



wwPDB EM Validation Summary Report ⓘ

May 21, 2022 – 07:23 am BST

PDB ID : 7R5S
EMDB ID : EMD-14336
Title : Structure of the human CCAN bound to alpha satellite DNA
Authors : Yatskevich, S.; Muir, K.W.; Bellini, D.; Zhang, Z.; Yang, J.; Tischer, T.;
Predin, M.; Dendooven, T.; McLaughlin, S.H.; Barford, D.
Deposited on : 2022-02-11
Resolution : 2.83 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

EMDB validation analysis : 0.0.1.dev8
MolProbity : 4.02b-467
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.28.1

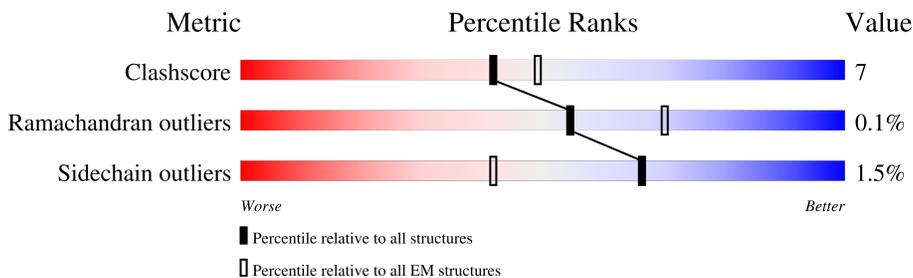
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 2.83 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.



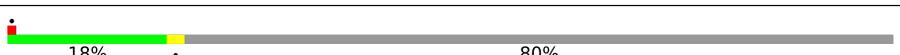
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

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 $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the EM map (all-atom inclusion $< 40\%$). The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	H	247	
2	I	756	
3	J	53	
4	K	269	
5	L	344	
6	M	180	
7	N	339	
8	O	300	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
9	P	288	
10	Q	268	
11	U	418	
12	R	177	
13	T	561	
14	W	88	
15	S	138	
16	X	81	
17	i	53	

2 Entry composition

There are 18 unique types of molecules in this entry. The entry contains 26871 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 Centromere protein H.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	H	204	1652	1036	286	319	11	0	0

- Molecule 2 is a protein called Centromere protein I.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	I	622	5014	3283	810	890	31	0	0

- Molecule 3 is a DNA chain called DNA (66-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
3	J	41	849	403	155	250	41	0	0

- Molecule 4 is a protein called Centromere protein K.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	K	233	1922	1220	318	374	10	0	0

- Molecule 5 is a protein called Centromere protein L.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
5	L	312	2502	1628	409	451	14	0	0

- Molecule 6 is a protein called Centromere protein M.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
6	M	172	1325	839	236	243	7	0	0

- Molecule 7 is a protein called Centromere protein N.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
7	N	318	2613	1678	453	472	10	0	0

- Molecule 8 is a protein called Centromere protein O.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
8	O	210	1642	1060	277	298	7	0	0

- Molecule 9 is a protein called Centromere protein P.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
9	P	224	1788	1141	310	329	8	0	0

- Molecule 10 is a protein called Centromere protein Q.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
10	Q	191	1526	953	258	304	11	0	0

- Molecule 11 is a protein called Centromere protein U.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
11	U	166	1365	861	242	257	5	0	0

- Molecule 12 is a protein called Centromere protein R.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
12	R	80	649	412	105	125	7	0	0

- Molecule 13 is a protein called Centromere protein T.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
13	T	112	915	586	163	159	7	0	0

- Molecule 14 is a protein called Centromere protein W.

Mol	Chain	Residues	Atoms					AltConf	Trace
14	W	88	Total	C	N	O	S	0	0
			704	445	143	112	4		

- Molecule 15 is a protein called Centromere protein S.

Mol	Chain	Residues	Atoms					AltConf	Trace
15	S	120	Total	C	N	O	S	0	0
			982	607	174	195	6		

- Molecule 16 is a protein called Centromere protein X.

Mol	Chain	Residues	Atoms					AltConf	Trace
16	X	74	Total	C	N	O	S	0	0
			590	378	104	107	1		

- Molecule 17 is a DNA chain called DNA (66-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
17	i	41	Total	C	N	O	P	0	0
			832	396	153	242	41		

- Molecule 18 is water.

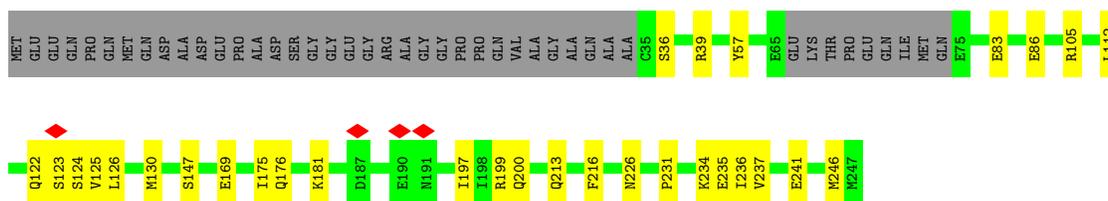
Mol	Chain	Residues	Atoms		AltConf
18	I	1	Total	O	0
			1	1	

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.

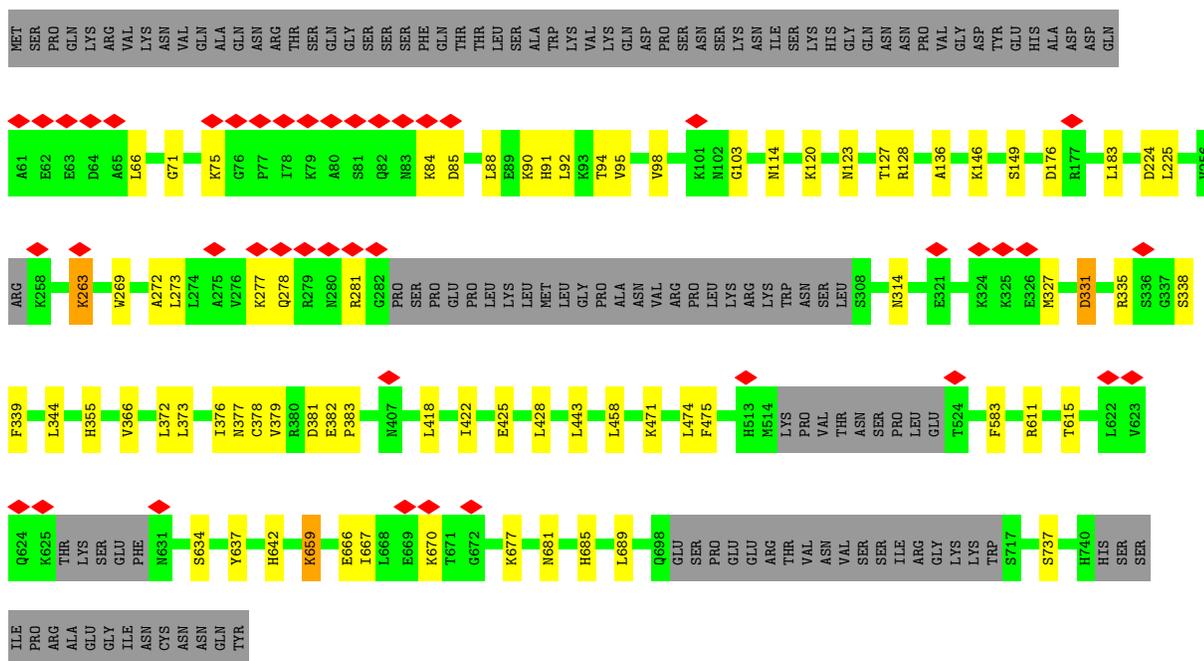
- Molecule 1: Centromere protein H

Chain H: 



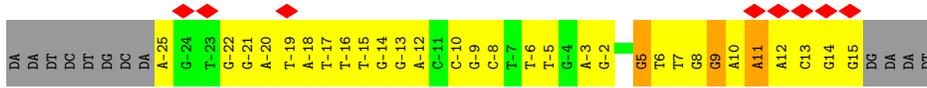
- Molecule 2: Centromere protein I

Chain I: 

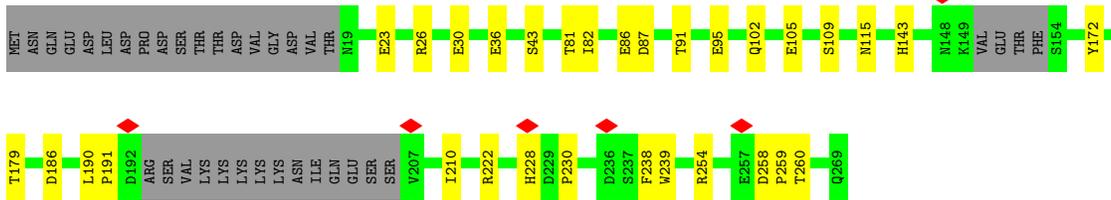
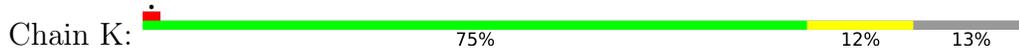


- Molecule 3: DNA (66-MER)

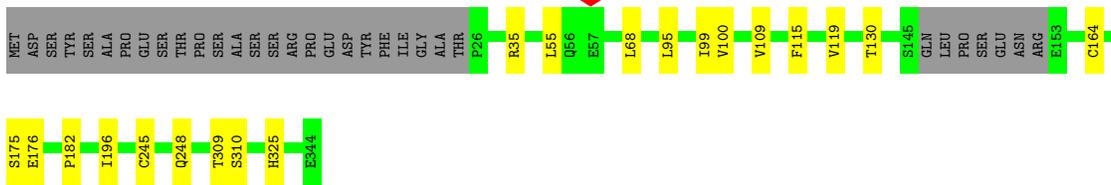
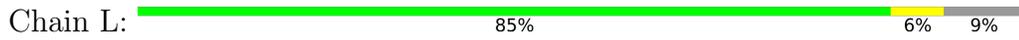
Chain J: 



• Molecule 4: Centromere protein K



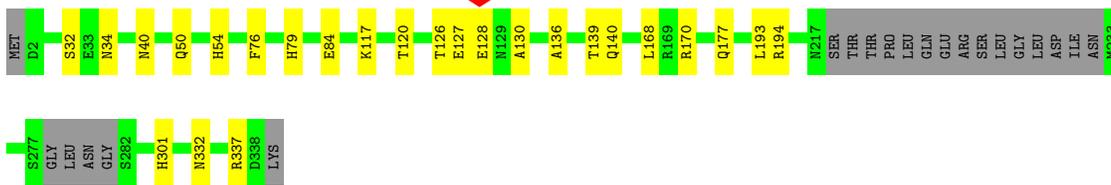
• Molecule 5: Centromere protein L



• Molecule 6: Centromere protein M

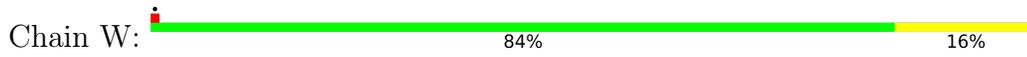


• Molecule 7: Centromere protein N

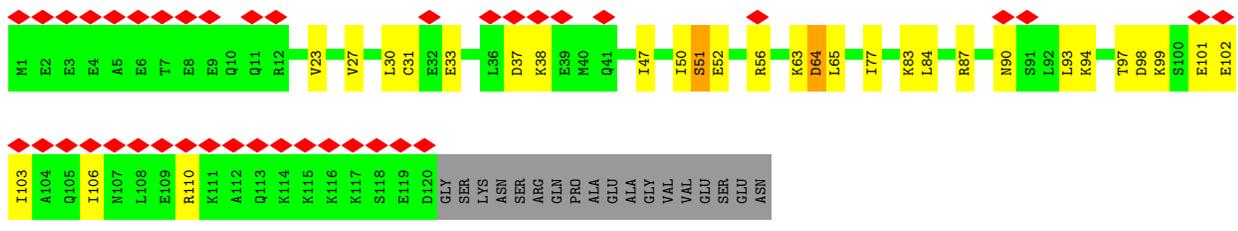


• Molecule 8: Centromere protein O

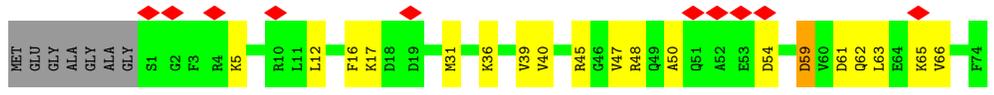




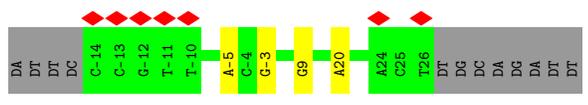
• Molecule 15: Centromere protein S



• Molecule 16: Centromere protein X



• Molecule 17: DNA (66-MER)



4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	123677	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	40	Depositor
Minimum defocus (nm)	1500	Depositor
Maximum defocus (nm)	3000	Depositor
Magnification	Not provided	
Image detector	GATAN K3 (6k x 4k)	Depositor
Maximum map value	0.039	Depositor
Minimum map value	-0.012	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.001	Depositor
Recommended contour level	0.007	Depositor
Map size (\AA)	368.496, 368.496, 368.496	wwPDB
Map dimensions	432, 432, 432	wwPDB
Map angles ($^\circ$)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (\AA)	0.853, 0.853, 0.853	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	H	0.24	0/1660	0.35	0/2218
2	I	0.24	0/5137	0.38	0/6955
3	J	0.83	6/952 (0.6%)	0.95	0/1470
4	K	0.24	0/1953	0.37	0/2634
5	L	0.29	0/2569	0.41	0/3485
6	M	0.23	0/1347	0.38	0/1827
7	N	0.24	0/2670	0.40	0/3606
8	O	0.24	0/1678	0.40	0/2280
9	P	0.24	0/1820	0.43	0/2451
10	Q	0.24	0/1538	0.43	0/2062
11	U	0.25	0/1383	0.42	0/1856
12	R	0.24	0/653	0.49	0/865
13	T	0.23	0/937	0.35	0/1263
14	W	0.23	0/711	0.37	0/944
15	S	0.25	0/991	0.41	0/1322
16	X	0.23	0/596	0.42	0/801
17	i	0.86	4/932 (0.4%)	0.90	0/1433
All	All	0.32	10/27527 (0.0%)	0.46	0/37472

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

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	J	15	DG	C1'-N9	-5.68	1.39	1.47
17	i	-3	DG	C1'-N9	-5.56	1.39	1.47
17	i	20	DA	C1'-N9	-5.54	1.39	1.47

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
17	i	9	DG	C1'-N9	-5.53	1.39	1.47
3	J	-25	DA	C1'-N9	-5.50	1.39	1.47

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	H	122	GLN	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	H	1652	0	1728	21	0
2	I	5014	0	5036	46	0
3	J	849	0	464	27	0
4	K	1922	0	1924	21	0
5	L	2502	0	2492	11	0
6	M	1325	0	1370	2	0
7	N	2613	0	2627	15	0
8	O	1642	0	1616	20	0
9	P	1788	0	1791	12	0
10	Q	1526	0	1586	41	0
11	U	1365	0	1396	45	0
12	R	649	0	673	27	0
13	T	915	0	924	8	0
14	W	704	0	789	10	0
15	S	982	0	987	26	0
16	X	590	0	623	14	0
17	i	832	0	460	0	0
18	I	1	0	0	1	0
All	All	26871	0	26486	299	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 299 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:J:5:DG:H2''	3:J:6:DT:C5	2.14	0.82
3:J:5:DG:H2''	3:J:6:DT:C7	2.12	0.80
2:I:66:LEU:HD21	2:I:98:VAL:HG11	1.65	0.79
3:J:5:DG:H2''	3:J:6:DT:H71	1.67	0.76
2:I:278:GLN:HA	2:I:281:ARG:HG2	1.66	0.76

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	H	200/247 (81%)	193 (96%)	6 (3%)	1 (0%)	29	51
2	I	610/756 (81%)	589 (97%)	21 (3%)	0	100	100
4	K	227/269 (84%)	224 (99%)	3 (1%)	0	100	100
5	L	308/344 (90%)	302 (98%)	6 (2%)	0	100	100
6	M	170/180 (94%)	169 (99%)	1 (1%)	0	100	100
7	N	312/339 (92%)	303 (97%)	9 (3%)	0	100	100
8	O	204/300 (68%)	201 (98%)	3 (2%)	0	100	100
9	P	220/288 (76%)	211 (96%)	8 (4%)	1 (0%)	29	51
10	Q	187/268 (70%)	180 (96%)	7 (4%)	0	100	100
11	U	164/418 (39%)	161 (98%)	3 (2%)	0	100	100
12	R	76/177 (43%)	72 (95%)	4 (5%)	0	100	100
13	T	110/561 (20%)	108 (98%)	2 (2%)	0	100	100
14	W	86/88 (98%)	82 (95%)	4 (5%)	0	100	100
15	S	118/138 (86%)	117 (99%)	1 (1%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
16	X	72/81 (89%)	70 (97%)	2 (3%)	0	100	100
All	All	3064/4454 (69%)	2982 (97%)	80 (3%)	2 (0%)	54	75

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	H	123	SER
9	P	112	HIS

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	H	188/224 (84%)	185 (98%)	3 (2%)	62	81
2	I	560/691 (81%)	553 (99%)	7 (1%)	69	84
4	K	221/260 (85%)	218 (99%)	3 (1%)	67	83
5	L	277/306 (90%)	275 (99%)	2 (1%)	84	91
6	M	151/158 (96%)	150 (99%)	1 (1%)	84	91
7	N	287/311 (92%)	282 (98%)	5 (2%)	60	80
8	O	177/263 (67%)	174 (98%)	3 (2%)	60	80
9	P	197/259 (76%)	193 (98%)	4 (2%)	55	76
10	Q	179/248 (72%)	177 (99%)	2 (1%)	73	86
11	U	152/379 (40%)	145 (95%)	7 (5%)	27	51
12	R	75/166 (45%)	74 (99%)	1 (1%)	69	84
13	T	100/461 (22%)	99 (99%)	1 (1%)	76	88
14	W	77/77 (100%)	77 (100%)	0	100	100
15	S	107/121 (88%)	105 (98%)	2 (2%)	57	77
16	X	65/67 (97%)	63 (97%)	2 (3%)	40	64
All	All	2813/3991 (70%)	2770 (98%)	43 (2%)	66	82

5 of 43 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
10	Q	130	LYS
11	U	392	PHE
10	Q	134	LEU
11	U	281	PHE
12	R	92	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 41 such sidechains are listed below:

Mol	Chain	Res	Type
10	Q	177	GLN
12	R	147	ASN
10	Q	181	ASN
11	U	403	HIS
13	T	528	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 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

There are no chain breaks in this entry.

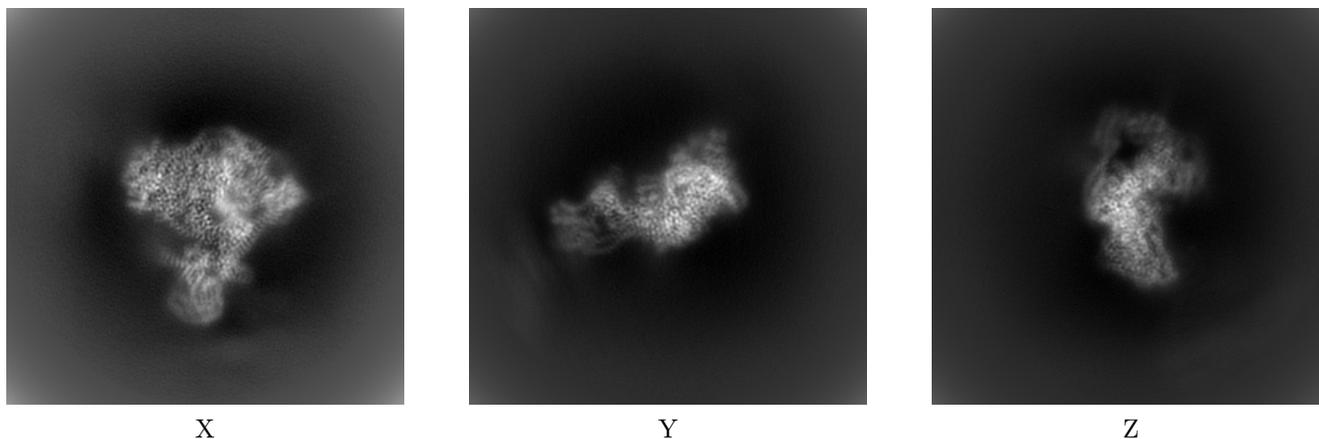
6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-14336. These allow visual inspection of the internal detail of the map and identification of artifacts.

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

6.1 Orthogonal projections [i](#)

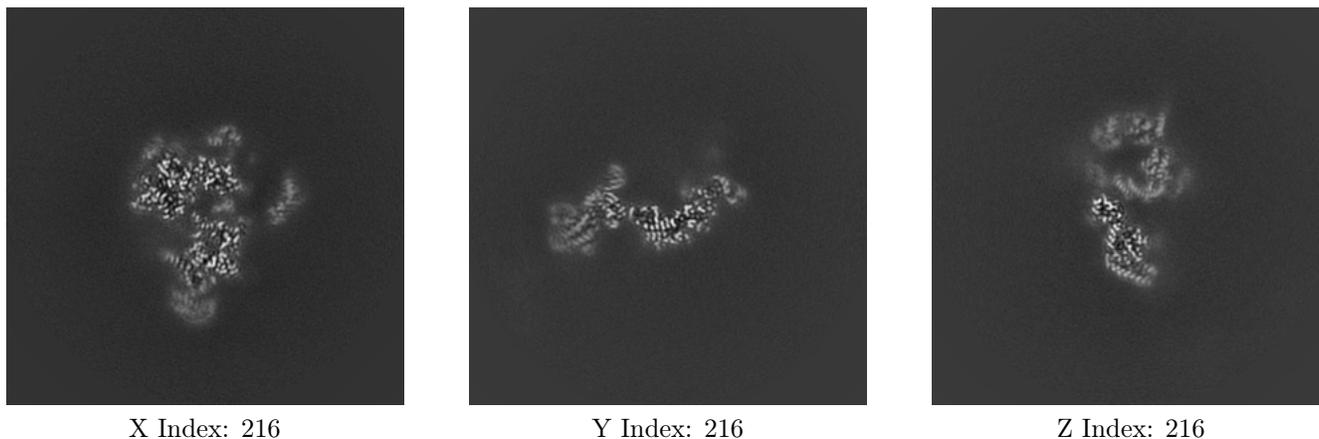
6.1.1 Primary map



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

6.2 Central slices [i](#)

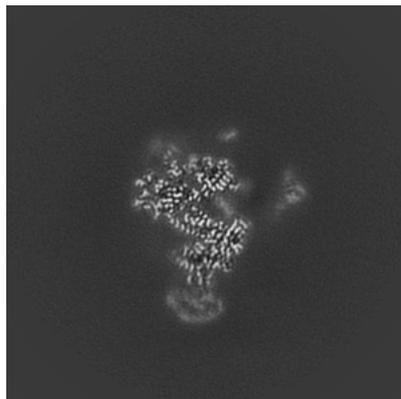
6.2.1 Primary map



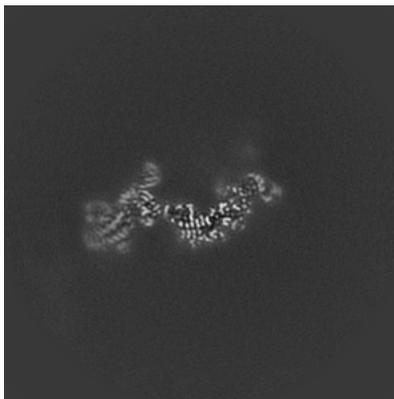
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices [i](#)

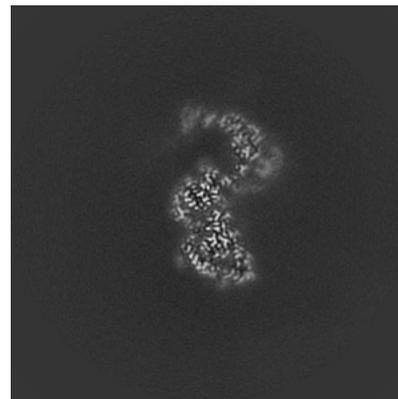
6.3.1 Primary map



X Index: 209



Y Index: 216



Z Index: 238

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

6.4 Orthogonal surface views [i](#)

6.4.1 Primary map



X



Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.007. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

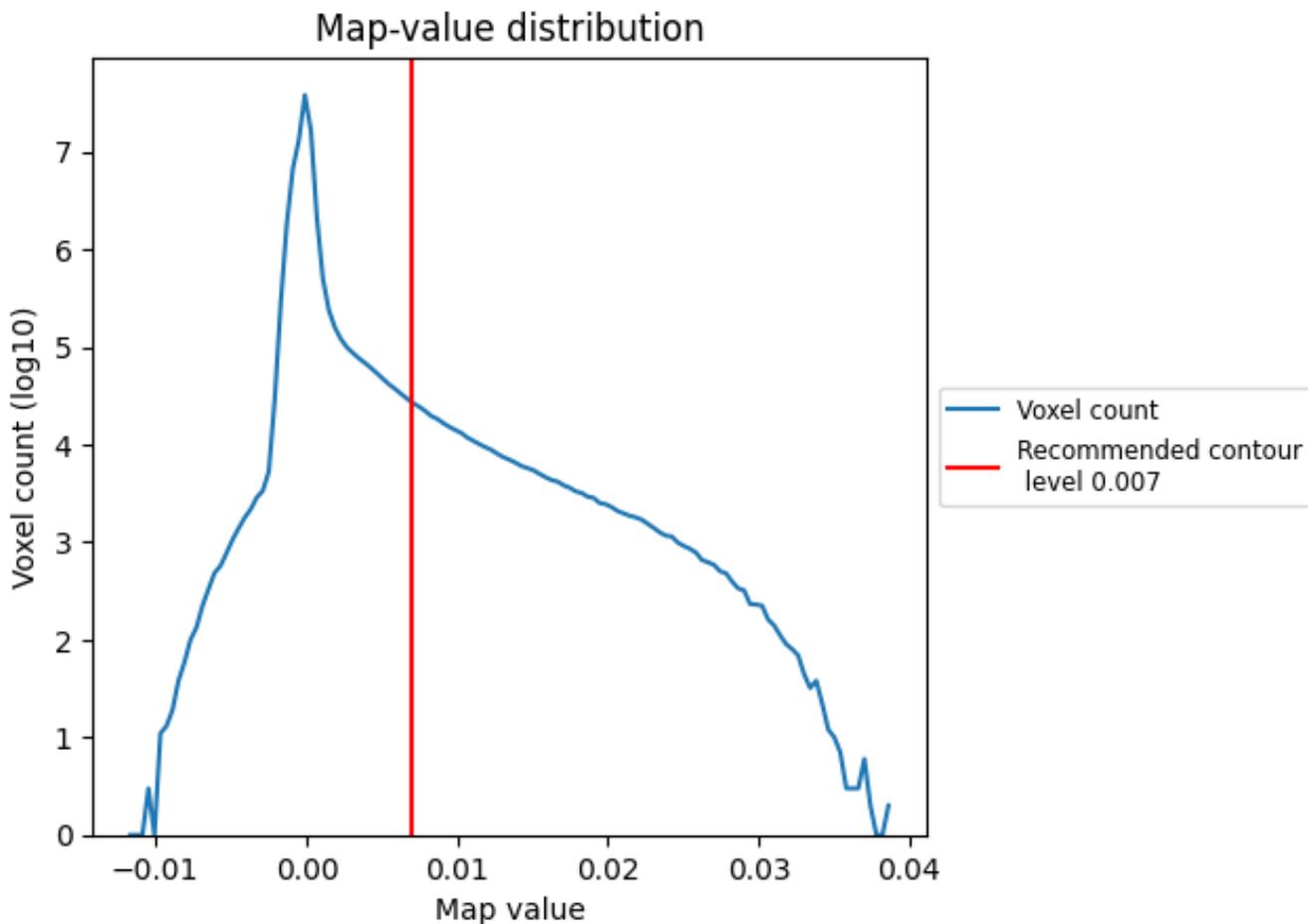
6.5 Mask visualisation

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

7 Map analysis [i](#)

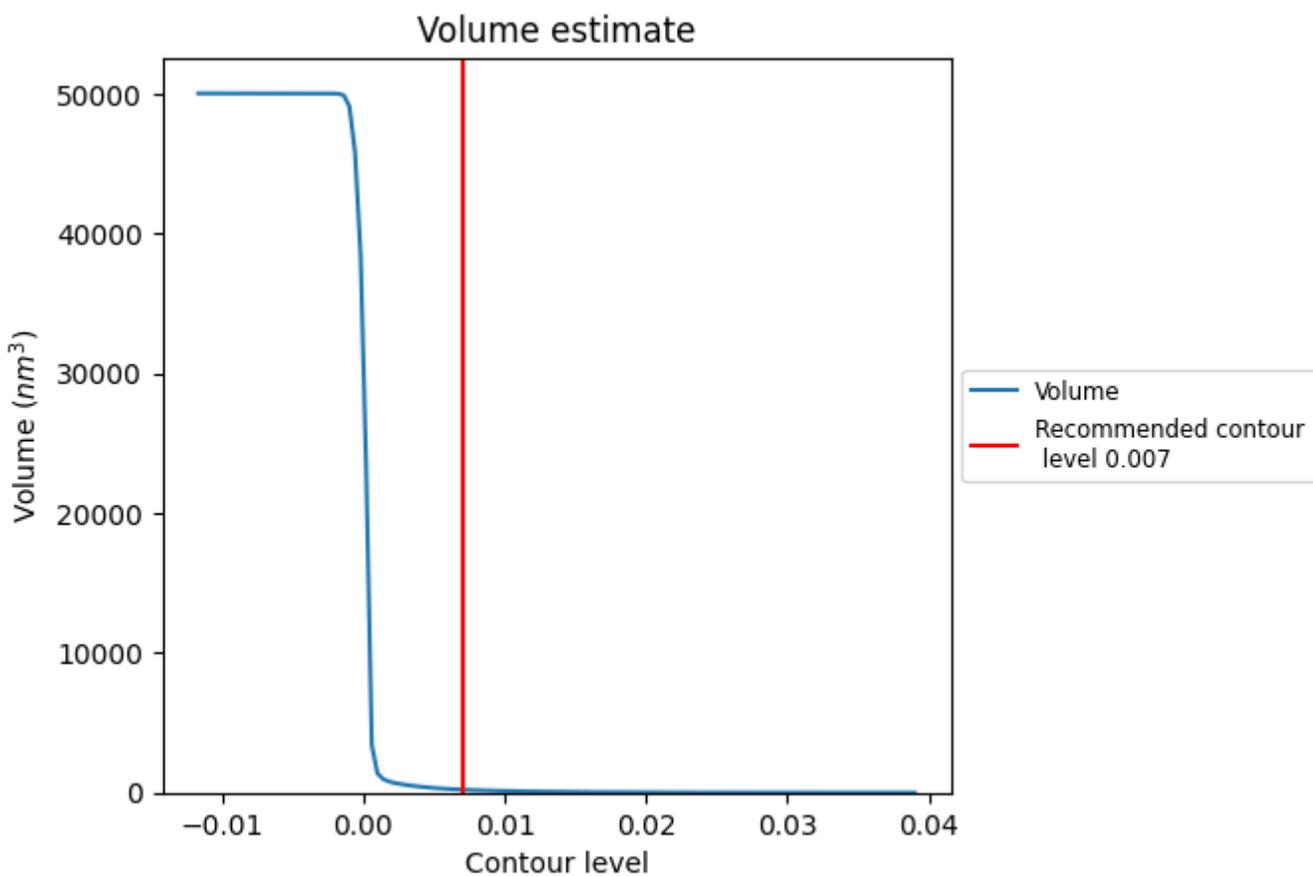
This section contains the results of statistical analysis of the map.

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.

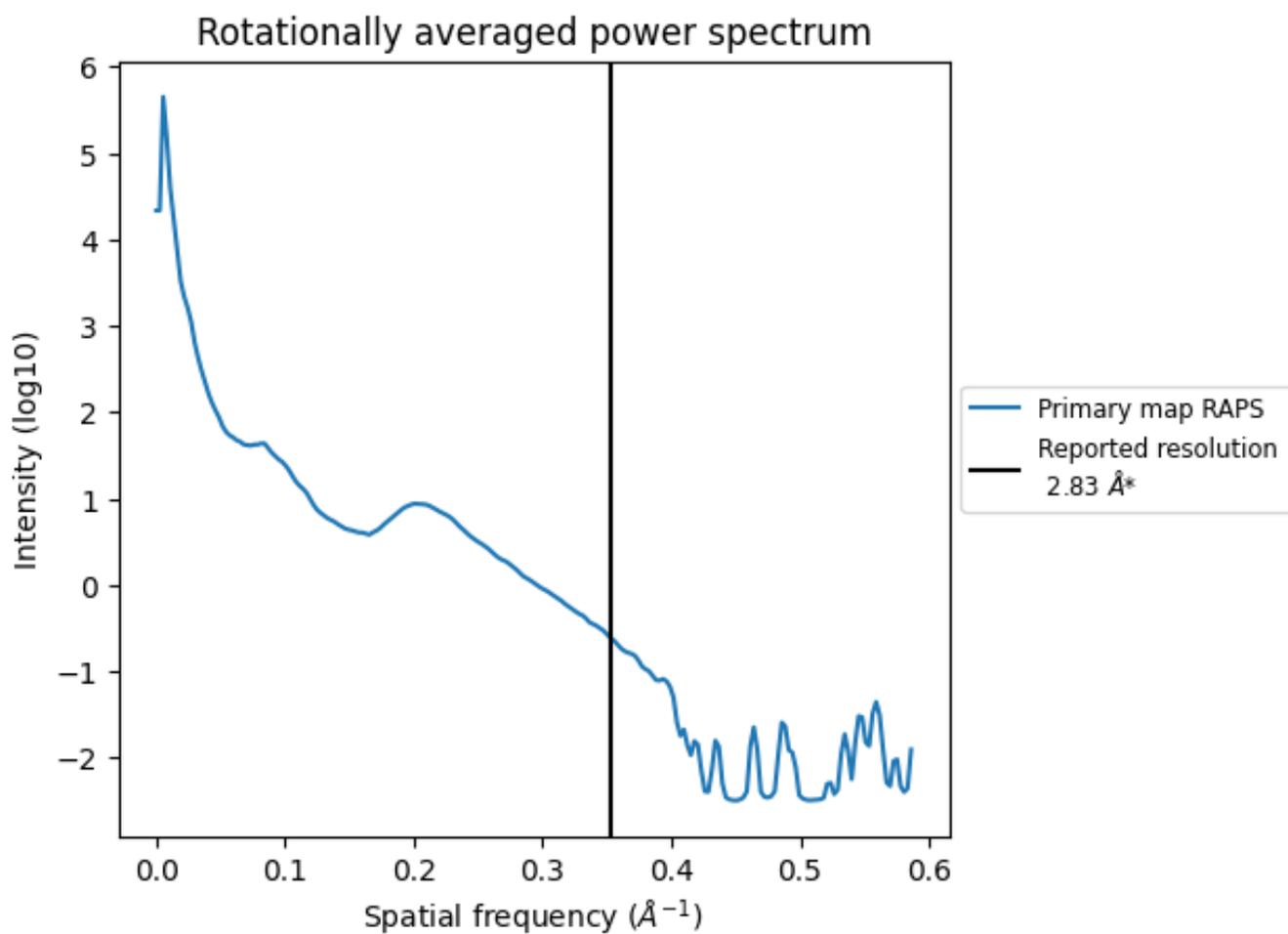
7.2 Volume estimate [i](#)



The volume at the recommended contour level is 212 nm³; this corresponds to an approximate mass of 192 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](#)

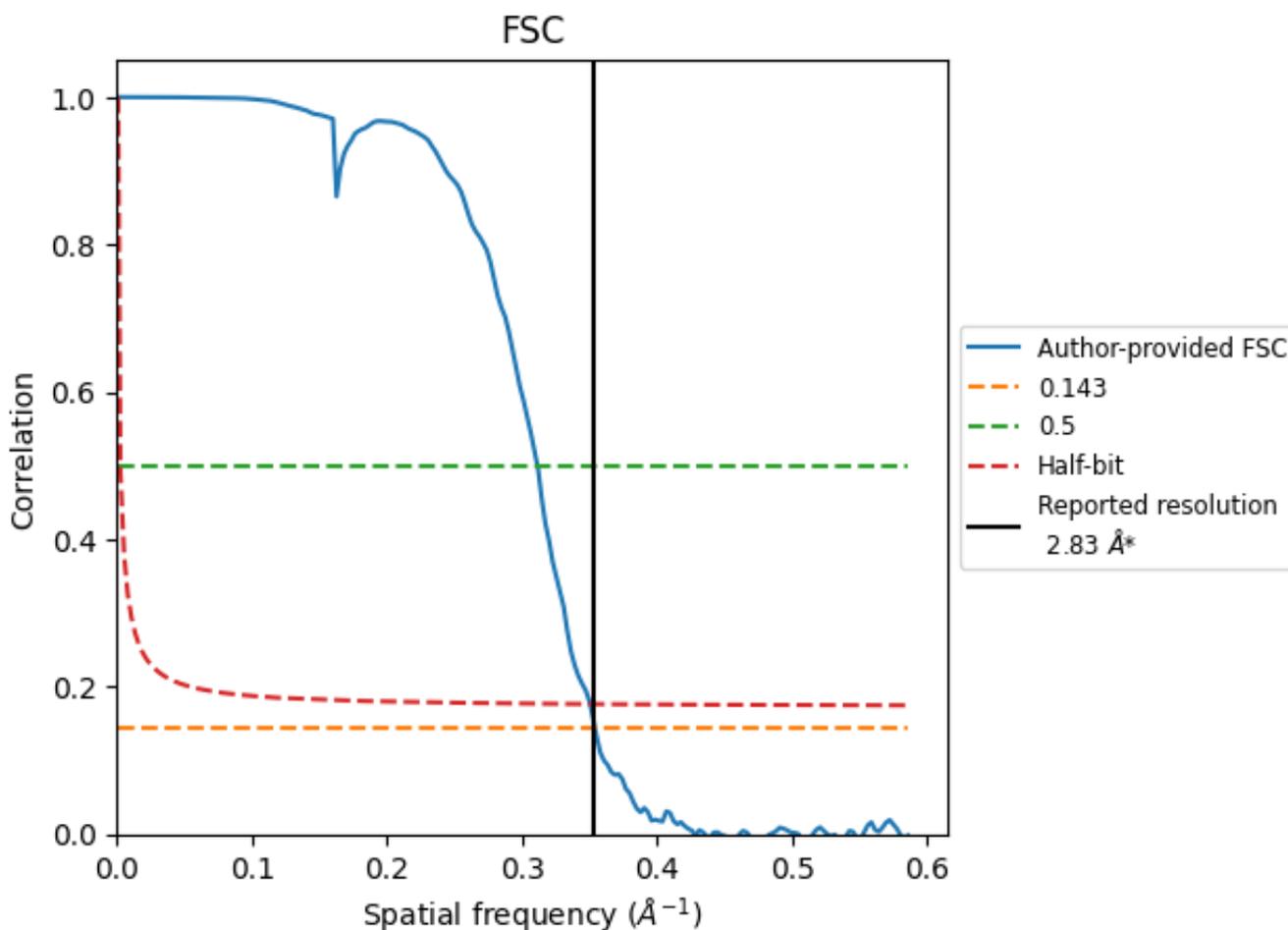


*Reported resolution corresponds to spatial frequency of 0.353\AA^{-1}

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 0.353 Å⁻¹

8.2 Resolution estimates [i](#)

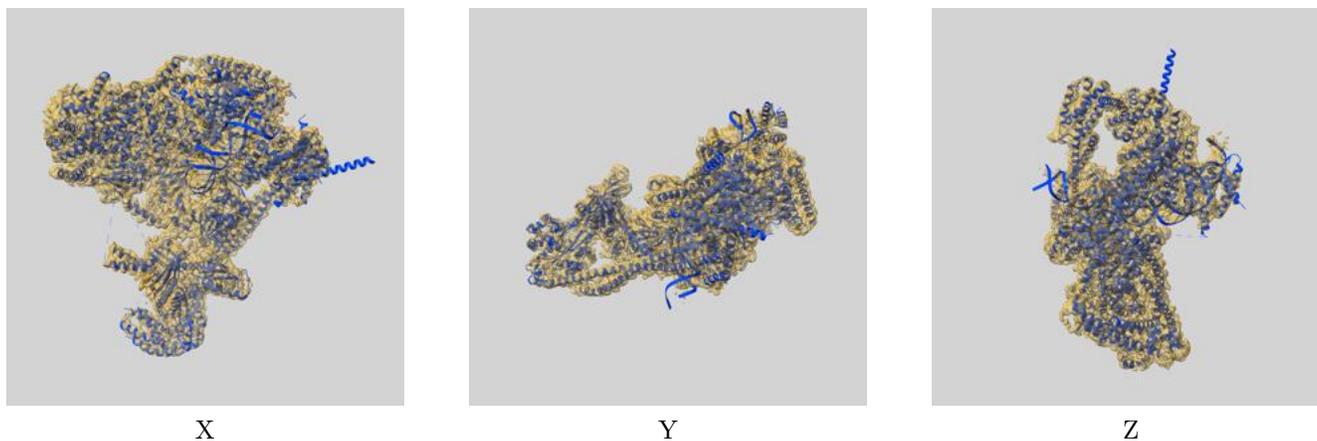
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	2.83	-	-
Author-provided FSC curve	2.82	3.21	2.85
Unmasked-calculated*	-	-	-

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

9 Map-model fit [i](#)

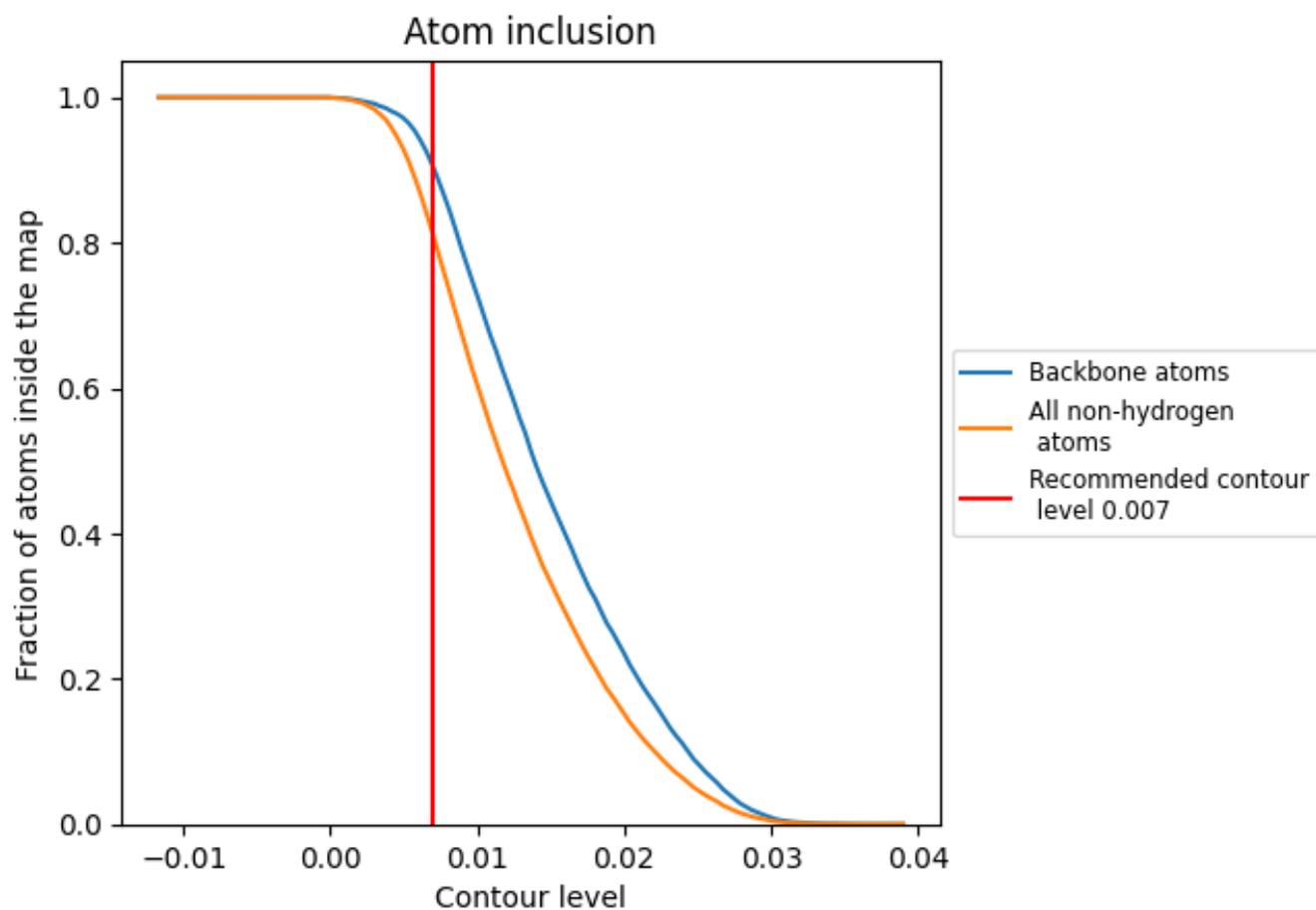
This section contains information regarding the fit between EMDB map EMD-14336 and PDB model 7R5S. Per-residue inclusion information can be found in section 3 on page 7.

9.1 Map-model overlay [i](#)



The images above show the 3D surface view of the map at the recommended contour level 0.007 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 Atom inclusion [i](#)



At the recommended contour level, 90% of all backbone atoms, 81% of all non-hydrogen atoms, are inside the map.