wwPDB EM Validation Summary Report (i)

May 5, 2024 – 06:53 PM EDT

PDB ID	:	6O96
EMDB ID	:	EMD-0652
Title	:	Dot1L bound to the H2BK120 Ubiquitinated nucleosome
Authors	:	Valencia-Sanchez, M.I.; De Ioannes, P.E.; Miao, W.; Vasilyev, N.; Chen, R.; Nudler, E.; Armache, J.-P.; Armache, K.-J.
Deposited on	:	2019-03-13
Resolution	:	3.50 Å(reported)
Based on initial models	:	1NW3, 1UBQ, 3TU4

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 (1)) were used in the production of this report:

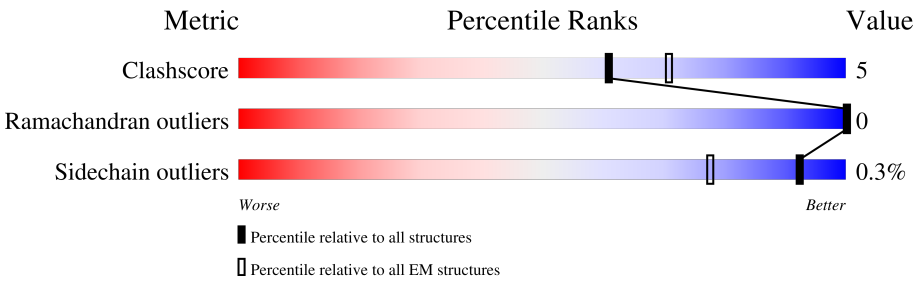
EMDB validation analysis	:	0.0.1.dev92
Mogul	:	1.8.5 (274361), CSD as541be (2020)
MolProbity	:	4.02b-467
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.13
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.50 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	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	135	61% 12% 27%
1	E	135	64% 8% 27%
2	B	103	• 72% 10% 18%
2	F	103	• 72% 10% 18%
3	C	130	• 71% 12% 18%
3	G	130	73% 9% 18%
4	D	123	60% 15% • 24%
4	H	123	66% 10% 24%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	I	146	91% 9%
6	J	146	88% 12%
7	K	331	17% 82% 15%
8	L	76	91% 9%

2 Entry composition (i)

There are 9 unique types of molecules in this entry. The entry contains 15194 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.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	98	801	506	153	139	3	0	0
1	E	98	801	506	153	139	3	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	102	ALA	GLY	engineered mutation	UNP P84233
E	102	ALA	GLY	engineered mutation	UNP P84233

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	B	84	663	419	129	114	1	0	0
2	F	84	663	419	129	114	1	0	0

- Molecule 3 is a protein called Histone H2A.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
3	C	107	811	510	158	143	0	0
3	G	107	815	513	159	143	0	0

- Molecule 4 is a protein called Histone H2B 1.1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	D	93	715	448	127	137	3	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	H	93	723	454	129	137	3	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	3	MET	-	expression tag	UNP P02281
D	32	THR	SER	engineered mutation	UNP P02281
D	120	CYS	LYS	engineered mutation	UNP P02281
H	3	MET	-	expression tag	UNP P02281
H	32	THR	SER	engineered mutation	UNP P02281
H	120	CYS	LYS	engineered mutation	UNP P02281

- Molecule 5 is a DNA chain called DNA (146-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
5	I	146	2975	1414	539	876	146	0	0

- Molecule 6 is a DNA chain called DNA (146-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
6	J	146	3011	1426	563	876	146	0	0

- Molecule 7 is a protein called Histone-lysine N-methyltransferase, H3 lysine-79 specific.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
7	K	321	2596	1654	446	484	12	0	0

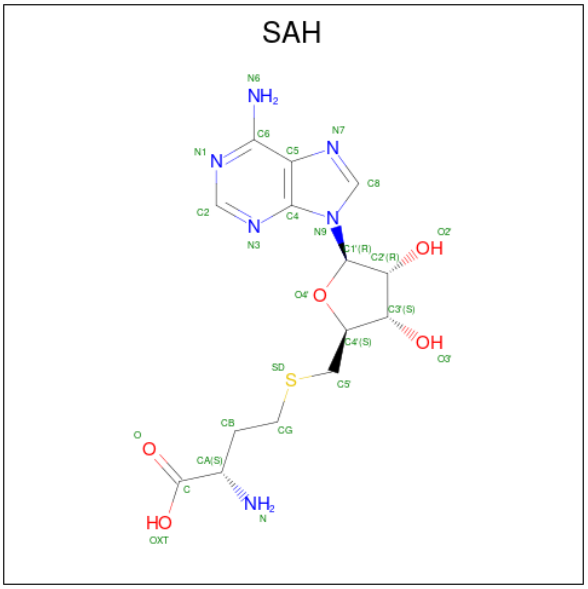
- Molecule 8 is a protein called Polyubiquitin-B.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
8	L	76	594	374	103	116	1	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
L	76	CYS	GLY	engineered mutation	UNP J3QS39

- Molecule 9 is S-ADENOSYL-L-HOMOCYSTEINE (three-letter code: SAH) (formula: $C_{14}H_{20}N_6O_5S$).



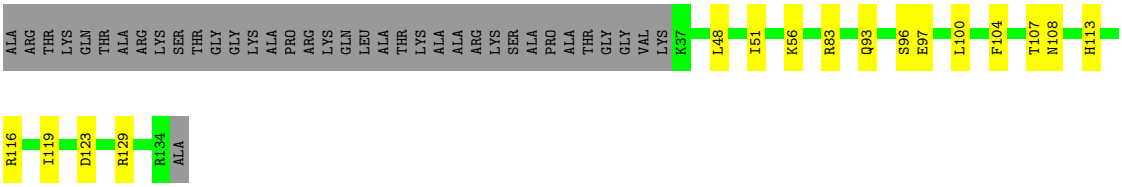
Mol	Chain	Residues	Atoms					AltConf
			Total	C	N	O	S	
9	K	1	26	14	6	5	1	0

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.

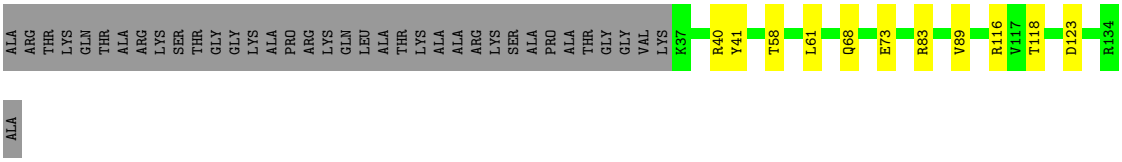
- Molecule 1: Histone H3.2

Chain A:



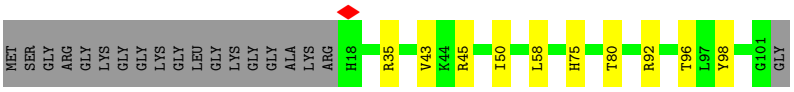
- Molecule 1: Histone H3.2

Chain E:



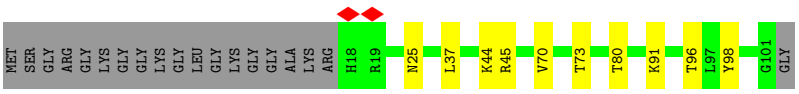
- Molecule 2: Histone H4

Chain B:



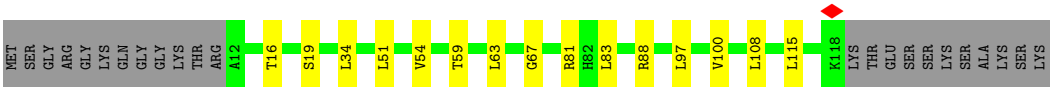
- Molecule 2: Histone H4

Chain F:



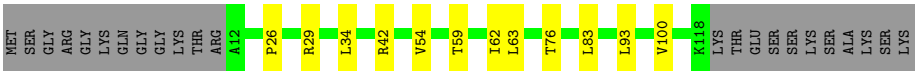
- Molecule 3: Histone H2A

Chain C:



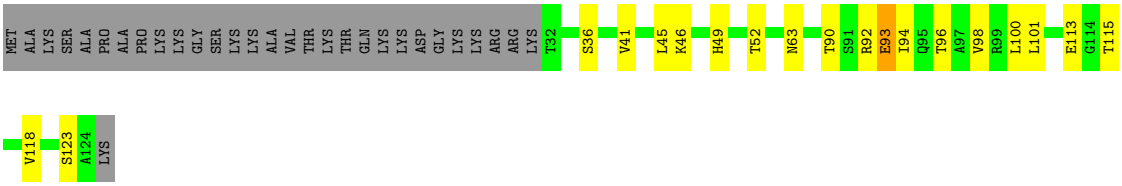
• Molecule 3: Histone H2A

Chain G:



• Molecule 4: Histone H2B 1.1

Chain D:

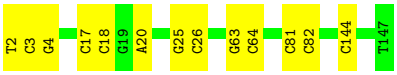


• Molecule 4: Histone H2B 1.1

Chain H:

• Molecule 5: DNA (146-MER)

Chain I:

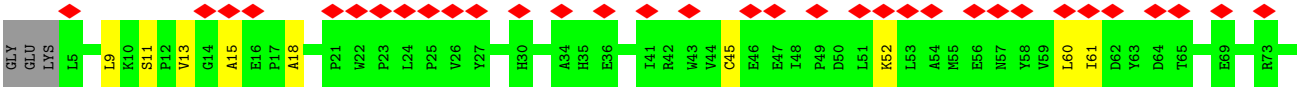


• Molecule 6: DNA (146-MER)

Chain J:

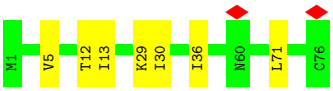
• Molecule 7: Histone-lysine N-methyltransferase, H3 lysine-79 specific

Chain K:



- Molecule 8: Polyubiquitin-B

Chain L:



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	211279	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 ($e^-/Å^2$)	1.1	Depositor
Minimum defocus (nm)	1700	Depositor
Maximum defocus (nm)	3200	Depositor
Magnification	130000	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	13.802	Depositor
Minimum map value	-8.897	Depositor
Average map value	-0.007	Depositor
Map value standard deviation	0.345	Depositor
Recommended contour level	1.2	Depositor
Map size (Å)	331.19998, 331.19998, 331.19998	wwPDB
Map dimensions	320, 320, 320	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.035, 1.035, 1.035	Depositor

5 Model quality (i)

5.1 Standard geometry (i)

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

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	0.58	0/813	0.61	0/1093
1	E	0.64	0/813	0.60	0/1093
2	B	0.71	0/670	0.65	0/899
2	F	0.70	0/670	0.63	0/899
3	C	0.61	0/821	0.66	0/1112
3	G	0.56	0/825	0.64	0/1116
4	D	0.68	0/726	0.61	0/982
4	H	0.62	0/734	0.57	0/990
5	I	0.94	0/3333	0.98	0/5137
6	J	0.94	0/3381	0.97	0/5221
7	K	0.29	0/2661	0.59	0/3608
8	L	0.27	0/600	0.59	0/809
All	All	0.73	0/16047	0.80	0/22959

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (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	831	12	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	801	0	831	10	0
2	B	663	0	699	13	0
2	F	663	0	699	8	0
3	C	811	0	849	15	0
3	G	815	0	860	13	0
4	D	715	0	717	15	0
4	H	723	0	739	11	0
5	I	2975	0	1640	14	0
6	J	3011	0	1640	14	0
7	K	2596	0	2510	35	0
8	L	594	0	611	12	0
9	K	26	0	19	2	0
All	All	15194	0	12645	129	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 5.

The worst 5 of 129 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)
6:J:15:DT:H2"	6:J:16:DC:O5'	1.70	0.91
8:L:30:ILE:HG22	8:L:36:ILE:CG2	2.10	0.80
7:K:122:PRO:O	7:K:125:LEU:HB2	1.82	0.79
7:K:189:ASP:O	7:K:193:LYS:HB2	1.87	0.74
8:L:30:ILE:HG22	8:L:36:ILE:HG22	1.70	0.73

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	96/135 (71%)	94 (98%)	2 (2%)	0	100	100
1	E	96/135 (71%)	93 (97%)	3 (3%)	0	100	100
2	B	82/103 (80%)	78 (95%)	4 (5%)	0	100	100
2	F	82/103 (80%)	79 (96%)	3 (4%)	0	100	100
3	C	105/130 (81%)	101 (96%)	4 (4%)	0	100	100
3	G	105/130 (81%)	102 (97%)	3 (3%)	0	100	100
4	D	91/123 (74%)	88 (97%)	3 (3%)	0	100	100
4	H	91/123 (74%)	89 (98%)	2 (2%)	0	100	100
7	K	317/331 (96%)	287 (90%)	30 (10%)	0	100	100
8	L	74/76 (97%)	73 (99%)	1 (1%)	0	100	100
All	All	1139/1389 (82%)	1084 (95%)	55 (5%)	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	84/110 (76%)	84 (100%)	0	100	100
1	E	84/110 (76%)	84 (100%)	0	100	100
2	B	67/79 (85%)	67 (100%)	0	100	100
2	F	67/79 (85%)	67 (100%)	0	100	100
3	C	81/102 (79%)	81 (100%)	0	100	100
3	G	82/102 (80%)	82 (100%)	0	100	100
4	D	77/103 (75%)	76 (99%)	1 (1%)	69	86
4	H	79/103 (77%)	79 (100%)	0	100	100
7	K	279/296 (94%)	277 (99%)	2 (1%)	84	93
8	L	66/69 (96%)	66 (100%)	0	100	100
All	All	966/1153 (84%)	963 (100%)	3 (0%)	92	97

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	93	GLU
7	K	190	ILE
7	K	316	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
7	K	315	HIS
7	K	324	ASN
8	L	60	ASN
8	L	31	GLN
3	G	89	ASN

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)

1 ligand is 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	$\#\|Z\| > 2$
9	SAH	K	401	-	24,28,28	1.08	2 (8%)	25,40,40	1.49	5 (20%)

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	SAH	K	401	-	-	2/11/31/31	0/3/3/3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
9	K	401	SAH	C5-C4	2.45	1.47	1.40
9	K	401	SAH	OXT-C	-2.18	1.23	1.30

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
9	K	401	SAH	N3-C2-N1	-3.53	123.16	128.68
9	K	401	SAH	OXT-C-O	-2.94	117.42	124.09
9	K	401	SAH	O4'-C1'-C2'	-2.61	103.11	106.93
9	K	401	SAH	C4-C5-N7	-2.55	106.74	109.40
9	K	401	SAH	OXT-C-CA	2.01	120.22	113.38

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
9	K	401	SAH	N-CA-CB-CG
9	K	401	SAH	C-CA-CB-CG

There are no ring outliers.

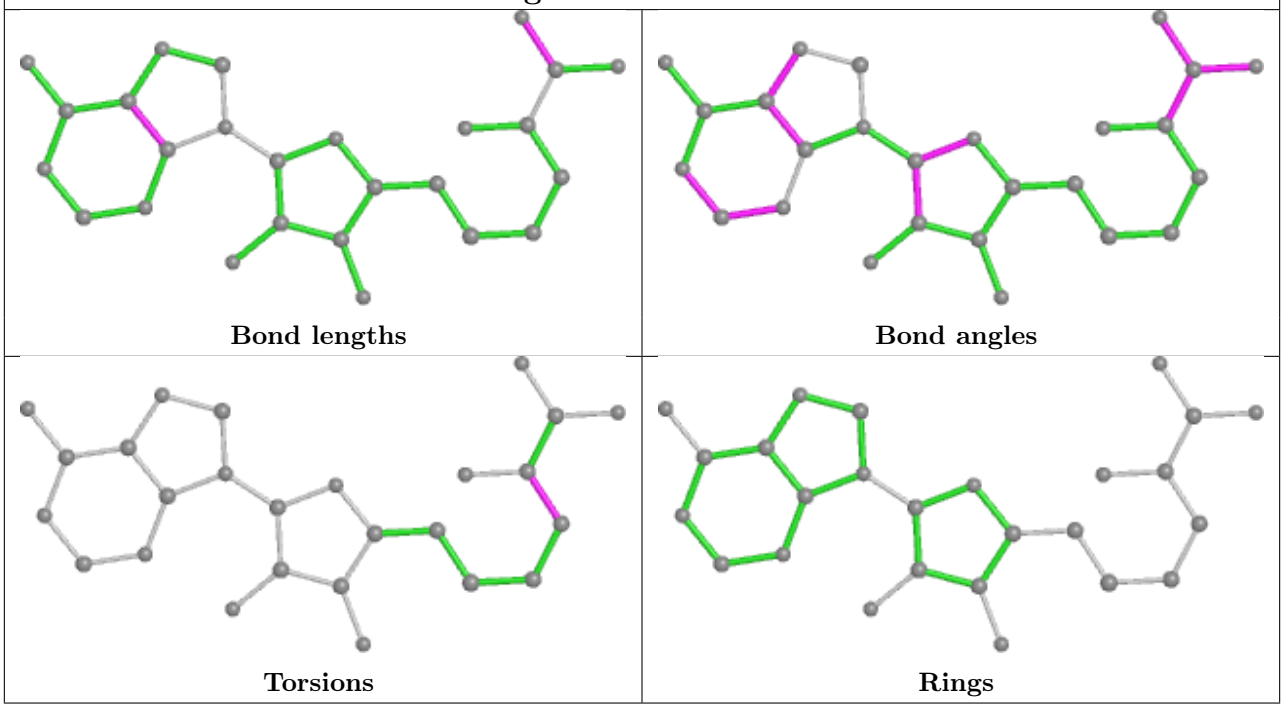
1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
9	K	401	SAH	2	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 then 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.

Ligand SAH K 401



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.

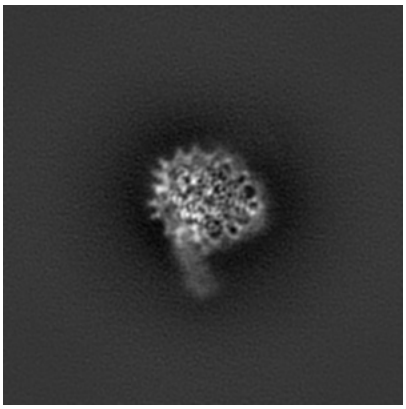
6 Map visualisation (i)

This section contains visualisations of the EMDB entry EMD-0652. These allow visual inspection of the internal detail of the map and identification of artifacts.

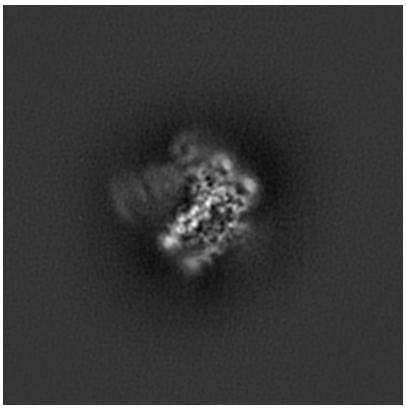
Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

6.1 Orthogonal projections (i)

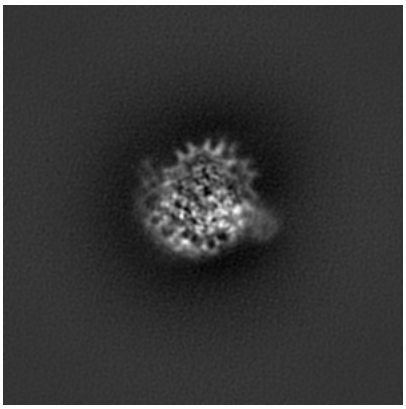
6.1.1 Primary map



X

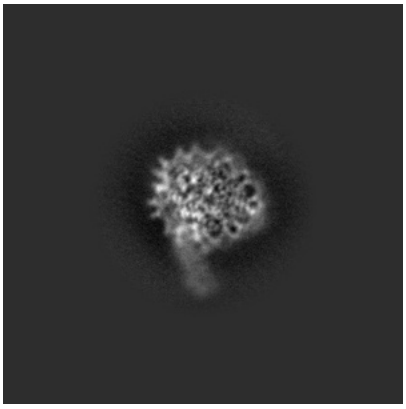


Y

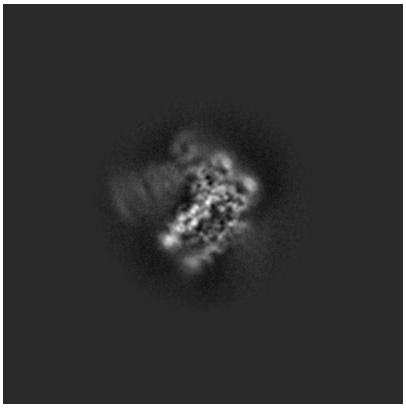


Z

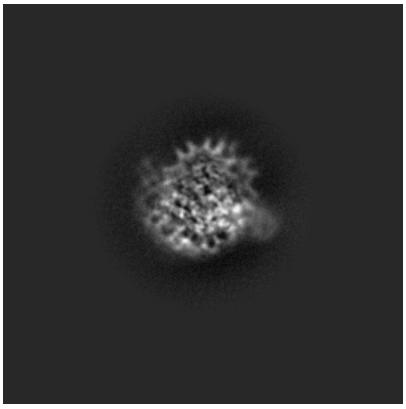
6.1.2 Raw map



X



Y

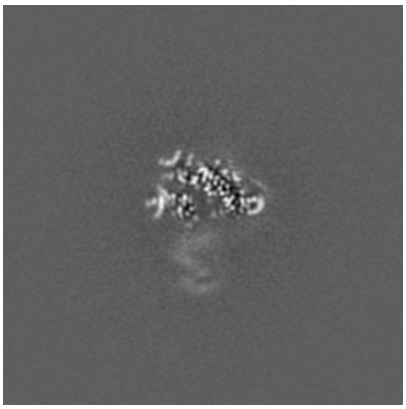


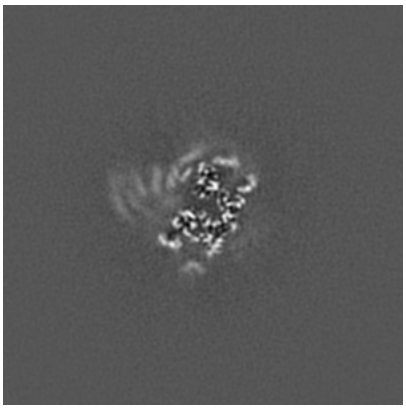
Z

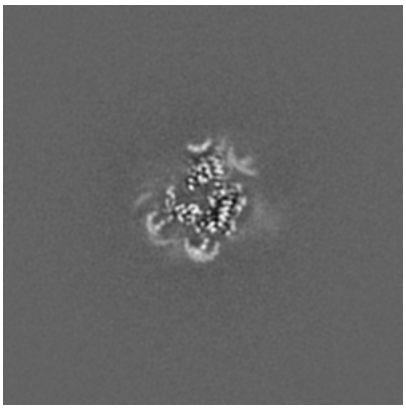
The images above show the map projected in three orthogonal directions.

6.2 Central slices (i)

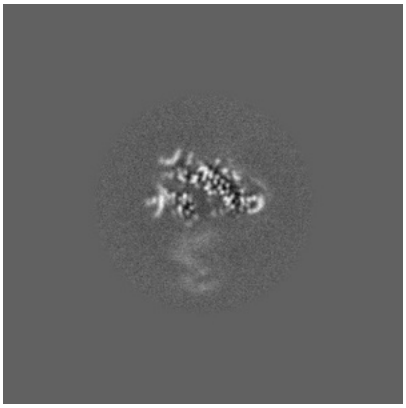
6.2.1 Primary map

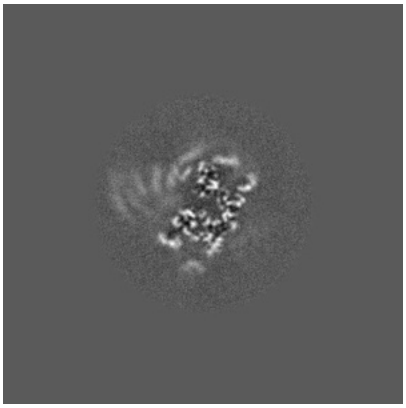

X Index: 160

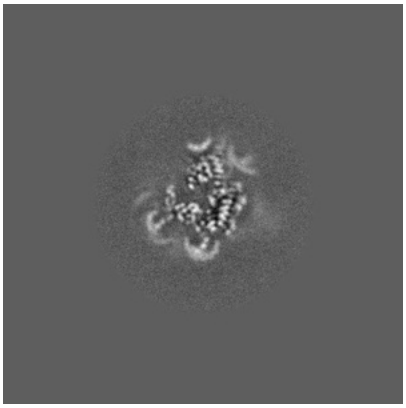

Y Index: 160


Z Index: 160

6.2.2 Raw map


X Index: 160

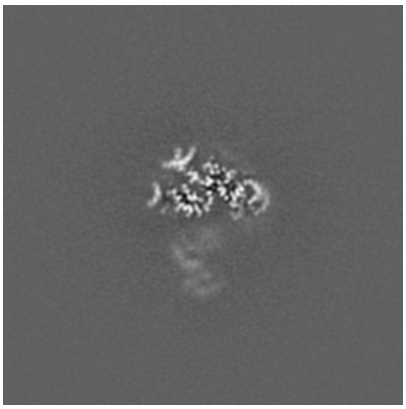

Y Index: 160

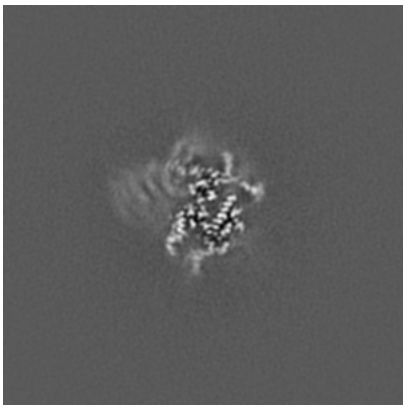

Z Index: 160

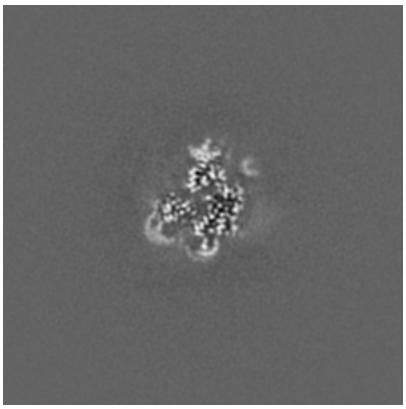
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

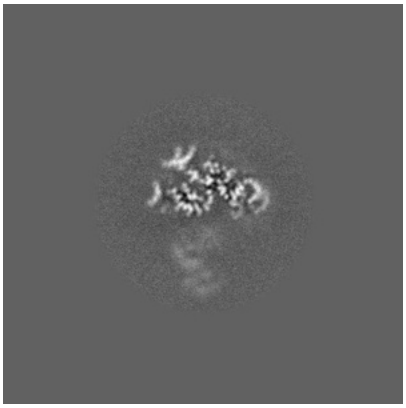
6.3.1 Primary map

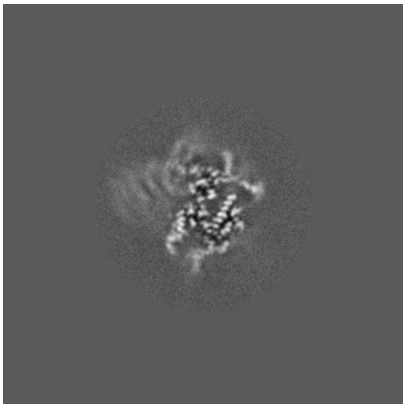

X Index: 165

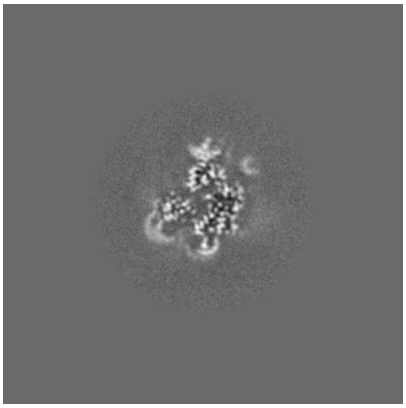

Y Index: 153


Z Index: 163

6.3.2 Raw map


X Index: 165

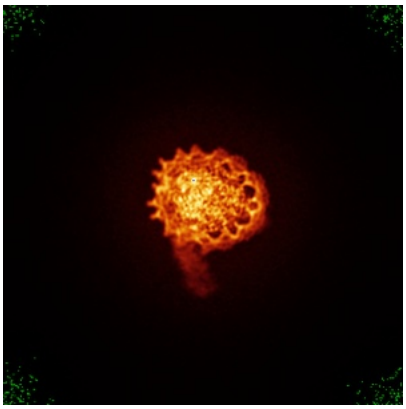

Y Index: 153

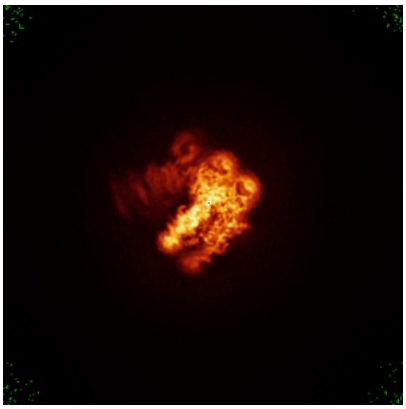

Z Index: 163

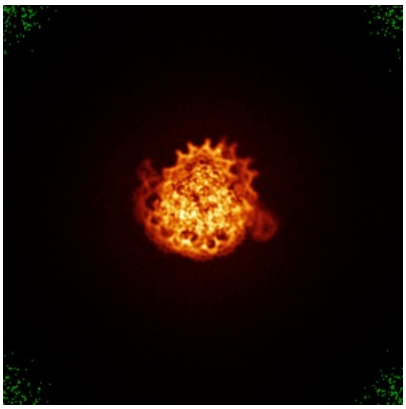
The images above show the largest variance slices of the map in three orthogonal directions.

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

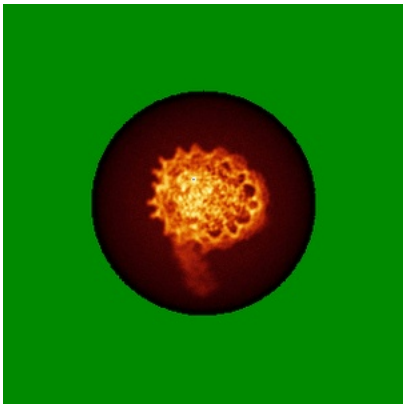
6.4.1 Primary map

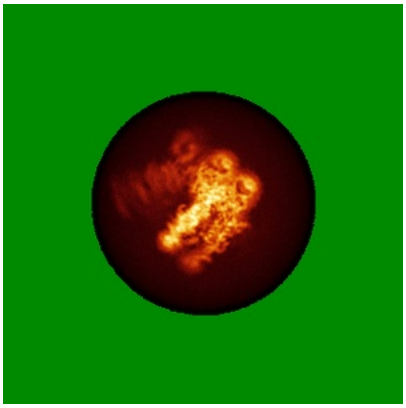

X

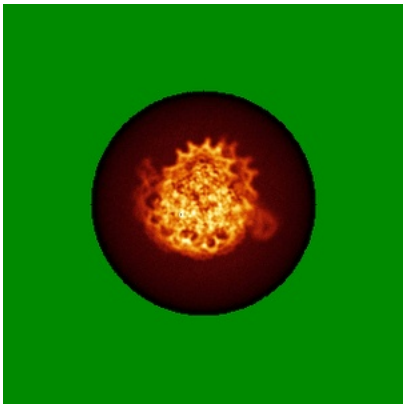

Y


Z

6.4.2 Raw map


X


Y


Z

The images above show the map standard deviation projections with false color in three orthogonal directions. Minimum values are shown in green, max in blue, and dark to light orange shades represent small to large values respectively.

6.5 Orthogonal surface views (i)

6.5.1 Primary map

X Y Z

The images above show the 3D surface view of the map at the recommended contour level 1.2. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

6.5.2 Raw map

X Y Z

These images show the 3D surface of the raw map. The raw map's contour level was selected so that its surface encloses the same volume as the primary map does at its recommended contour level.

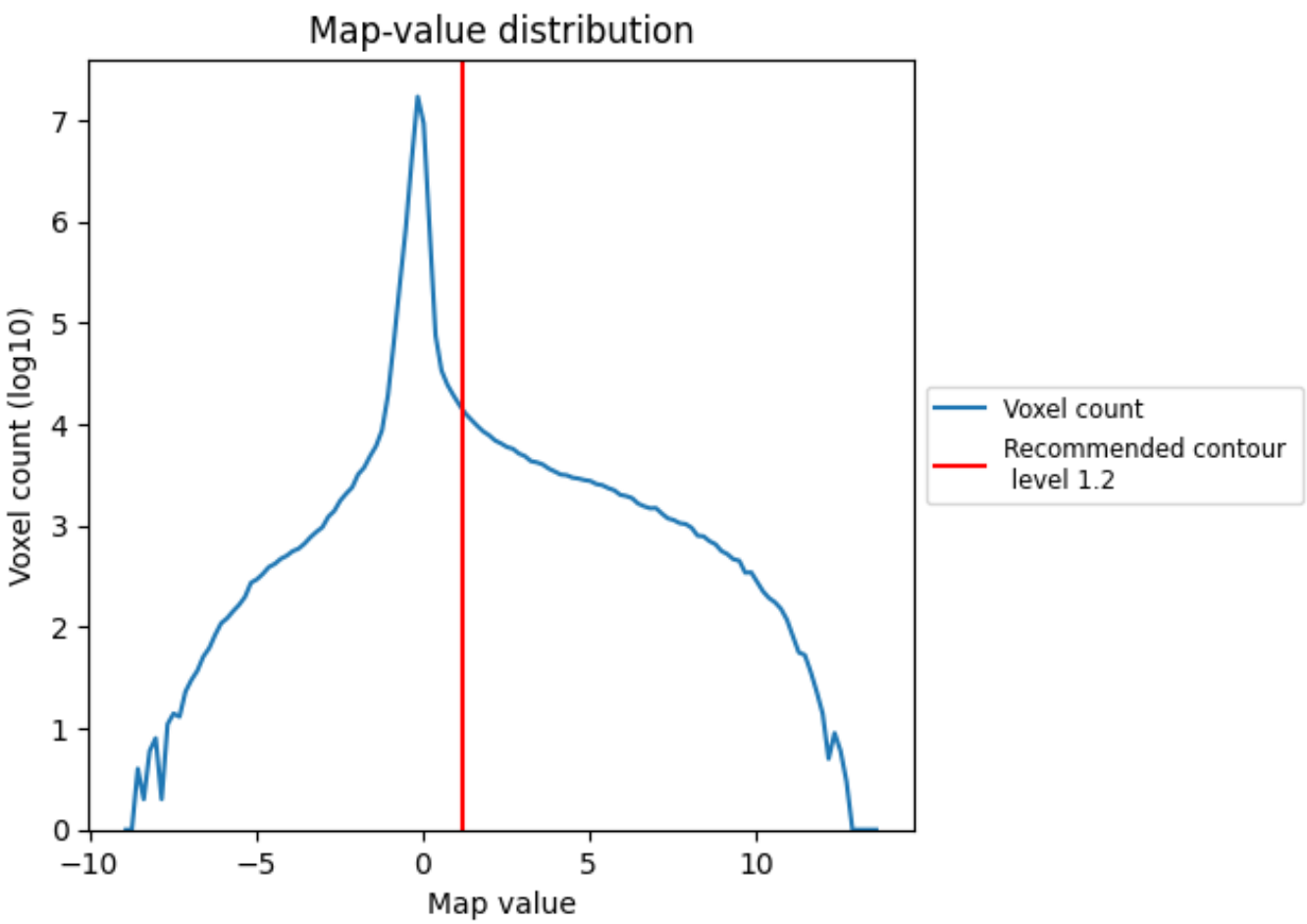
6.6 Mask visualisation (i)

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

7 Map analysis (i)

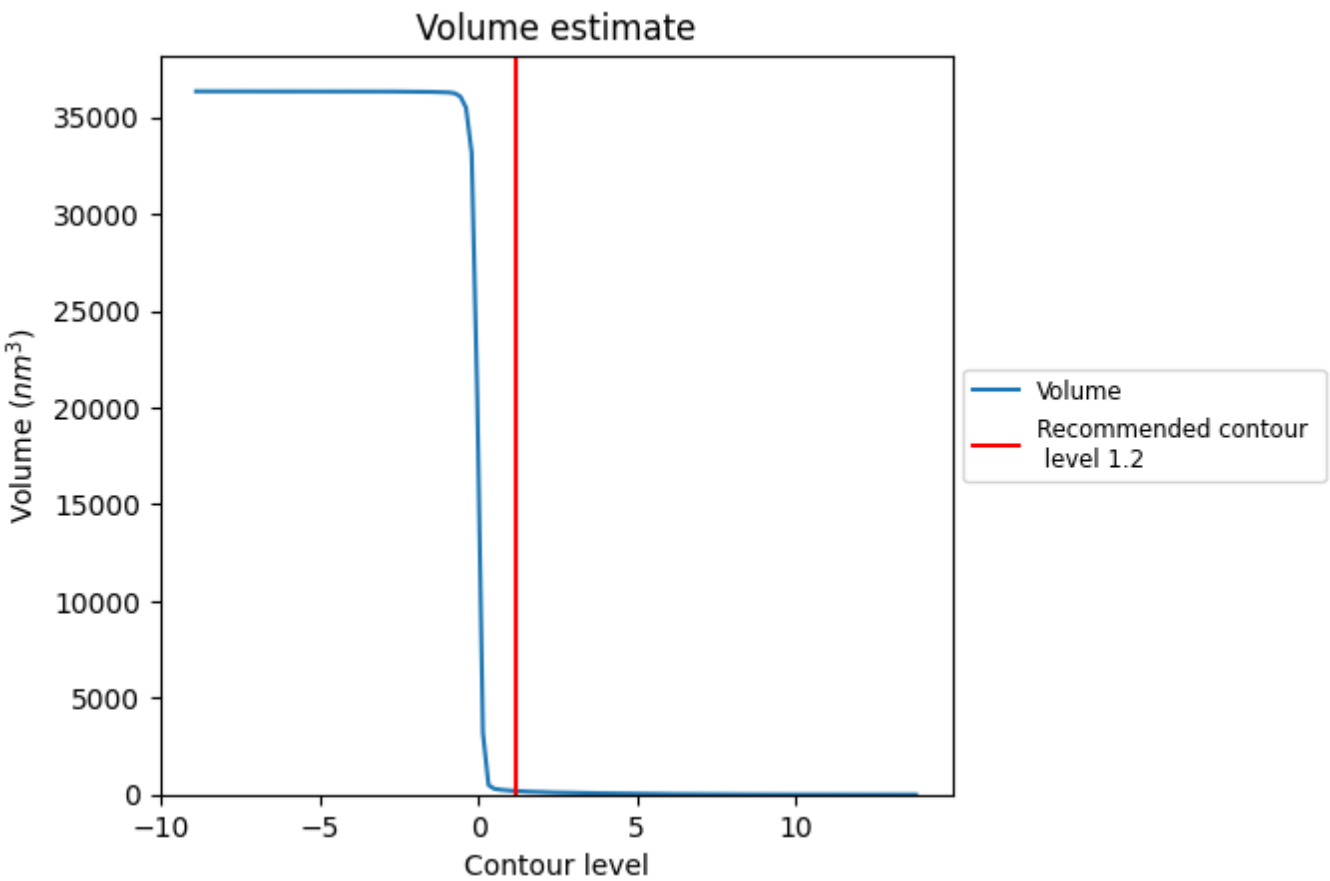
This section contains the results of statistical analysis of the map.

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.

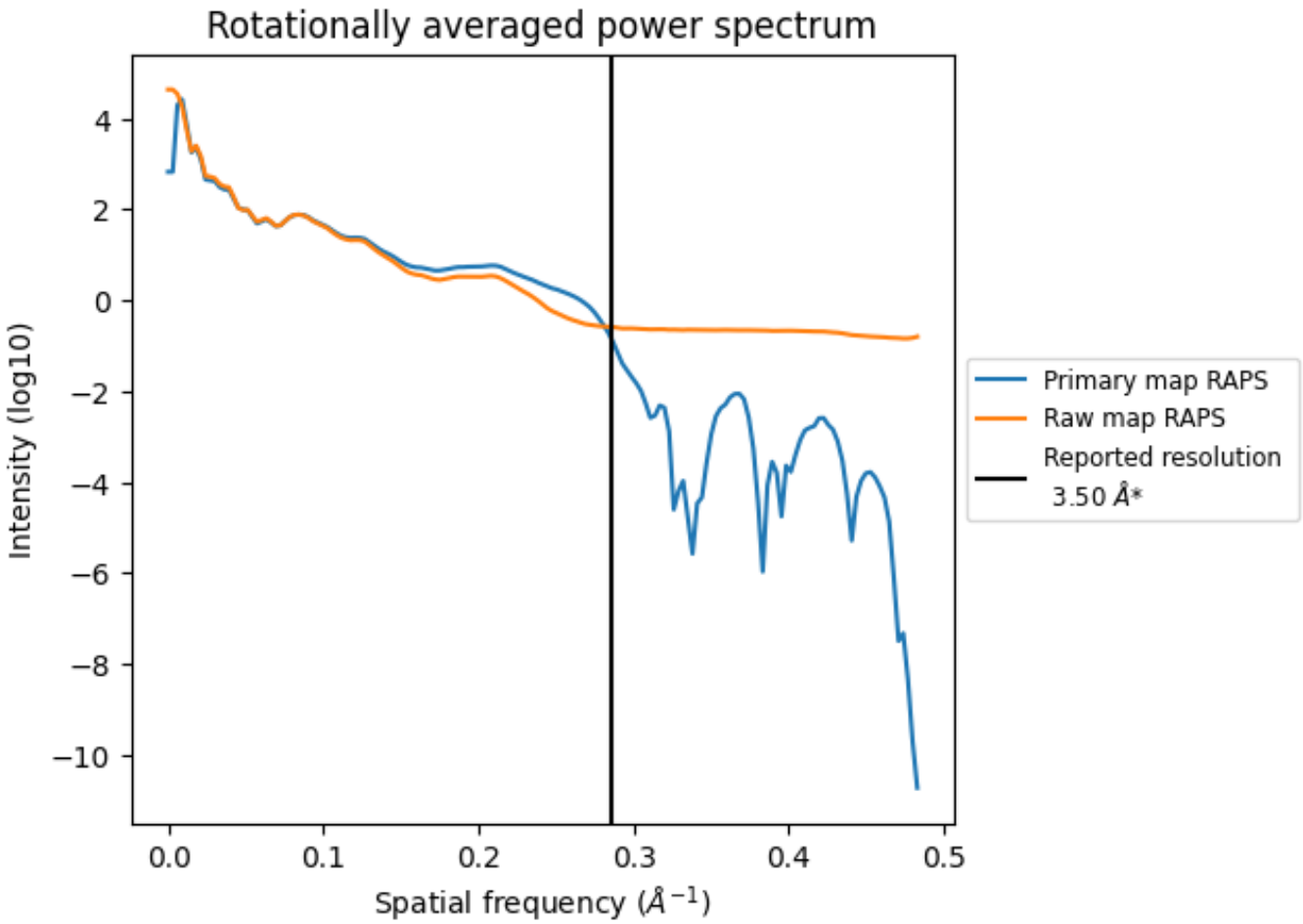
7.2 Volume estimate (i)



The volume at the recommended contour level is 183 nm^3; this corresponds to an approximate mass of 165 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)

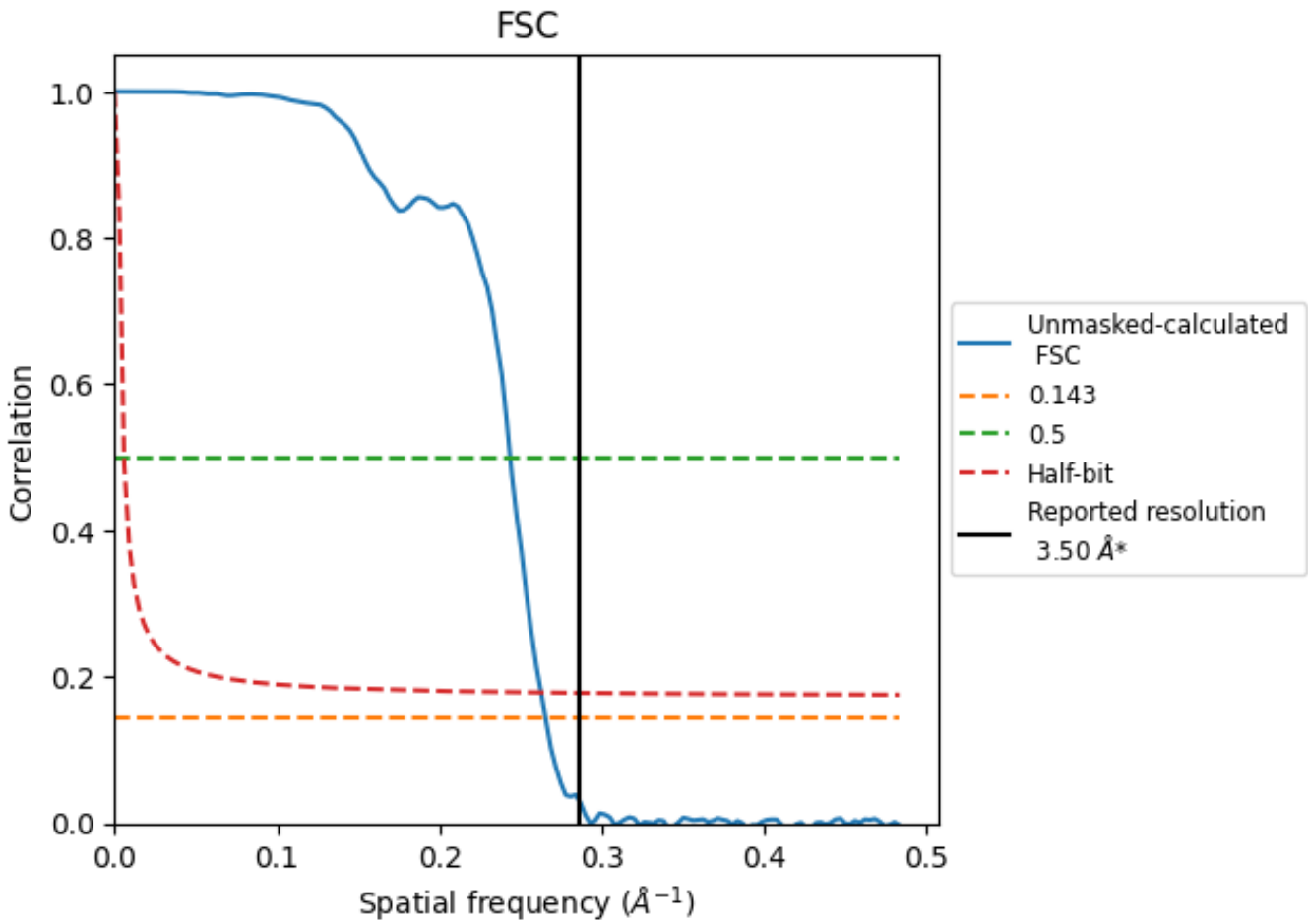


*Reported resolution corresponds to spatial frequency of 0.286 Å^{-1}

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 0.286 Å^{-1}

8.2 Resolution estimates (i)

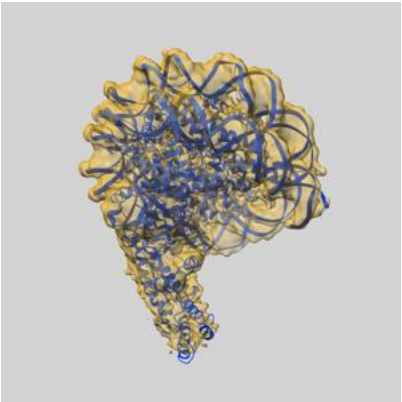
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.50	-	-
Author-provided FSC curve	-	-	-
Unmasked-calculated*	3.77	4.11	3.81

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

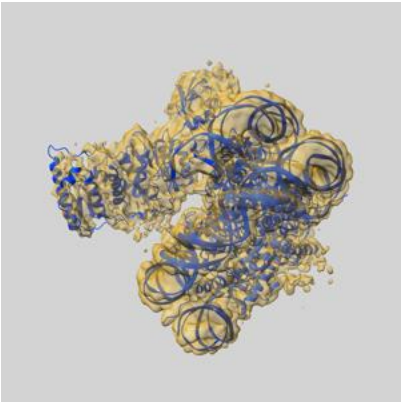
9 Map-model fit (i)

This section contains information regarding the fit between EMDB map EMD-0652 and PDB model 6O96. Per-residue inclusion information can be found in section 3 on page 7.

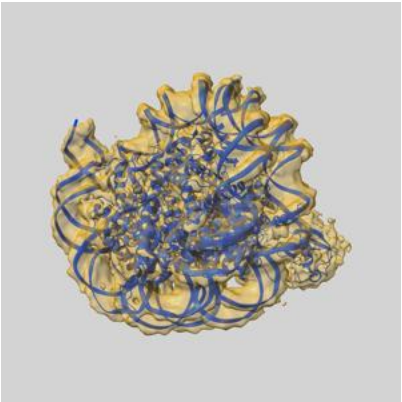
9.1 Map-model overlay (i)



X



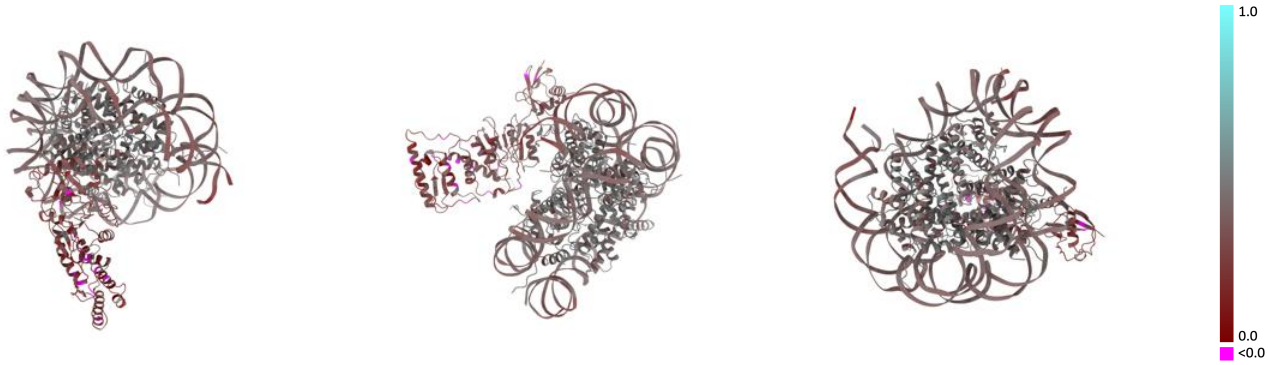
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 1.2 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.

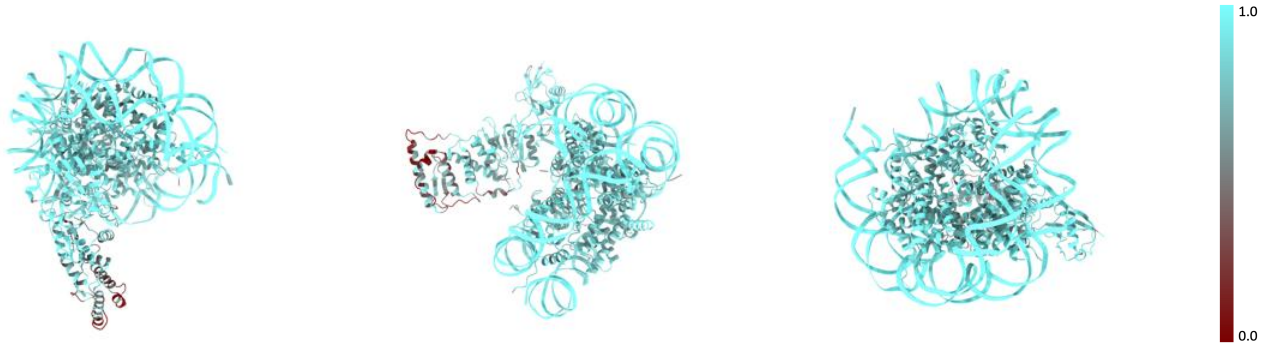
9.2 Q-score mapped to coordinate model (i)



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.

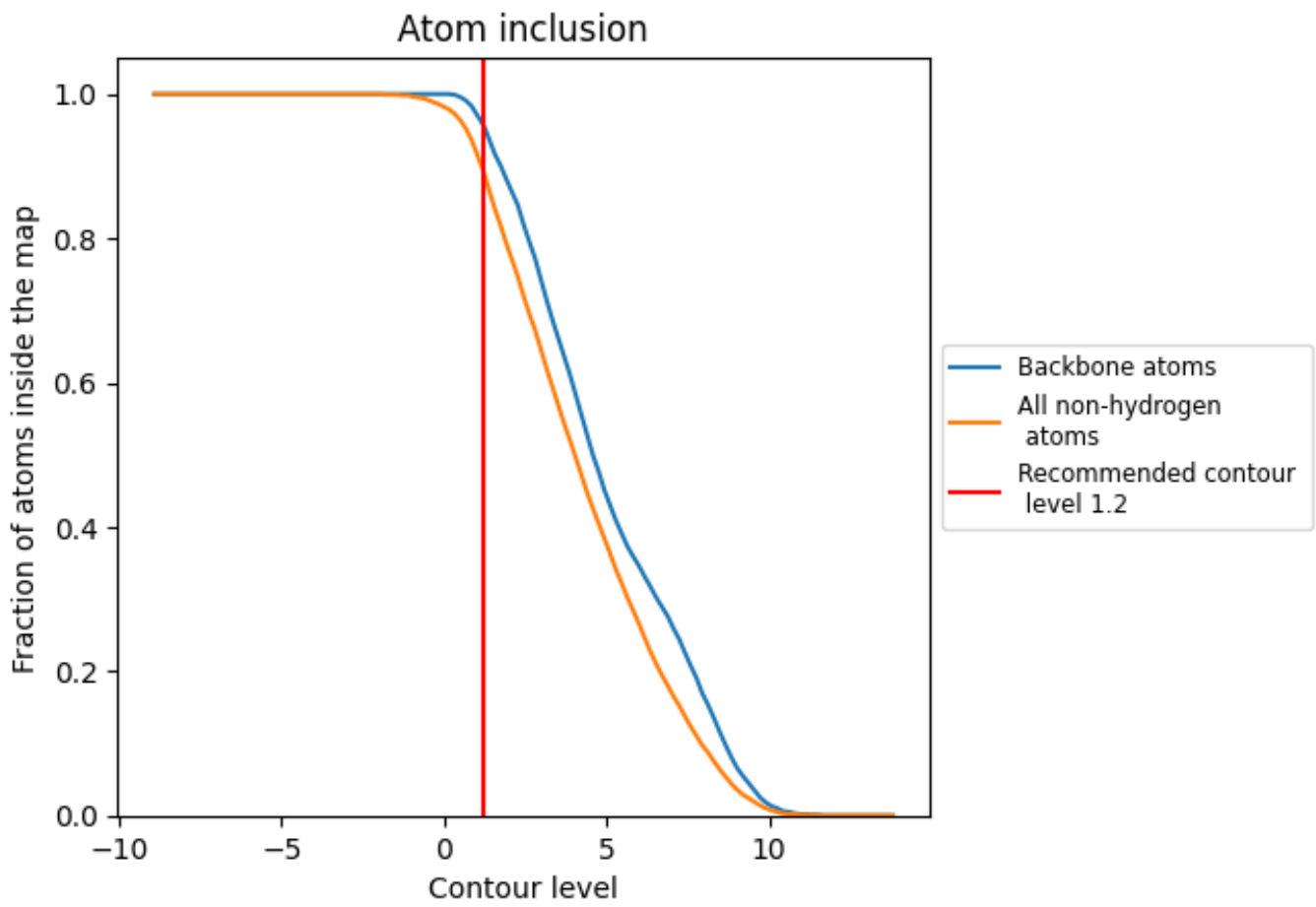
9.3 Atom inclusion mapped to coordinate model (i)



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 (1.2).

9.4 Atom inclusion (i)



At the recommended contour level, 96% 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 (1.2) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.8960	0.3800
A	0.8790	0.4470
B	0.8730	0.4560
C	0.8730	0.4510
D	0.9010	0.4420
E	0.8860	0.4480
F	0.8840	0.4480
G	0.8700	0.4460
H	0.8870	0.4340
I	0.9890	0.3790
J	0.9930	0.3790
K	0.7050	0.2550
L	0.8940	0.2690