recommended contour level (0.17) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.7270	0.4840
A	0.8220	0.5320
B	0.7700	0.5080
C	0.8110	0.5280
D	0.6770	0.4530
E	0.7370	0.5040
H	0.6420	0.4500