wwPDB EM Validation Summary Report

Oct 6, 2024 – 01:43 AM JST

PDB ID	:	7BY0
EMDB ID	:	EMD-30239
Title	:	The cryo-EM structure of CENP-A nucleosome in complex with the phosphorylated CENP-C
Authors	:	Ariyoshi, M.; Makino, F.; Fukagawa, T.
Deposited on	:	2020-04-21
Resolution	:	4.50 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:

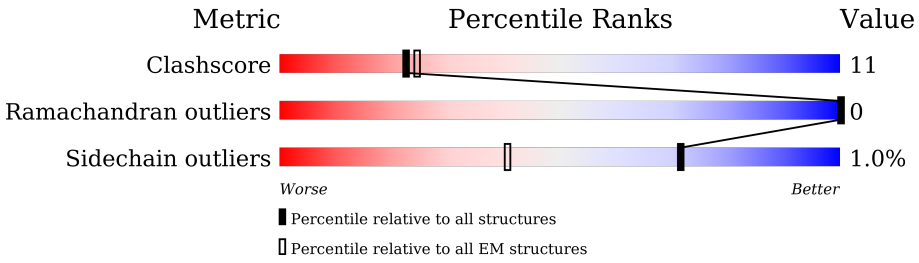
EMDB validation analysis	:	0.0.1.dev113
Mogul	:	1.8.5 (274361), CSD as541be (2020)
MolProbity	:	4.02b-467
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
MapQ	:	1.9.13
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 4.50 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.



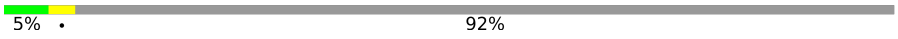
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	210492	15764
Ramachandran outliers	207382	16835
Sidechain outliers	206894	16415

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	141	55% 18% 27%
1	E	141	6% 45% 27% 27%
2	B	102	53% 24% 24%
2	F	102	52% 25% 24%
3	C	130	63% 20% 17%
3	G	130	62% 22% 17%
4	D	126	60% 14% 25%
4	H	126	58% 17% 25%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	I	145	
6	J	145	
7	K	664	
7	L	664	

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 12693 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 H3.2,Histone H3-like centromeric protein A.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	103	Total	C	N	O	S	0	0
			854	544	167	140	3		
1	E	103	Total	C	N	O	S	0	0
			854	544	167	140	3		

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	78	Total	C	N	O	S	0	0
			629	397	123	108	1		
2	F	78	Total	C	N	O	S	0	0
			629	397	123	108	1		

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

Mol	Chain	Residues	Atoms				AltConf	Trace
3	C	108	Total	C	N	O	0	0
			832	525	164	143		
3	G	108	Total	C	N	O	0	0
			835	526	165	144		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
4	D	94	Total	C	N	O	S	0	0
			732	459	133	138	2		
4	H	94	Total	C	N	O	S	0	0
			736	462	134	138	2		

- Molecule 5 is a DNA chain called DNA (145-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
5	I	143	Total	C	N	O	P	0	0
			2914	1384	530	857	143		

- Molecule 6 is a DNA chain called DNA (145-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
6	J	143	Total	C	N	O	P	0	0
			2949	1396	551	859	143		

- Molecule 7 is a protein called Maltose Binding Protein tag, linker,CENP-C.

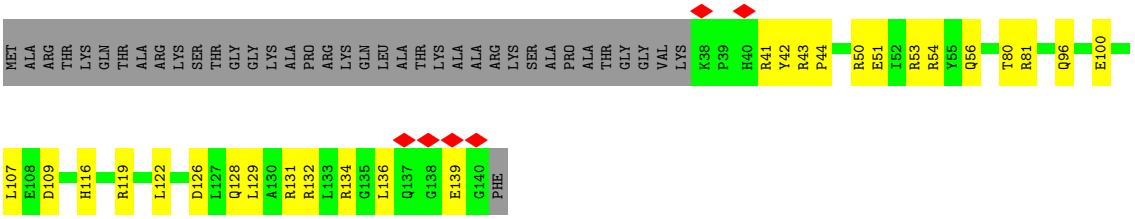
Mol	Chain	Residues	Atoms					AltConf	Trace
7	K	51	Total	C	N	O	P	0	0
			363	223	75	64	1		
7	L	51	Total	C	N	O	P	0	0
			366	226	75	64	1		

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.

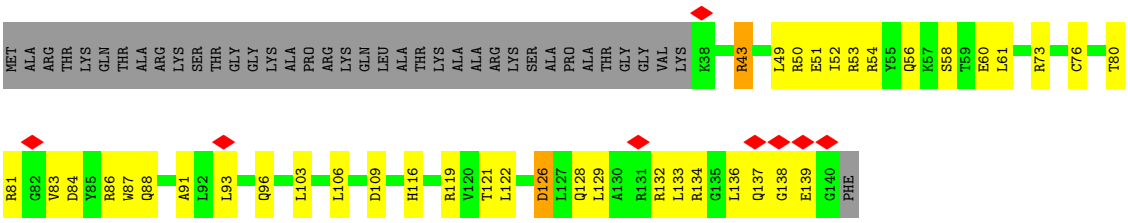
• Molecule 1: Histone H3.2,Histone H3-like centromeric protein A

Chain A:



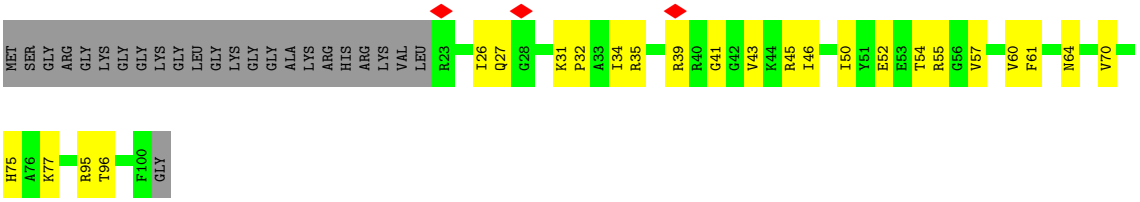
• Molecule 1: Histone H3.2,Histone H3-like centromeric protein A

Chain E:



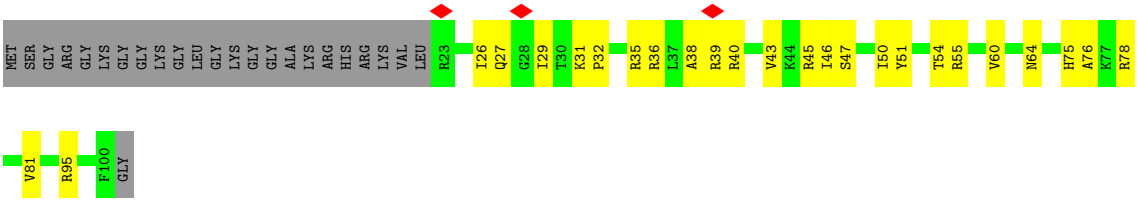
• Molecule 2: Histone H4

Chain B:



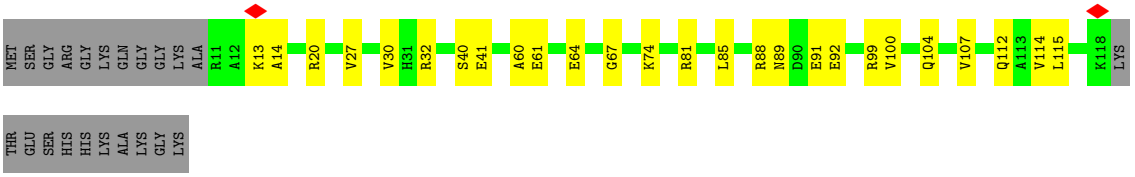
• Molecule 2: Histone H4

Chain F:



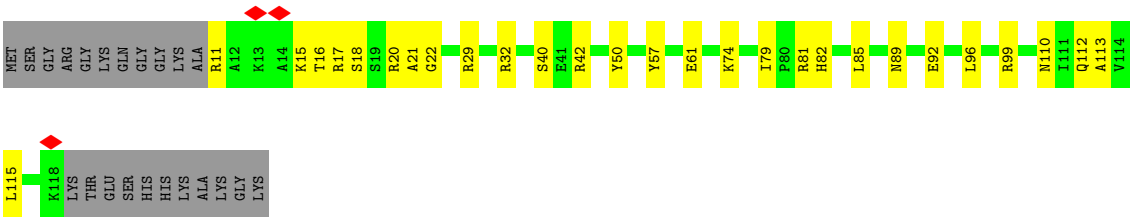
• Molecule 3: Histone H2A type 1-B/E

Chain C:



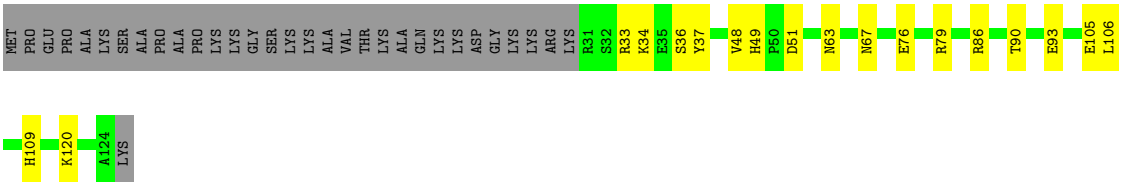
• Molecule 3: Histone H2A type 1-B/E

Chain G:



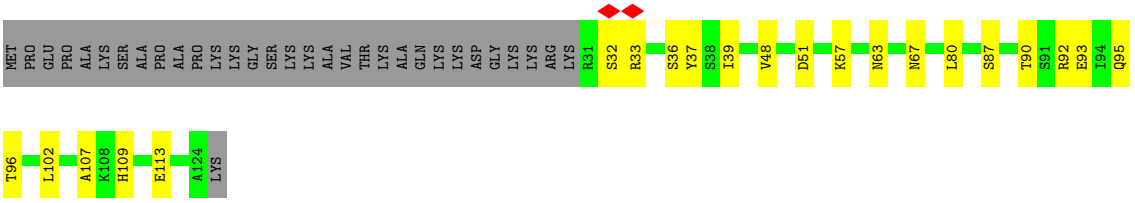
• Molecule 4: Histone H2B type 1-J

Chain D:



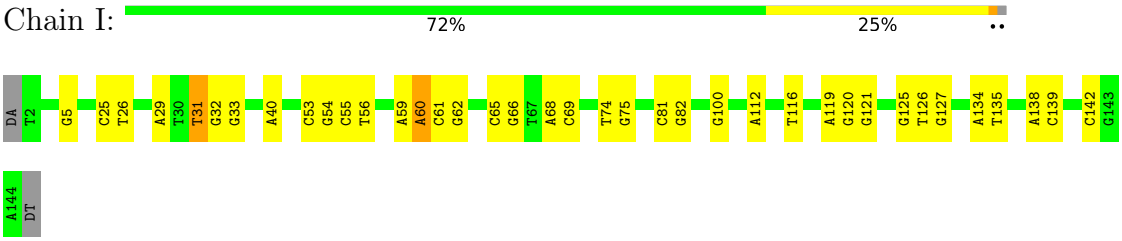
• Molecule 4: Histone H2B type 1-J

Chain H:



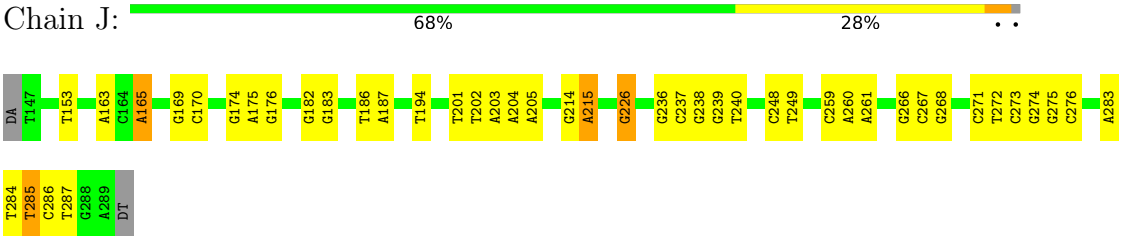
• Molecule 5: DNA (145-MER)

Chain I:



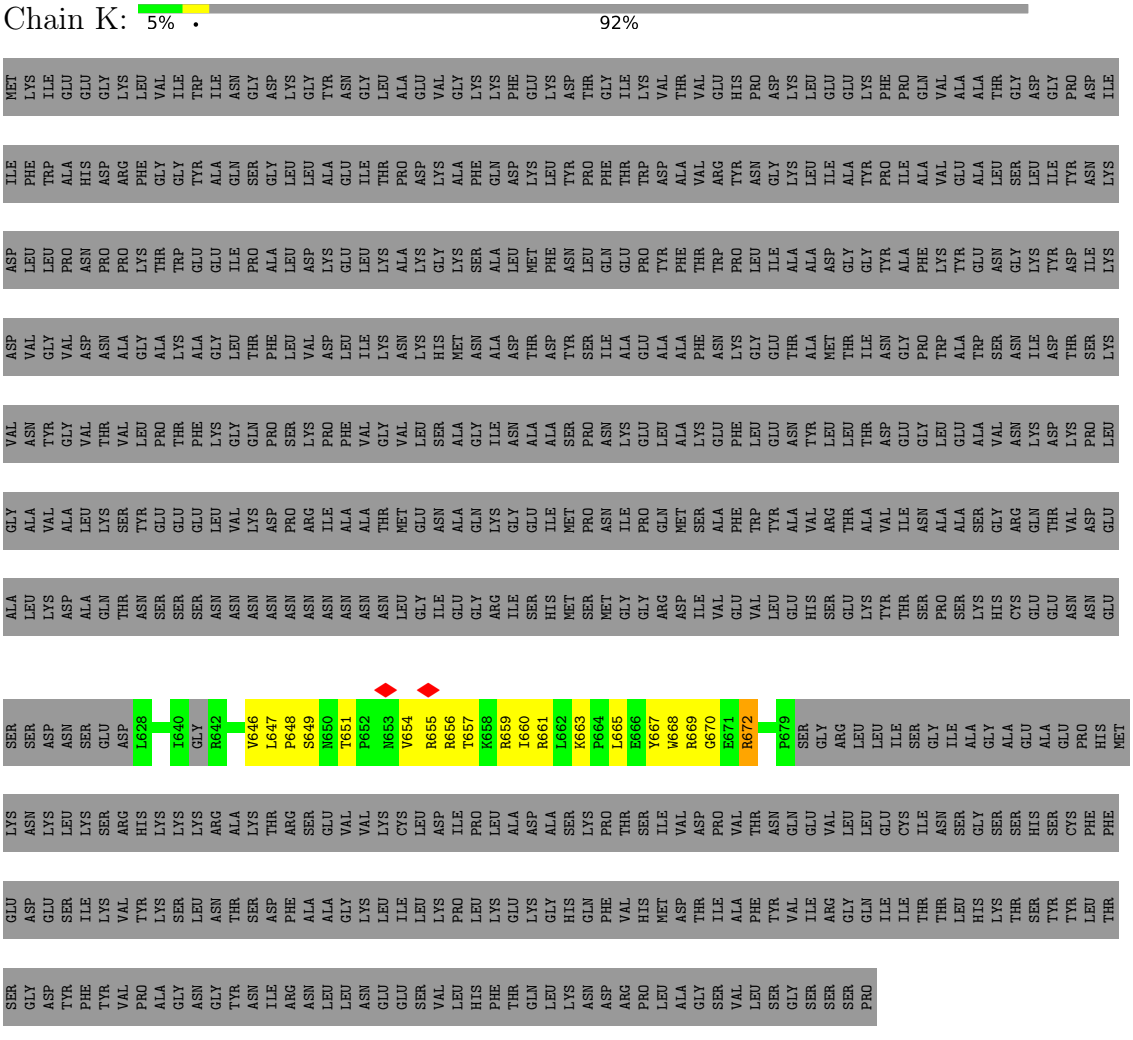
• Molecule 6: DNA (145-MER)

Chain J:



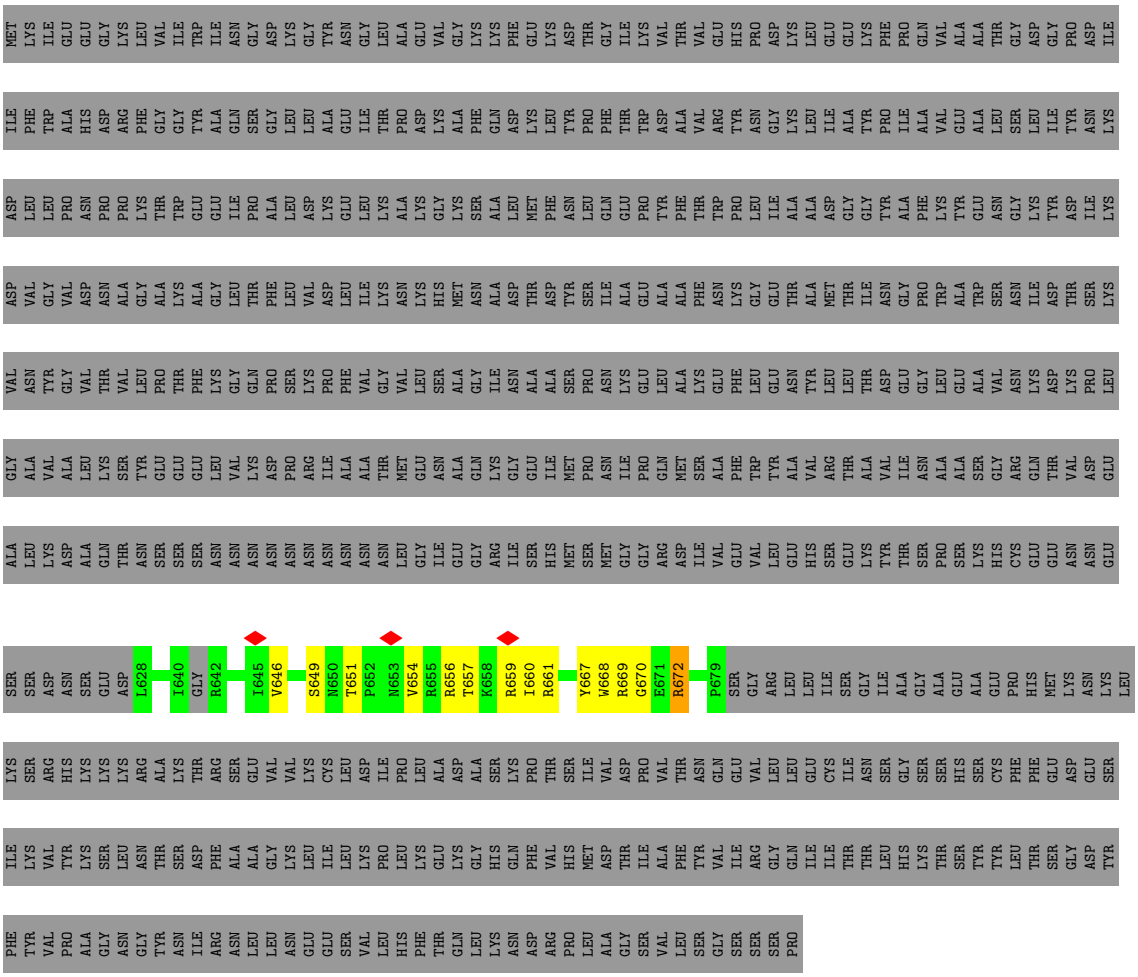
• Molecule 7: Maltose Binding Protein tag, linker,CENP-C

Chain K:



• Molecule 7: Maltose Binding Protein tag, linker,CENP-C

Chain L: 6% • 92%



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C2	Depositor
Number of particles used	38542	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	JEOL CRYO ARM 200	Depositor
Voltage (kV)	200	Depositor
Electron dose ($e^-/Å^2$)	1.2	Depositor
Minimum defocus (nm)	-500	Depositor
Maximum defocus (nm)	-7000	Depositor
Magnification	45454	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	0.062	Depositor
Minimum map value	-0.015	Depositor
Average map value	0.001	Depositor
Map value standard deviation	0.004	Depositor
Recommended contour level	0.02	Depositor
Map size (Å)	233.20001, 233.20001, 233.20001	wwPDB
Map dimensions	212, 212, 212	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.1, 1.1, 1.1	Depositor

5 Model quality

5.1 Standard geometry

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

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	0.33	0/870	0.57	0/1172
1	E	0.35	0/870	0.56	0/1172
2	B	0.35	0/636	0.59	0/852
2	F	0.34	0/636	0.54	0/852
3	C	0.31	0/842	0.53	0/1135
3	G	0.33	0/845	0.54	0/1139
4	D	0.32	0/743	0.50	0/1000
4	H	0.34	0/747	0.48	0/1004
5	I	0.91	1/3265 (0.0%)	1.12	5/5032 (0.1%)
6	J	0.92	1/3311 (0.0%)	1.08	5/5113 (0.1%)
7	K	0.29	0/355	0.59	0/478
7	L	0.30	0/358	0.61	0/482
All	All	0.68	2/13478 (0.0%)	0.88	10/19431 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	J	165	DA	C1'-N9	-7.02	1.37	1.47
5	I	60	DA	N9-C4	-5.26	1.34	1.37

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	I	32	DG	O4'-C4'-C3'	-7.26	101.60	104.50
5	I	55	DC	O4'-C4'-C3'	-6.00	102.10	104.50
6	J	286	DC	O4'-C4'-C3'	-5.76	102.20	104.50
6	J	163	DA	O4'-C4'-C3'	-5.63	102.25	104.50
6	J	285	DT	C3'-C2'-C1'	-5.42	95.99	102.50

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	854	0	891	35	0
1	E	854	0	891	40	0
2	B	629	0	670	18	0
2	F	629	0	670	24	0
3	C	832	0	893	24	0
3	G	835	0	897	26	0
4	D	732	0	747	26	0
4	H	736	0	758	19	0
5	I	2914	0	1605	31	0
6	J	2949	0	1606	38	0
7	K	363	0	306	41	0
7	L	366	0	315	18	0
All	All	12693	0	10249	233	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 11.

The worst 5 of 233 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:131:ARG:NH1	7:K:668:TRP:HA	1.61	1.15
3:C:91:GLU:HG3	7:K:665:LEU:HB3	1.09	1.05
3:C:92:GLU:OE1	7:K:659:ARG:NH1	1.95	0.98
3:G:92:GLU:OE2	7:L:659:ARG:NH1	2.05	0.90
3:C:91:GLU:HG3	7:K:665:LEU:CB	2.00	0.87

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	101/141 (72%)	99 (98%)	2 (2%)	0	100	100
1	E	101/141 (72%)	98 (97%)	3 (3%)	0	100	100
2	B	76/102 (74%)	74 (97%)	2 (3%)	0	100	100
2	F	76/102 (74%)	74 (97%)	2 (3%)	0	100	100
3	C	106/130 (82%)	103 (97%)	3 (3%)	0	100	100
3	G	106/130 (82%)	102 (96%)	4 (4%)	0	100	100
4	D	92/126 (73%)	92 (100%)	0	0	100	100
4	H	92/126 (73%)	88 (96%)	4 (4%)	0	100	100
7	K	46/664 (7%)	43 (94%)	3 (6%)	0	100	100
7	L	46/664 (7%)	43 (94%)	3 (6%)	0	100	100
All	All	842/2326 (36%)	816 (97%)	26 (3%)	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	87/113 (77%)	86 (99%)	1 (1%)	70	80
1	E	87/113 (77%)	84 (97%)	3 (3%)	32	53
2	B	65/79 (82%)	65 (100%)	0	100	100
2	F	65/79 (82%)	64 (98%)	1 (2%)	60	75

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	C	84/100 (84%)	84 (100%)	0	100	100
3	G	85/100 (85%)	85 (100%)	0	100	100
4	D	79/105 (75%)	79 (100%)	0	100	100
4	H	80/105 (76%)	80 (100%)	0	100	100
7	K	26/562 (5%)	25 (96%)	1 (4%)	28	50
7	L	27/562 (5%)	26 (96%)	1 (4%)	29	51
All	All	685/1918 (36%)	678 (99%)	7 (1%)	71	81

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

Mol	Chain	Res	Type
1	E	126	ASP
2	F	31	LYS
7	L	672	ARG
7	K	672	ARG
1	E	81	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
3	G	68	ASN
4	H	95	GLN
3	C	104	GLN
4	D	63	ASN
4	D	95	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)

2 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
7	TPO	L	651	7	8,10,11	1.04	0	10,14,16	1.80	1 (10%)
7	TPO	K	651	7	8,10,11	1.03	0	10,14,16	1.78	1 (10%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
7	TPO	L	651	7	-	2/9/11/13	-
7	TPO	K	651	7	-	1/9/11/13	-

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
7	L	651	TPO	P-OG1-CB	-5.20	107.50	123.21
7	K	651	TPO	P-OG1-CB	-5.15	107.64	123.21

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
7	K	651	TPO	CB-OG1-P-O1P
7	L	651	TPO	CB-OG1-P-O1P
7	L	651	TPO	CB-OG1-P-O3P

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no oligosaccharides in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

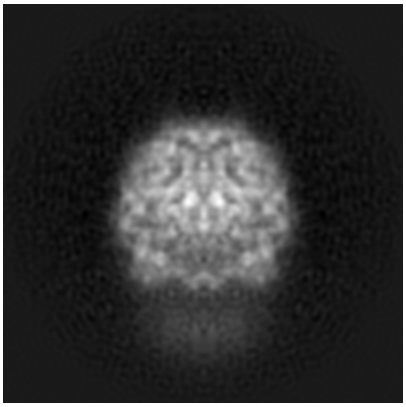
6 Map visualisation

This section contains visualisations of the EMDB entry EMD-30239. These allow visual inspection of the internal detail of the map and identification of artifacts.

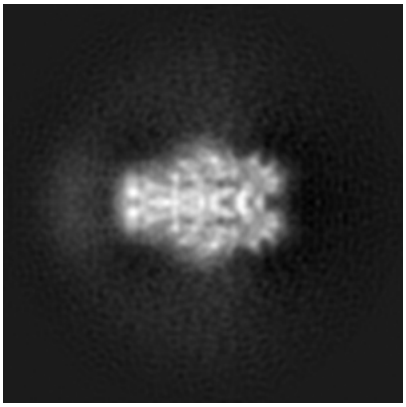
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections

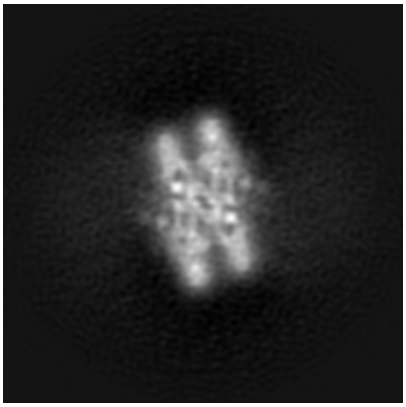
6.1.1 Primary map



X



Y



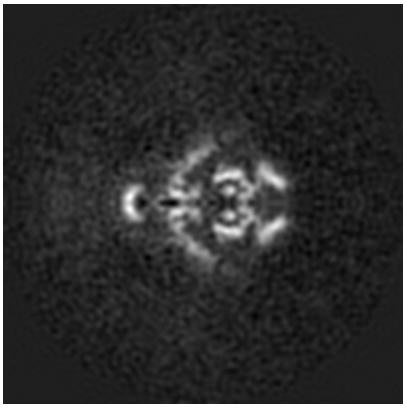
Z

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

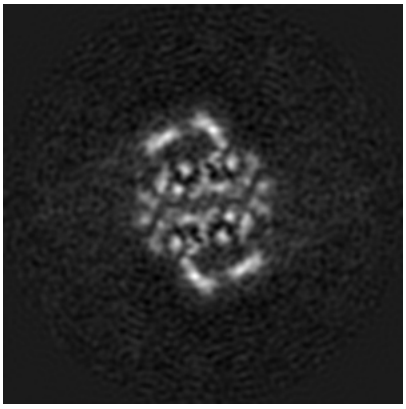
6.2 Central slices

6.2.1 Primary map

X Index: 106



Y Index: 106

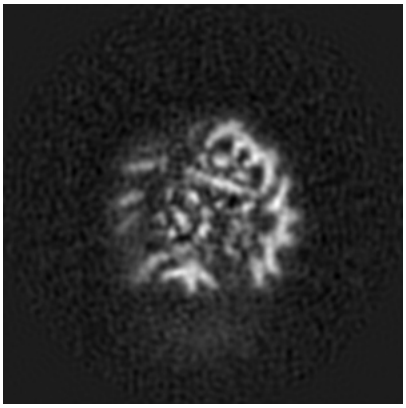


Z Index: 106

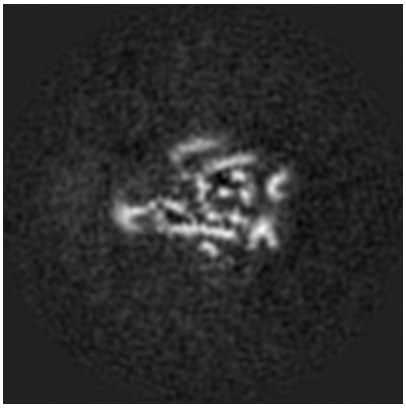
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

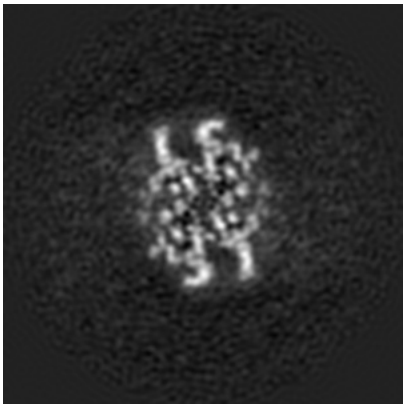
6.3.1 Primary map



X Index: 111



Y Index: 114

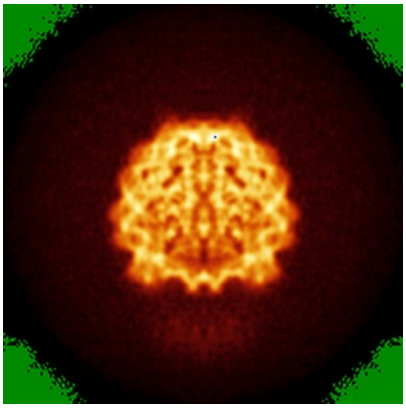


Z Index: 112

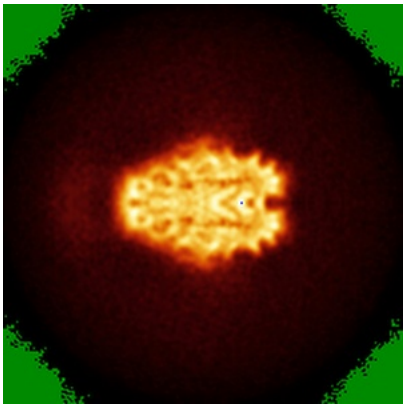
The images above show the largest variance slices of the map in three orthogonal directions.

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

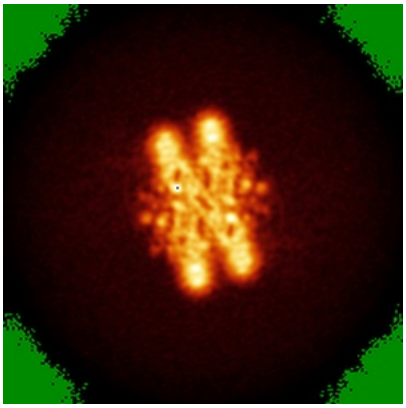
6.4.1 Primary map



X



Y

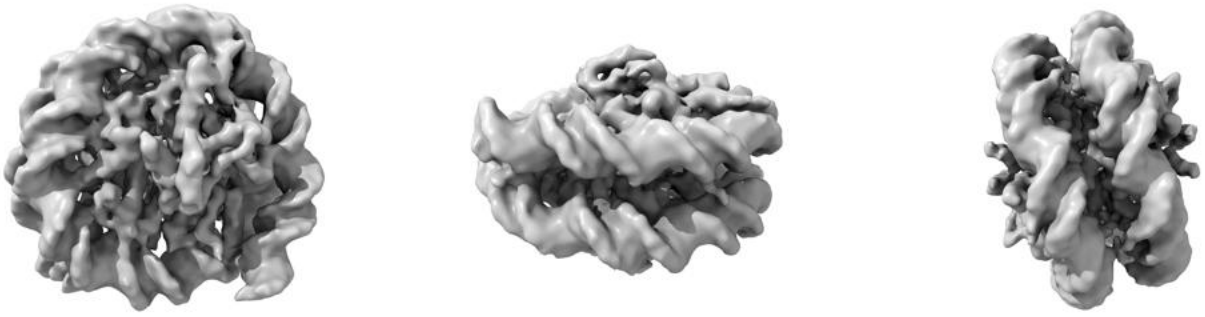


Z

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

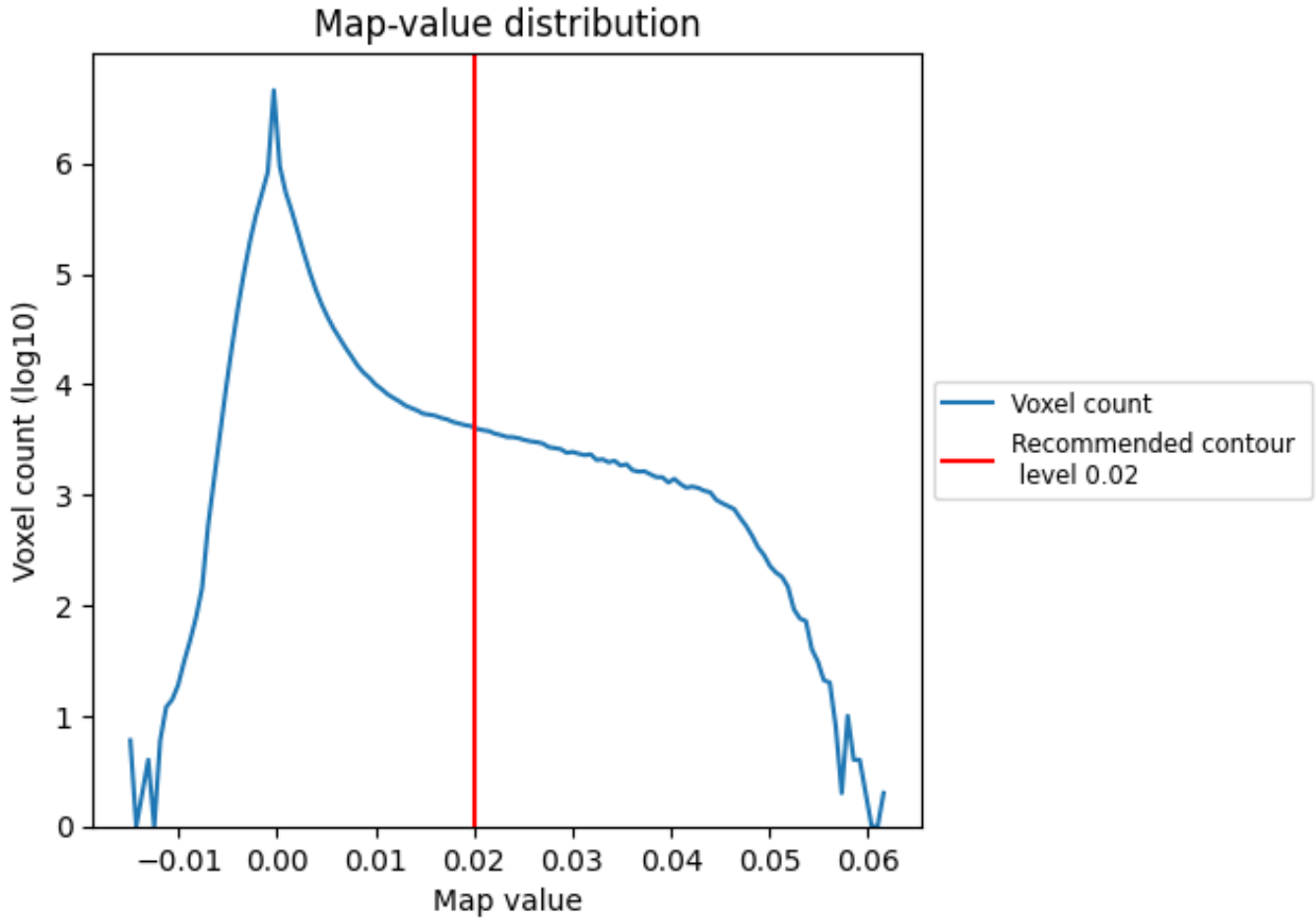
6.6 Mask visualisation (i)

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

7 Map analysis

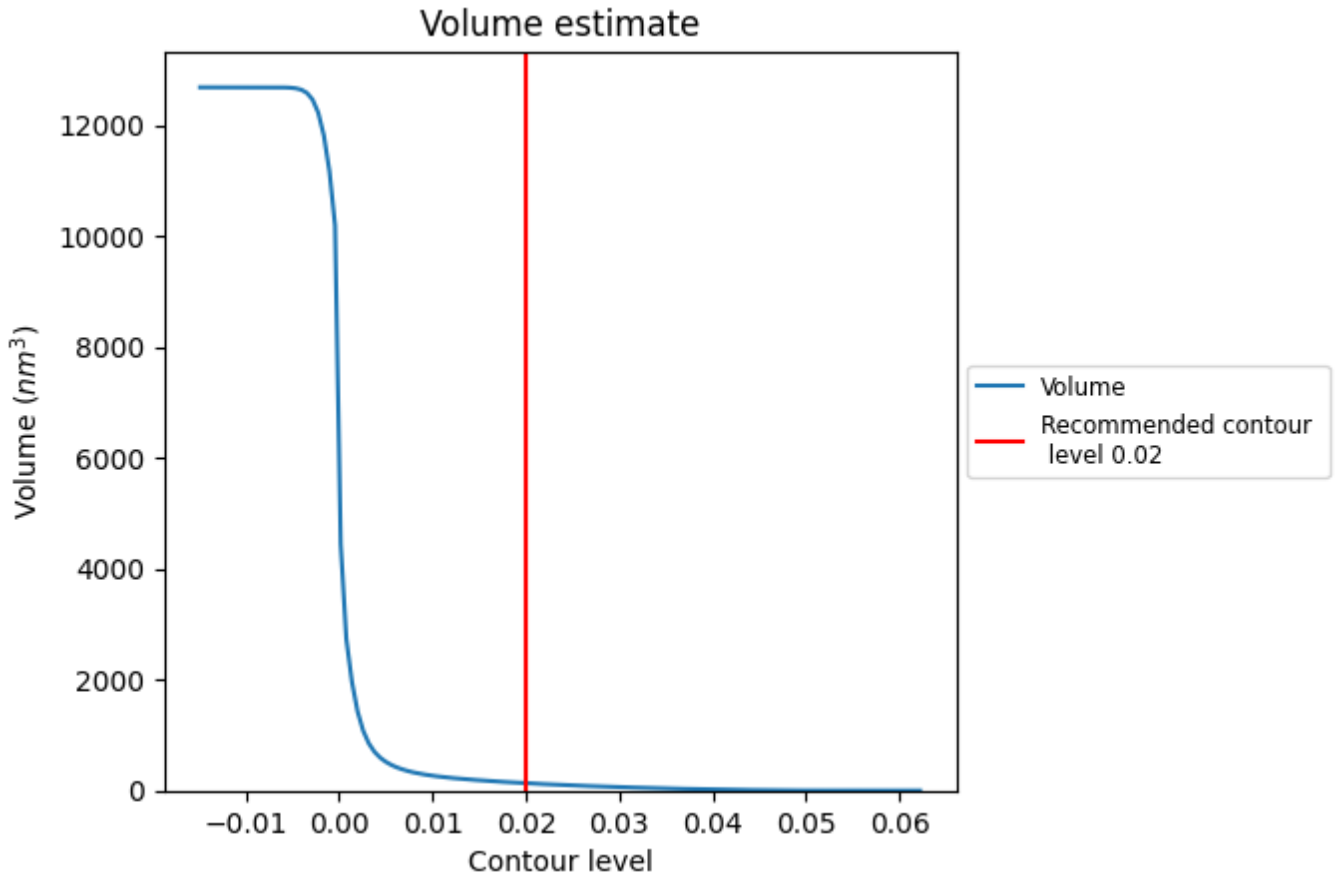
This section contains the results of statistical analysis of the map.

7.1 Map-value distribution



The map-value distribution is plotted in 128 intervals along the x-axis. The y-axis is logarithmic. A spike in this graph at zero usually indicates that the volume has been masked.

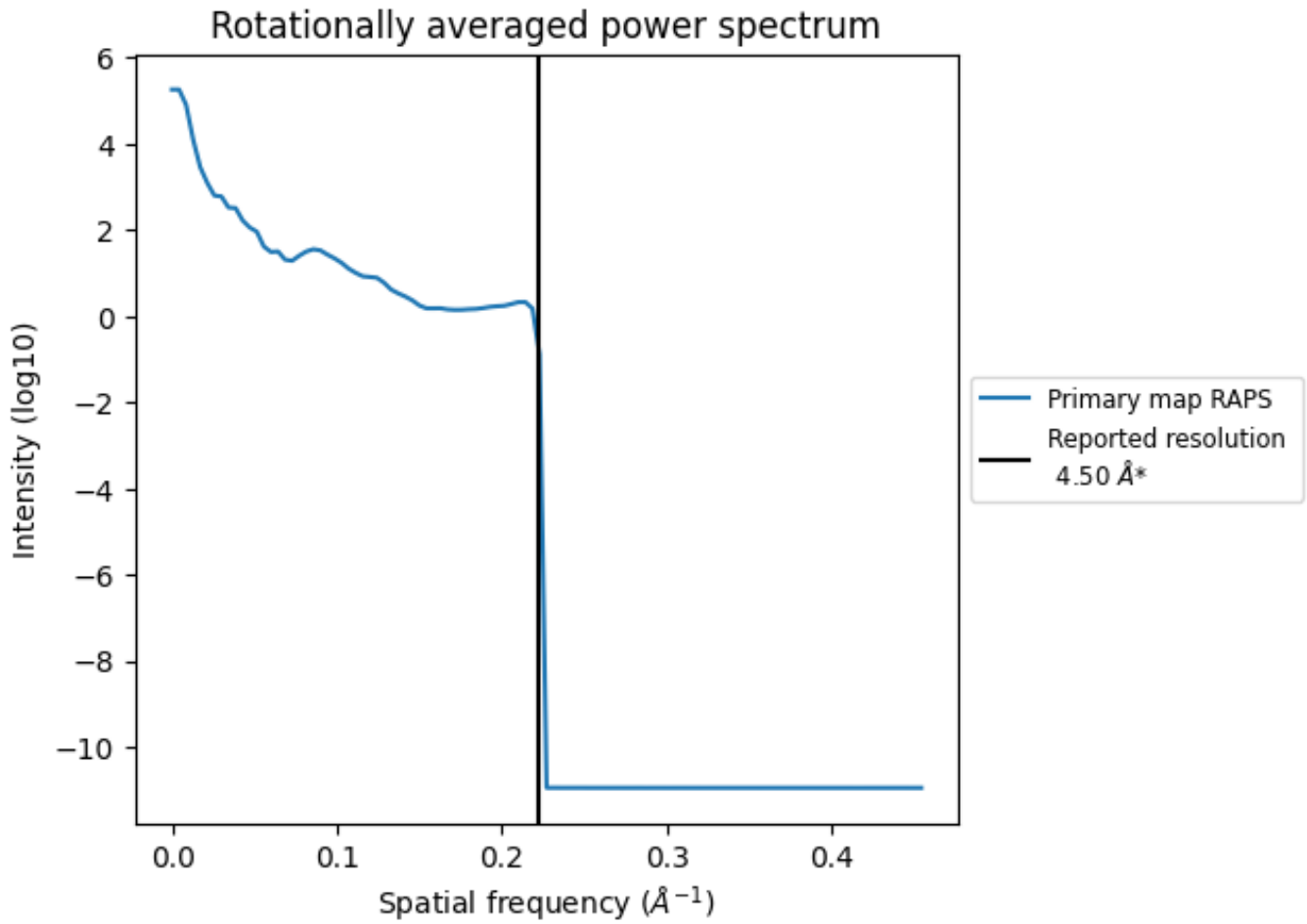
7.2 Volume estimate (i)



The volume at the recommended contour level is 133 nm^3; this corresponds to an approximate mass of 120 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)

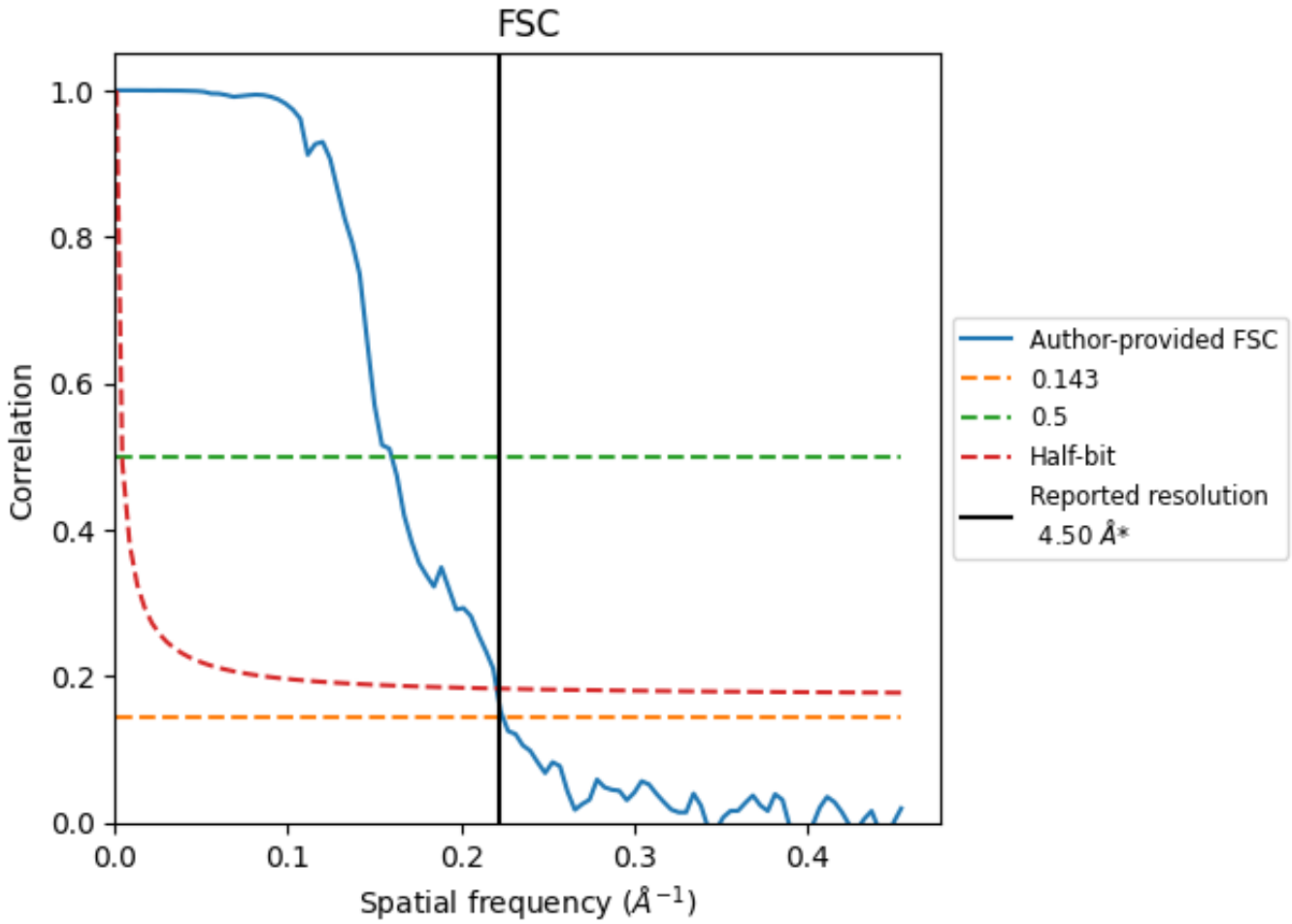


*Reported resolution corresponds to spatial frequency of 0.222 Å^{-1}

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 0.222 Å^{-1}

8.2 Resolution estimates (i)

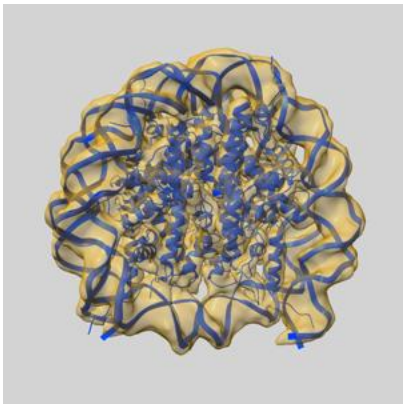
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	4.50	-	-
Author-provided FSC curve	4.46	6.25	4.53
Unmasked-calculated*	-	-	-

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

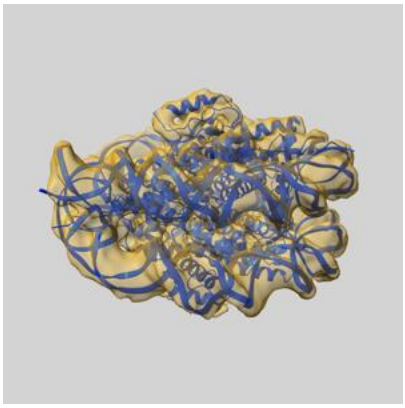
9 Map-model fit

This section contains information regarding the fit between EMDB map EMD-30239 and PDB model 7BY0. Per-residue inclusion information can be found in section 3 on page 6.

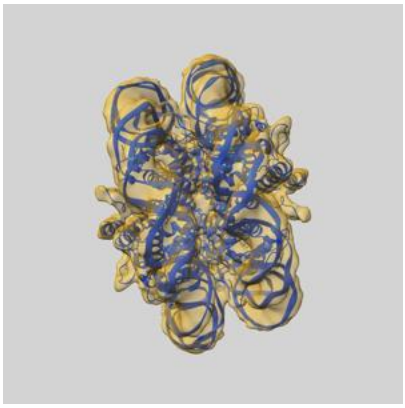
9.1 Map-model overlay



X



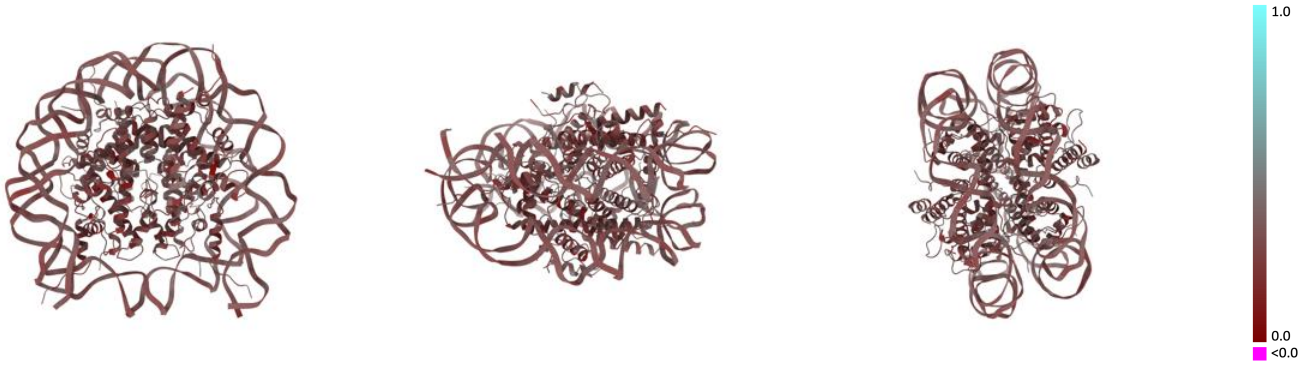
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.02 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.

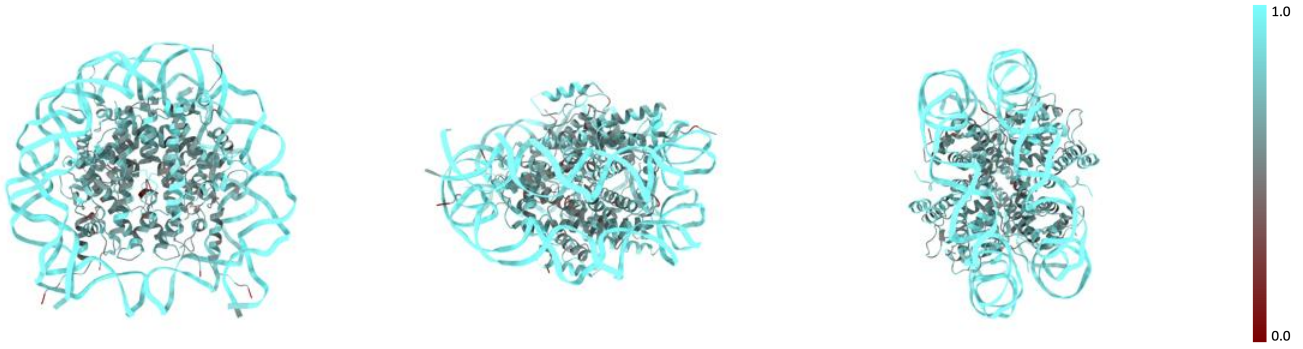
9.2 Q-score mapped to coordinate model (i)



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.

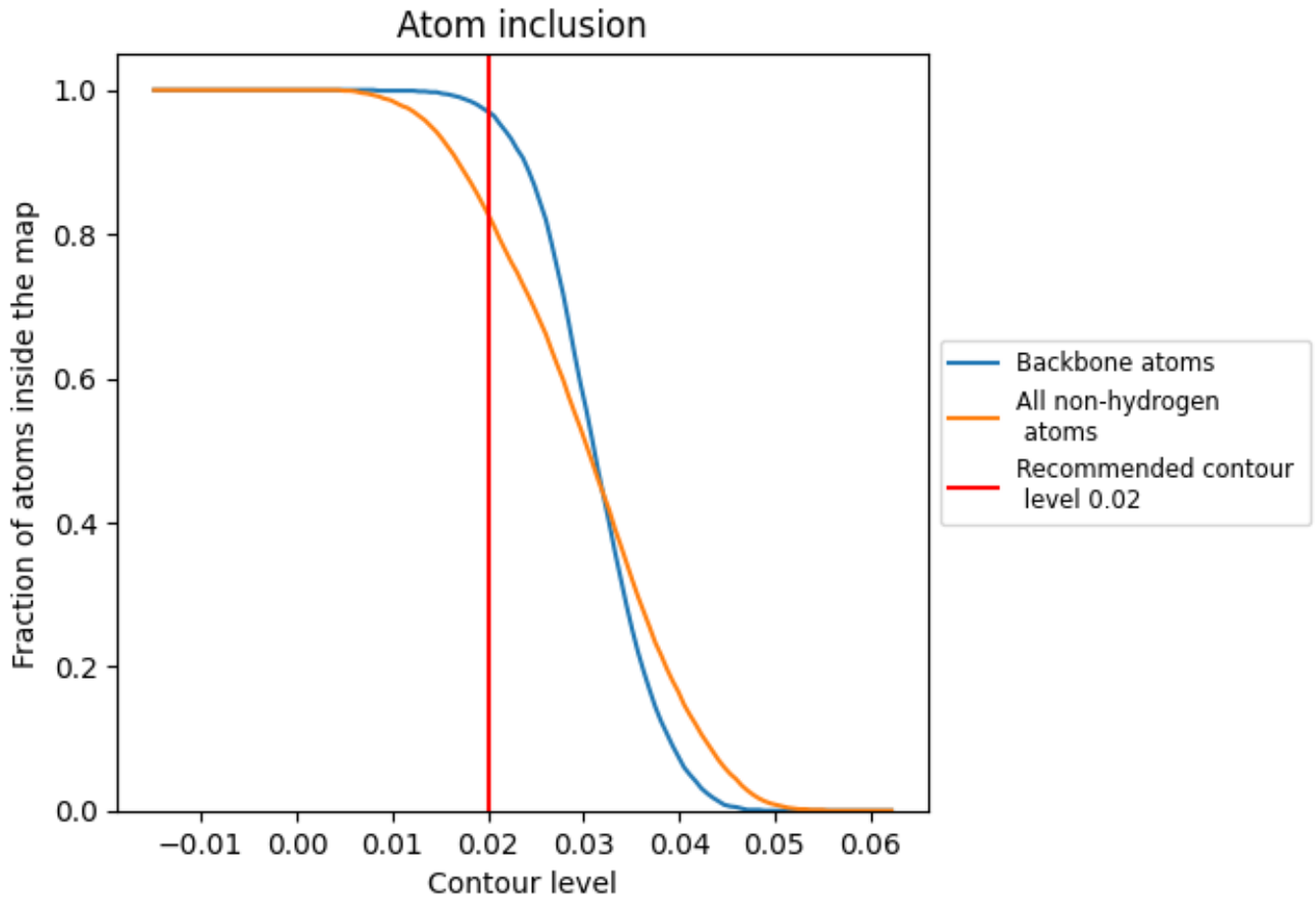
9.3 Atom inclusion mapped to coordinate model (i)



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 (0.02).

9.4 Atom inclusion (i)



At the recommended contour level, 97% of all backbone atoms, 83% 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 (0.02) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.8280	0.2970
A	0.6720	0.2680
B	0.6970	0.2710
C	0.7220	0.2900
D	0.7530	0.2950
E	0.6720	0.2690
F	0.7080	0.2850
G	0.7030	0.2870
H	0.7290	0.2920
I	0.9590	0.3090
J	0.9590	0.3070
K	0.7590	0.3340
L	0.7470	0.3350