



wwPDB EM Validation Summary Report ⓘ

Nov 21, 2022 – 09:26 PM JST

PDB ID : 7E8D
EMDB ID : EMD-31015
Title : NSD2 E1099K mutant bound to nucleosome
Authors : Sengoku, T.; Sato, K.; Nishizawa, T.; Nureki, O.; Ogata, K.
Deposited on : 2021-03-01
Resolution : 2.80 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.dev43
Mogul : 1.8.5 (274361), CSD as541be (2020)
MolProbity : 4.02b-467
buster-report : 1.1.7 (2018)
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ : 1.9.9
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.31.3

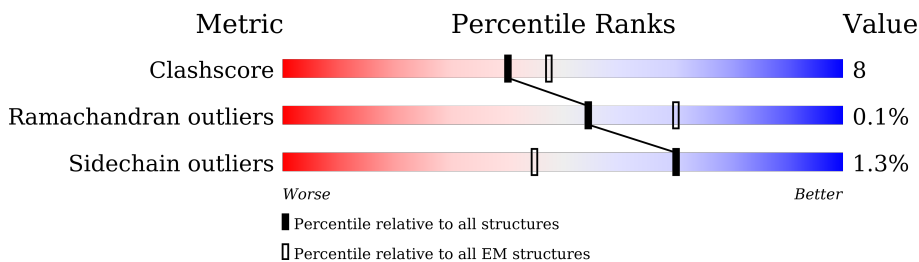
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 2.80 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.



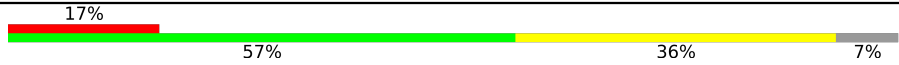


Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

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	A	135	
1	E	135	
2	B	102	
2	F	102	
3	C	129	
3	G	129	
4	D	125	
4	H	125	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	I	185	
6	J	185	
7	K	254	

2 Entry composition [i](#)

There are 9 unique types of molecules in this entry. The entry contains 14852 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	801	505	155	137	4	0	0
1	E	104	845	532	163	145	5	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	36	MET	LYS	variant	UNP P68431
E	36	MET	LYS	variant	UNP P68431

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	B	79	633	399	124	109	1	0	0
2	F	78	622	393	120	108	1	0	0

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

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
3	C	109	833	526	163	144	0	0
3	G	109	841	532	165	144	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	94	736	462	134	138	2	0	0

Continued on next page...

Continued from previous page...

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

- Molecule 5 is a DNA chain called DNA (185-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
5	I	172	Total	C	N	O	P	0	0
			3504	1663	635	1034	172		

- Molecule 6 is a DNA chain called DNA (185-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
6	J	172	Total	C	N	O	P	0	0
			3548	1677	669	1030	172		

- Molecule 7 is a protein called Histone-lysine N-methyltransferase NSD2.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	K	218	Total	C	N	O	S	1	0
			1714	1076	295	322	21		

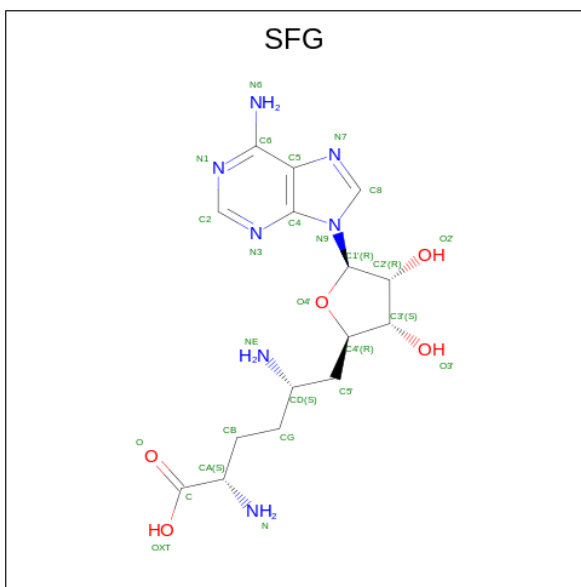
There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
K	1099	LYS	GLU	engineered mutation	UNP O96028

- Molecule 8 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		AltConf
8	K	3	Total	Zn	0
			3	3	

- Molecule 9 is SINEFUNGIN (three-letter code: SFG) (formula: C₁₅H₂₃N₇O₅) (labeled as "Ligand of Interest" by depositor).

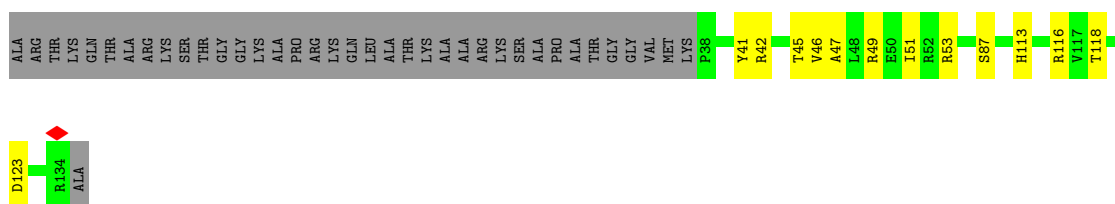


Mol	Chain	Residues	Atoms				AltConf
			Total	C	N	O	
9	K	1	27	15	7	5	0

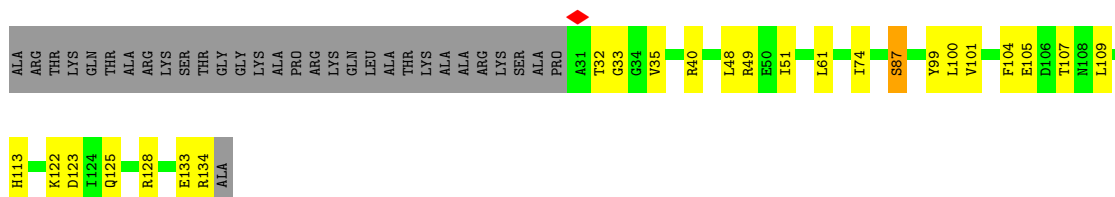
3 Residue-property plots [i](#)

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.

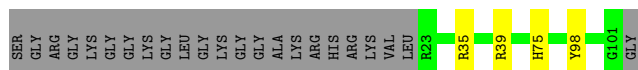
- Molecule 1: Histone H3.1



- Molecule 1: Histone H3.1



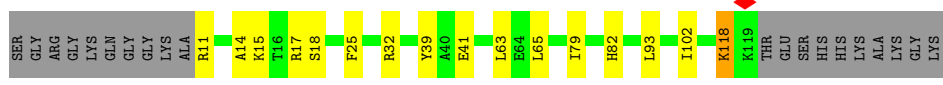
- Molecule 2: Histone H4



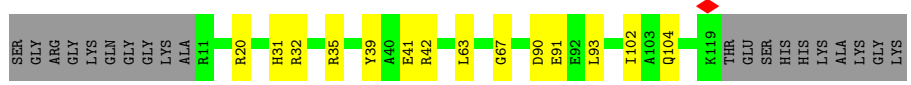
- Molecule 2: Histone H4



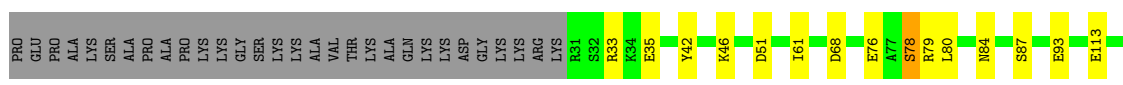
- Molecule 3: Histone H2A type 1



• Molecule 3: Histone H2A type 1



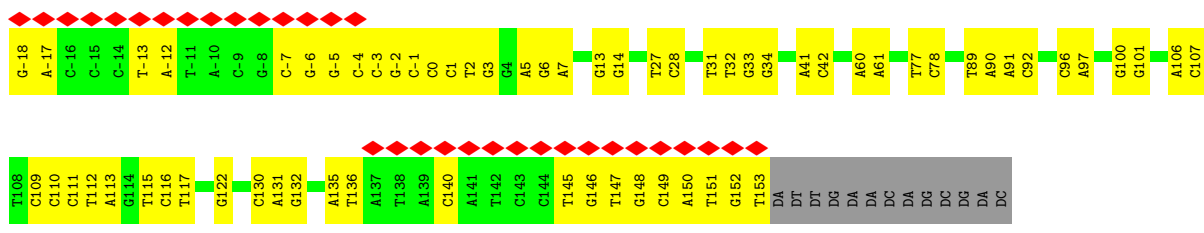
• Molecule 4: Histone H2B type 1-J



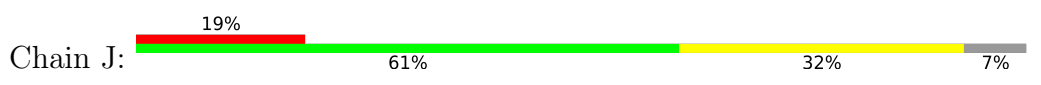
• Molecule 4: Histone H2B type 1-J

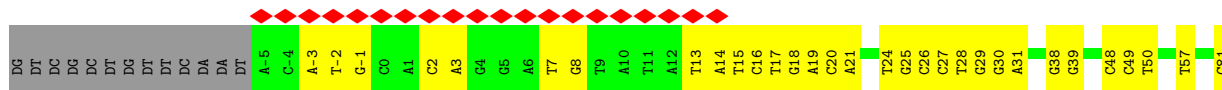


• Molecule 5: DNA (185-MER)

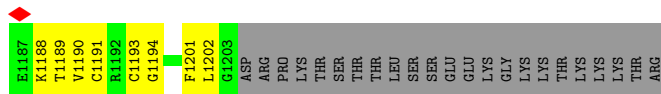
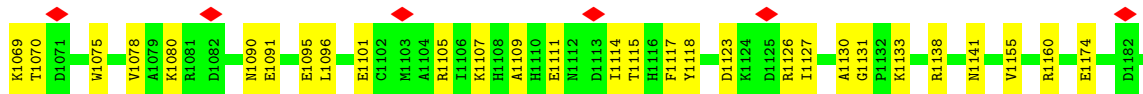
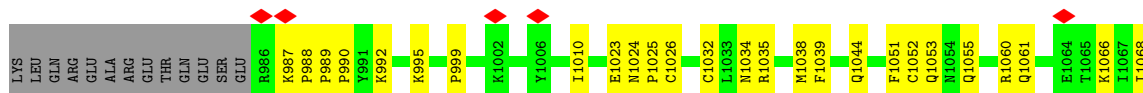


• Molecule 6: DNA (185-MER)





• Molecule 7: Histone-lysine N-methyltransferase NSD2



4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	136108	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 ($e^-/\text{\AA}^2$)	50	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K3 BIOQUANTUM (6k x 4k)	Depositor
Maximum map value	0.072	Depositor
Minimum map value	-0.031	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.002	Depositor
Recommended contour level	0.01	Depositor
Map size (Å)	239.04, 239.04, 239.04	wwPDB
Map dimensions	288, 288, 288	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	0.83, 0.83, 0.83	Depositor

5 Model quality

5.1 Standard geometry

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

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

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	A	0.56	0/813	0.47	0/1090
1	E	0.55	0/857	0.47	0/1149
2	B	0.57	0/640	0.51	0/857
2	F	0.60	0/629	0.48	0/843
3	C	0.49	0/843	0.48	0/1138
3	G	0.50	0/851	0.47	0/1146
4	D	0.50	0/747	0.46	0/1004
4	H	0.54	0/756	0.48	0/1015
5	I	1.04	0/3925	0.97	0/6050
6	J	1.06	0/3985	0.94	0/6154
7	K	0.38	0/1757	0.46	0/2374
All	All	0.82	0/15803	0.77	0/22820

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
3	C	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	C	15	LYS	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	A	801	0	839	8	0
1	E	845	0	887	21	0
2	B	633	0	673	3	0
2	F	622	0	660	10	0
3	C	833	0	888	13	0
3	G	841	0	910	14	0
4	D	736	0	758	12	0
4	H	745	0	771	17	0
5	I	3504	0	1930	50	0
6	J	3548	0	1928	46	0
7	K	1714	0	1645	43	0
8	K	3	0	0	0	0
9	K	27	0	22	3	0
All	All	14852	0	11911	202	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 8.

The worst 5 of 202 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:C:11:ARG:NH1	5:I:117:DT:O2	2.07	0.87
5:I:131:DA:H2'	5:I:132:DG:C8	2.18	0.79
6:J:111:DC:H2''	6:J:112:DG:C8	2.23	0.72
6:J:18:DG:H2''	6:J:19:DA:C8	2.26	0.70
7:K:992:LYS:HB2	7:K:1133:LYS:HG3	1.73	0.69

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	95/135 (70%)	89 (94%)	6 (6%)	0	100	100
1	E	102/135 (76%)	100 (98%)	2 (2%)	0	100	100
2	B	77/102 (76%)	75 (97%)	2 (3%)	0	100	100
2	F	76/102 (74%)	75 (99%)	1 (1%)	0	100	100
3	C	107/129 (83%)	100 (94%)	6 (6%)	1 (1%)	17	46
3	G	107/129 (83%)	102 (95%)	5 (5%)	0	100	100
4	D	92/125 (74%)	89 (97%)	3 (3%)	0	100	100
4	H	93/125 (74%)	90 (97%)	3 (3%)	0	100	100
7	K	217/254 (85%)	196 (90%)	21 (10%)	0	100	100
All	All	966/1236 (78%)	916 (95%)	49 (5%)	1 (0%)	54	81

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	17	ARG

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	85/110 (77%)	83 (98%)	2 (2%)	49	81
1	E	89/110 (81%)	88 (99%)	1 (1%)	73	92
2	B	65/78 (83%)	65 (100%)	0	100	100
2	F	64/78 (82%)	63 (98%)	1 (2%)	62	88
3	C	83/98 (85%)	82 (99%)	1 (1%)	71	92
3	G	85/98 (87%)	85 (100%)	0	100	100
4	D	80/104 (77%)	77 (96%)	3 (4%)	33	67

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
4	H	81/104 (78%)	79 (98%)	2 (2%)	47	80
7	K	190/226 (84%)	189 (100%)	1 (0%)	88	96
All	All	822/1006 (82%)	811 (99%)	11 (1%)	70	91

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	F	92	ARG
4	H	87	SER
7	K	1115	THR
4	H	120	LYS
4	D	113	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 3 are monoatomic - leaving 1 for Mogul analysis.

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
9	SFG	K	1304	-	25,29,29	0.95	1 (4%)	23,42,42	1.40	3 (13%)

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
9	SFG	K	1304	-	-	2/13/33/33	0/3/3/3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
9	K	1304	SFG	C5-C4	2.27	1.46	1.40

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
9	K	1304	SFG	C3'-C2'-C1'	3.45	106.17	100.98
9	K	1304	SFG	N3-C2-N1	-3.11	123.82	128.68
9	K	1304	SFG	C4-C5-N7	-2.47	106.82	109.40

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
9	K	1304	SFG	O-C-CA-N
9	K	1304	SFG	OXT-C-CA-N

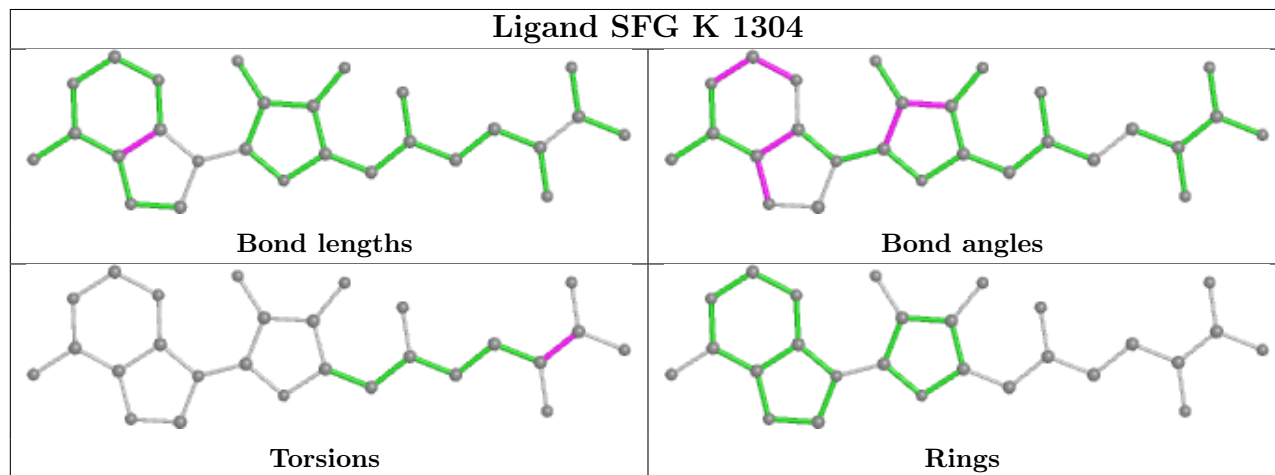
There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
9	K	1304	SFG	3	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less than 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier.

Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.



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.

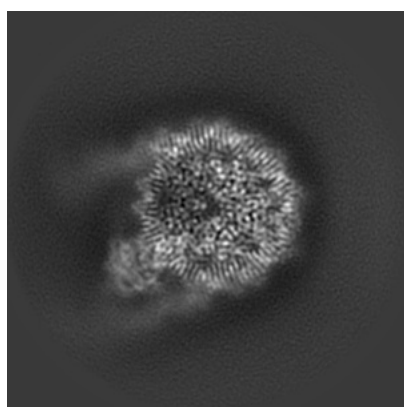
6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-31015. These allow visual inspection of the internal detail of the map and identification of artifacts.

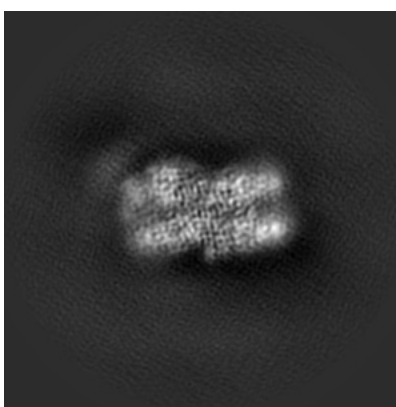
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections [i](#)

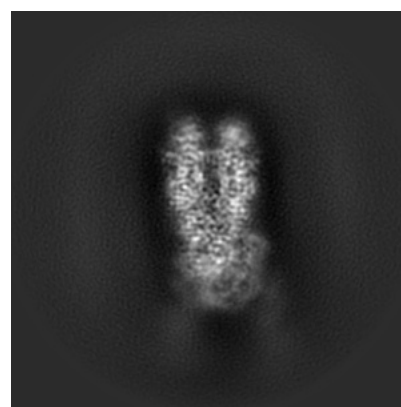
6.1.1 Primary map



X



Y

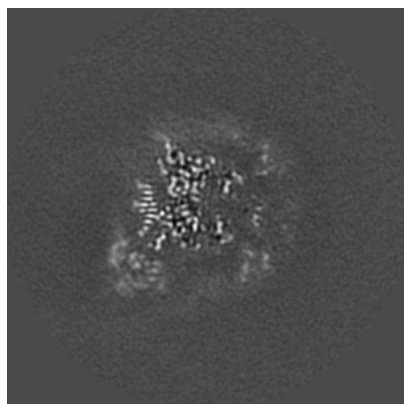


Z

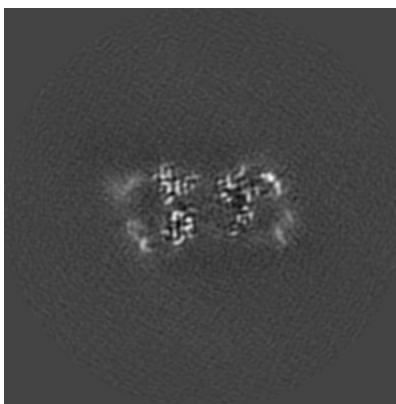
The images above show the map projected in three orthogonal directions.

6.2 Central slices [i](#)

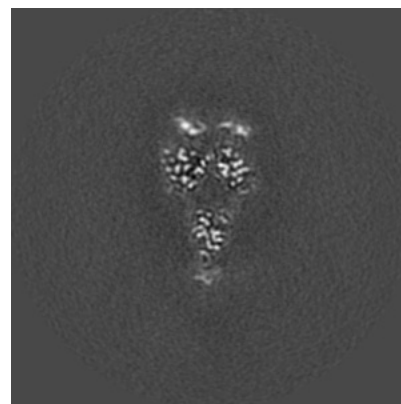
6.2.1 Primary map



X Index: 144



Y Index: 144

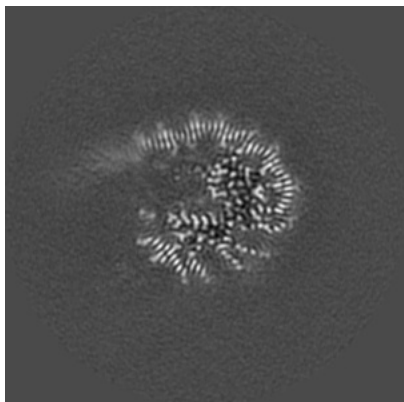


Z Index: 144

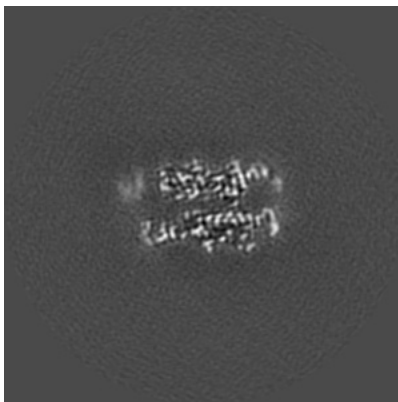
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices [i](#)

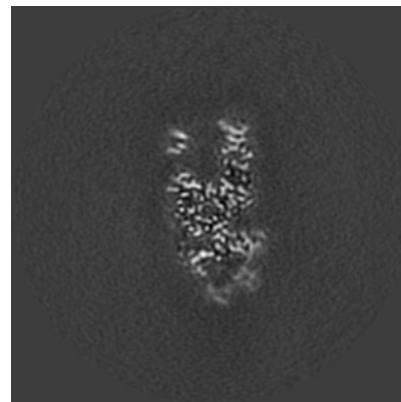
6.3.1 Primary map



X Index: 130



Y Index: 166

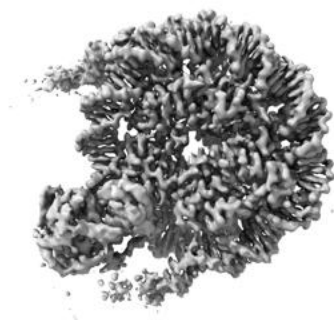


Z Index: 120

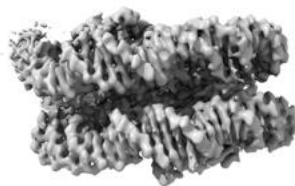
The images above show the largest variance slices of the map in three orthogonal directions.

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

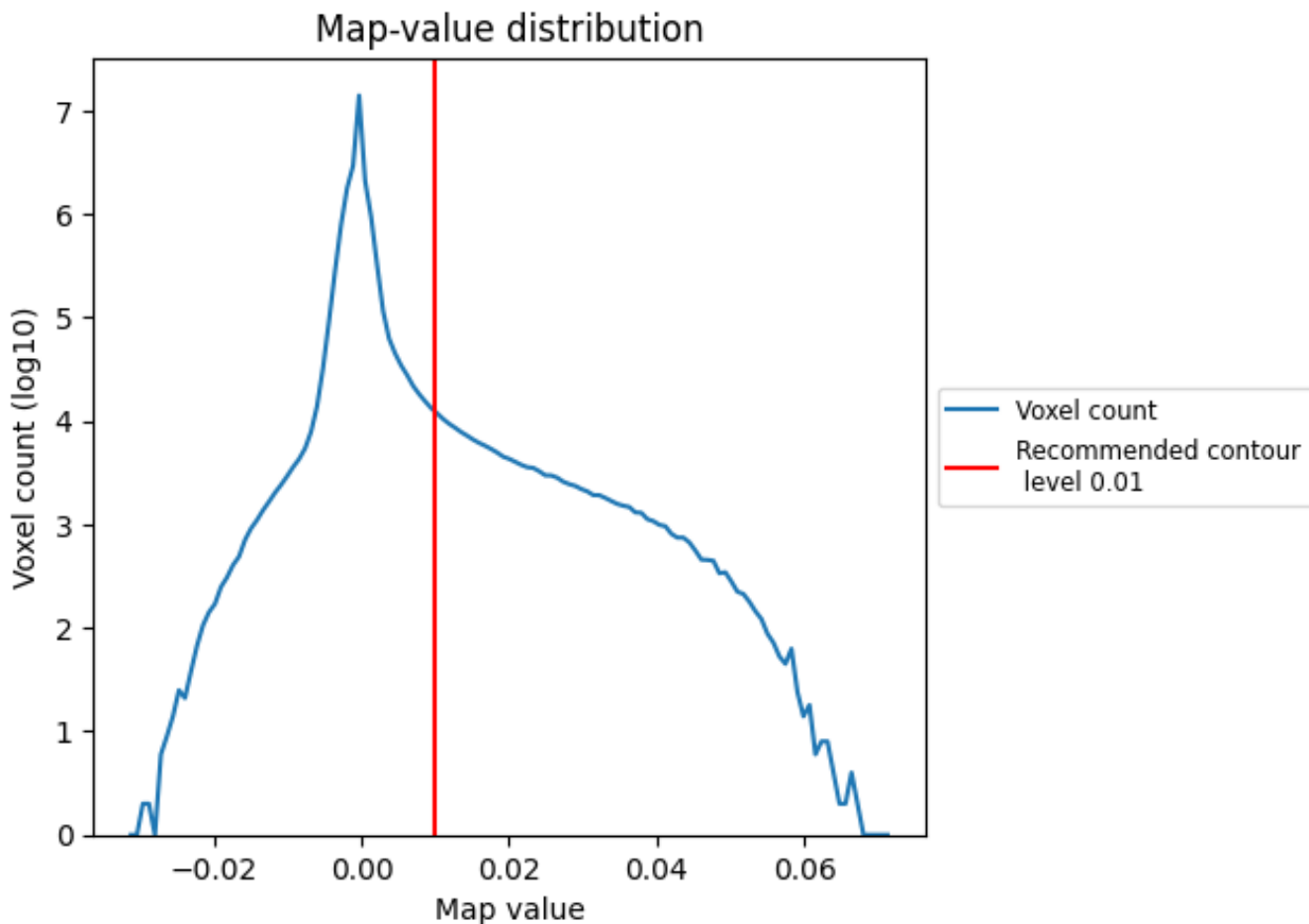
6.5 Mask visualisation

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

7 Map analysis [i](#)

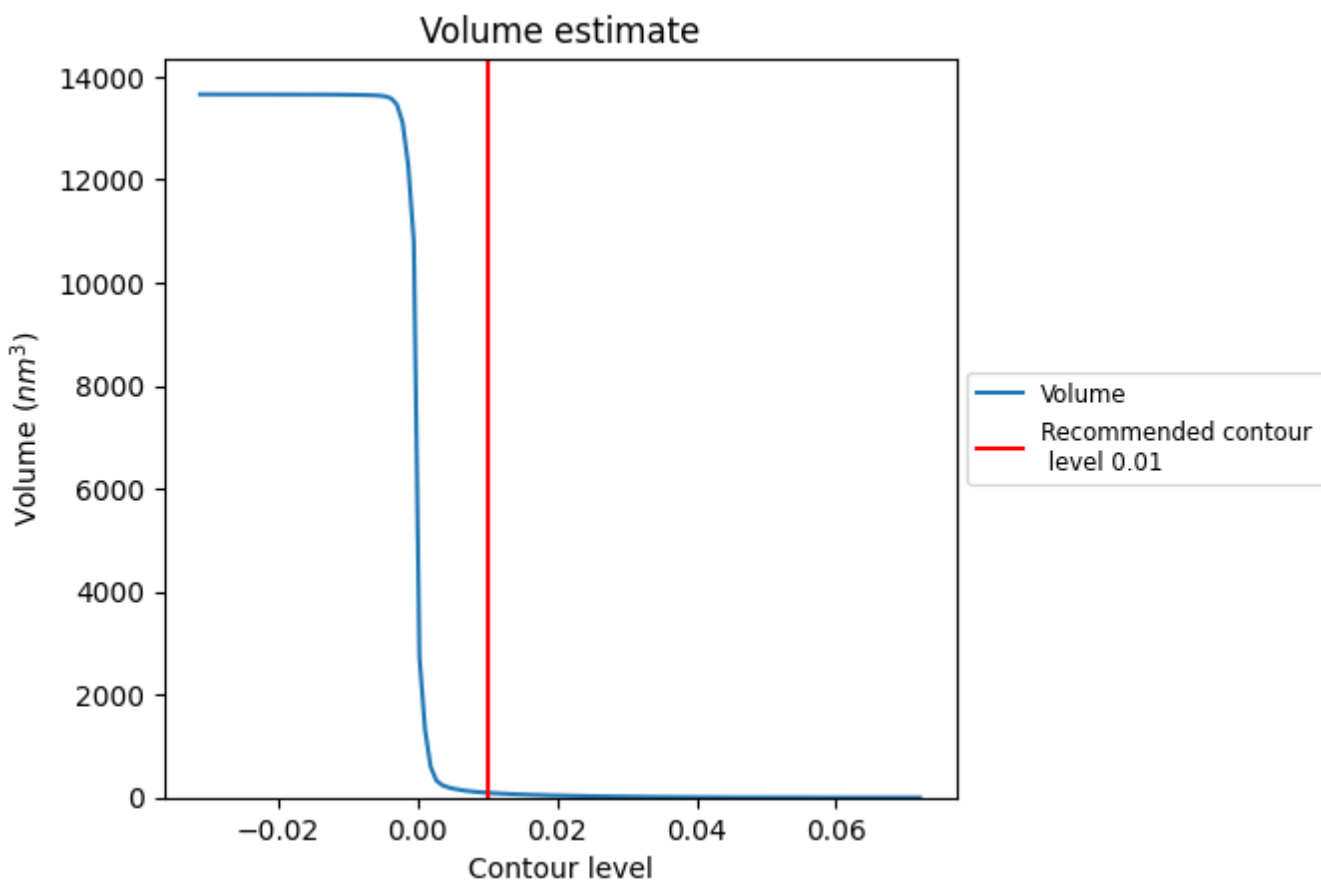
This section contains the results of statistical analysis of the map.

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.

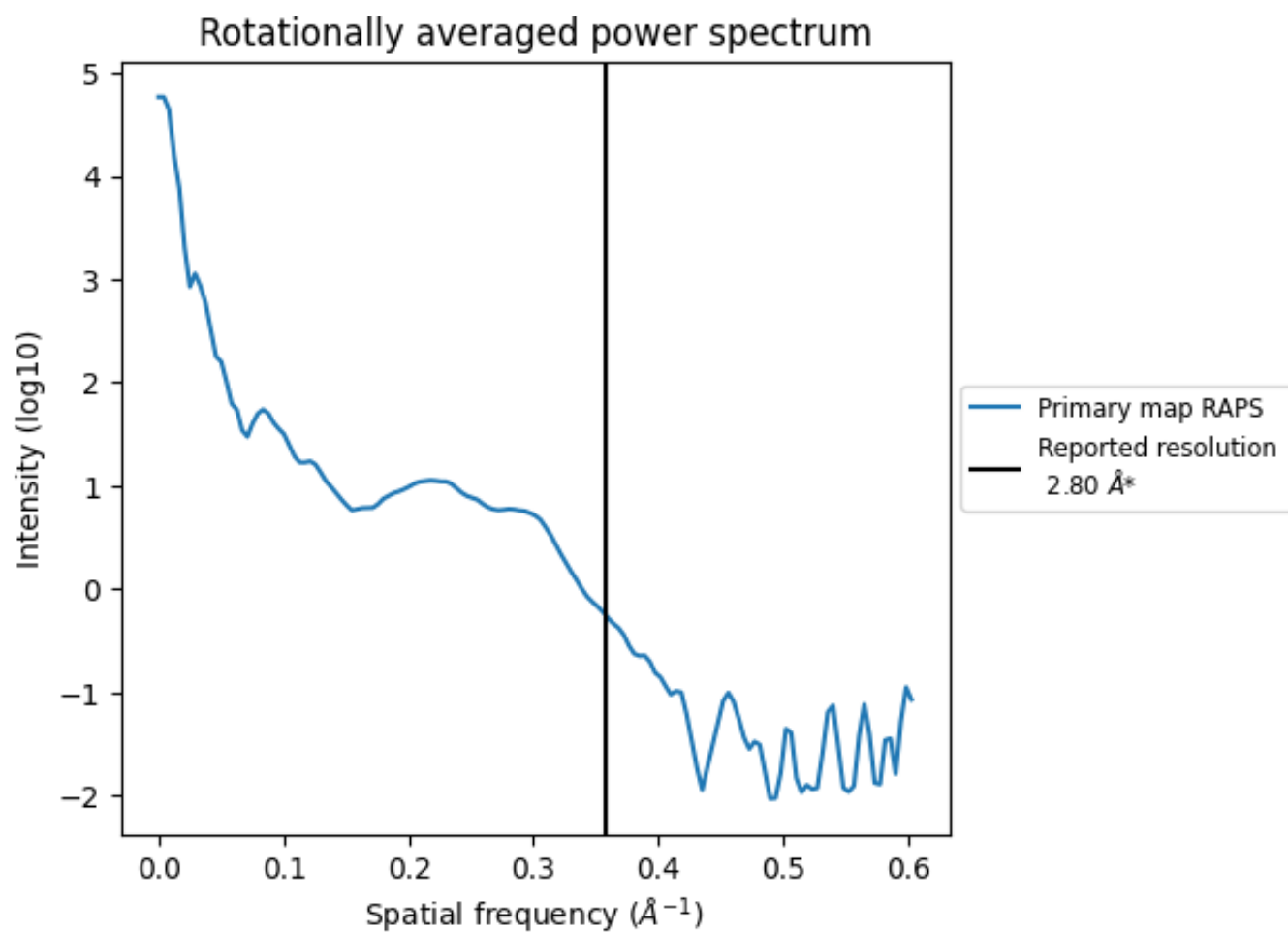
7.2 Volume estimate [\(i\)](#)



The volume at the recommended contour level is 93 nm^3 ; this corresponds to an approximate mass of 84 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](#)

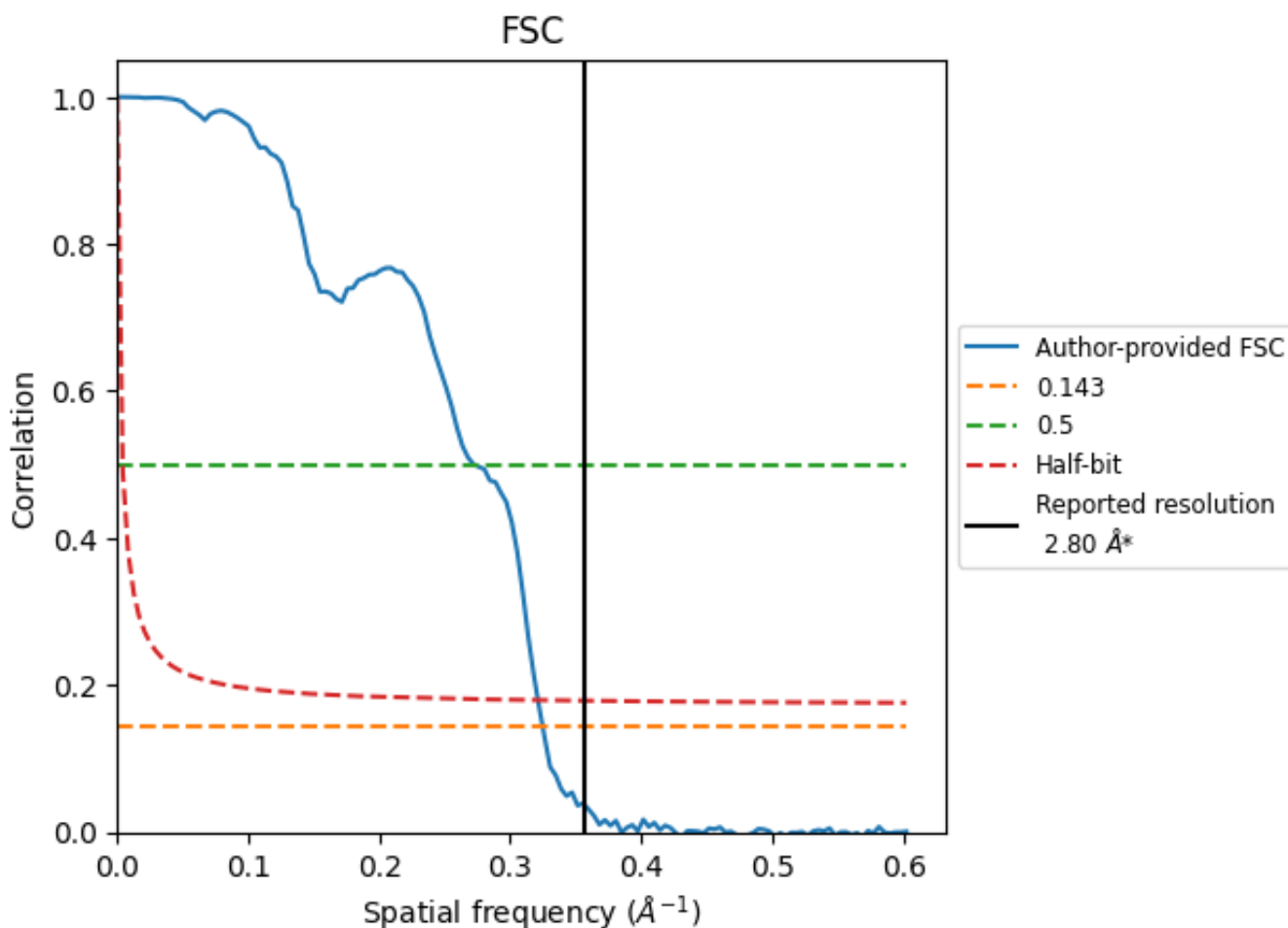


*Reported resolution corresponds to spatial frequency of 0.357\AA^{-1}

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 0.357 Å⁻¹

8.2 Resolution estimates [i](#)

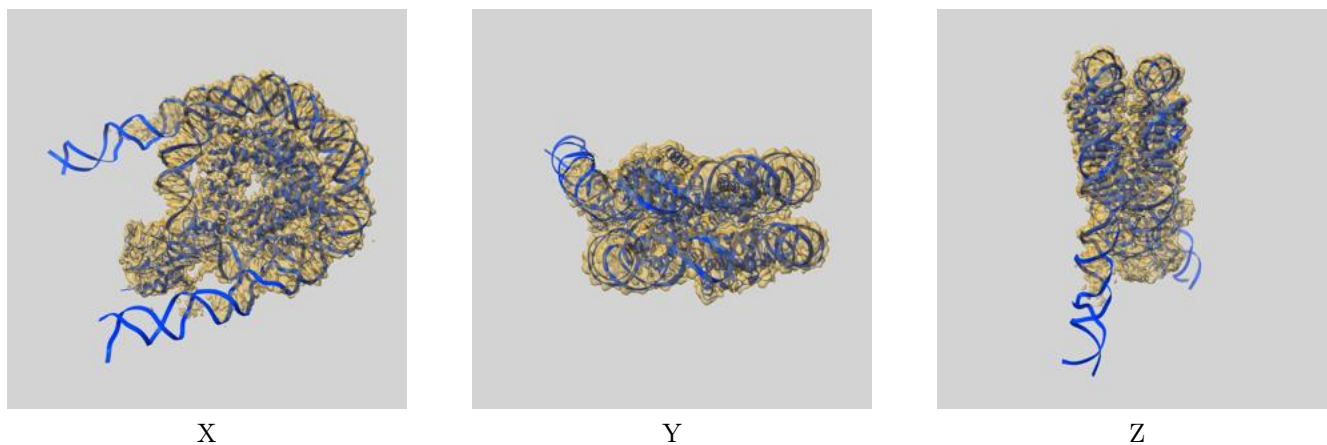
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	2.80	-	-
Author-provided FSC curve	3.08	3.68	3.11
Unmasked-calculated*	-	-	-

*Resolution estimate based on FSC curve calculated by comparison of deposited half-maps. The value from author-provided FSC intersecting FSC 0.143 CUT-OFF 3.08 differs from the reported value 2.8 by more than 10 %

9 Map-model fit [i](#)

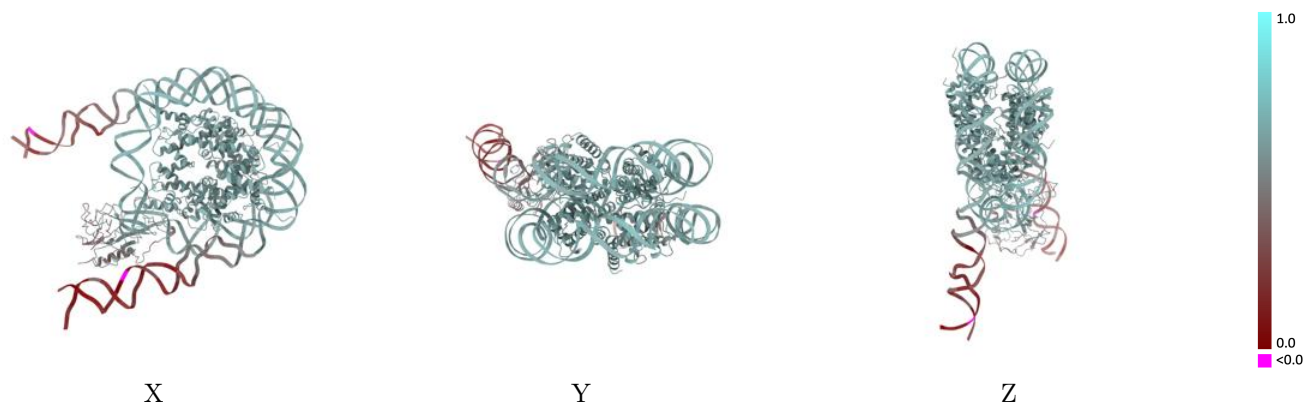
This section contains information regarding the fit between EMDB map EMD-31015 and PDB model 7E8D. Per-residue inclusion information can be found in section 3 on page 7.

9.1 Map-model overlay [i](#)



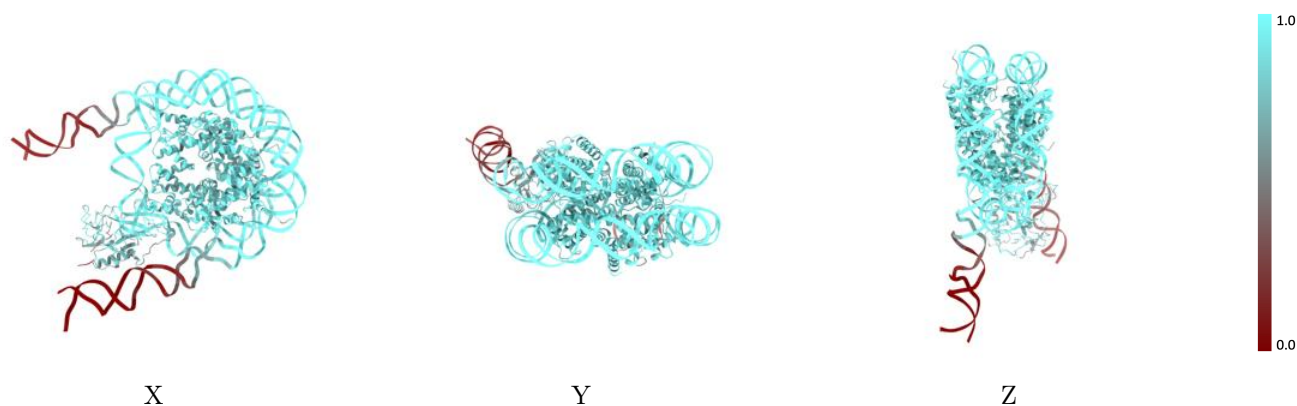
The images above show the 3D surface view of the map at the recommended contour level 0.01 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

9.2 Q-score mapped to coordinate model [i](#)



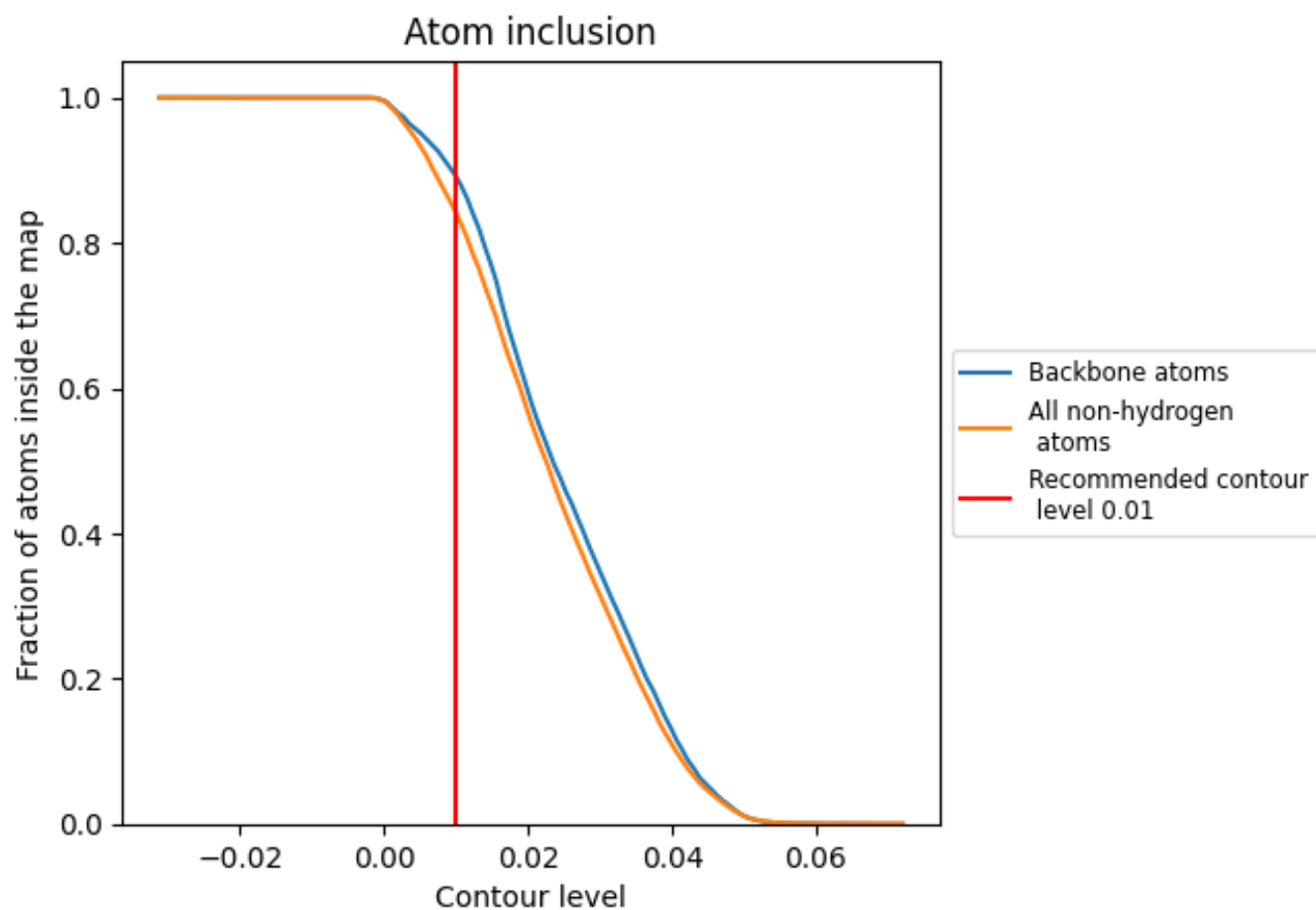
The images above show the model with each residue coloured according to its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions 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](#)



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 (0.01).

























9.4 Atom inclusion [i](#)



At the recommended contour level, 89% of all backbone atoms, 84% 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 (0.01) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	 0.8415	 0.5560
A	 0.9351	 0.6210
B	 0.9654	 0.6320
C	 0.9220	 0.6110
D	 0.9275	 0.6110
E	 0.9275	 0.6190
F	 0.9666	 0.6260
G	 0.9203	 0.6180
H	 0.9366	 0.6100
I	 0.7865	 0.5160
J	 0.7771	 0.5200
K	 0.7647	 0.4940

