wwPDB EM Validation Summary Report ⓘ

Apr 23, 2024 – 12:22 pm BST

PDB ID	:	7AT8
EMDB ID	:	EMD-11910
Title	:	Histone H3 recognition by nucleosome-bound PRC2 subunit EZH2.
Authors	:	Finogenova, K.; Benda, C.; Schaefer, I.B.; Poepsel, S.; Strauss, M.; Mueller, J.
Deposited on	:	2020-10-29
Resolution	:	4.40 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:

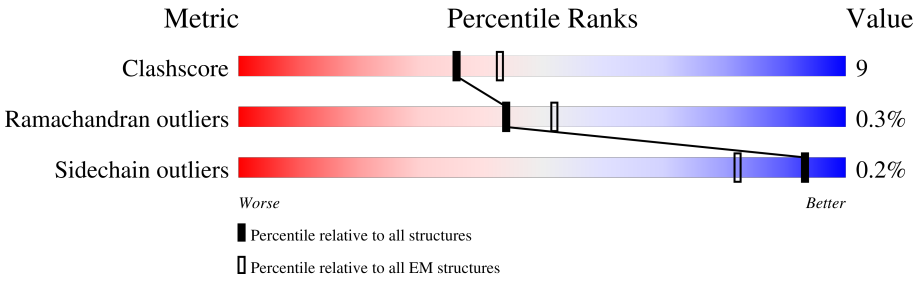
EMDB validation analysis	:	0.0.1.dev92
Mogul	:	1.8.4, CSD as541be (2020)
MolProbity	:	4.02b-467
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.13
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 4.40 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.



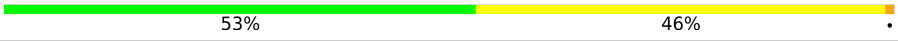
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

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	764	5% 34% 6% 59%
2	C	739	12% 5% 83%
3	D	135	75% 9% 16%
3	H	135	64% 7% 29%
4	E	102	61% 18% 22%
4	I	102	67% 10% 24%
5	F	129	73% 7% 20%
5	J	129	70% 12% 19%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	G	122	
6	K	122	
7	T	156	
8	U	156	

2 Entry composition (i)

There are 10 unique types of molecules in this entry. The entry contains 15951 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 Isoform 2 of Histone-lysine N-methyltransferase EZH2,Isoform 2 of Histone-lysine N-methyltransferase EZH2,Histone-lysine N-methyltransferase EZH2,Isoform 2 of Histone-lysine N-methyltransferase EZH2,Isoform 2 of Histone-lysine N-methyltransferase EZH2.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	311	Total	C	N	O	S	0	0
			2470	1556	434	450	30		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	PRO	ARG	conflict	UNP Q15910-2

- Molecule 2 is a protein called Polycomb protein SUZ12.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	C	124	Total	C	N	O	S	0	0
			1024	647	177	188	12		

- Molecule 3 is a protein called Histone H3.2.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	D	113	Total	C	N	O	S	0	0
			911	573	178	157	3		
3	H	96	Total	C	N	O	S	0	0
			791	498	152	138	3		

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	102	ALA	GLY	conflict	UNP P84233
H	102	ALA	GLY	conflict	UNP P84233

- Molecule 4 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	E	80	Total 638	C 401	N 125	O 111	S 1	0	0
4	I	78	Total 619	C 391	N 120	O 107	S 1	0	0

- Molecule 5 is a protein called Histone H2A.

Mol	Chain	Residues	Atoms				AltConf	Trace
5	F	103	Total 795	C 501	N 155	O 139	0	0
5	J	105	Total 809	C 510	N 158	O 141	0	0

- Molecule 6 is a protein called Histone H2B 1.1.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	G	95	Total 745	C 469	N 134	O 140	S 2	0	0
6	K	93	Total 726	C 457	N 130	O 137	S 2	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	29	THR	SER	conflict	UNP P02281
K	29	THR	SER	conflict	UNP P02281

- Molecule 7 is a DNA chain called Widom601 DNA plus linker.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	T	156	Total 3176	C 1512	N 573	O 936	P 155	0	0

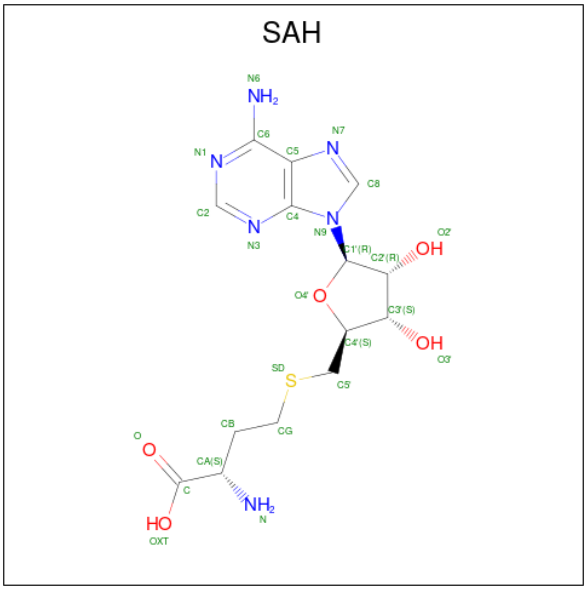
- Molecule 8 is a DNA chain called Widom601 DNA plus linker.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	U	156	Total 3214	C 1524	N 603	O 932	P 155	0	0

- Molecule 9 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	AltConf
9	A	7	Total Zn 7 7	0

- Molecule 10 is S-ADENOSYL-L-HOMOCYSTEINE (three-letter code: SAH) (formula: $C_{14}H_{20}N_6O_5S$).



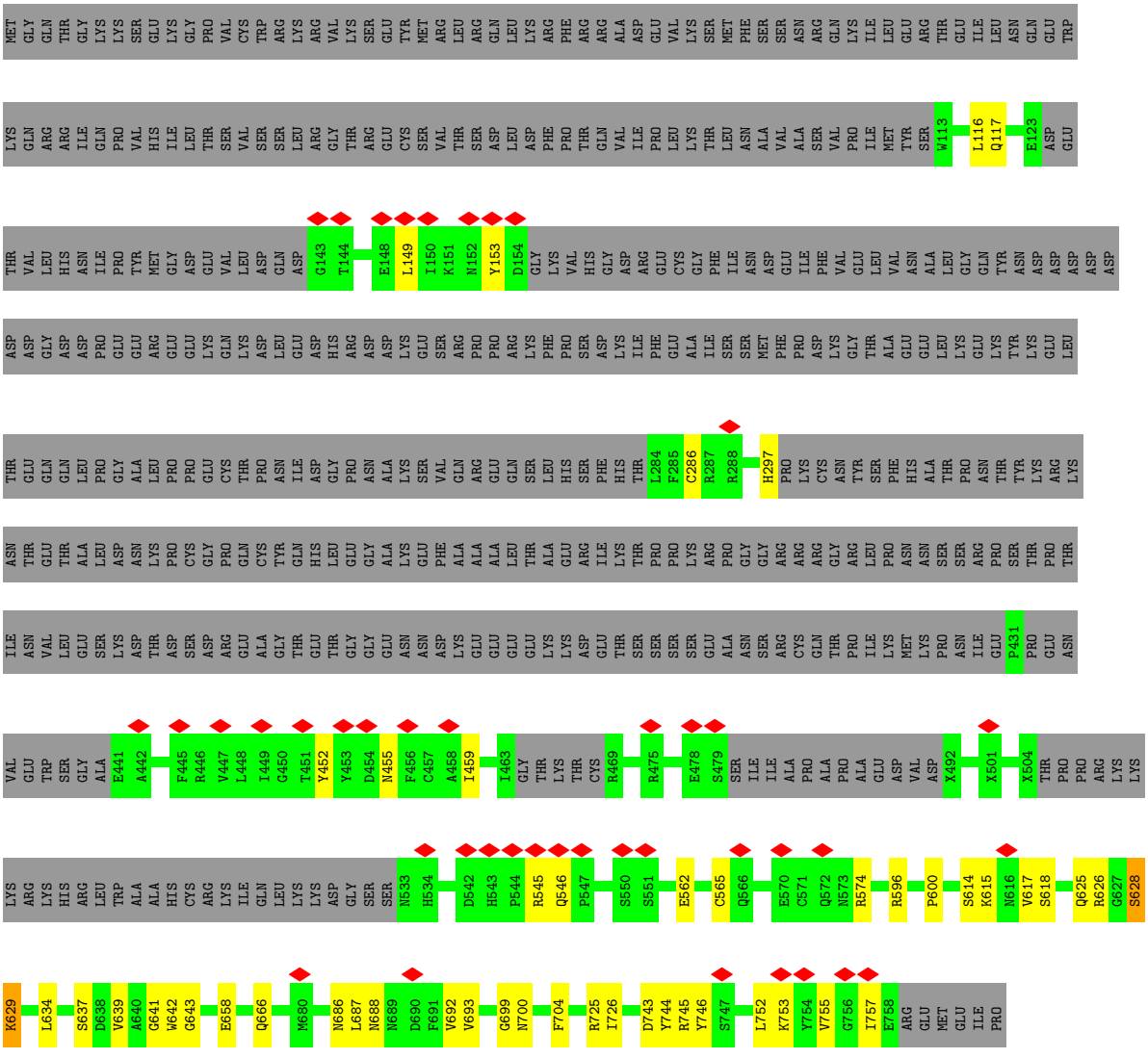
Mol	Chain	Residues	Atoms	AltConf
10	A	1	Total C N O S 26 14 6 5 1	0

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.

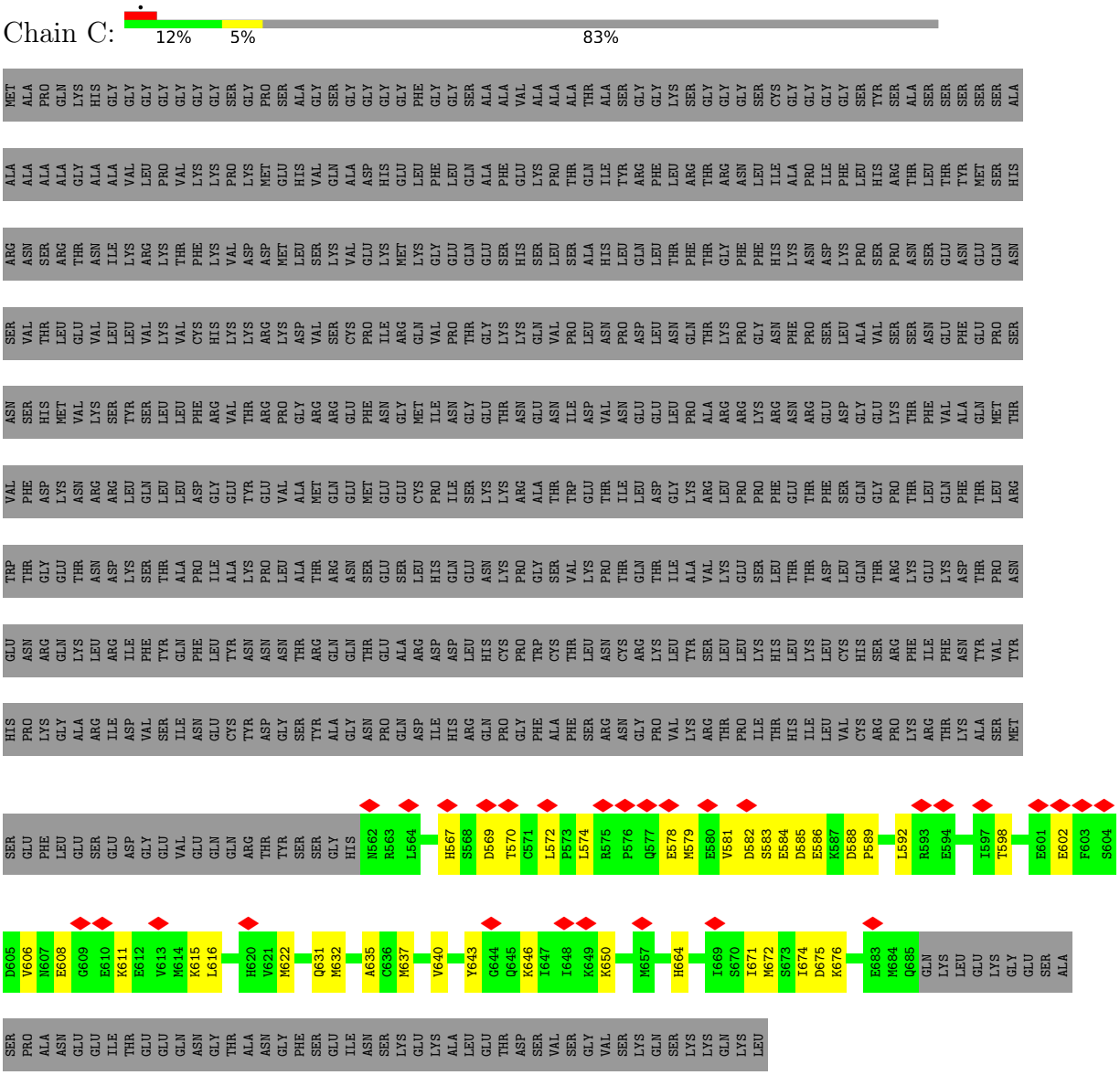
• Molecule 1: Isoform 2 of Histone-lysine N-methyltransferase EZH2,Isoform 2 of Histone-lysine N-methyltransferase EZH2,Histone-lysine N-methyltransferase EZH2,Isoform 2 of Histone-lysine N-methyltransferase EZH2,Isoform 2 of Histone-lysine N-methyltransferase EZH2

Chain A:



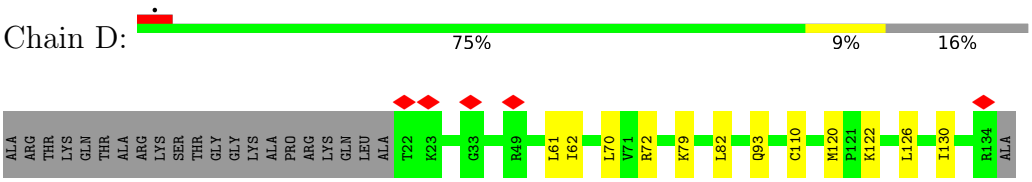
• Molecule 2: Polycomb protein SUZ12

Chain C:



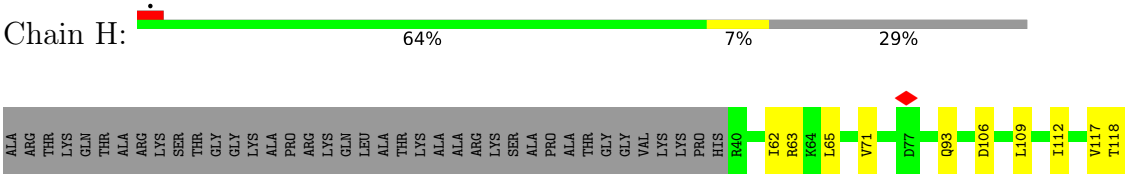
• Molecule 3: Histone H3.2

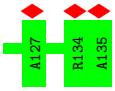
Chain D:



• Molecule 3: Histone H3.2

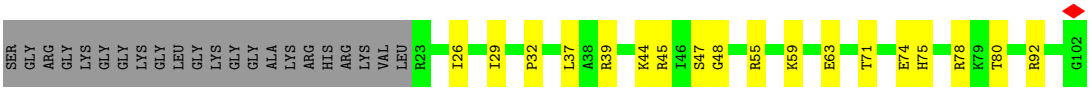
Chain H:





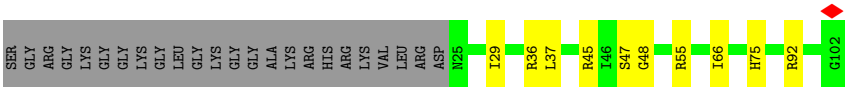
• Molecule 4: Histone H4

Chain E:



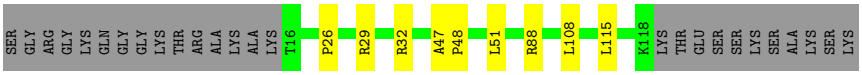
• Molecule 4: Histone H4

Chain I:



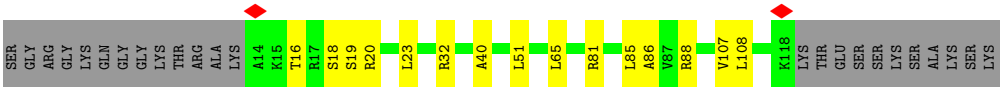
• Molecule 5: Histone H2A

Chain F:



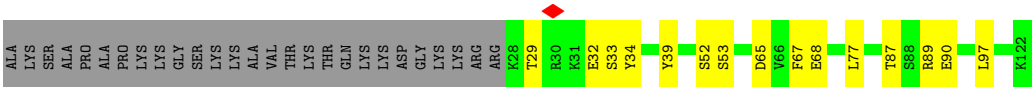
• Molecule 5: Histone H2A

Chain J:



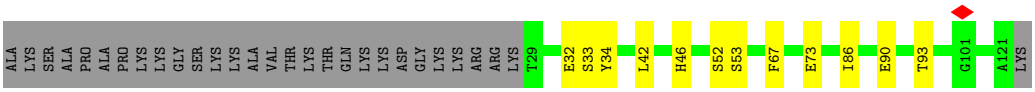
• Molecule 6: Histone H2B 1.1

Chain G:



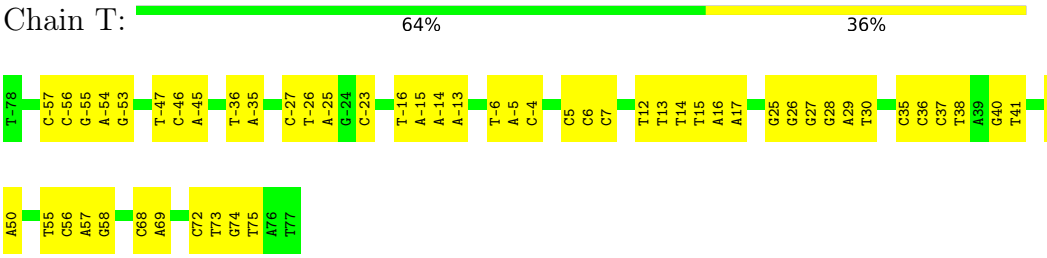
• Molecule 6: Histone H2B 1.1

Chain K:



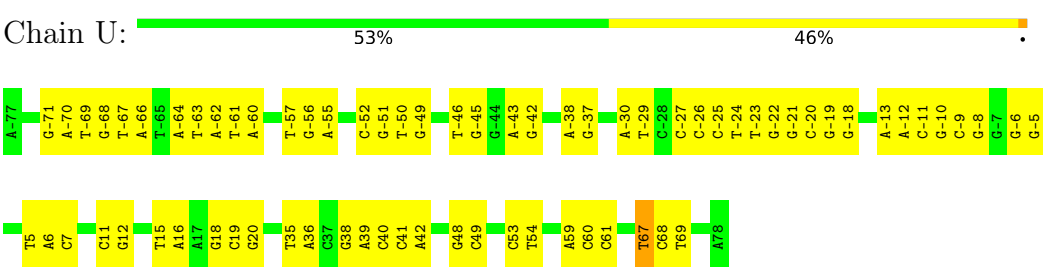
• Molecule 7: Widom601 DNA plus linker

Chain T:



• Molecule 8: Widom601 DNA plus linker

Chain U:



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	45849	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 ($e^-/Å^2$)	52.96	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	0.162	Depositor
Minimum map value	-0.073	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.002	Depositor
Recommended contour level	0.0328	Depositor
Map size (Å)	544.752, 544.752, 544.752	wwPDB
Map dimensions	312, 312, 312	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.746, 1.746, 1.746	Depositor

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, 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	0.28	0/2459	0.40	0/3308
2	C	0.33	0/1044	0.50	0/1402
3	D	0.27	0/924	0.44	0/1239
3	H	0.24	0/801	0.41	0/1073
4	E	0.26	0/645	0.45	0/862
4	I	0.25	0/626	0.46	0/837
5	F	0.25	0/805	0.42	0/1088
5	J	0.24	0/819	0.44	0/1106
6	G	0.25	0/756	0.41	0/1015
6	K	0.26	0/737	0.41	0/993
7	T	0.54	0/3558	0.94	0/5485
8	U	0.56	0/3610	0.95	1/5575 (0.0%)
All	All	0.41	0/16784	0.71	1/23983 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
8	U	67	DT	O4'-C1'-N1	5.96	112.17	108.00

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

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	2470	0	2293	72	0
2	C	1024	0	1006	28	0
3	D	911	0	966	9	0
3	H	791	0	831	12	0
4	E	638	0	676	17	0
4	I	619	0	659	9	0
5	F	795	0	846	10	0
5	J	809	0	864	11	0
6	G	745	0	773	18	0
6	K	726	0	747	10	0
7	T	3176	0	1755	44	0
8	U	3214	0	1753	50	0
9	A	7	0	0	0	0
10	A	26	0	19	1	0
All	All	15951	0	13188	247	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 9.

The worst 5 of 247 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:615:LYS:O	1:A:626:ARG:NH1	1.76	1.14
1:A:615:LYS:NZ	1:A:628:SER:HA	1.64	1.11
1:A:615:LYS:C	1:A:626:ARG:NH1	2.07	1.07
1:A:615:LYS:CE	1:A:628:SER:HB3	1.86	1.06
1:A:615:LYS:C	1:A:626:ARG:HH11	1.58	1.05

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	286/764 (37%)	271 (95%)	13 (4%)	2 (1%)	22 62
2	C	122/739 (16%)	113 (93%)	7 (6%)	2 (2%)	9 45
3	D	111/135 (82%)	109 (98%)	2 (2%)	0	100 100
3	H	94/135 (70%)	93 (99%)	1 (1%)	0	100 100
4	E	78/102 (76%)	77 (99%)	1 (1%)	0	100 100
4	I	76/102 (74%)	76 (100%)	0	0	100 100
5	F	101/129 (78%)	100 (99%)	1 (1%)	0	100 100
5	J	103/129 (80%)	102 (99%)	1 (1%)	0	100 100
6	G	93/122 (76%)	93 (100%)	0	0	100 100
6	K	91/122 (75%)	91 (100%)	0	0	100 100
All	All	1155/2479 (47%)	1125 (97%)	26 (2%)	4 (0%)	44 76

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	608	GLU
2	C	581	VAL
1	A	628	SER
1	A	755	VAL

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	264/672 (39%)	263 (100%)	1 (0%)	91 94
2	C	116/646 (18%)	115 (99%)	1 (1%)	78 88
3	D	95/110 (86%)	95 (100%)	0	100 100
3	H	83/110 (76%)	83 (100%)	0	100 100
4	E	65/78 (83%)	65 (100%)	0	100 100
4	I	63/78 (81%)	63 (100%)	0	100 100
5	F	82/101 (81%)	82 (100%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
5	J	83/101 (82%)	83 (100%)	0	100 100
6	G	81/102 (79%)	81 (100%)	0	100 100
6	K	79/102 (78%)	79 (100%)	0	100 100
All	All	1011/2100 (48%)	1009 (100%)	2 (0%)	93 96

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	629	LYS
2	C	586	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	625	GLN
1	A	748	GLN
2	C	620	HIS

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 8 ligands modelled in this entry, 7 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	$\#\|Z\| > 2$
10	SAH	A	1008	-	24,28,28	1.22	3 (12%)	25,40,40	1.77	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
10	SAH	A	1008	-	-	0/11/31/31	0/3/3/3

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
10	A	1008	SAH	C2-N3	4.12	1.38	1.32
10	A	1008	SAH	C2-N1	2.49	1.38	1.33
10	A	1008	SAH	OXT-C	-2.17	1.23	1.30

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
10	A	1008	SAH	N3-C2-N1	-5.50	120.08	128.68
10	A	1008	SAH	C5'-SD-CG	-3.51	91.73	102.27
10	A	1008	SAH	C3'-C2'-C1'	2.91	105.36	100.98
10	A	1008	SAH	OXT-C-O	-2.58	118.23	124.09
10	A	1008	SAH	OXT-C-CA	2.35	121.39	113.38

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

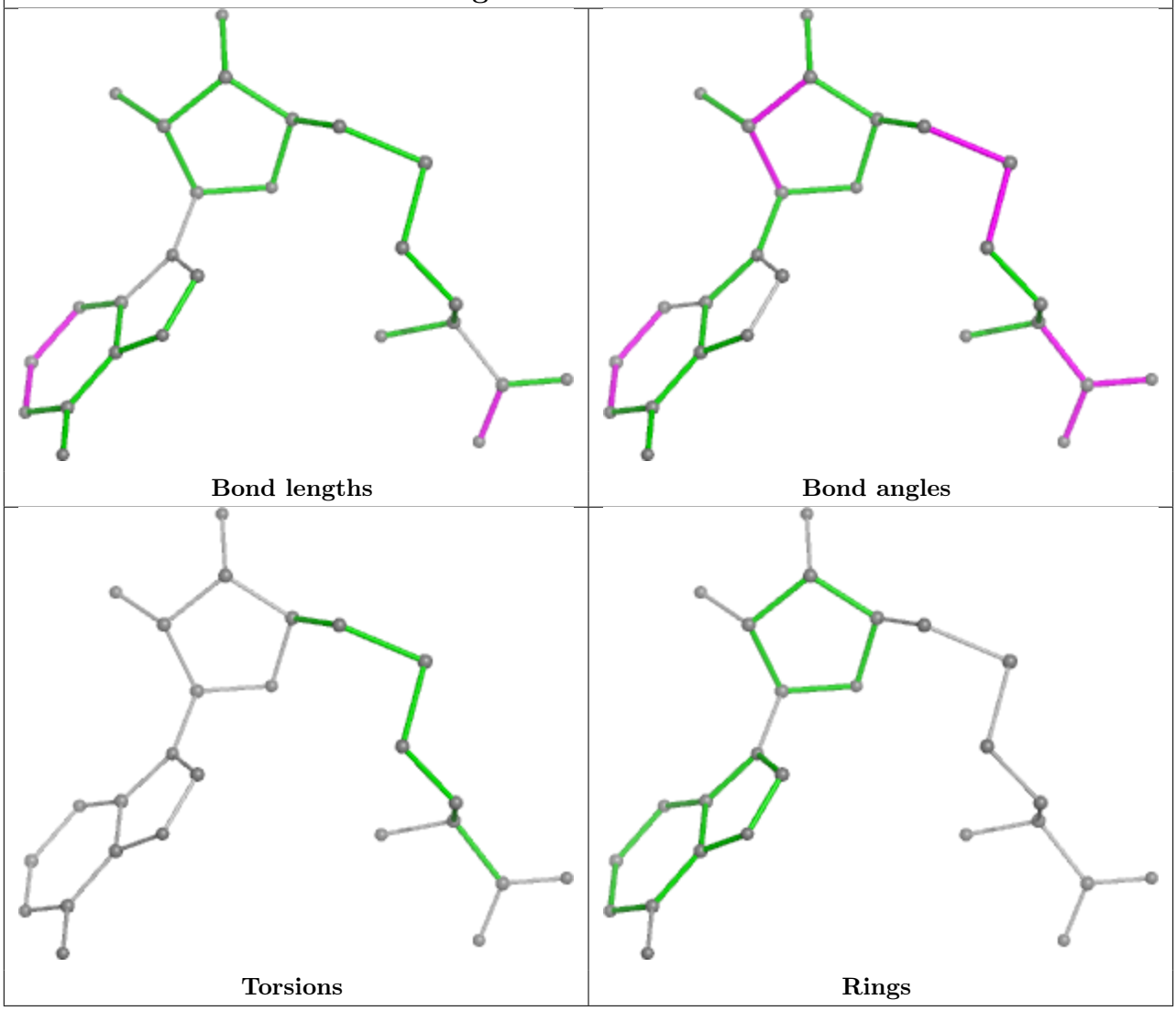
1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
10	A	1008	SAH	1	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 A 1008



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.

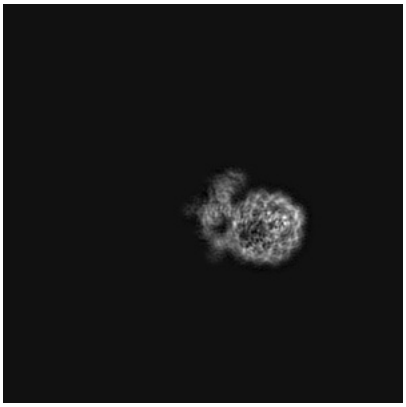
6 Map visualisation

This section contains visualisations of the EMDB entry EMD-11910. These allow visual inspection of the internal detail of the map and identification of artifacts.

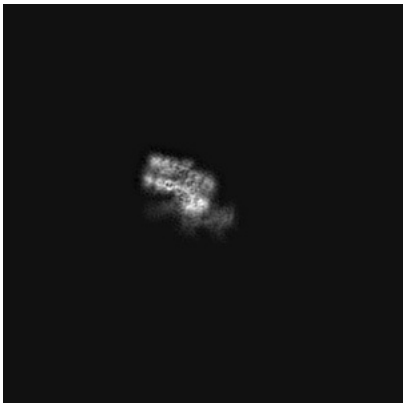
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections

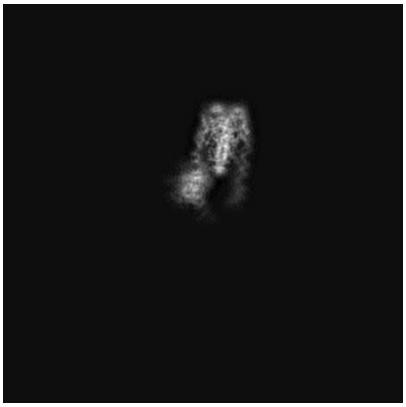
6.1.1 Primary map



X



Y

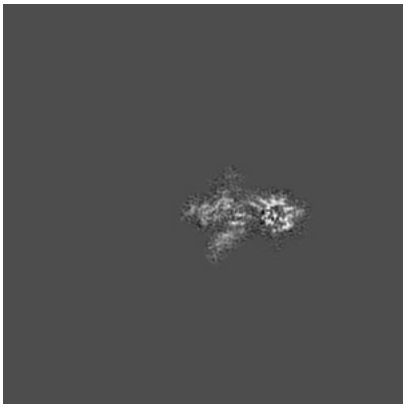


Z

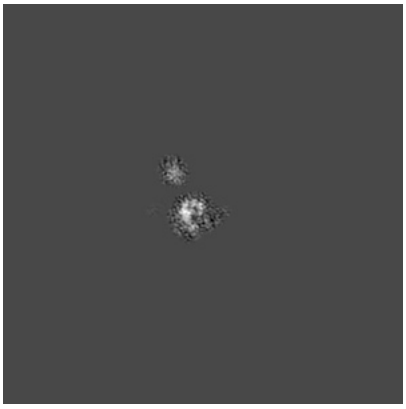
The images above show the map projected in three orthogonal directions.

6.2 Central slices

6.2.1 Primary map



X Index: 156



Y Index: 156

Z Index: 156

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: 168

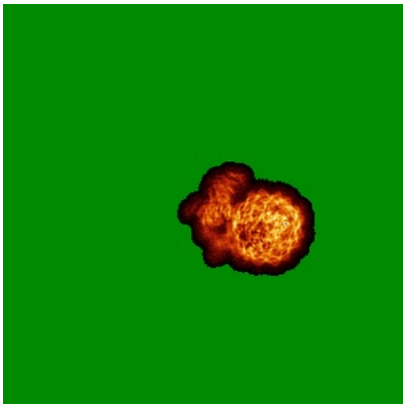
Y Index: 213

Z Index: 145

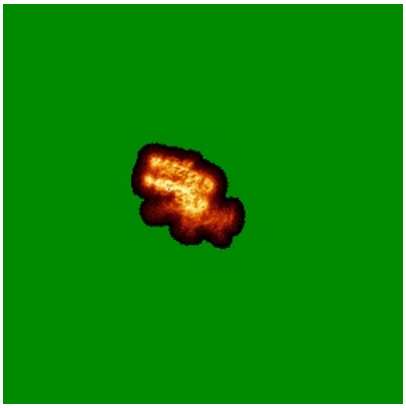
The images above show the largest variance slices of the map in three orthogonal directions.

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

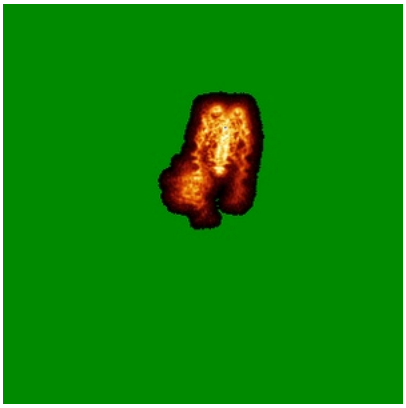
6.4.1 Primary map



X



Y

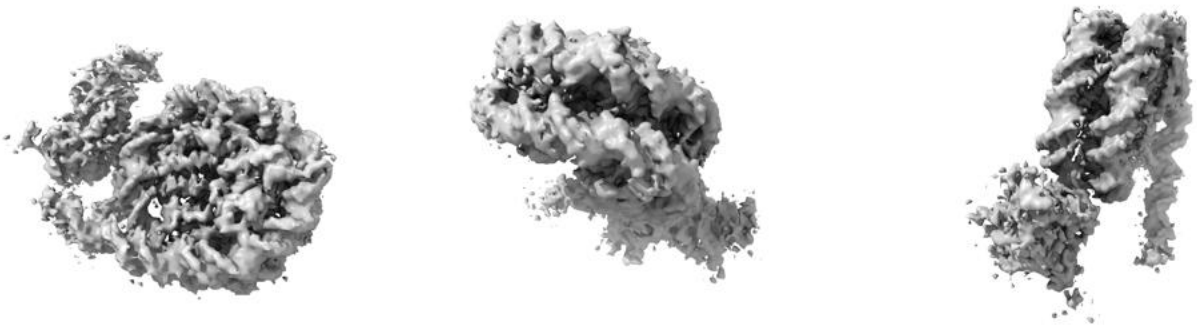


Z

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 0.0328. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

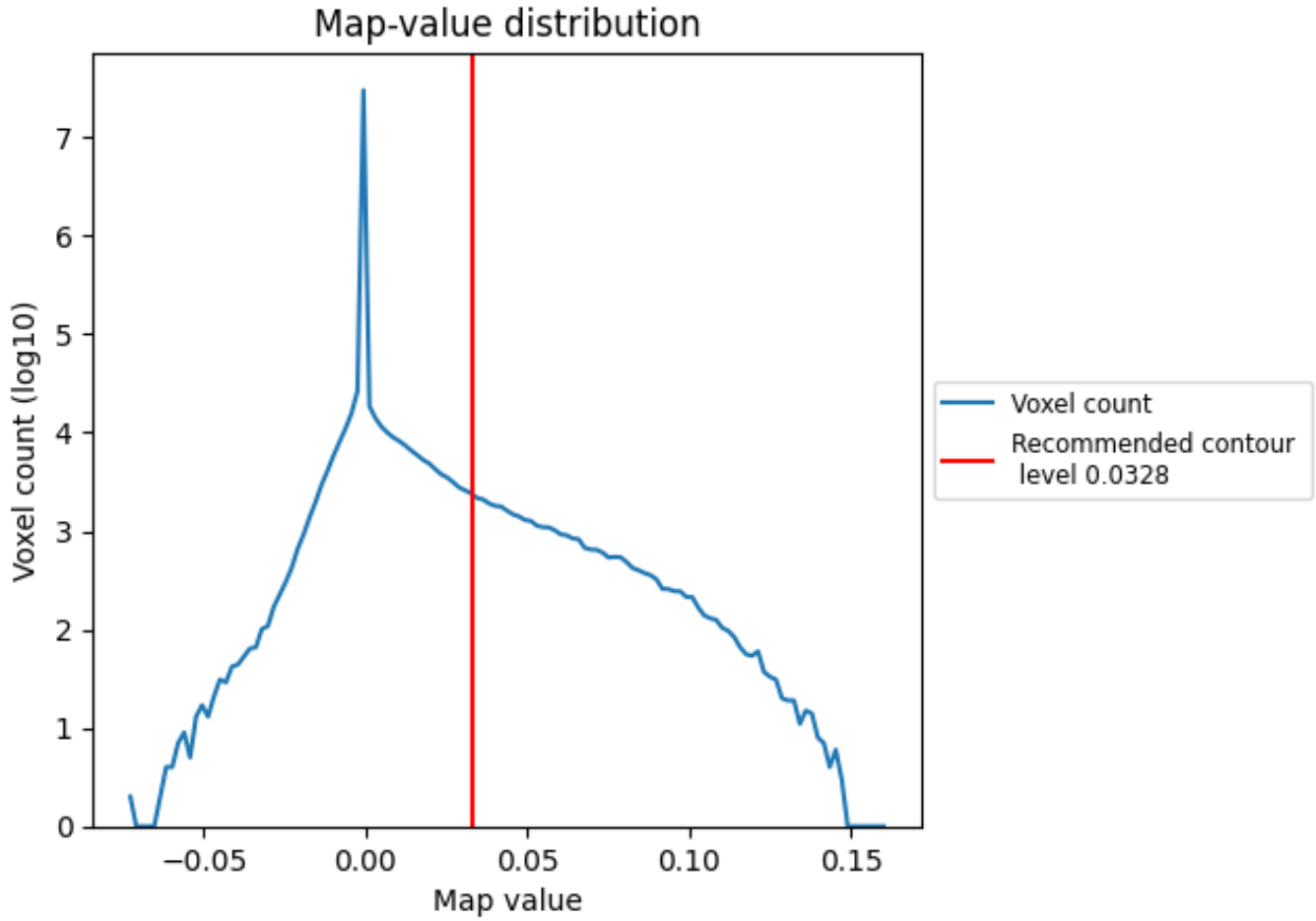
6.6 Mask visualisation (i)

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

7 Map analysis

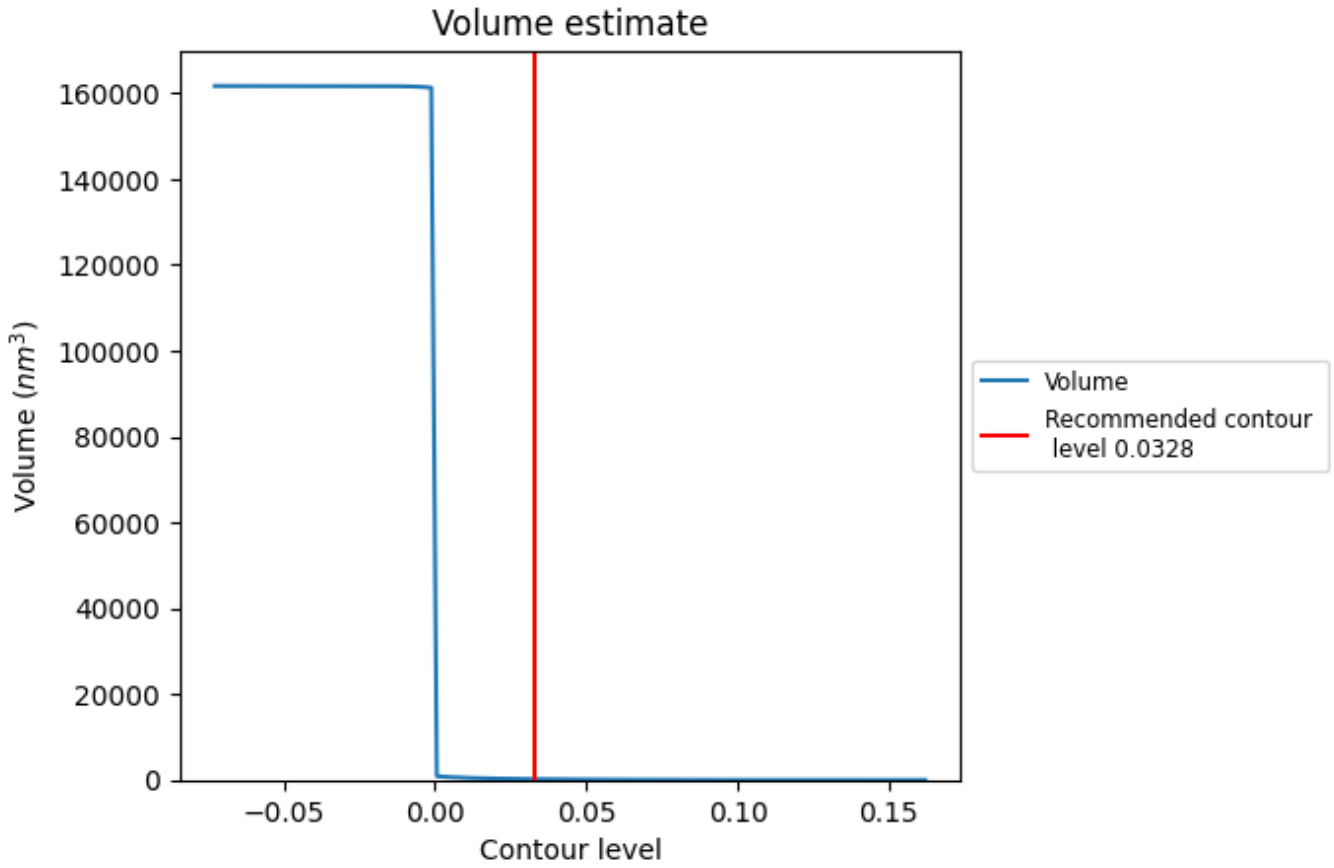
This section contains the results of statistical analysis of the map.

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.

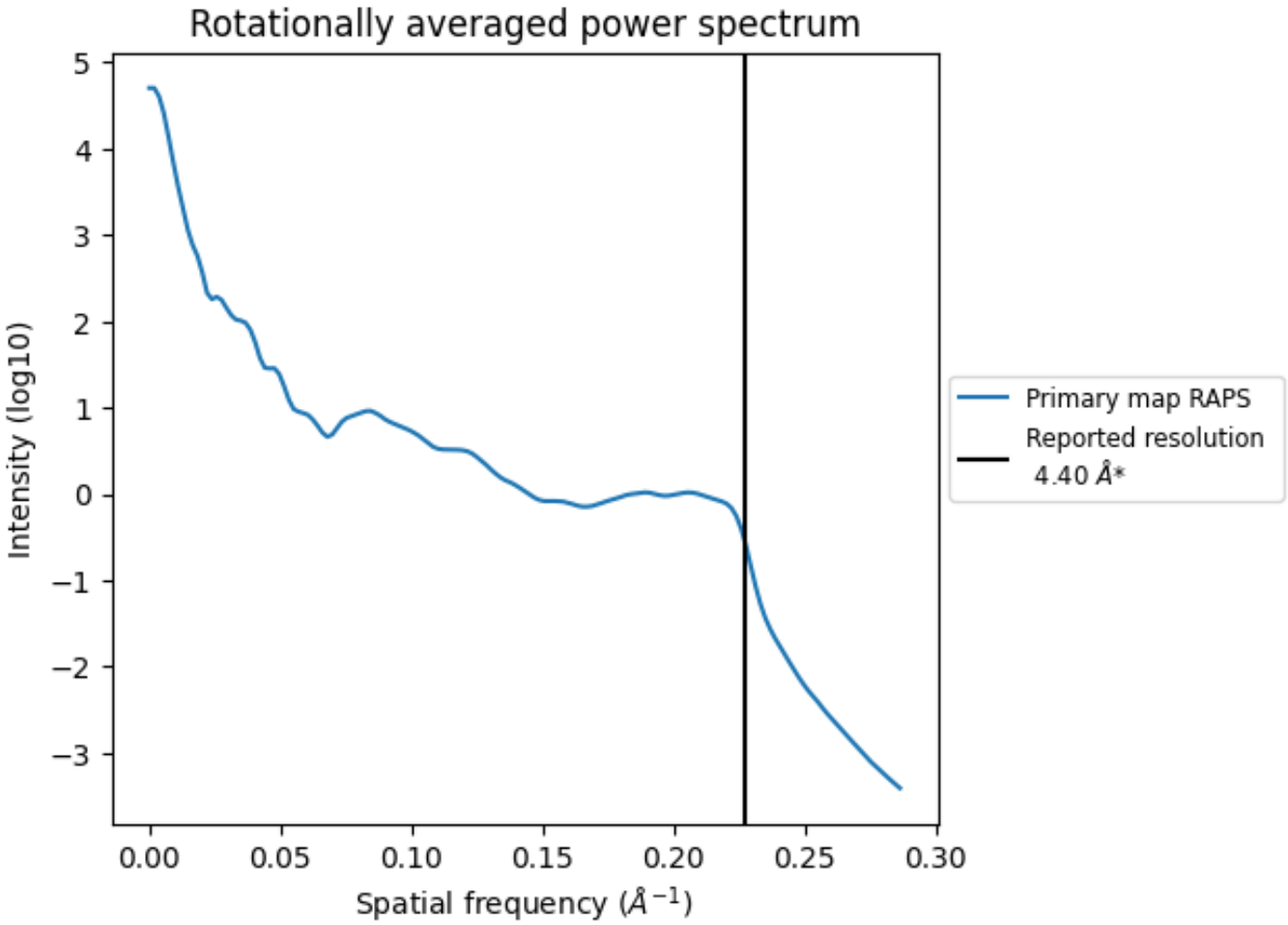
7.2 Volume estimate (i)



The volume at the recommended contour level is 193 nm^3; this corresponds to an approximate mass of 174 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 0.227 Å^{-1}

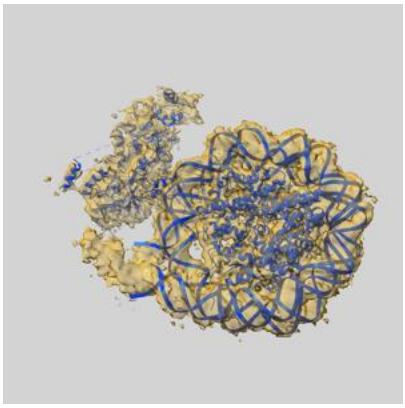
8 Fourier-Shell correlation

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

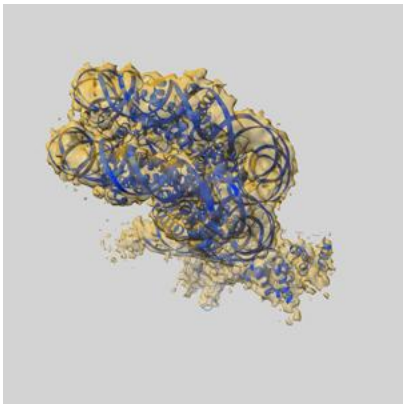
9 Map-model fit

This section contains information regarding the fit between EMDB map EMD-11910 and PDB model 7AT8. Per-residue inclusion information can be found in section 3 on page 7.

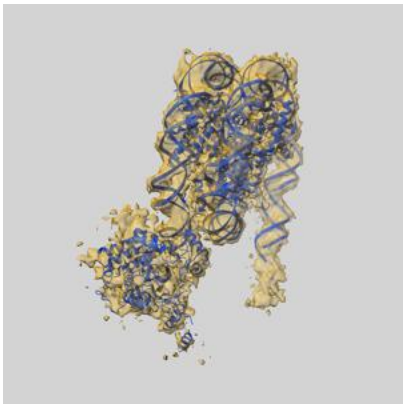
9.1 Map-model overlay



X



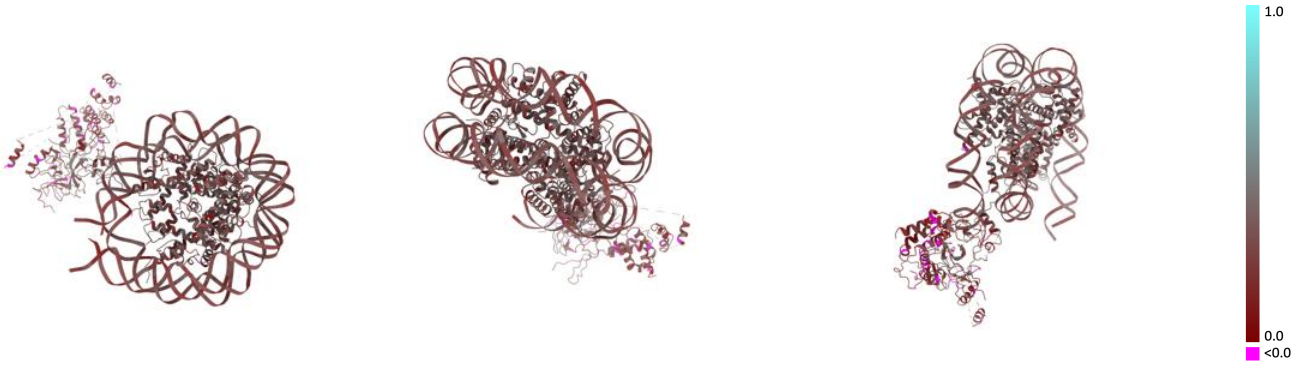
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.0328 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.

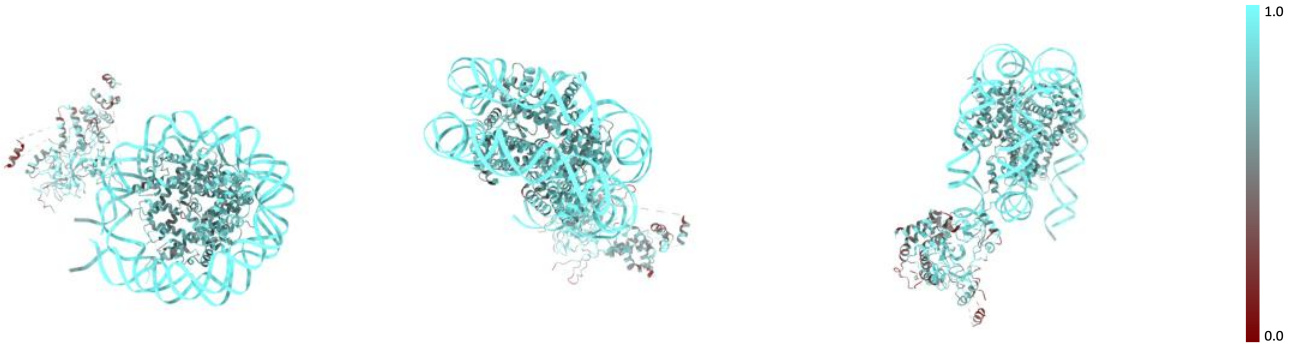
9.2 Q-score mapped to coordinate model (i)



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.

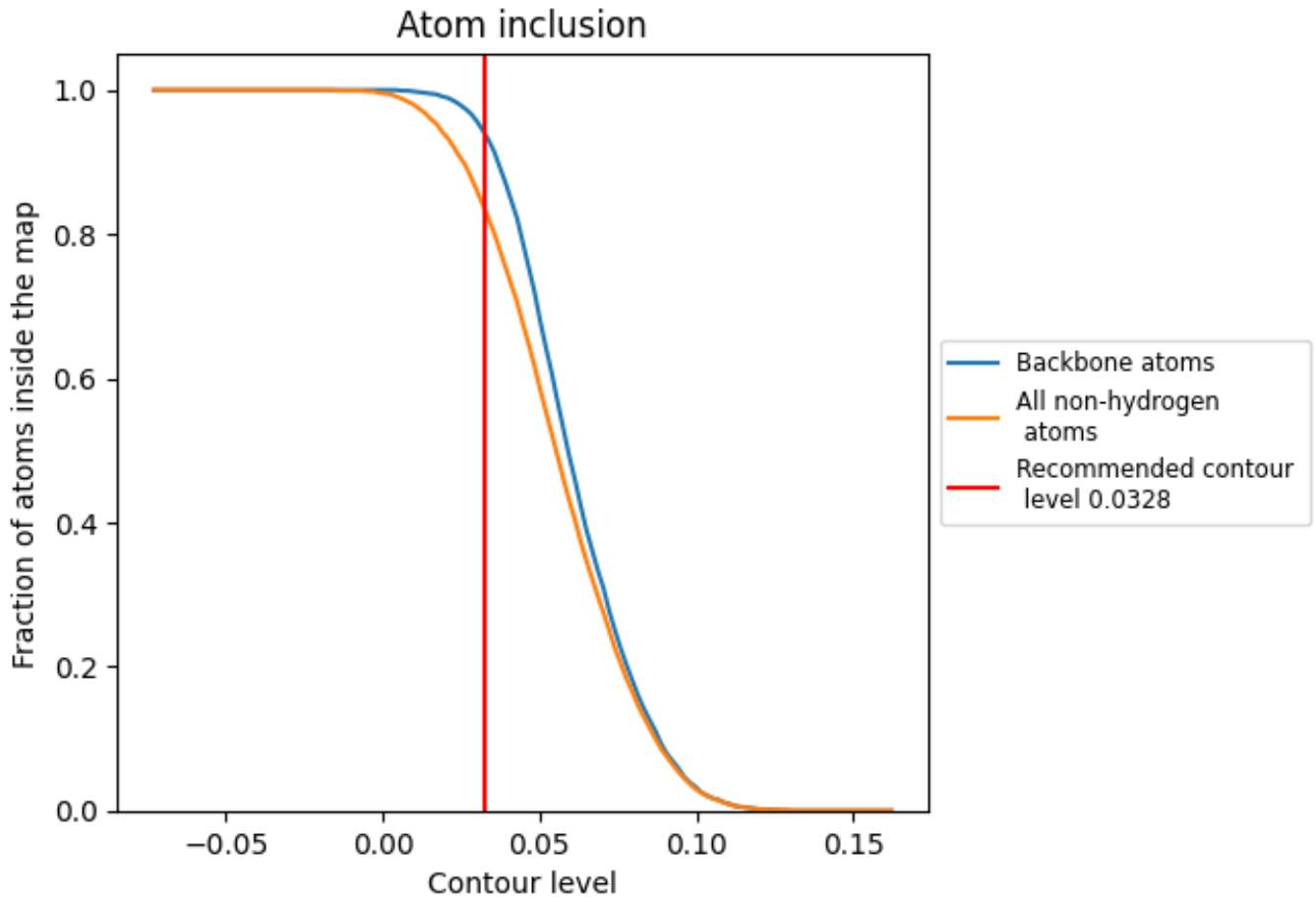
9.3 Atom inclusion mapped to coordinate model (i)



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 (0.0328).

9.4 Atom inclusion (i)



At the recommended contour level, 94% 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 (0.0328) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.8320	0.2900
A	0.7250	0.2590
C	0.6050	0.1670
D	0.7820	0.3250
E	0.8200	0.3400
F	0.8050	0.3250
G	0.7980	0.3150
H	0.7540	0.3080
I	0.8120	0.3250
J	0.7910	0.3270
K	0.8110	0.3280
T	0.9460	0.2870
U	0.9400	0.2930