wwPDB EM Validation Summary Report

Dec 29, 2024 – 10:09 PM EST

PDB ID	:	7YG1
EMDB ID	:	EMD-33804
Title	:	Cryo-EM structure of the C-terminal domain of the human sodium-chloride cotransporter
Authors	:	Nan, J.; Yang, X.M.; Shan, Z.Y.; Yuan, Y.F.; Zhang, Y.Q.
Deposited on	:	2022-07-09
Resolution	:	3.77 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:

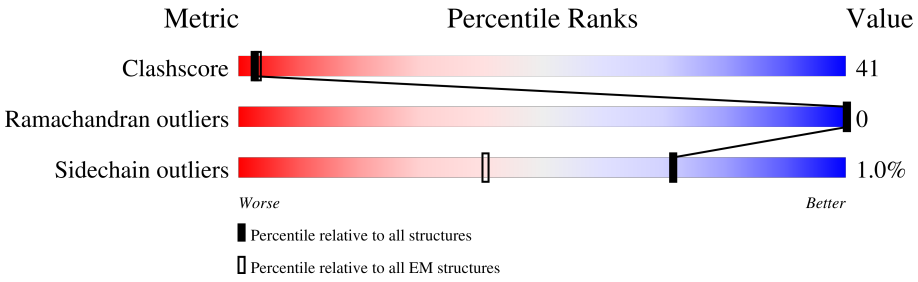
EMDB validation analysis	:	0.0.1.dev113
MolProbity	:	4.02b-467
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
MapQ	:	1.9.13
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	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 3.77 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 >=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	1053	
1	B	1053	

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 6208 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 Solute carrier family 12 member 3.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	385	Total	C	N	O	S	0	0
			3109	1984	559	547	19		
1	B	385	Total	C	N	O	S	0	0
			3099	1977	558	545	19		

There are 70 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	GLN	deletion	UNP P55017
A	264	GLY	ALA	engineered mutation	UNP P55017
A	1022	LEU	-	expression tag	UNP P55017
A	1023	GLU	-	expression tag	UNP P55017
A	1024	GLY	-	expression tag	UNP P55017
A	1025	SER	-	expression tag	UNP P55017
A	1026	ASP	-	expression tag	UNP P55017
A	1027	GLU	-	expression tag	UNP P55017
A	1028	VAL	-	expression tag	UNP P55017
A	1029	ASP	-	expression tag	UNP P55017
A	1030	ALA	-	expression tag	UNP P55017
A	1031	GLY	-	expression tag	UNP P55017
A	1032	SER	-	expression tag	UNP P55017
A	1033	HIS	-	expression tag	UNP P55017
A	1034	HIS	-	expression tag	UNP P55017
A	1035	HIS	-	expression tag	UNP P55017
A	1036	HIS	-	expression tag	UNP P55017
A	1037	HIS	-	expression tag	UNP P55017
A	1038	HIS	-	expression tag	UNP P55017
A	1039	HIS	-	expression tag	UNP P55017
A	1040	HIS	-	expression tag	UNP P55017
A	1041	HIS	-	expression tag	UNP P55017
A	1042	HIS	-	expression tag	UNP P55017
A	1043	GLY	-	expression tag	UNP P55017
A	1044	SER	-	expression tag	UNP P55017
A	1045	VAL	-	expression tag	UNP P55017

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	1046	GLU	-	expression tag	UNP P55017
A	1047	ASP	-	expression tag	UNP P55017
A	1048	TYR	-	expression tag	UNP P55017
A	1049	LYS	-	expression tag	UNP P55017
A	1050	ASP	-	expression tag	UNP P55017
A	1051	ASP	-	expression tag	UNP P55017
A	1052	ASP	-	expression tag	UNP P55017
A	1053	ASP	-	expression tag	UNP P55017
A	1054	LYS	-	expression tag	UNP P55017
B	?	-	GLN	deletion	UNP P55017
B	264	GLY	ALA	engineered mutation	UNP P55017
B	1022	LEU	-	expression tag	UNP P55017
B	1023	GLU	-	expression tag	UNP P55017
B	1024	GLY	-	expression tag	UNP P55017
B	1025	SER	-	expression tag	UNP P55017
B	1026	ASP	-	expression tag	UNP P55017
B	1027	GLU	-	expression tag	UNP P55017
B	1028	VAL	-	expression tag	UNP P55017
B	1029	ASP	-	expression tag	UNP P55017
B	1030	ALA	-	expression tag	UNP P55017
B	1031	GLY	-	expression tag	UNP P55017
B	1032	SER	-	expression tag	UNP P55017
B	1033	HIS	-	expression tag	UNP P55017
B	1034	HIS	-	expression tag	UNP P55017
B	1035	HIS	-	expression tag	UNP P55017
B	1036	HIS	-	expression tag	UNP P55017
B	1037	HIS	-	expression tag	UNP P55017
B	1038	HIS	-	expression tag	UNP P55017
B	1039	HIS	-	expression tag	UNP P55017
B	1040	HIS	-	expression tag	UNP P55017
B	1041	HIS	-	expression tag	UNP P55017
B	1042	HIS	-	expression tag	UNP P55017
B	1043	GLY	-	expression tag	UNP P55017
B	1044	SER	-	expression tag	UNP P55017
B	1045	VAL	-	expression tag	UNP P55017
B	1046	GLU	-	expression tag	UNP P55017
B	1047	ASP	-	expression tag	UNP P55017
B	1048	TYR	-	expression tag	UNP P55017
B	1049	LYS	-	expression tag	UNP P55017
B	1050	ASP	-	expression tag	UNP P55017
B	1051	ASP	-	expression tag	UNP P55017
B	1052	ASP	-	expression tag	UNP P55017

Continued on next page...

Continued from previous page...

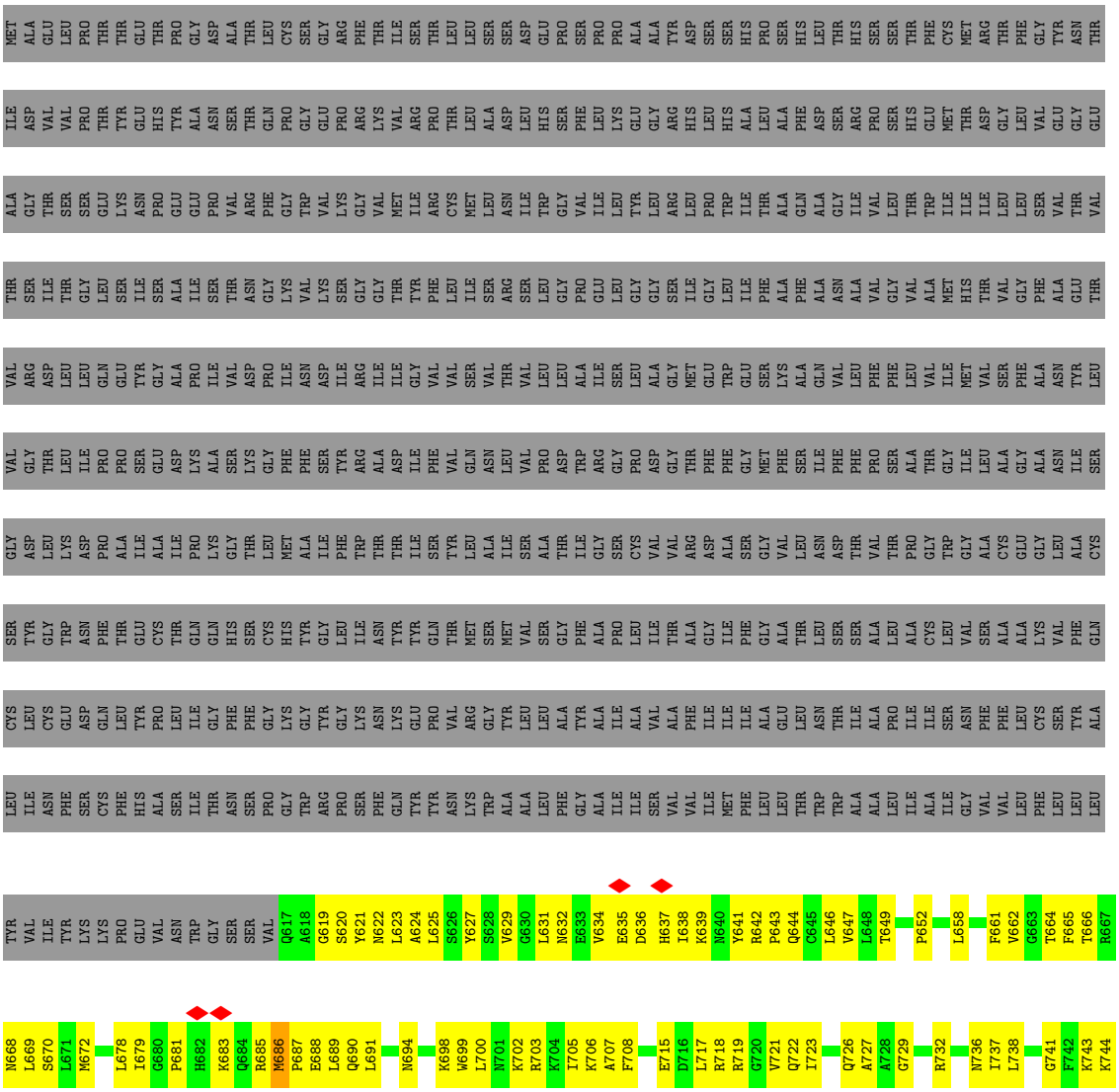
Chain	Residue	Modelled	Actual	Comment	Reference
B	1053	ASP	-	expression tag	UNP P55017
B	1054	LYS	-	expression tag	UNP P55017

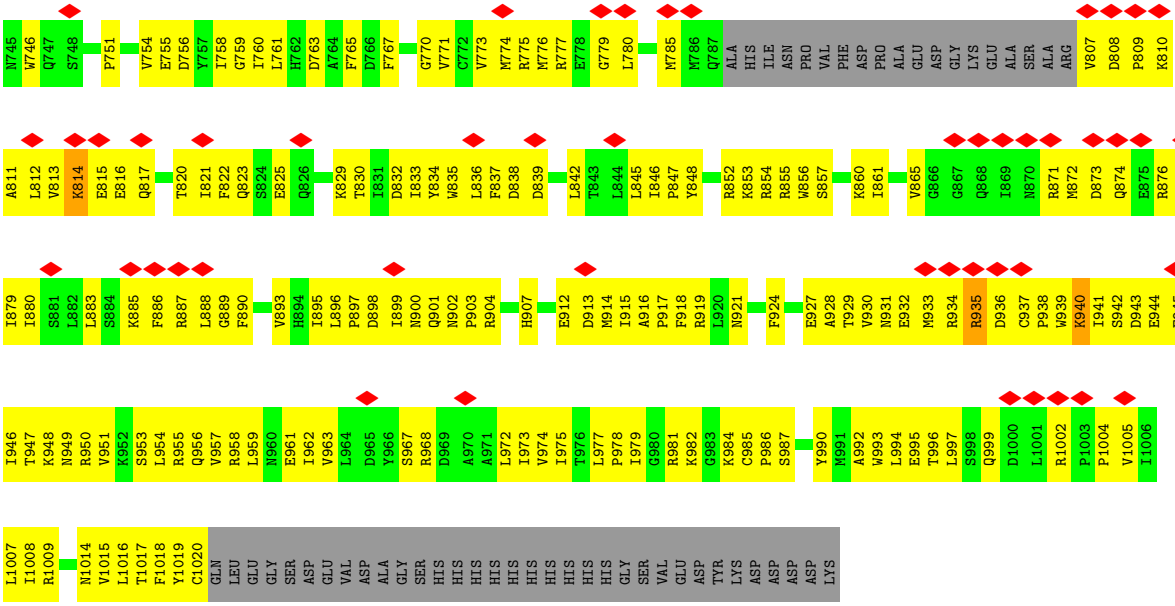
3 Residue-property plots (i)

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: Solute carrier family 12 member 3

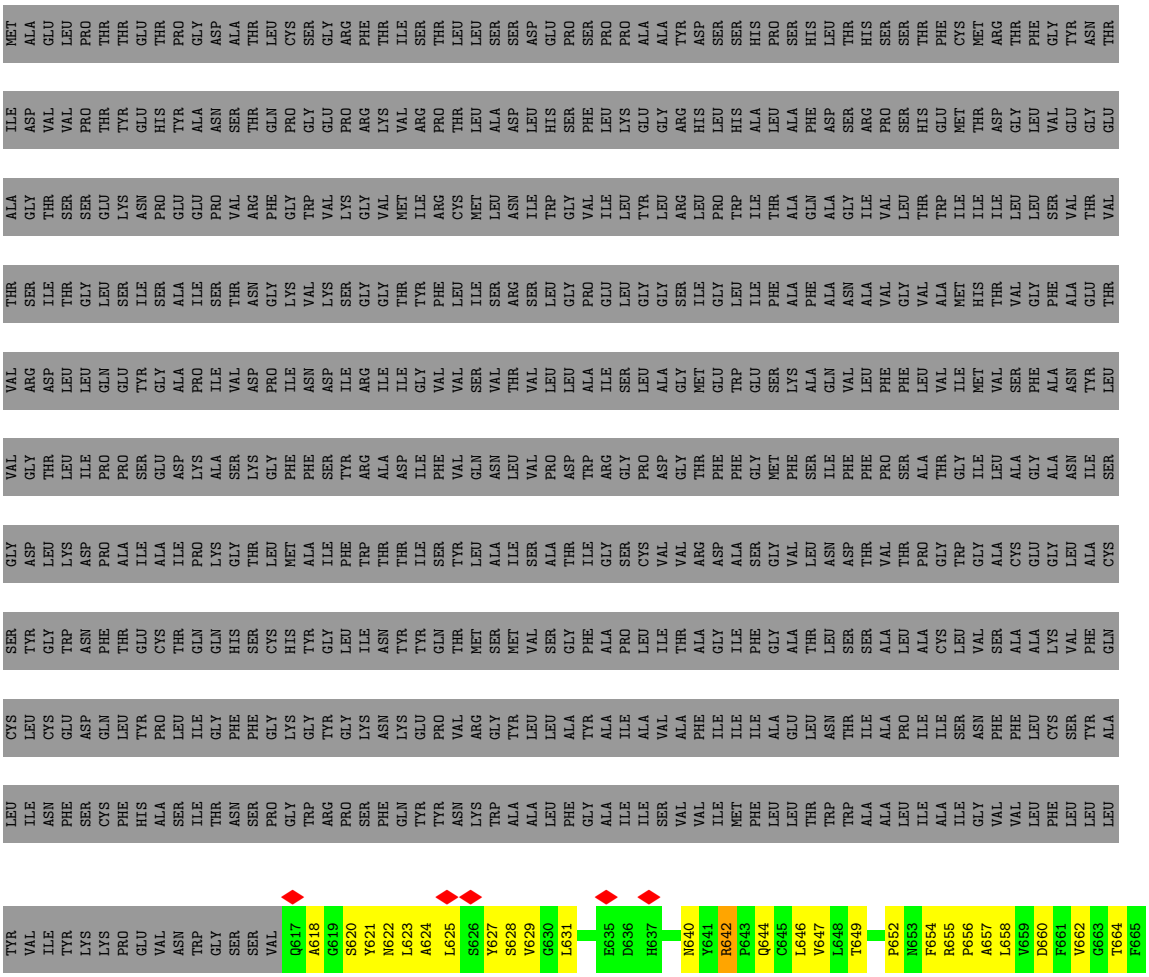
Chain A:

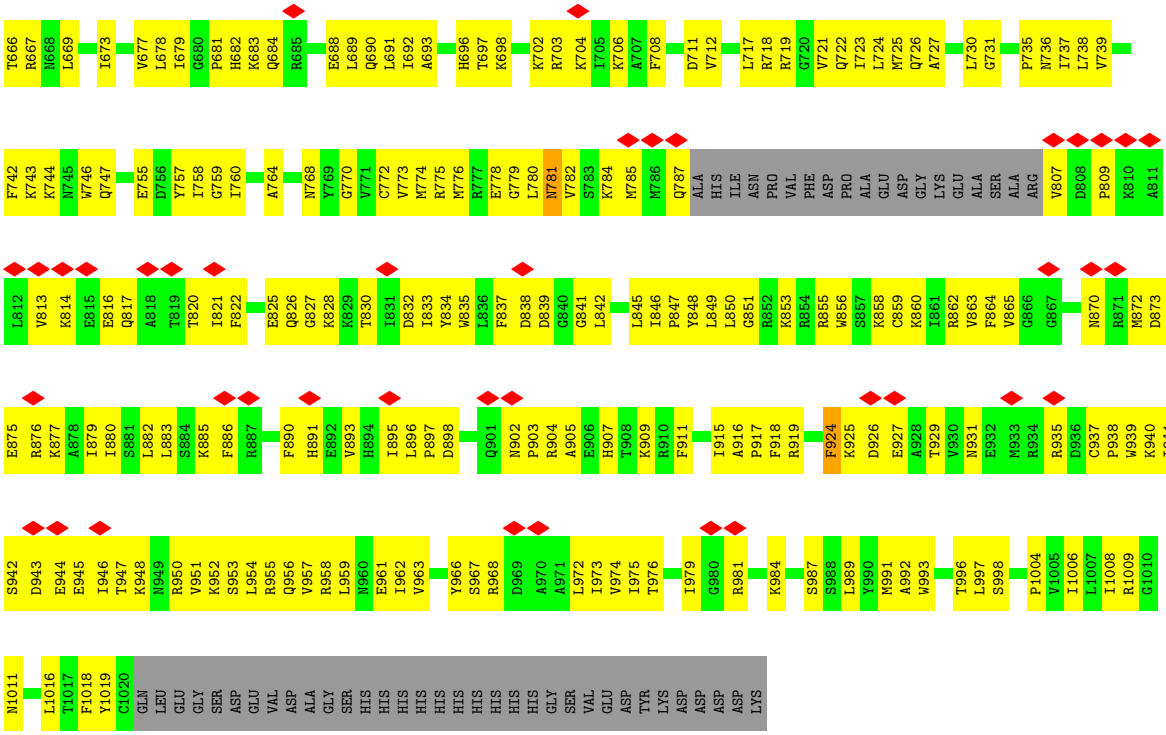




• Molecule 1: Solute carrier family 12 member 3

Chain B:




T666 R667 N668 L669 I673 V677 L678 I679 G680 P681 H682 K683 Q684 R685 E688 L689 Q690 L691 I692 A693 H696 T697 K698 K702 R703 K704 I705 K706 A707 F708 D711 V712 L717 R718 R719 G720 V721 Q722 I723 L724 M725 Q726 A727 L730 G731 P735 N736 I737 L738 V739
F742 K743 K744 N745 W746 Q747 E755 D756 Y757 I758 G759 I760 A764 N768 Y769 G770 V771 C772 V773 M774 R775 M776 R777 E778 G779 L780 N781 V782 S783 K784 M785 M786 Q787 ALA HIS ILE ASN PRO VAL PHE ASP PRO ALA GLU ASP GLY LYS GLU ALA SER ALA ARG V807 D808 P809 K810 A811
L812 V813 K814 E815 E816 Q817 A818 T819 T820 I821 F822 E825 Q826 G827 K828 K829 T830 I831 D832 I833 Y834 W835 L836 F837 D838 D839 G840 G841 L842 L845 I846 P847 Y848 L849 L850 C851 R852 K853 R854 R855 W856 S857 K858 C859 K860 I861 R862 V863 F864 V865 G866 G867 N870 R871 M872 D873 Q874
E875 R876 K877 A878 I879 I880 S881 L882 L883 S884 K885 F886 R887 F890 H891 E892 V893 H894 I895 L896 P897 D898 Q901 N902 P903 R904 A905 E906 H907 T908 K909 R910 F911 I915 A916 P917 F918 R919 F924 K925 D926 E927 A928 T929 V930 N931 E932 M933 R934 R935 D936 C937 P938 W939 K940 I941
S942 D943 E944 E945 I946 T947 K948 N949 R950 V951 K952 S953 L954 R955 Q956 V957 R958 L959 N960 E961 I962 V963 Y966 S967 R968 D969 A970 A971 L972 I973 V974 I975 T976 I979 G980 R981 K984 S987 S988 L989 Y990 M991 A992 W993 T996 L997 S998 P1004 V1005 I1006 L1007 I1008 R1009 G1010
N1011 L1016 T1017 F1018 Y1019 C1020 GLN LEU GLU GLY SER ASP GLU VAL ASP ALA GLY SER HIS HIS HIS HIS HIS HIS HIS HIS HIS HIS GLY SER VAL GLU ASP TYR LYS ASP ASP ASP LYS

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	79225	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	NONE	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	52.8	Depositor
Minimum defocus (nm)	1500	Depositor
Maximum defocus (nm)	2300	Depositor
Magnification	81000	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	0.721	Depositor
Minimum map value	-0.434	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.023	Depositor
Recommended contour level	0.12	Depositor
Map size (Å)	267.776, 267.776, 267.776	wwPDB
Map dimensions	256, 256, 256	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.046, 1.046, 1.046	Depositor

5 Model quality

5.1 Standard geometry

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$\|Z\|>5$	RMSZ	#$\|Z\|>5$
1	A	0.28	0/3173	0.51	0/4281
1	B	0.27	0/3162	0.56	0/4266
All	All	0.27	0/6335	0.53	0/8547

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

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	3109	0	3177	274	0
1	B	3099	0	3166	256	0
All	All	6208	0	6343	511	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 41.

The worst 5 of 511 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:1005:VAL:HG13	1:A:1007:LEU:HD23	1.35	1.02
1:A:700:LEU:HG	1:A:705:ILE:HG21	1.45	0.99

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:641:TYR:H	1:A:1017:THR:HG21	1.32	0.92
1:A:924:PHE:HB3	1:A:927:GLU:HB3	1.53	0.91
1:B:946:ILE:HG12	1:B:989:LEU:HD13	1.54	0.90

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	381/1053 (36%)	318 (84%)	63 (16%)	0	100	100
1	B	381/1053 (36%)	337 (88%)	44 (12%)	0	100	100
All	All	762/2106 (36%)	655 (86%)	107 (14%)	0	100	100

There are no Ramachandran outliers to report.

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	340/890 (38%)	336 (99%)	4 (1%)	67	79
1	B	338/890 (38%)	335 (99%)	3 (1%)	75	83
All	All	678/1780 (38%)	671 (99%)	7 (1%)	71	81

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	940	LYS
1	B	642	ARG
1	B	924	PHE
1	B	781	ASN
1	A	935	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	999	GLN
1	A	1012	GLN
1	A	1014	ASN
1	A	676	HIS
1	A	644	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.

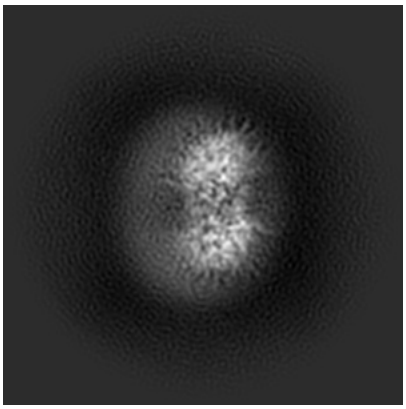
6 Map visualisation

This section contains visualisations of the EMDB entry EMD-33804. These allow visual inspection of the internal detail of the map and identification of artifacts.

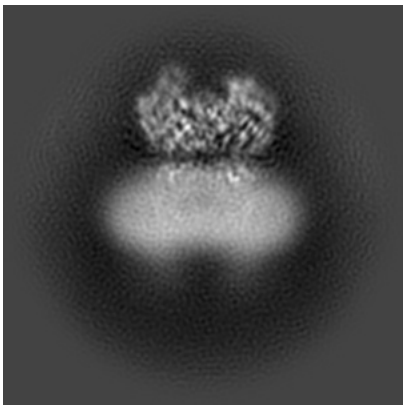
Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

6.1 Orthogonal projections

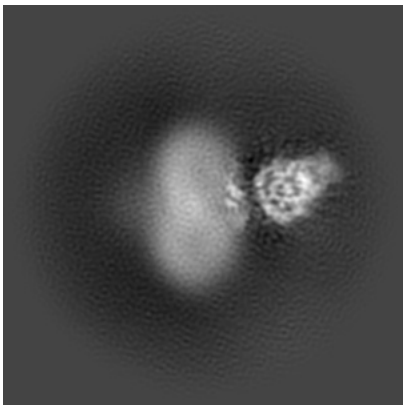
6.1.1 Primary map



X

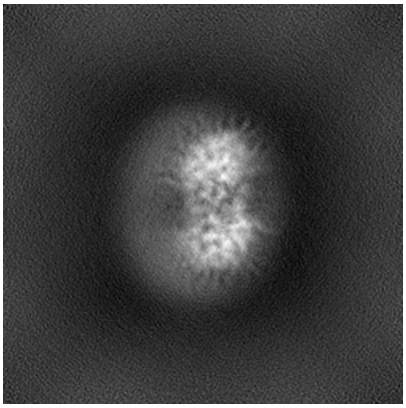


Y

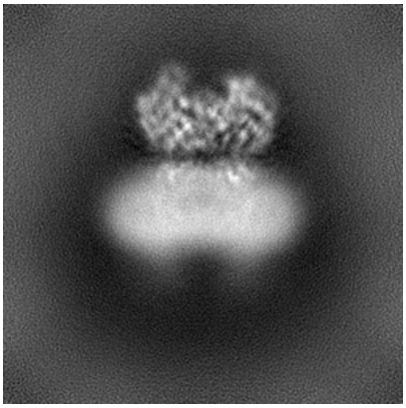


Z

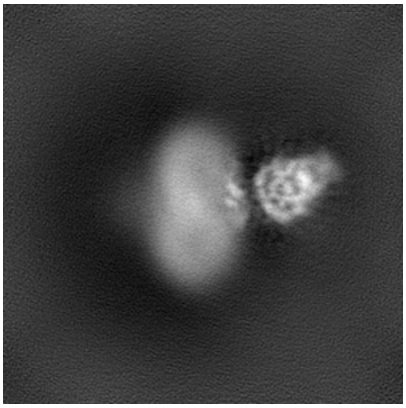
6.1.2 Raw map



X



Y

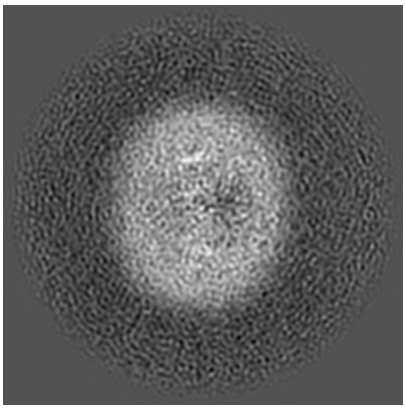


Z

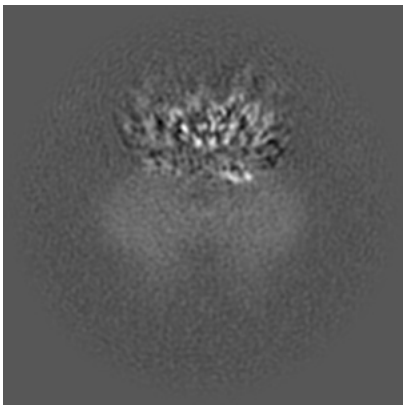
The images above show the map projected in three orthogonal directions.

6.2 Central slices (i)

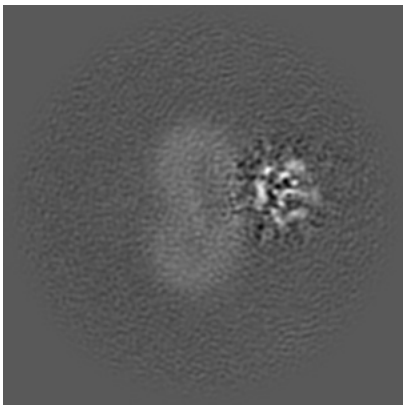
6.2.1 Primary map



X Index: 128

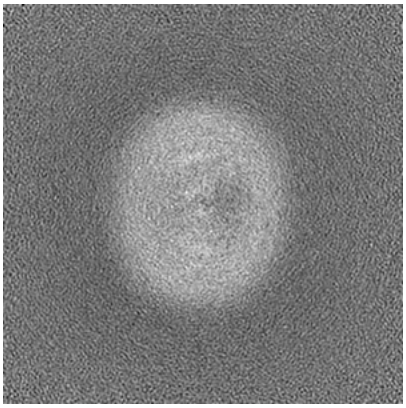


Y Index: 128

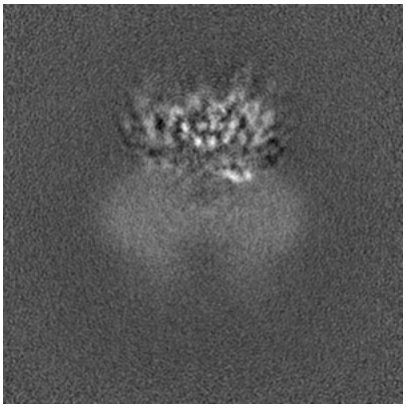


Z Index: 128

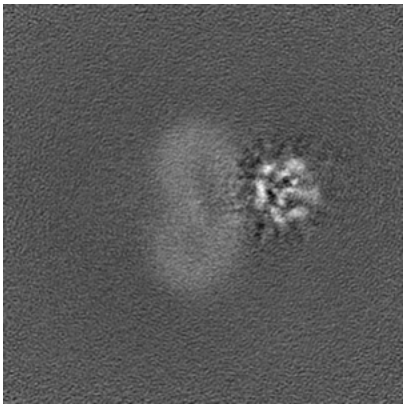
6.2.2 Raw map



X Index: 128



Y Index: 128

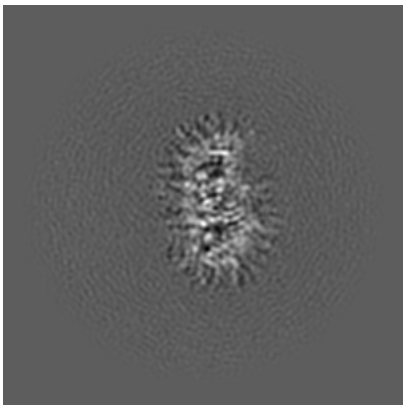


Z Index: 128

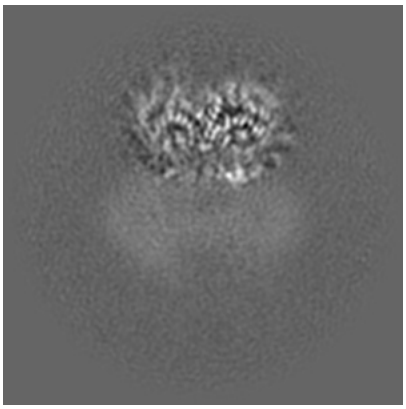
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

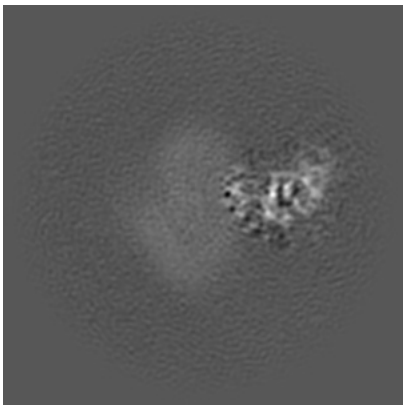
6.3.1 Primary map



X Index: 177

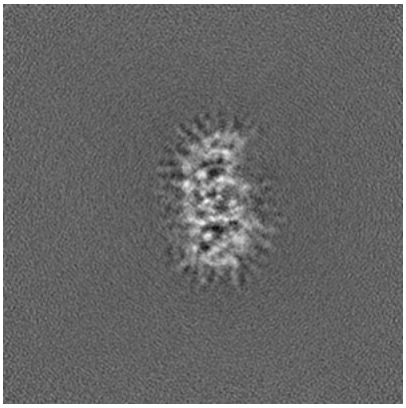


Y Index: 138

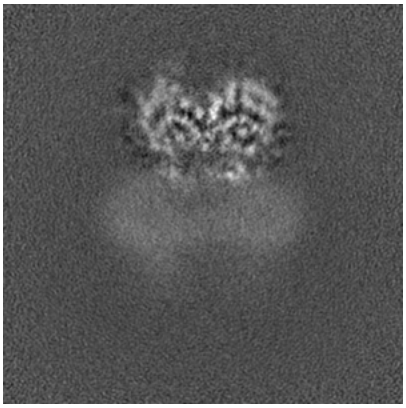


Z Index: 153

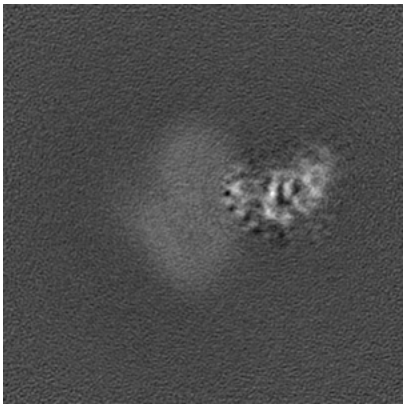
6.3.2 Raw map



X Index: 178



Y Index: 140

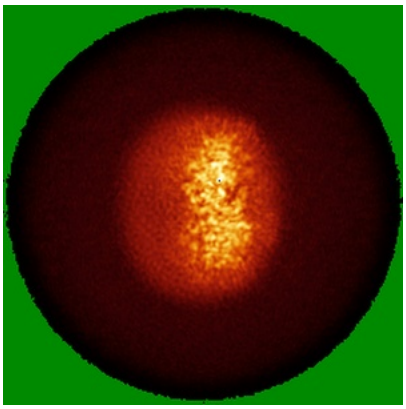


Z Index: 153

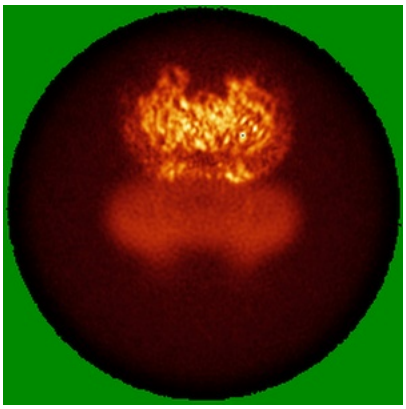
The images above show the largest variance slices of the map in three orthogonal directions.

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

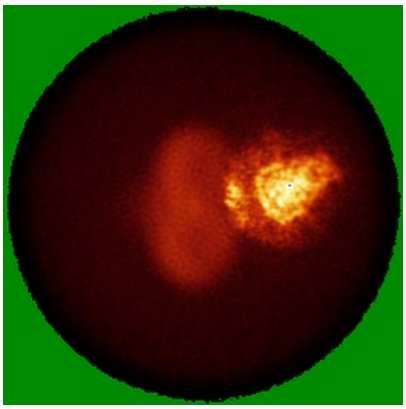
6.4.1 Primary map



X

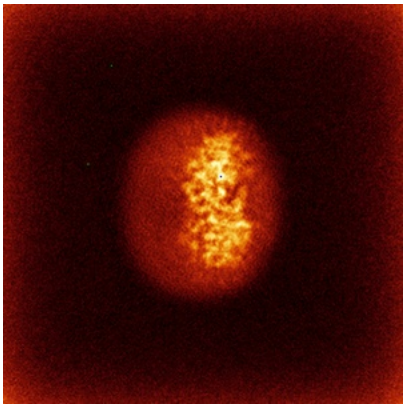


Y

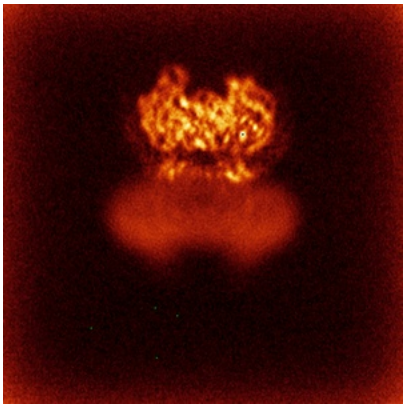


Z

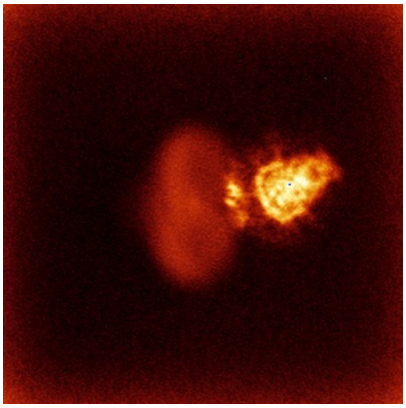
6.4.2 Raw map



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.

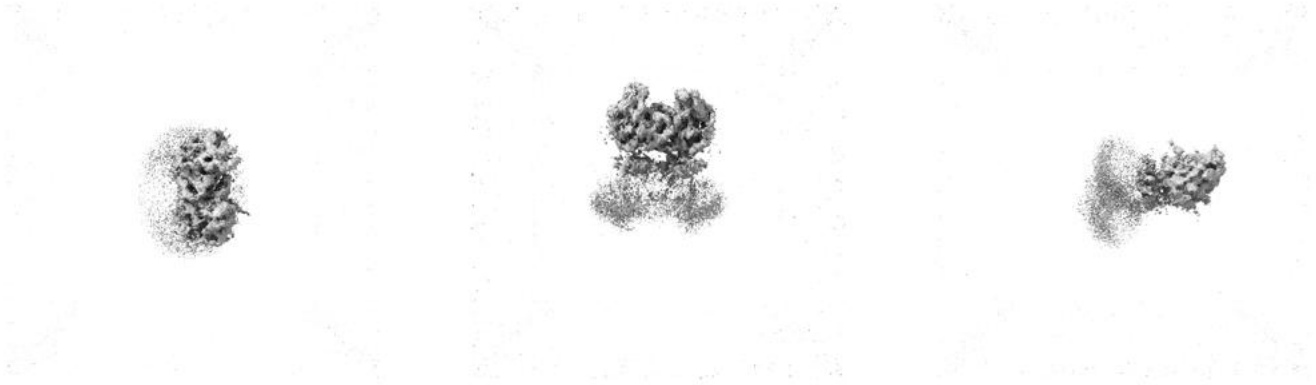
6.5 Orthogonal surface views (i)

6.5.1 Primary map

X Y Z

The images above show the 3D surface view of the map at the recommended contour level 0.12. 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.

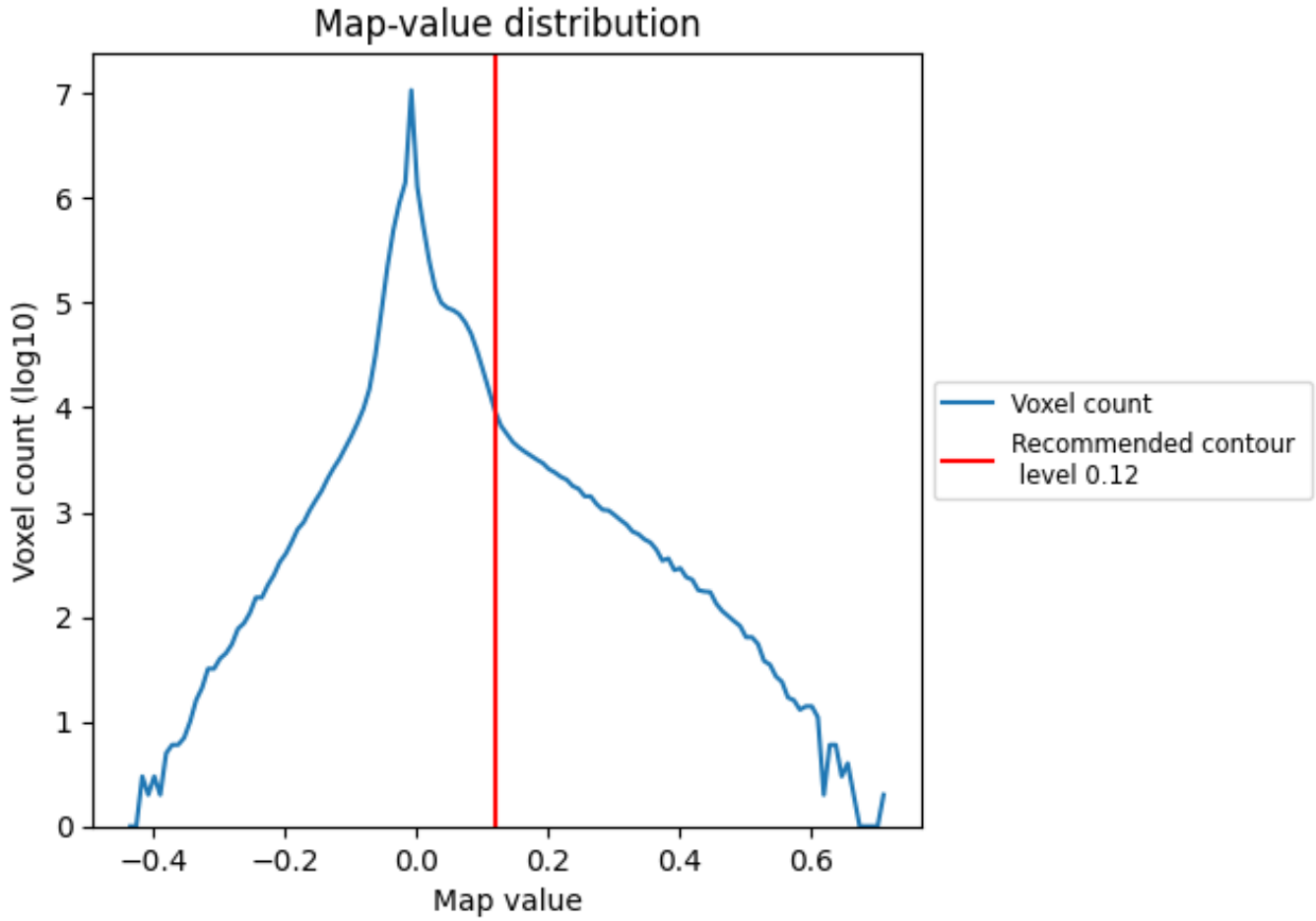
6.6 Mask visualisation (i)

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

7 Map analysis

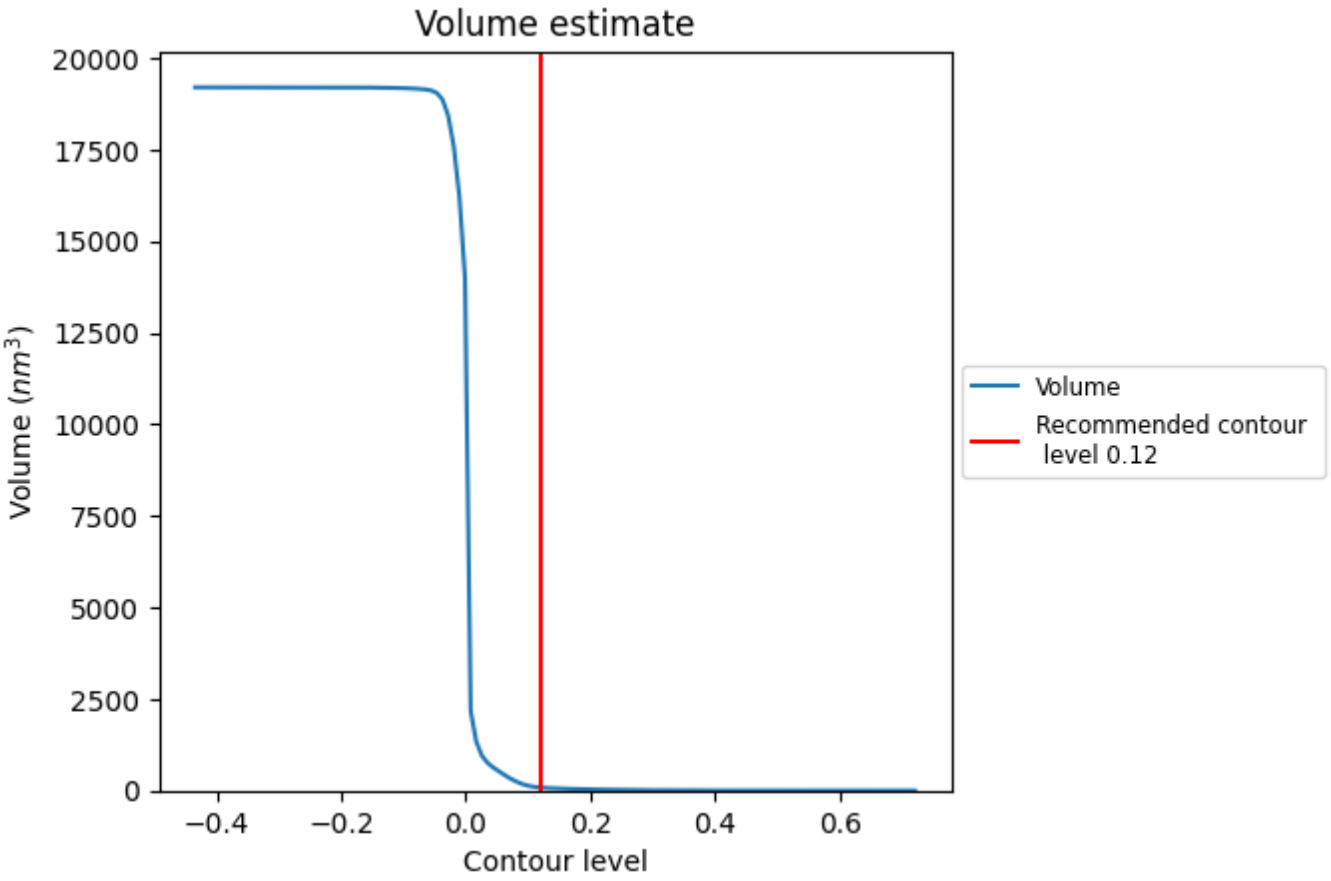
This section contains the results of statistical analysis of the map.

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.

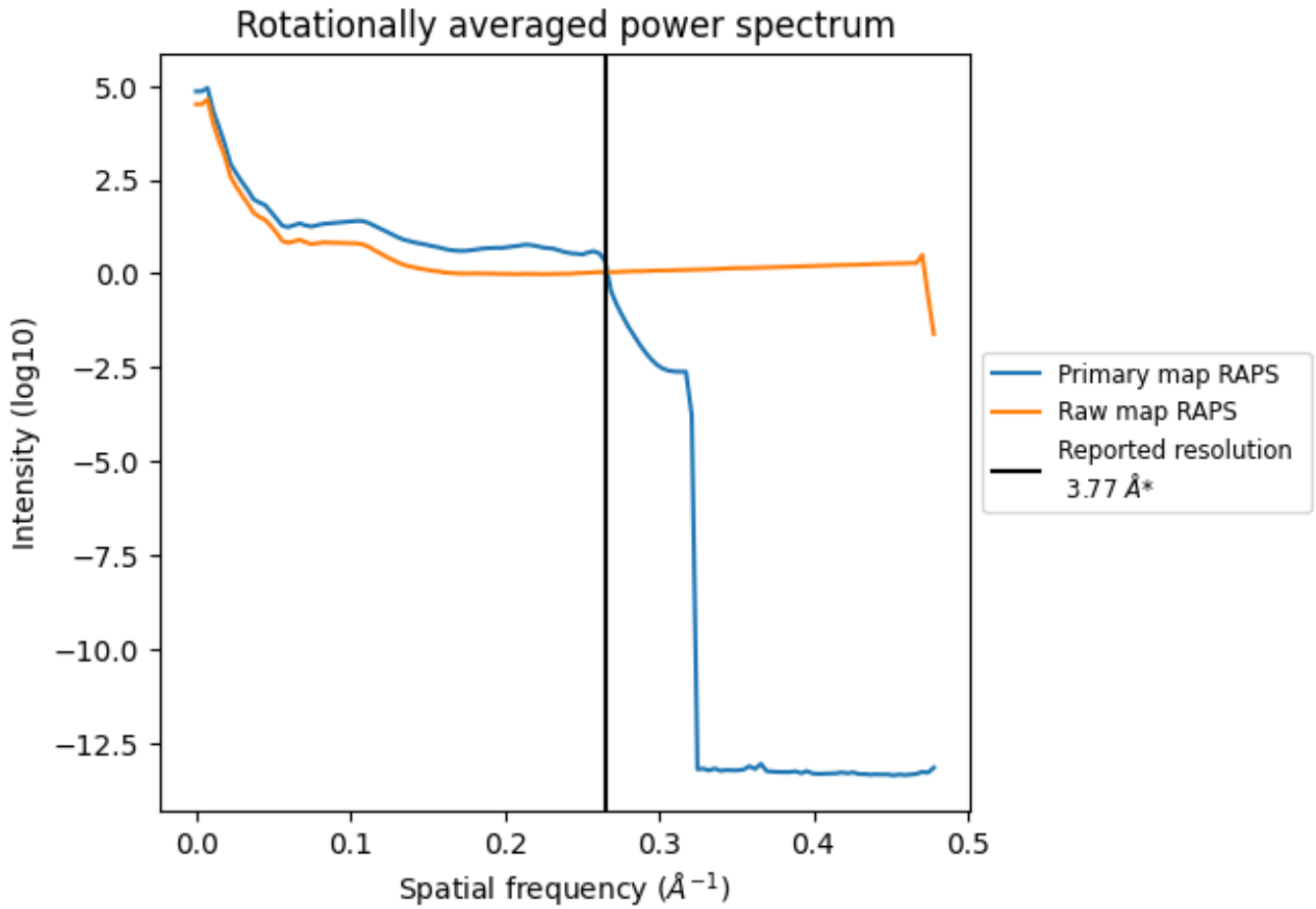
7.2 Volume estimate (i)



The volume at the recommended contour level is 83 nm^3; this corresponds to an approximate mass of 75 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)

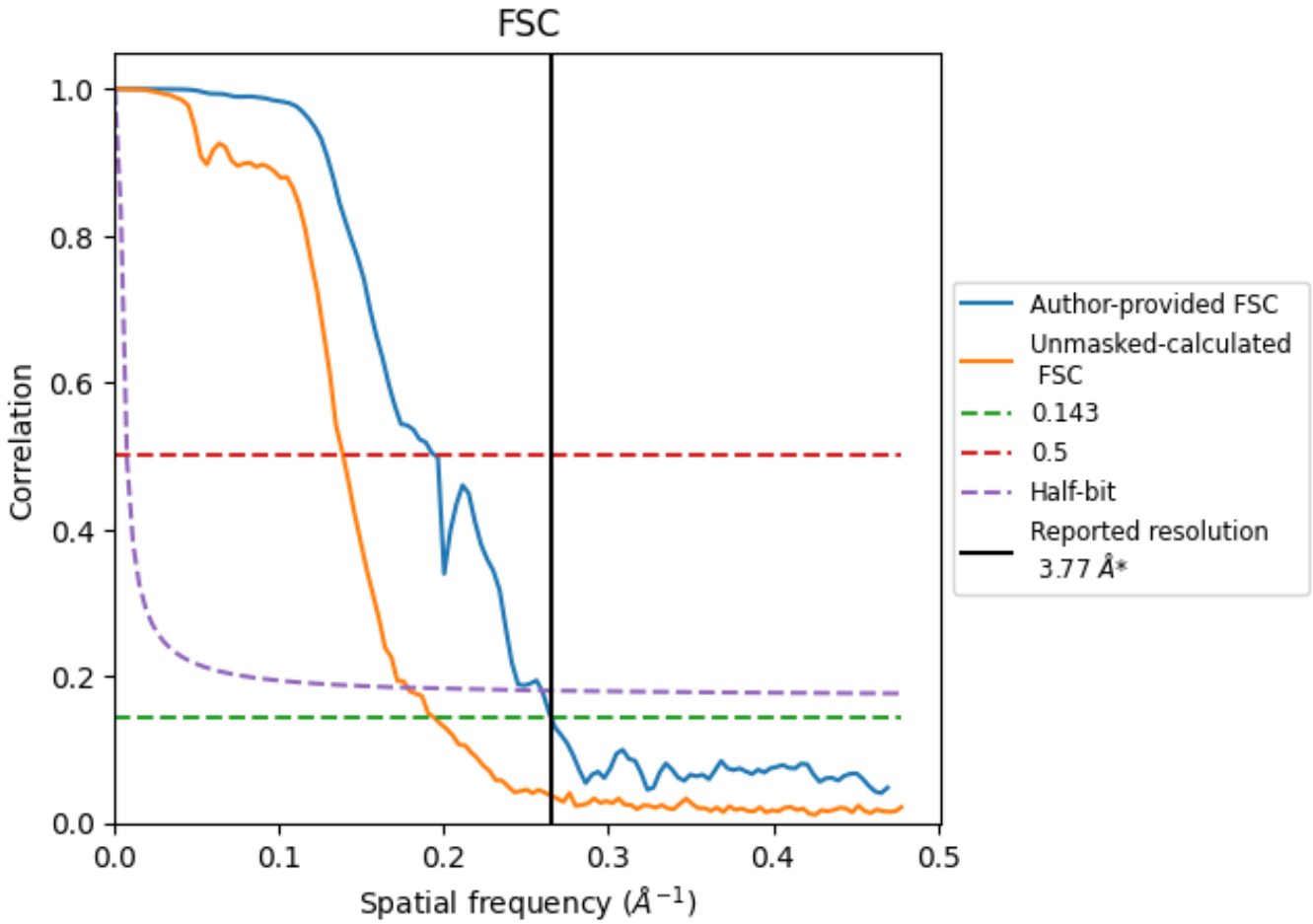


*Reported resolution corresponds to spatial frequency of 0.265 $Å^{-1}$

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 0.265 $Å^{-1}$

8.2 Resolution estimates (i)

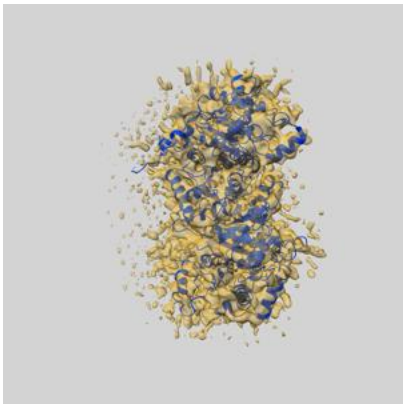
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.77	-	-
Author-provided FSC curve	3.77	5.13	3.85
Unmasked-calculated*	5.15	7.20	5.62

*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 5.15 differs from the reported value 3.77 by more than 10 %

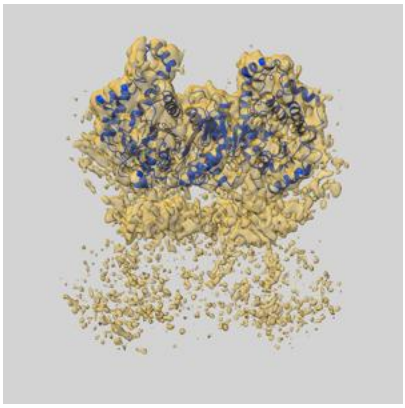
9 Map-model fit

This section contains information regarding the fit between EMDB map EMD-33804 and PDB model 7YG1. Per-residue inclusion information can be found in section 3 on page 6.

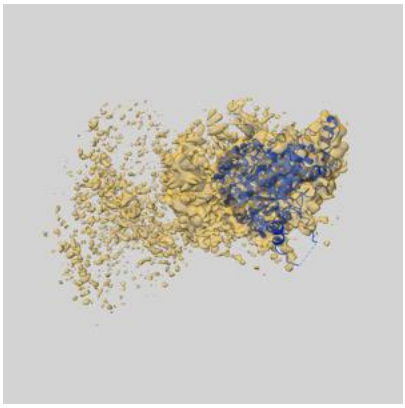
9.1 Map-model overlay



X



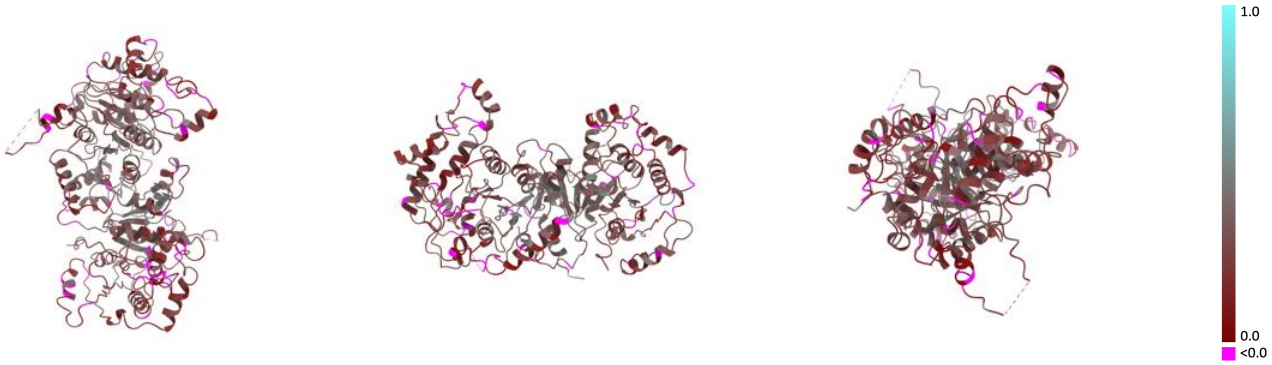
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.12 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.

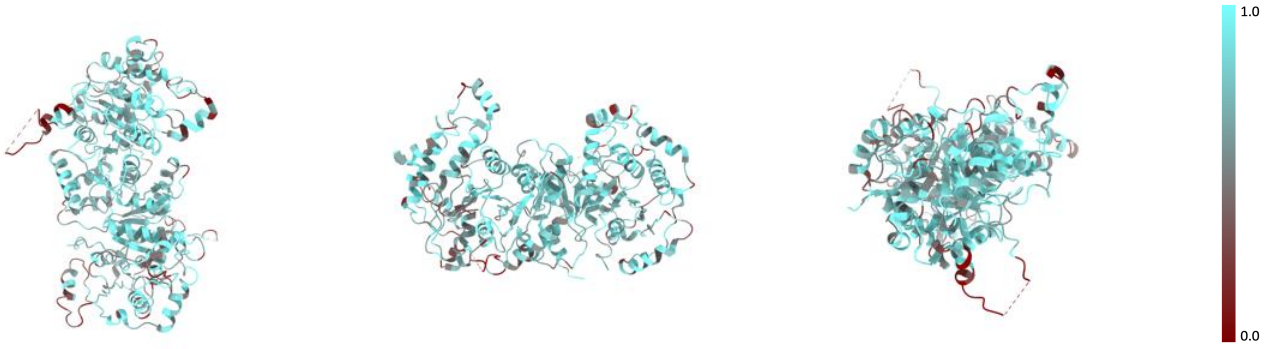
9.2 Q-score mapped to coordinate model (i)



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.

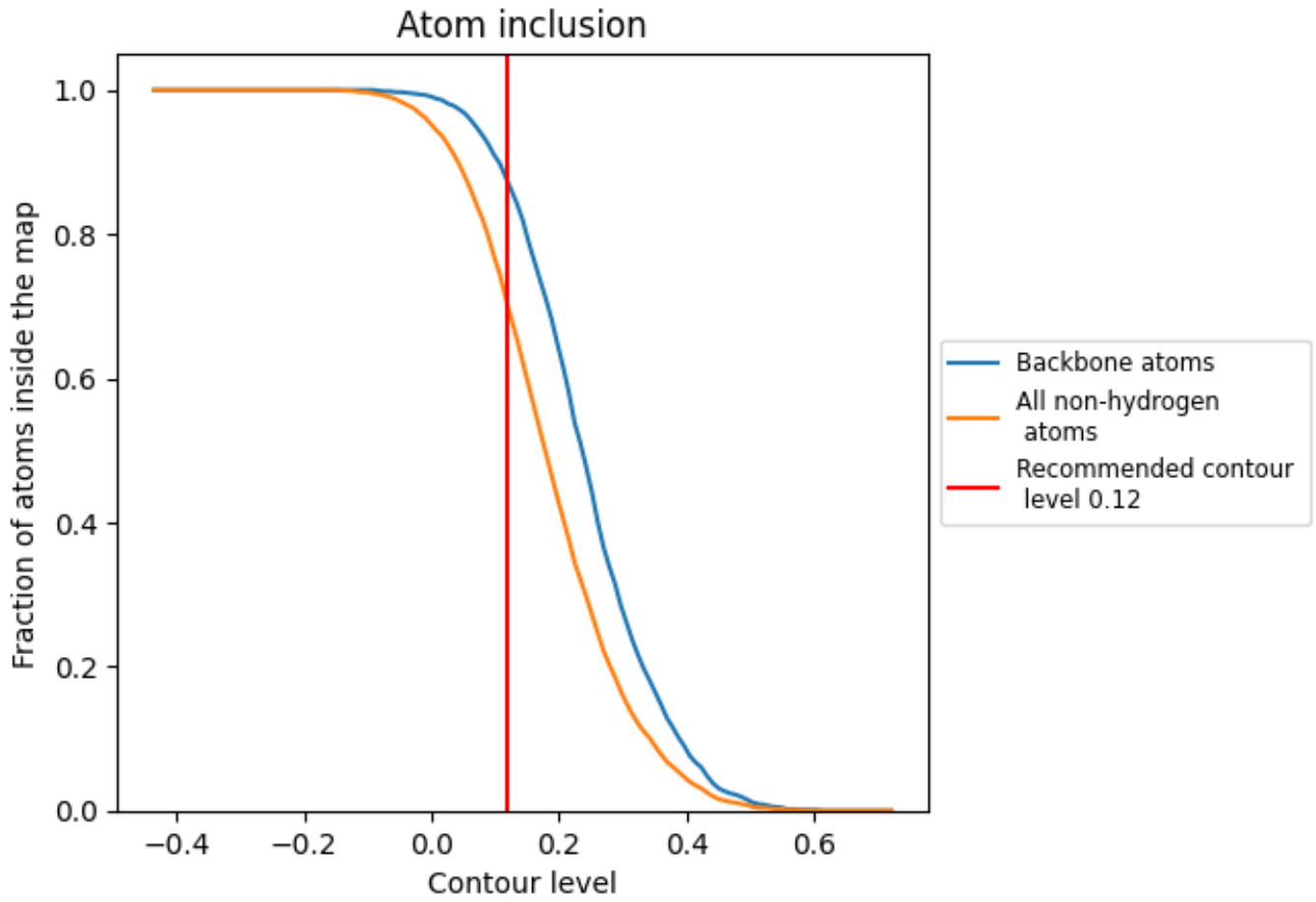
9.3 Atom inclusion mapped to coordinate model (i)



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 (0.12).

9.4 Atom inclusion (i)



At the recommended contour level, 87% of all backbone atoms, 70% 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 (0.12) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.6990	0.2440
A	0.6850	0.2340
B	0.7140	0.2540