wwPDB EM Validation Summary Report (i)

Feb 24, 2024 – 12:40 PM EST

PDB ID	:	7K5Y
EMDB ID	:	EMD-22684
Title	:	Cryo-EM structure of a chromatosome containing human linker histone H1.4
Authors	:	Zhou, B.-R.; Bai, Y.
Deposited on	:	2020-09-17
Resolution	:	2.76 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 (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:

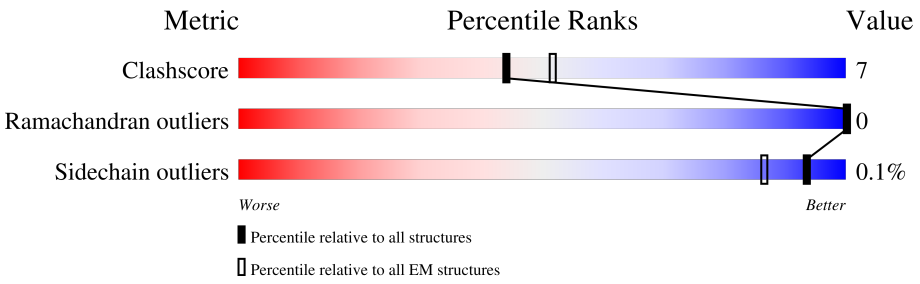
EMDB validation analysis	:	0.0.1.dev70
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.13
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.76 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	136	58% 13% 29%
1	E	136	62% 10% 29%
2	B	103	72% 9% 19%
2	F	103	72% 9% 19%
3	C	130	79% 5% 16%
3	G	130	77% 7% 16%
4	D	126	70% 6% 25%
4	H	126	68% 7% 25%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	I	197	12% 50% 50%
6	J	197	13% 50% 50%
7	M	270	72% 12% 16%
7	N	270	68% 16% 16%
8	U	219	5% 30% 5% 66%

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 18258 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 Histone H3.1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	97	799	505	153	137	4	0	0
1	E	97	799	505	153	137	4	0	0

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	B	83	662	418	129	114	1	0	0
2	F	83	662	418	129	114	1	0	0

- Molecule 3 is a protein called Histone H2A type 1-B/E.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
3	C	109	840	529	166	145	0	0
3	G	109	840	529	166	145	0	0

- Molecule 4 is a protein called Histone H2B type 1-J.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	D	95	745	468	136	139	2	0	0
4	H	95	745	468	136	139	2	0	0

- Molecule 5 is a DNA chain called DNA (197-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
5	I	197	4015	1904	733	1181	197	0	0

- Molecule 6 is a DNA chain called DNA (197-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
6	J	197	4062	1920	762	1183	197	0	0

- Molecule 7 is a protein called scFv20.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
7	M	227	1777	1120	290	358	9	0	0
			Total	C	N	O	S		
7	N	227	1777	1120	290	358	9	0	0

- Molecule 8 is a protein called Histone H1.4.

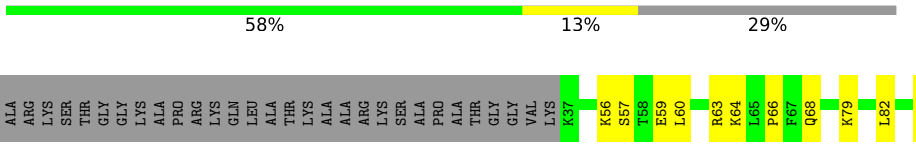
Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
8	U	75	535	336	97	102	0	0

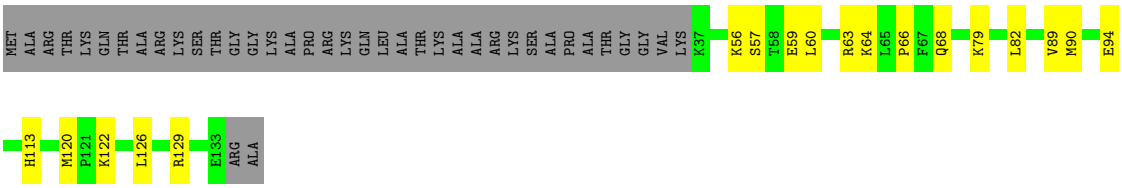
3 Residue-property plots (i)

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: Histone H3.1

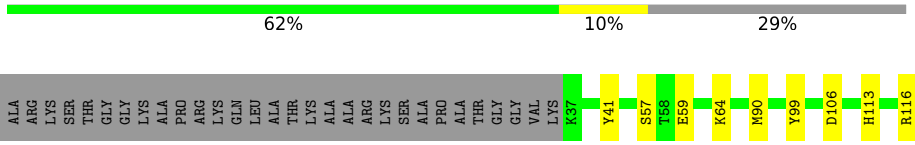
Chain A:

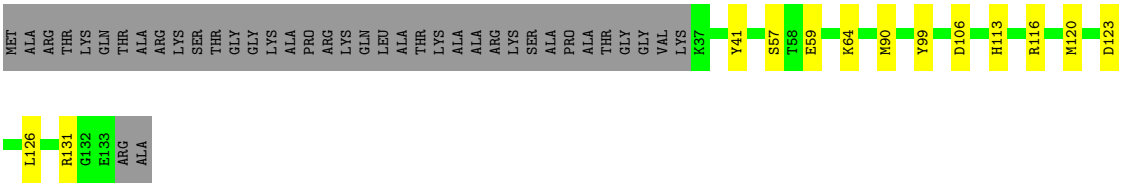




• Molecule 1: Histone H3.1

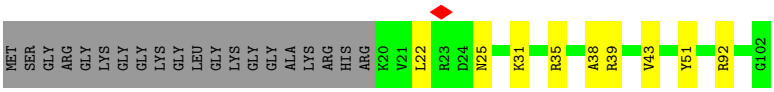
Chain E:





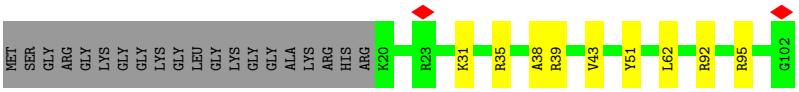
• Molecule 2: Histone H4

Chain B:



• Molecule 2: Histone H4

Chain F:



• Molecule 3: Histone H2A type 1-B/E

Chain C: 79% 5% 16%

MET SER GLY ARG GLY LYS GLN GLY GLY LYS A10 R17 G28 E64 L65 I79 H82 K118 LYS THR GLU SER HIS HIS LYS ALA LYS GLY LYS

• Molecule 3: Histone H2A type 1-B/E

Chain G: 77% 7% 16%

MET SER GLY ARG GLY LYS GLN GLY GLY LYS A10 K15 R20 R32 E61 E64 L65 I79 H82 D90 K118 LYS THR GLU SER HIS HIS LYS ALA LYS GLY LYS

• Molecule 4: Histone H2B type 1-J

Chain D: 70% 6% 25%

MET PRO GLU PRO ALA LYS SER ALA PRO ALA PRO LYS LYS GLY SER LYS LYS ALA VAL THR LYS ALA GLN LYS LYS ASP GLY LYS LYS ARG K30 R31 I39 K43 Q95 R99 L100 L101 E113 K116 A124 LYS

• Molecule 4: Histone H2B type 1-J

Chain H: 68% 7% 25%

MET PRO GLU PRO ALA LYS SER ALA PRO ALA PRO LYS LYS GLY SER LYS LYS ALA VAL THR LYS ALA GLN LYS LYS ASP GLY LYS LYS ARG K30 E35 V48 S56 E76 R79 A97 L101 K108 H109 A124 LYS

• Molecule 5: DNA (197-MER)

Chain I: 12% 50% 50%

G1 G2 G3 C4 T5 G6 G7 A8 C9 C10 C11 T12 A13 T14 A15 C16 G17 C18 G19 G20 G23 C24 C35 C36 C37 G38 G39 T40 G41 C42 C43 G44 A45 G46 G47 C51 T52 C53 A54 A55 C72 T73 A74 C79 G80 A84 A85 A86 C87 G88 T93 A94

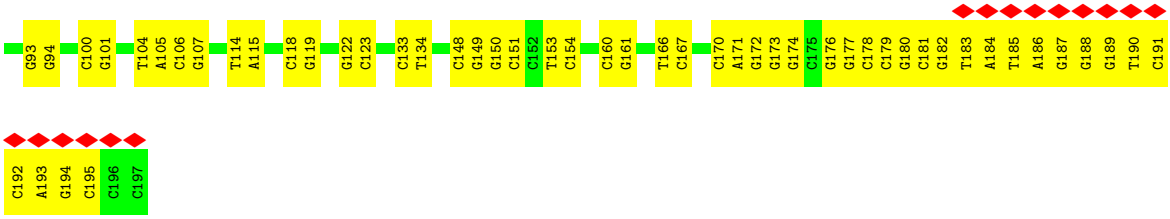
C103 C104 C105 C106 C107 G108 C109 G110 T111 T114 A115 A116 G124 G125 G126 G127 A128 T129 T133 C134 C135 C144 A145 C148 A149 G153 T154 C155 A160 T161 C165 A166 T167 C168 C169 T170 G171 T172 G173 C174 A175 T176 T181 G182 A183 A184 C185 A186 G187 C188 G189

A190 C191 C192 A193 C194 C195 C196 C197

• Molecule 6: DNA (197-MER)

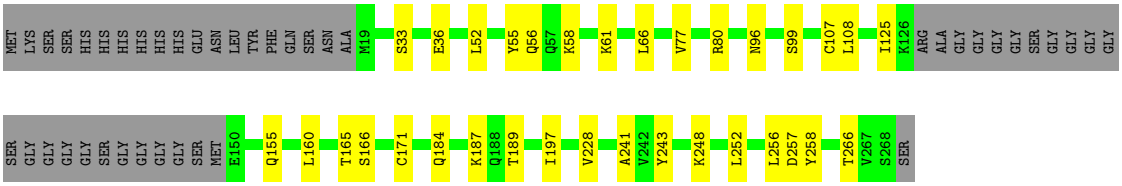
Chain J: 13% 50% 50%

G1 G2 G3 G4 T5 G6 G7 T8 C9 G10 C11 T12 G13 T14 T15 C16 A17 A18 T19 A20 C21 A22 G30 A31 T32 G33 A37 T38 A39 G43 A44 G48 T49 G50 C51 C52 T53 G62 G63 G64 A65 G66 A69 T70 C73 C74 T75 T76 G77 G78 T83 A84



• Molecule 7: scFv20

Chain M:

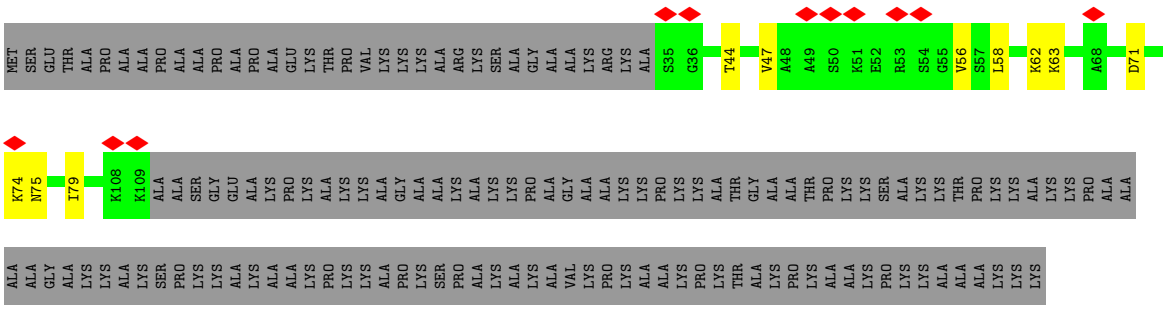


• Molecule 7: scFv20

Chain N:

• Molecule 8: Histone H1.4

Chain U:



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	112385	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING ONLY	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	40	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	4.576	Depositor
Minimum map value	-2.113	Depositor
Average map value	-0.002	Depositor
Map value standard deviation	0.105	Depositor
Recommended contour level	0.35	Depositor
Map size (Å)	322.56, 322.56, 322.56	wwPDB
Map dimensions	320, 320, 320	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.008, 1.008, 1.008	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	0.28	0/811	0.42	0/1088
1	E	0.29	0/811	0.43	0/1088
2	B	0.28	0/669	0.44	0/894
2	F	0.27	0/669	0.46	0/894
3	C	0.27	0/850	0.41	0/1146
3	G	0.27	0/850	0.43	0/1146
4	D	0.27	0/756	0.41	0/1015
4	H	0.27	0/756	0.40	0/1015
5	I	0.59	0/4499	0.95	1/6935 (0.0%)
6	J	0.58	0/4561	0.93	0/7044
7	M	0.29	0/1819	0.46	0/2464
7	N	0.29	0/1819	0.46	0/2464
8	U	0.25	0/538	0.46	0/718
All	All	0.45	0/19408	0.74	1/27911 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	I	176	DT	O4'-C1'-N1	5.01	111.51	108.00

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	799	0	838	10	0
1	E	799	0	838	10	0
2	B	662	0	709	6	0
2	F	662	0	709	8	0
3	C	840	0	902	4	0
3	G	840	0	902	9	0
4	D	745	0	771	4	0
4	H	745	0	771	9	0
5	I	4015	0	2208	64	0
6	J	4062	0	2209	68	0
7	M	1777	0	1721	21	0
7	N	1777	0	1721	29	0
8	U	535	0	588	6	0
All	All	18258	0	14887	228	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 7.

The worst 5 of 228 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:G:90:ASP:OD2	7:N:251:ARG:NH2	2.11	0.83
1:A:120:MET:SD	1:A:122:LYS:NZ	2.63	0.72
7:M:187:LYS:HB2	7:M:197:ILE:HD11	1.72	0.71
1:E:99:TYR:HD1	2:F:95:ARG:HH12	1.38	0.71
8:U:71:ASP:OD2	8:U:74:LYS:NZ	2.22	0.70

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	95/136 (70%)	95 (100%)	0	0	100	100
1	E	95/136 (70%)	94 (99%)	1 (1%)	0	100	100
2	B	81/103 (79%)	80 (99%)	1 (1%)	0	100	100
2	F	81/103 (79%)	80 (99%)	1 (1%)	0	100	100
3	C	107/130 (82%)	106 (99%)	1 (1%)	0	100	100
3	G	107/130 (82%)	107 (100%)	0	0	100	100
4	D	93/126 (74%)	92 (99%)	1 (1%)	0	100	100
4	H	93/126 (74%)	92 (99%)	1 (1%)	0	100	100
7	M	223/270 (83%)	217 (97%)	6 (3%)	0	100	100
7	N	223/270 (83%)	219 (98%)	4 (2%)	0	100	100
8	U	73/219 (33%)	71 (97%)	2 (3%)	0	100	100
All	All	1271/1749 (73%)	1253 (99%)	18 (1%)	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	85/111 (77%)	84 (99%)	1 (1%)	71	82
1	E	85/111 (77%)	85 (100%)	0	100	100
2	B	68/79 (86%)	68 (100%)	0	100	100
2	F	68/79 (86%)	68 (100%)	0	100	100
3	C	85/100 (85%)	85 (100%)	0	100	100
3	G	85/100 (85%)	85 (100%)	0	100	100
4	D	81/105 (77%)	81 (100%)	0	100	100
4	H	81/105 (77%)	81 (100%)	0	100	100
7	M	200/225 (89%)	200 (100%)	0	100	100
7	N	200/225 (89%)	200 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
8	U	57/146 (39%)	57 (100%)	0	100	100
All	All	1095/1386 (79%)	1094 (100%)	1 (0%)	93	96

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	129	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	113	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)

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.

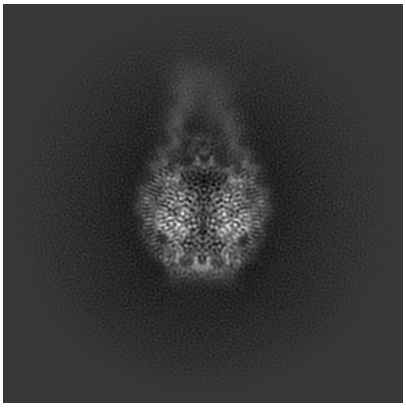
6 Map visualisation (i)

This section contains visualisations of the EMDB entry EMD-22684. These allow visual inspection of the internal detail of the map and identification of artifacts.

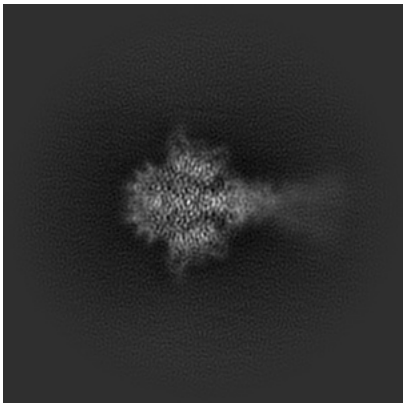
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections (i)

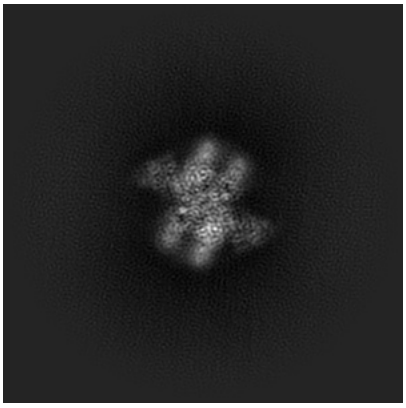
6.1.1 Primary map



X



Y

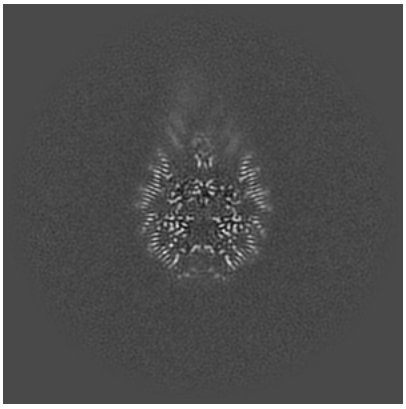


Z

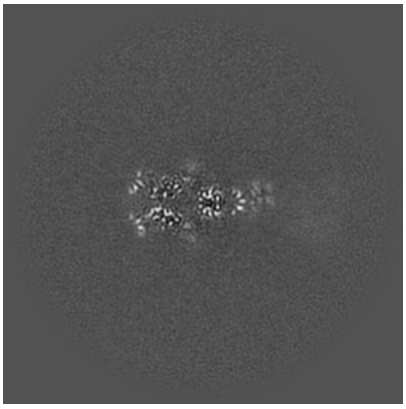
The images above show the map projected in three orthogonal directions.

6.2 Central slices (i)

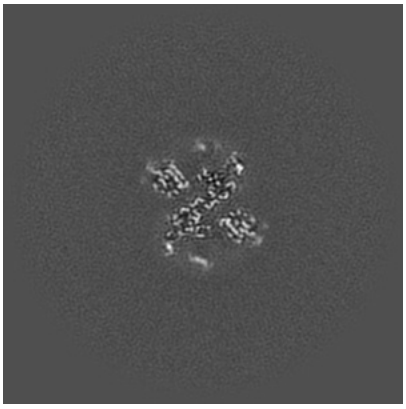
6.2.1 Primary map



X Index: 160



Y Index: 160

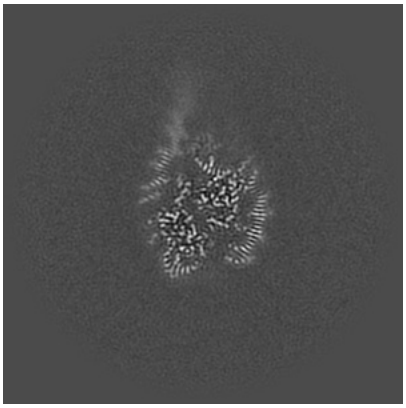


Z Index: 160

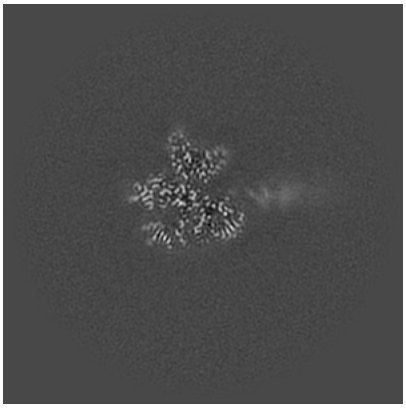
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

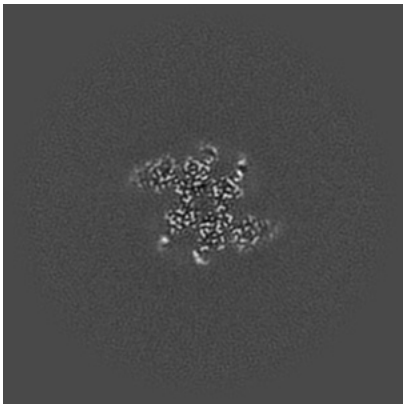
6.3.1 Primary map



X Index: 165



Y Index: 139

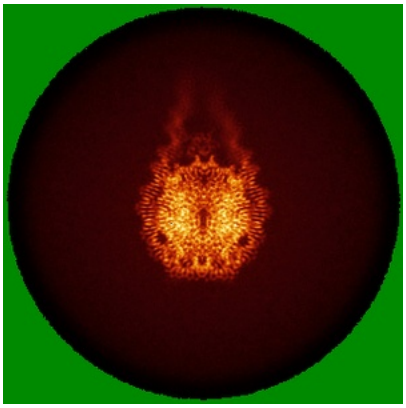


Z Index: 140

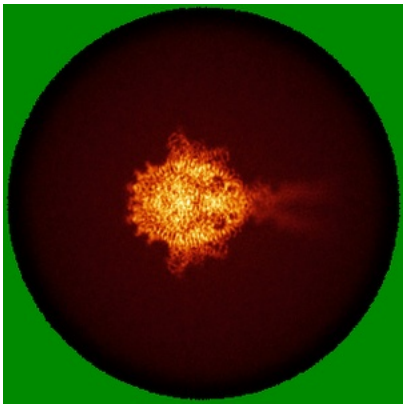
The images above show the largest variance slices of the map in three orthogonal directions.

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

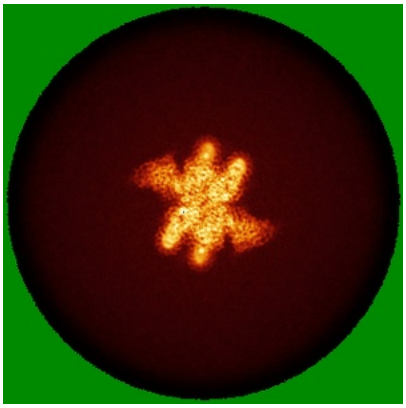
6.4.1 Primary map



X



Y

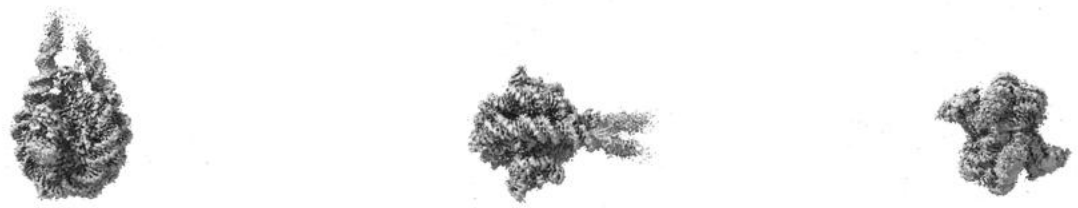


Z

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 0.35. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

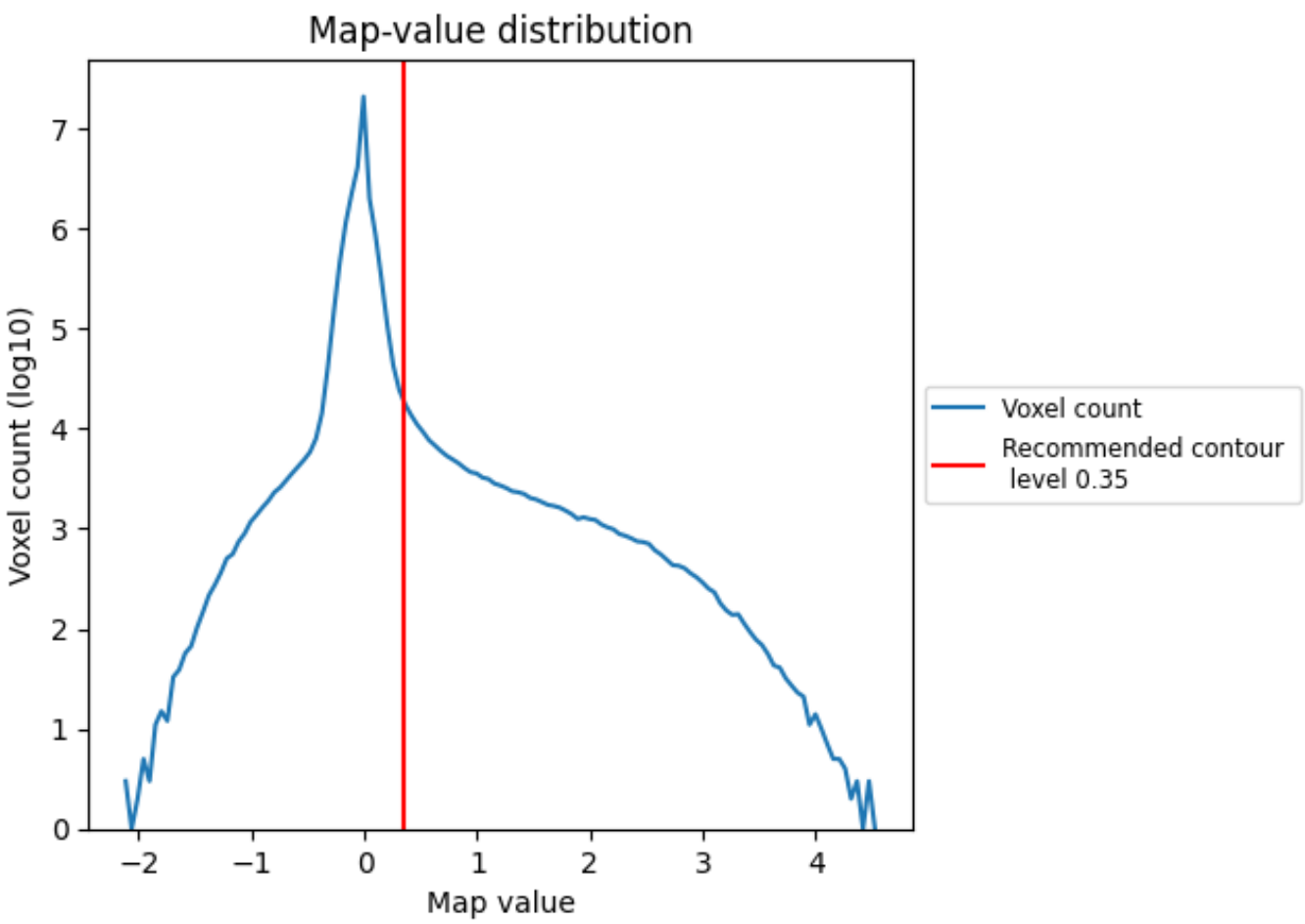
6.6 Mask visualisation (i)

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

7 Map analysis (i)

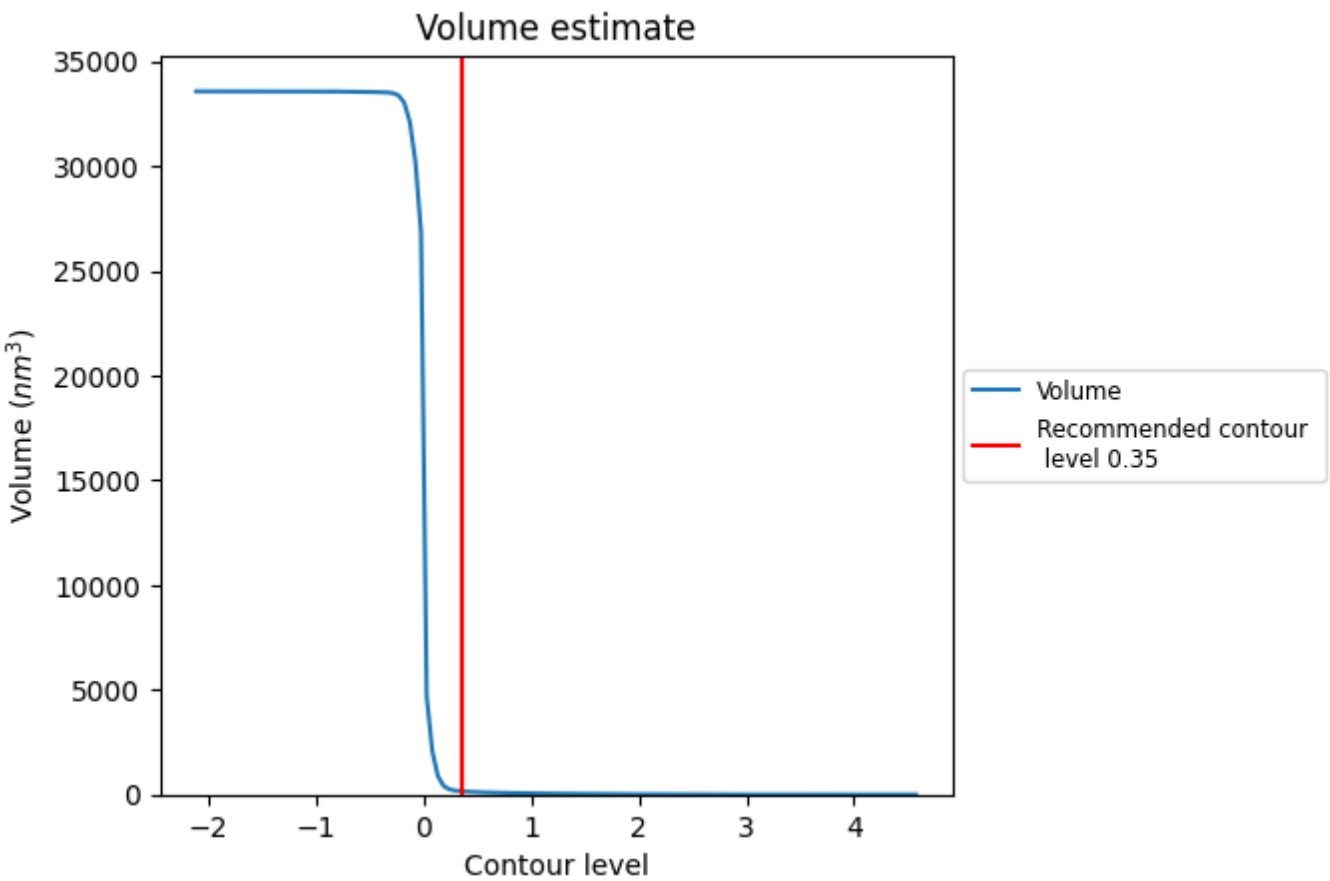
This section contains the results of statistical analysis of the map.

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.

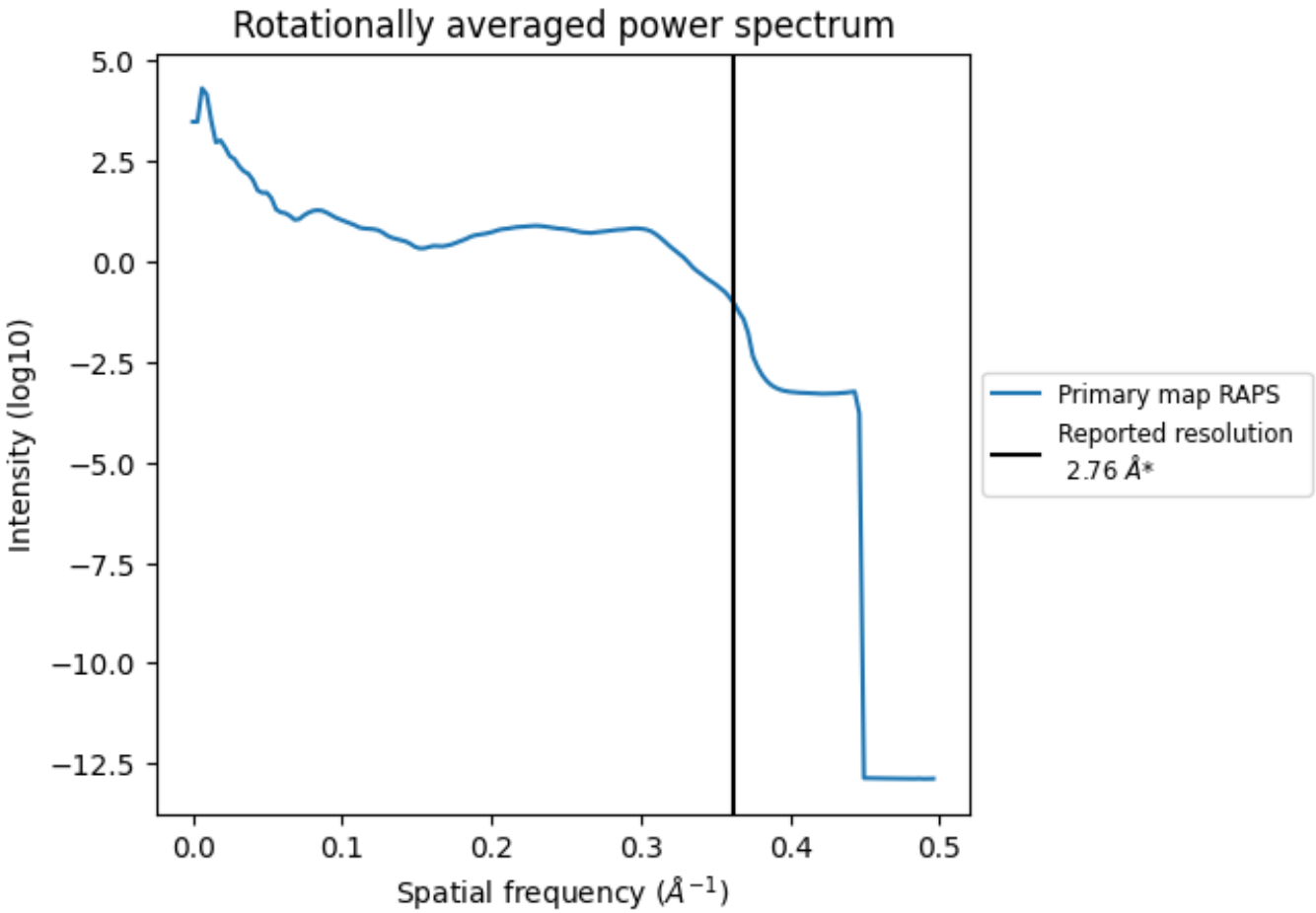
7.2 Volume estimate (i)



The volume at the recommended contour level is 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 0.362 Å^{-1}

8 Fourier-Shell correlation (i)

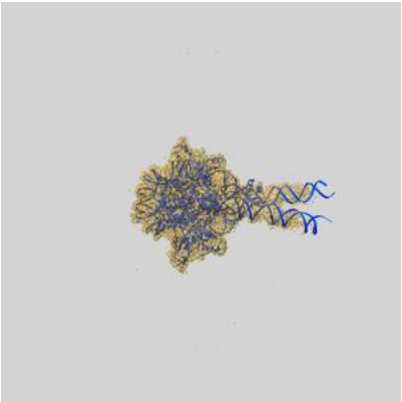
This section was not generated. No FSC curve or half-maps provided.

9 Map-model fit (i)

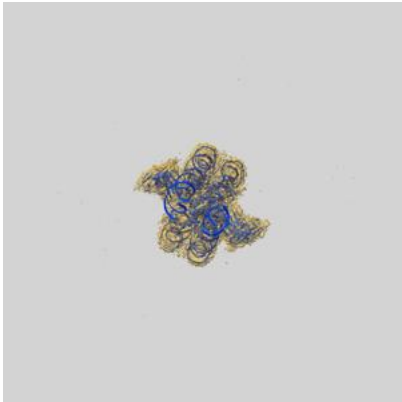
This section contains information regarding the fit between EMDB map EMD-22684 and PDB model 7K5Y. Per-residue inclusion information can be found in section 3 on page 6.

9.1 Map-model overlay (i)

X



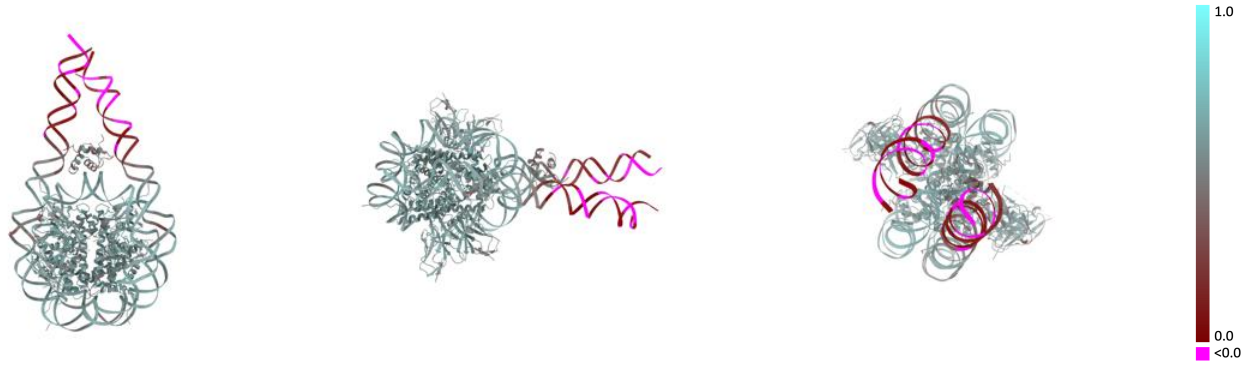
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.35 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.

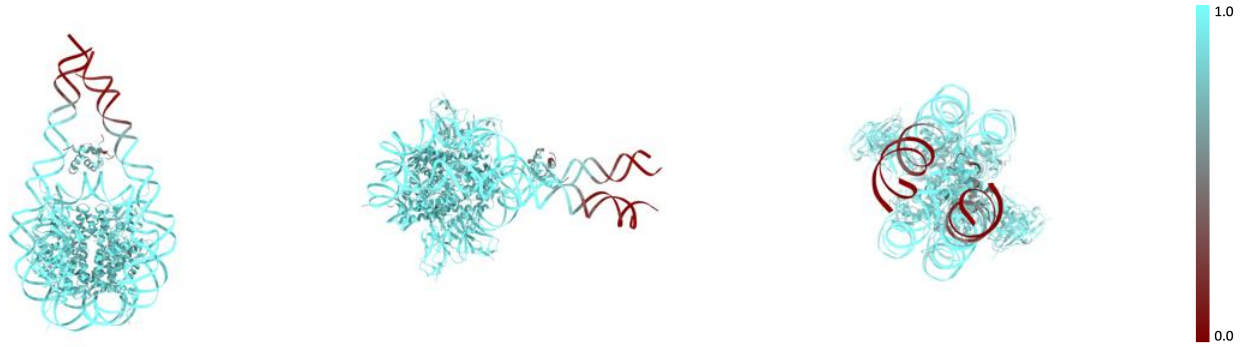
9.2 Q-score mapped to coordinate model (i)



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.

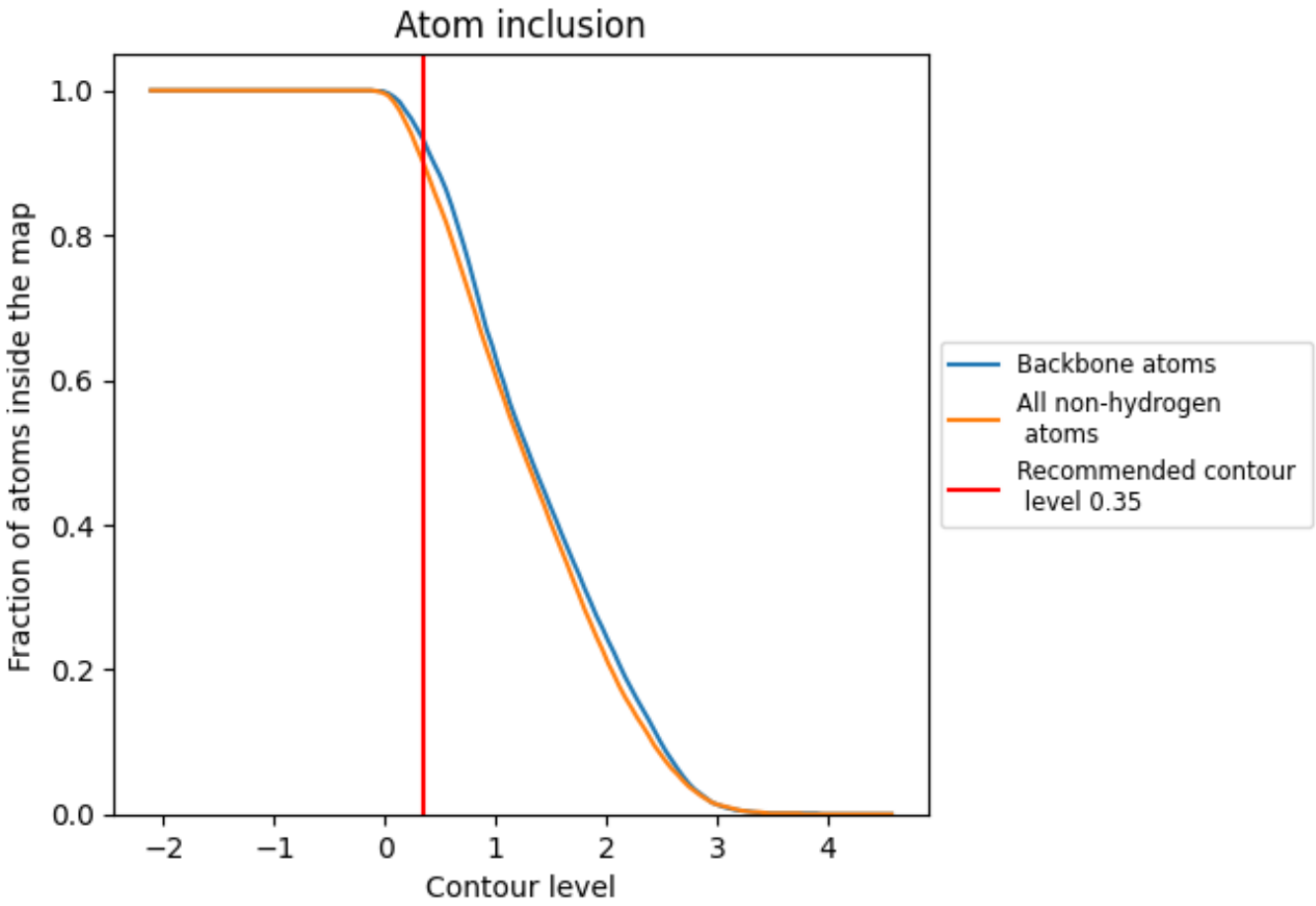
9.3 Atom inclusion mapped to coordinate model (i)



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 (0.35).

9.4 Atom inclusion (i)



At the recommended contour level, 93% of all backbone atoms, 90% 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 (0.35) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.8990	0.5270
A	0.9550	0.5970
B	0.9510	0.5950
C	0.9680	0.6100
D	0.9370	0.5890
E	0.9600	0.5980
F	0.9430	0.5960
G	0.9640	0.6050
H	0.9500	0.5900
I	0.8510	0.4690
J	0.8500	0.4540
M	0.9460	0.5670
N	0.9520	0.5700
U	0.6940	0.4470