



# wwPDB EM Validation Summary Report ⓘ

Oct 21, 2024 – 10:56 AM EDT

PDB ID : 8VWV  
EMDB ID : EMD-43601  
Title : OGG1 bound to a nucleosome containing SoxoG at SHL4 (composite map)  
Authors : Weaver, T.M.; Ling, J.A.; Freudenthal, B.D.  
Deposited on : 2024-02-02  
Resolution : 3.60 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](mailto: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 : 2022.3.0, CSD as543be (2022)  
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

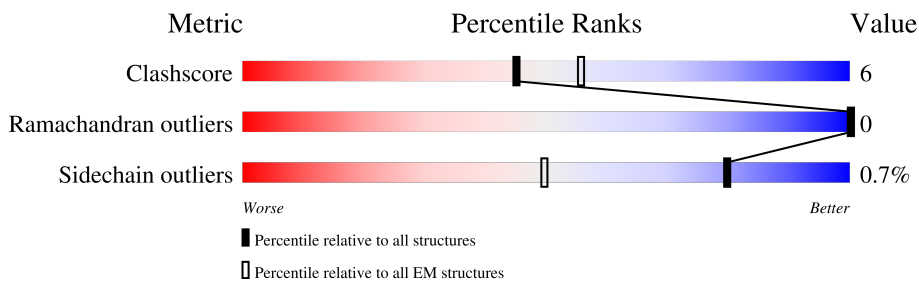
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 3.60 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  $\geq 3$ , 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	147	 <p>6% 65% 35%</p>
6	J	147	 <p>7% 59% 40%</p>
7	K	345	 <p>42% 79% 10% 11%</p>

## 2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 14474 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	97	800	505	155	138	2	0	0
1	E	97	800	505	155	138	2	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	110	ALA	CYS	engineered mutation	UNP Q71DI3
E	110	ALA	CYS	engineered mutation	UNP Q71DI3

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	B	82	652	412	127	112	1	0	0
2	F	82	657	416	128	112	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	106	816	517	158	141	0	0
3	G	110	846	535	166	145	0	0

- Molecule 4 is a protein called Histone H2B type 1-C/E/F/G/I.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	D	93	725	455	130	138	2	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					AltConf	Trace
4	H	91	Total	C	N	O	S	0	0
			708	446	125	135	2		

- Molecule 5 is a DNA chain called 601 I strand (damaged strand).

Mol	Chain	Residues	Atoms					AltConf	Trace
5	I	147	Total	C	N	O	P	0	0
			2997	1423	545	882	147		

- Molecule 6 is a DNA chain called 601 J strand (non-damaged).

Mol	Chain	Residues	Atoms					AltConf	Trace
6	J	147	Total	C	N	O	P	0	0
			3031	1435	566	883	147		

- Molecule 7 is a protein called N-glycosylase/DNA lyase.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	K	308	Total	C	N	O	S	0	0
			2442	1548	439	444	11		

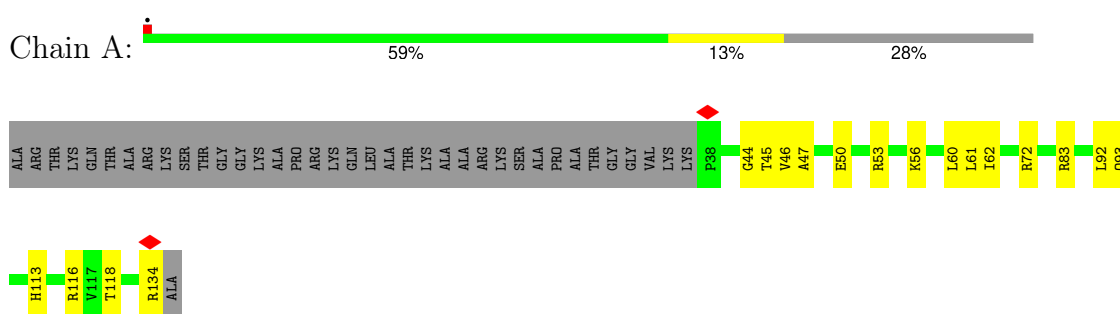
There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
K	249	GLN	LYS	engineered mutation	UNP O15527

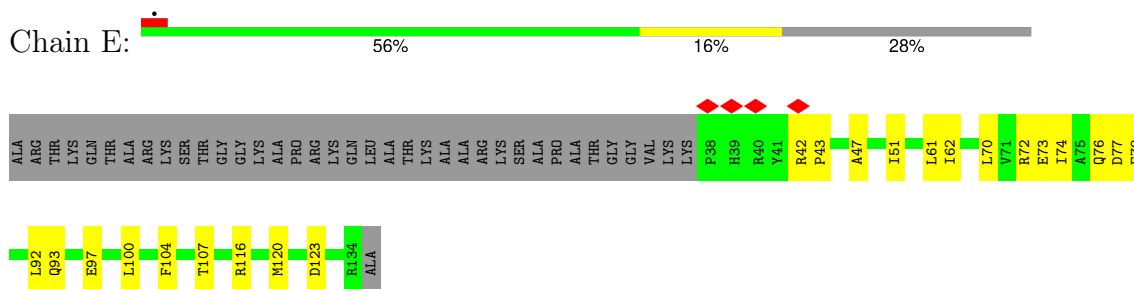
### 3 Residue-property plots [i](#)

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.

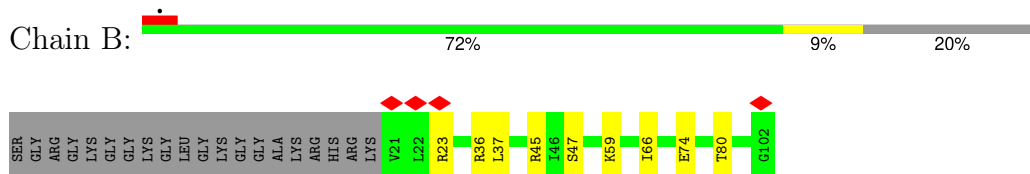
- Molecule 1: Histone H3.2



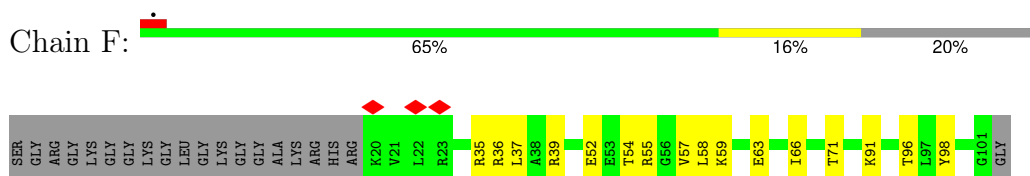
- Molecule 1: Histone H3.2



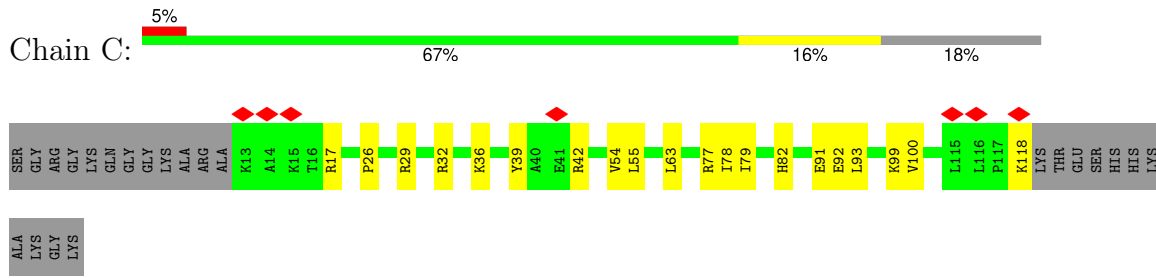
- Molecule 2: Histone H4



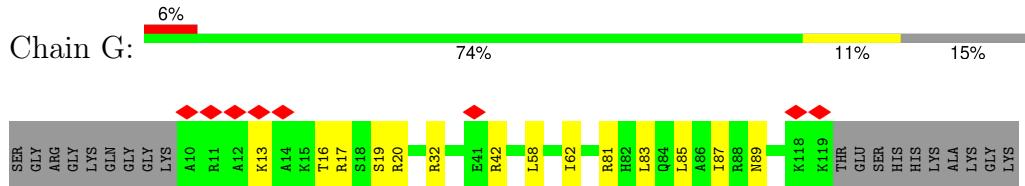
- Molecule 2: Histone H4



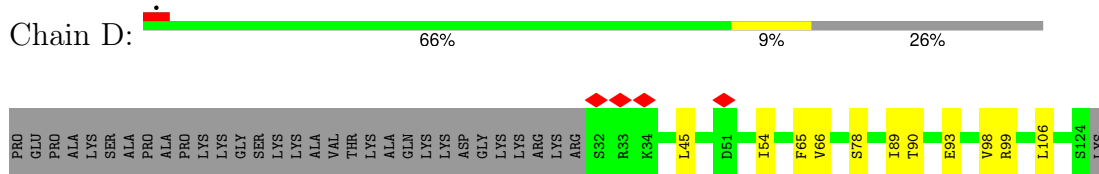
- Molecule 3: Histone H2A type 1



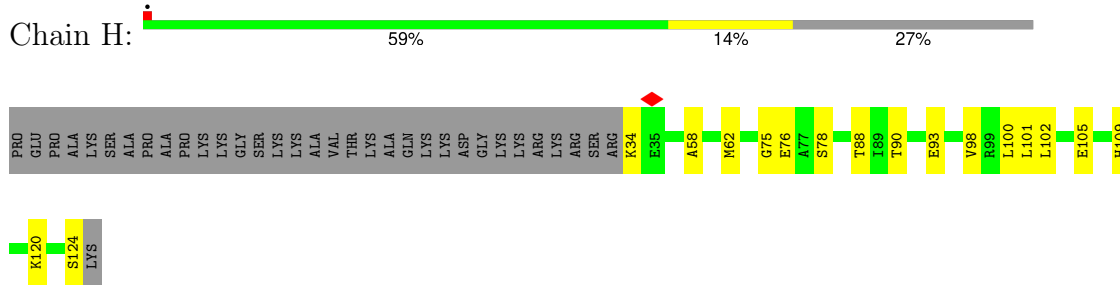
• Molecule 3: Histone H2A type 1



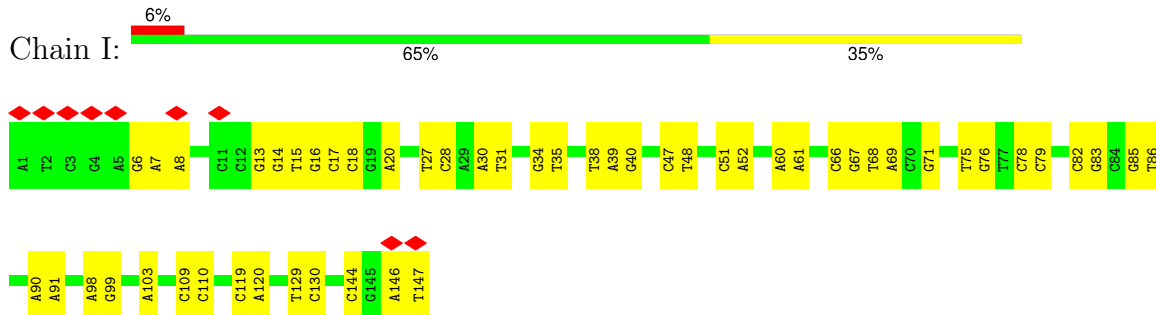
• Molecule 4: Histone H2B type 1-C/E/F/G/I



• Molecule 4: Histone H2B type 1-C/E/F/G/I

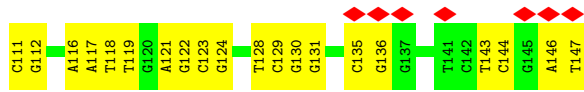
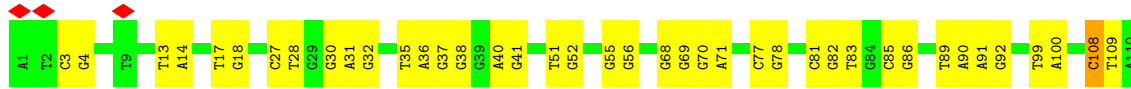


• Molecule 5: 601 I strand (damaged strand)

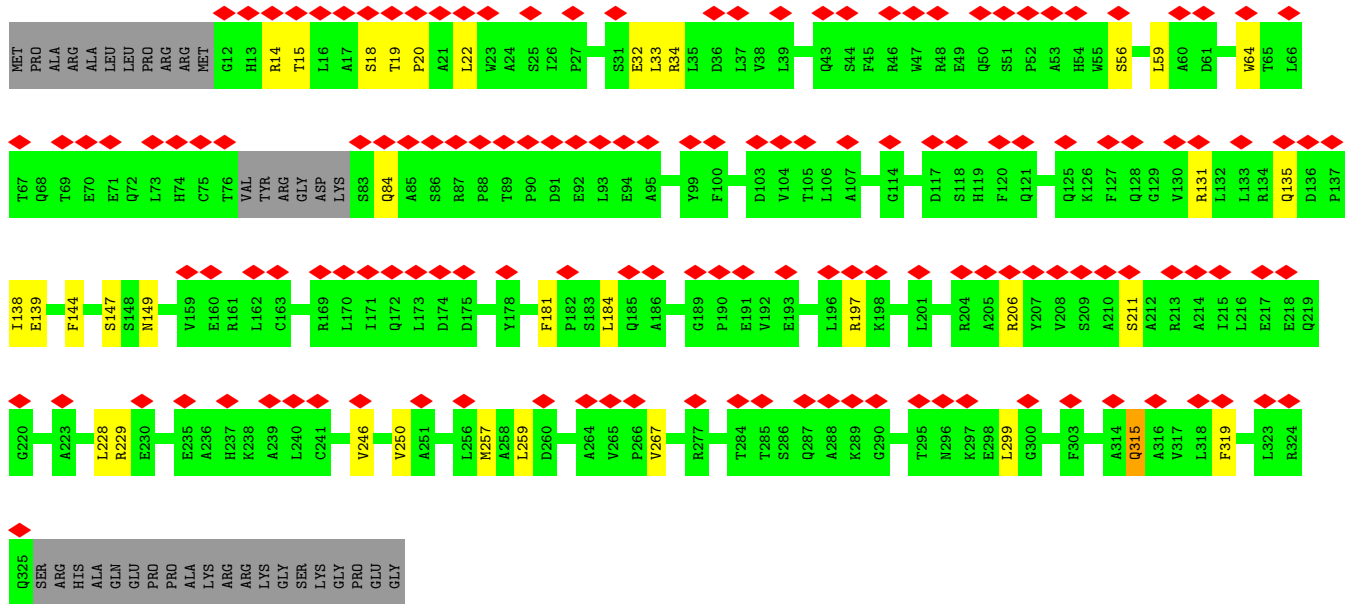
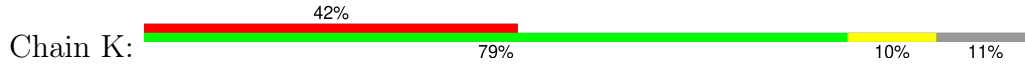


• Molecule 6: 601 J strand (non-damaged)





• Molecule 7: N-glycosylase/DNA lyase





## 4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	19114	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 ( $e^-/\text{\AA}^2$ )	60	Depositor
Minimum defocus (nm)	500	Depositor
Maximum defocus (nm)	2500	Depositor
Magnification	Not provided	
Image detector	GATAN K3 BIOQUANTUM (6k x 4k)	Depositor
Maximum map value	33.684	Depositor
Minimum map value	-20.127	Depositor
Average map value	0.000	Depositor
Map value standard deviation	1.004	Depositor
Recommended contour level	8	Depositor
Map size ( $\text{\AA}$ )	320.4, 320.4, 320.4	wwPDB
Map dimensions	600, 600, 600	wwPDB
Map angles ( $^\circ$ )	90.0, 90.0, 90.0	wwPDB
Pixel spacing ( $\text{\AA}$ )	0.534, 0.534, 0.534	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: 8OG

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	0.30	0/812	0.62	0/1089
1	E	0.30	0/812	0.63	0/1089
2	B	0.32	0/659	0.61	0/883
2	F	0.29	0/664	0.60	0/889
3	C	0.31	0/826	0.60	0/1114
3	G	0.30	0/856	0.58	0/1153
4	D	0.32	0/736	0.58	0/990
4	H	0.31	0/719	0.56	0/968
5	I	0.54	0/3331	0.94	0/5131
6	J	0.54	0/3403	0.92	1/5255 (0.0%)
7	K	0.43	0/2509	0.64	0/3416
All	All	0.44	0/15327	0.78	1/21977 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	J	108	DC	C3'-C2'-C1'	-5.14	96.33	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	800	0	839	18	0
1	E	800	0	839	17	0
2	B	652	0	696	10	0
2	F	657	0	706	13	0
3	C	816	0	879	18	0
3	G	846	0	915	9	0
4	D	725	0	743	10	0
4	H	708	0	725	11	0
5	I	2997	0	1650	38	0
6	J	3031	0	1651	40	0
7	K	2442	0	2374	21	0
All	All	14474	0	12017	161	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 6.

The worst 5 of 161 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:E:100:LEU:HD11	2:F:58:LEU:HD12	1.75	0.69
1:E:92:LEU:HD11	2:F:66:ILE:HD11	1.74	0.68
3:C:39:TYR:HB3	4:D:78:SER:HB2	1.75	0.66
3:C:63:LEU:HD12	4:D:45:LEU:HD13	1.77	0.66
7:K:32:GLU:HG3	7:K:131:ARG:HH21	1.60	0.66

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%)	91 (96%)	4 (4%)	0	100	100
1	E	95/135 (70%)	93 (98%)	2 (2%)	0	100	100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	80/102 (78%)	80 (100%)	0	0	100	100
2	F	80/102 (78%)	75 (94%)	5 (6%)	0	100	100
3	C	104/129 (81%)	102 (98%)	2 (2%)	0	100	100
3	G	108/129 (84%)	104 (96%)	4 (4%)	0	100	100
4	D	91/125 (73%)	91 (100%)	0	0	100	100
4	H	89/125 (71%)	86 (97%)	3 (3%)	0	100	100
7	K	304/345 (88%)	291 (96%)	13 (4%)	0	100	100
All	All	1046/1327 (79%)	1013 (97%)	33 (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	84/109 (77%)	84 (100%)	0	100	100
1	E	84/109 (77%)	84 (100%)	0	100	100
2	B	67/78 (86%)	67 (100%)	0	100	100
2	F	68/78 (87%)	68 (100%)	0	100	100
3	C	83/98 (85%)	82 (99%)	1 (1%)	67	82
3	G	85/98 (87%)	85 (100%)	0	100	100
4	D	80/105 (76%)	80 (100%)	0	100	100
4	H	78/105 (74%)	78 (100%)	0	100	100
7	K	259/288 (90%)	254 (98%)	5 (2%)	52	73
All	All	888/1068 (83%)	882 (99%)	6 (1%)	80	90

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
7	K	211	SER

Continued on next page...

*Continued from previous page...*

Mol	Chain	Res	Type
7	K	228	LEU
7	K	315	GLN
7	K	197	ARG
3	C	36	LYS

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](#)

1 non-standard protein/DNA/RNA residue 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	$\# Z  > 2$
5	8OG	I	34	5	22,25,26	0.80	0	26,37,40	0.79	0

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
5	8OG	I	34	5	-	1/7/21/22	0/3/3/3

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	I	34	8OG	C3'-C4'-C5'-O5'

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	I	34	8OG	2	0

## 5.5 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

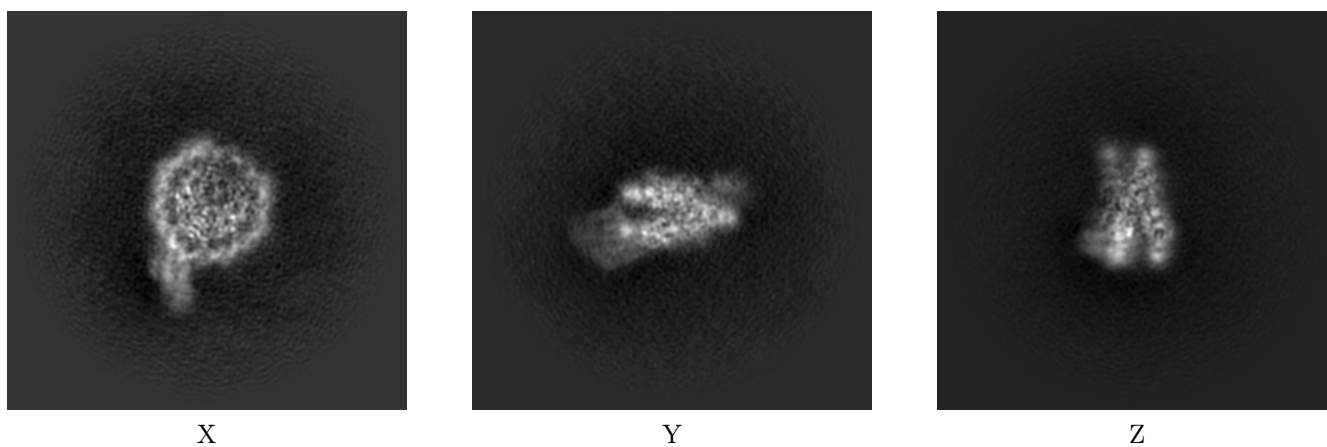
## 6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-43601. These allow visual inspection of the internal detail of the map and identification of artifacts.

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

### 6.1 Orthogonal projections [i](#)

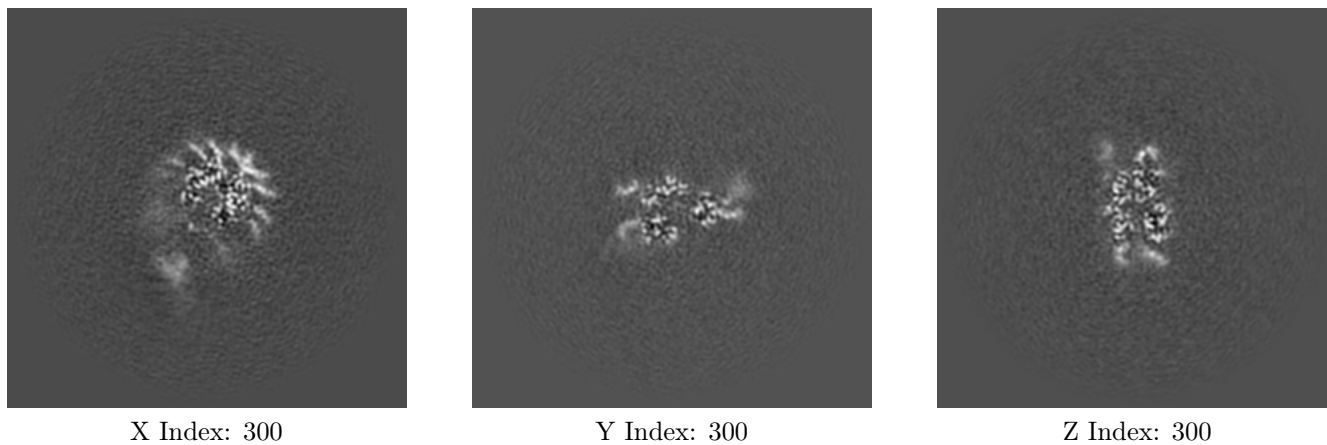
#### 6.1.1 Primary map



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

### 6.2 Central slices [i](#)

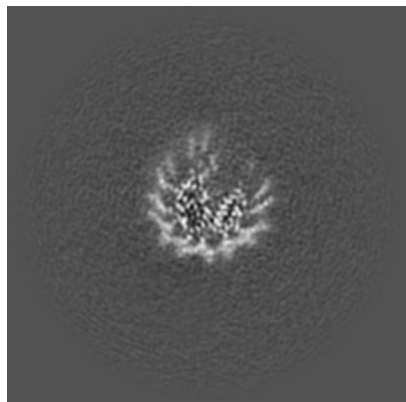
#### 6.2.1 Primary map



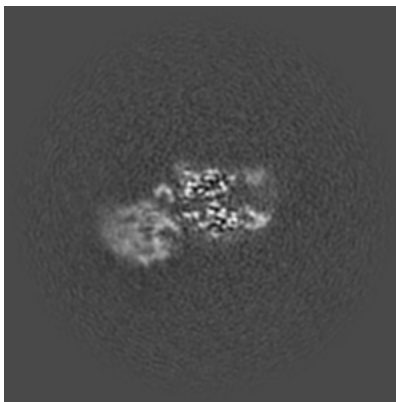
The images above show central slices of the map in three orthogonal directions.

## 6.3 Largest variance slices [i](#)

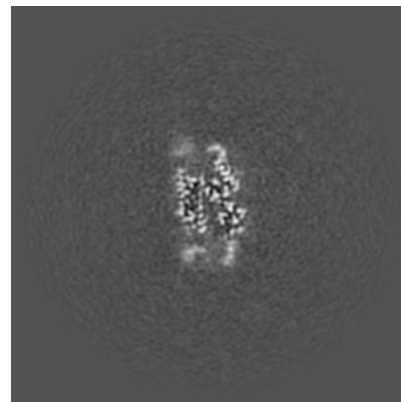
### 6.3.1 Primary map



X Index: 323



Y Index: 263

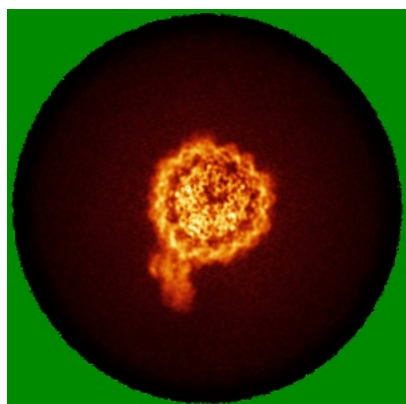


Z Index: 288

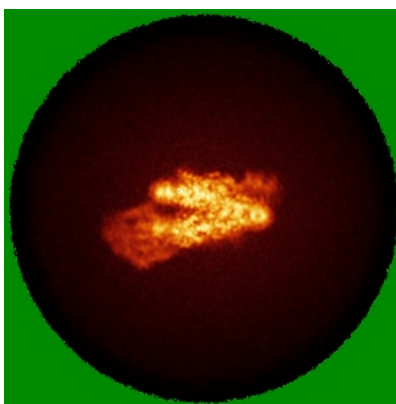
The images above show the largest variance slices of the map in three orthogonal directions.

## 6.4 Orthogonal standard-deviation projections (False-color) [i](#)

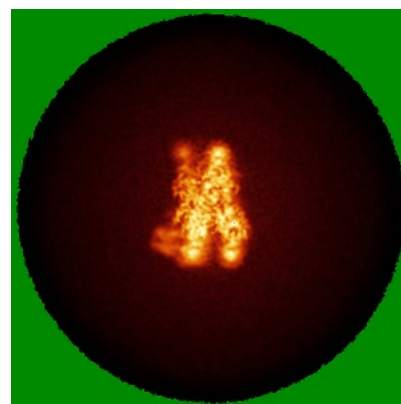
### 6.4.1 Primary map



X



Y



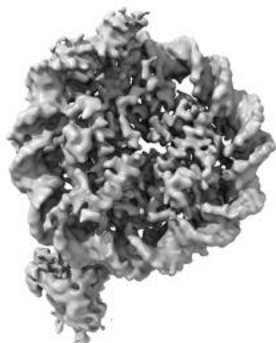
Z

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.

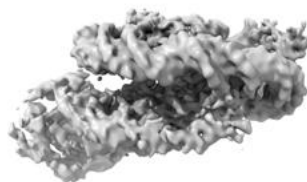


## 6.5 Orthogonal surface views [i](#)

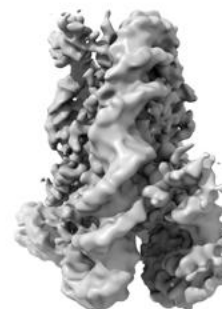
### 6.5.1 Primary map



X



Y



Z

The images above show the 3D surface view of the map at the recommended contour level 8.0. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

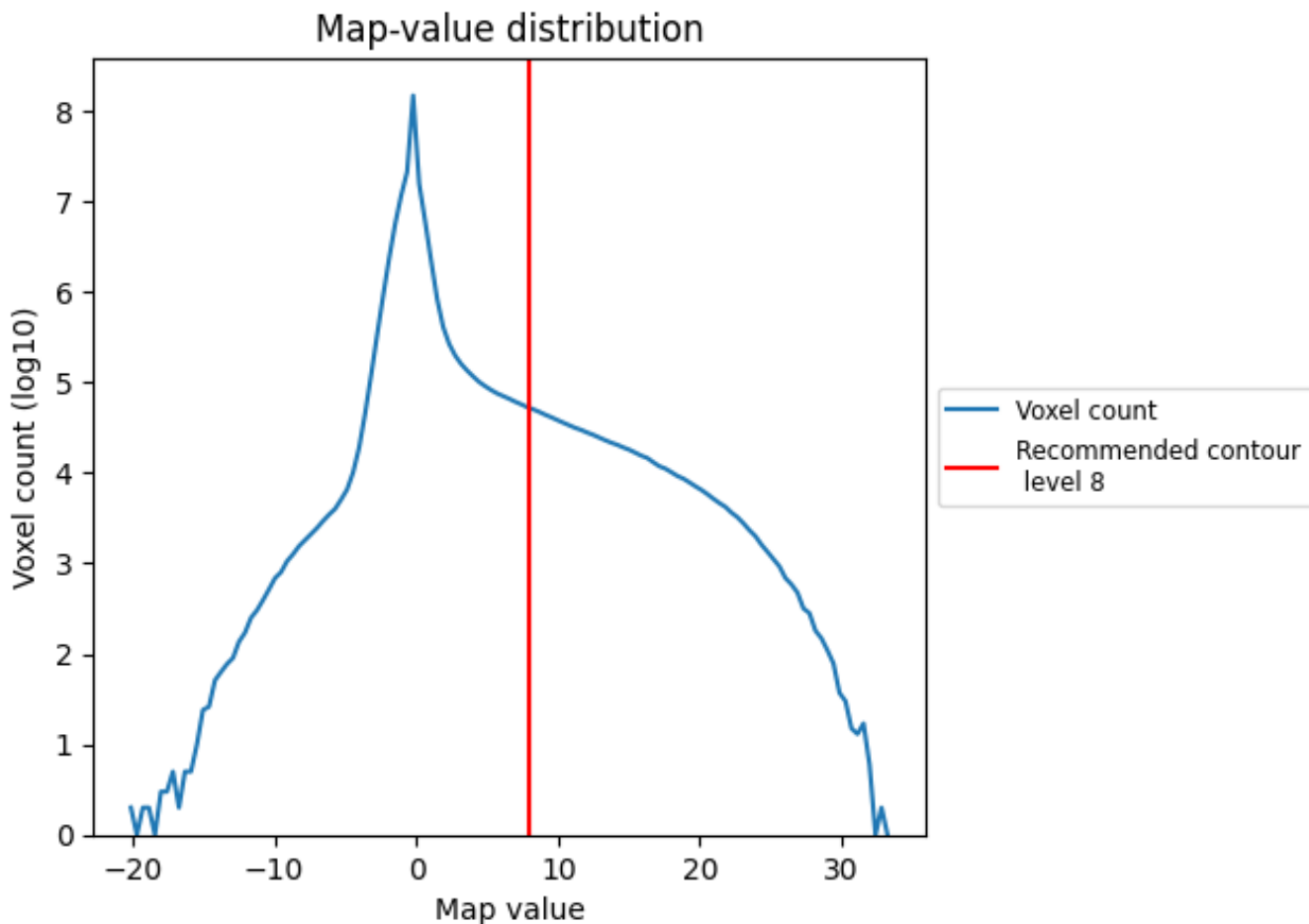
## 6.6 Mask visualisation [i](#)

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

## 7 Map analysis [i](#)

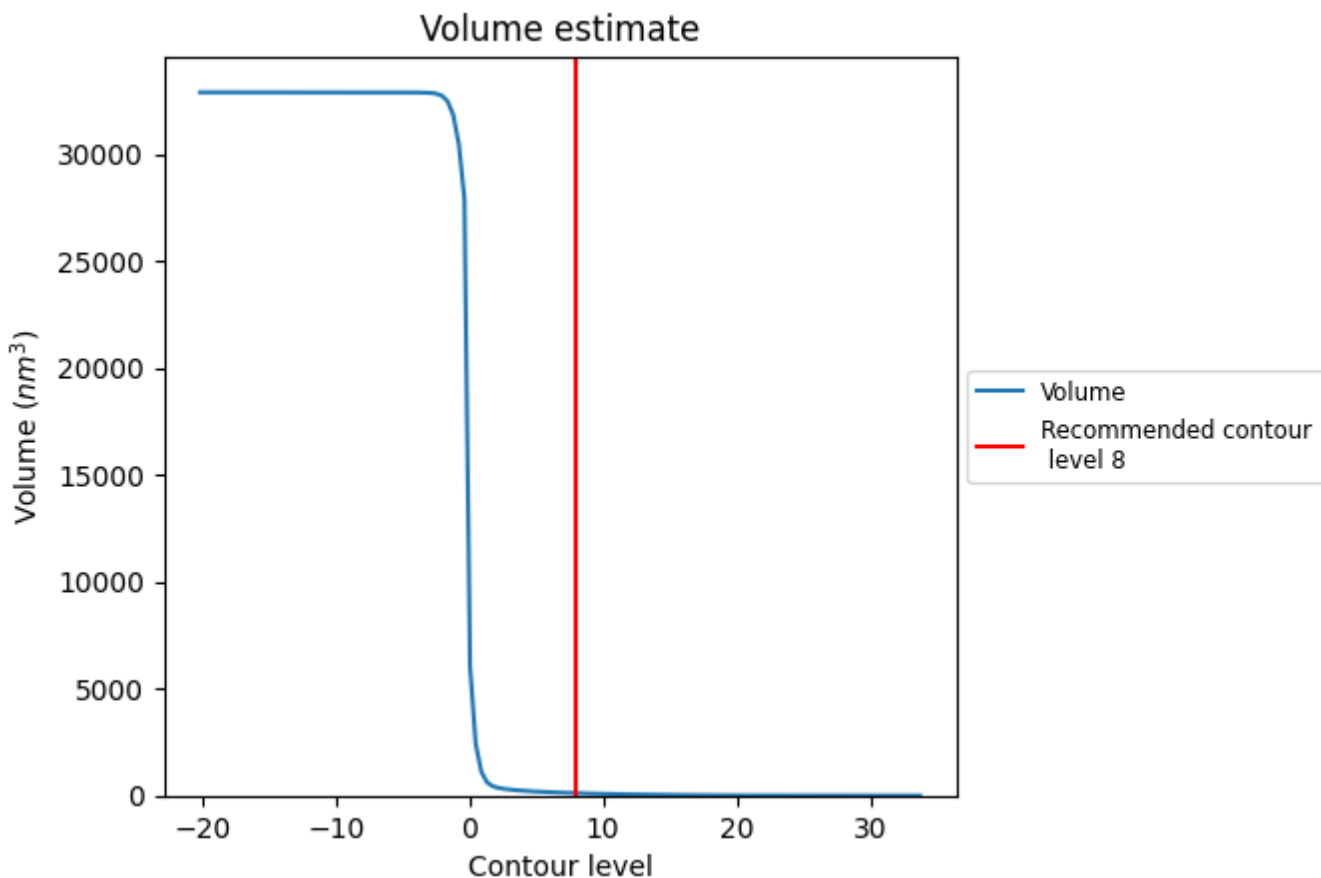
This section contains the results of statistical analysis of the map.

### 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.

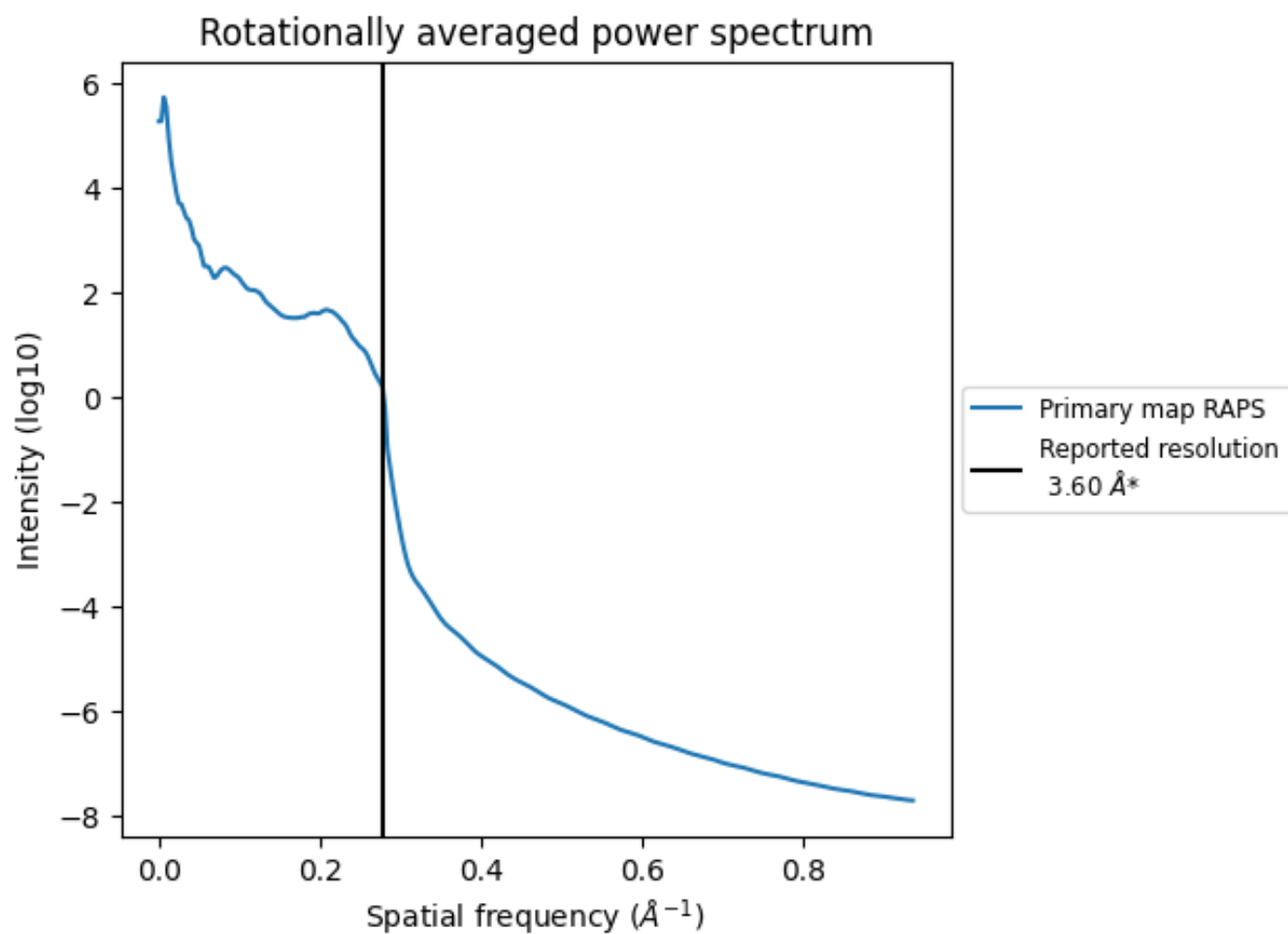
## 7.2 Volume estimate [\(i\)](#)



The volume at the recommended contour level is 112 nm<sup>3</sup>; this corresponds to an approximate mass of 101 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  $0.278 \text{\AA}^{-1}$

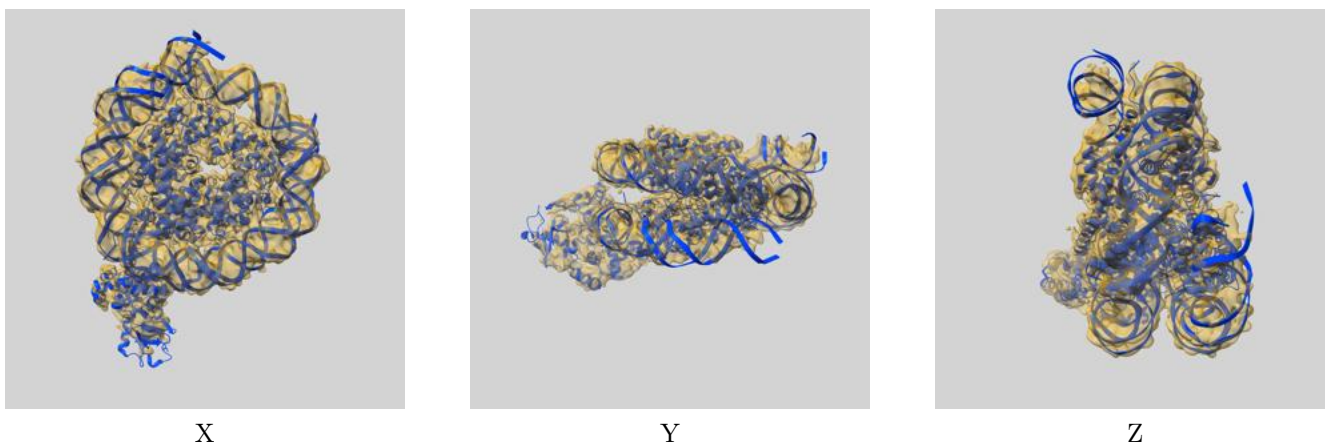
## 8 Fourier-Shell correlation

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

## 9 Map-model fit [i](#)

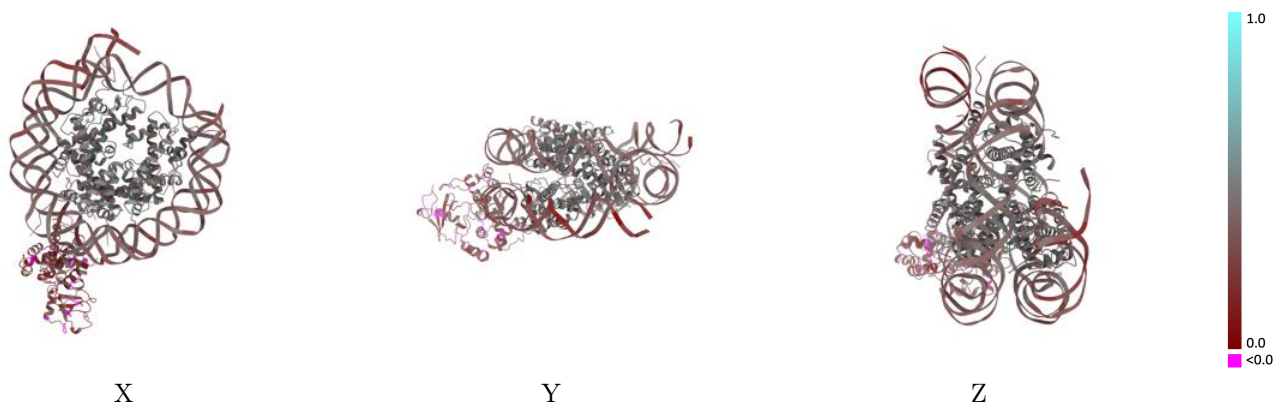
This section contains information regarding the fit between EMDB map EMD-43601 and PDB model 8VWV. Per-residue inclusion information can be found in section 3 on page 6.

### 9.1 Map-model overlay [i](#)



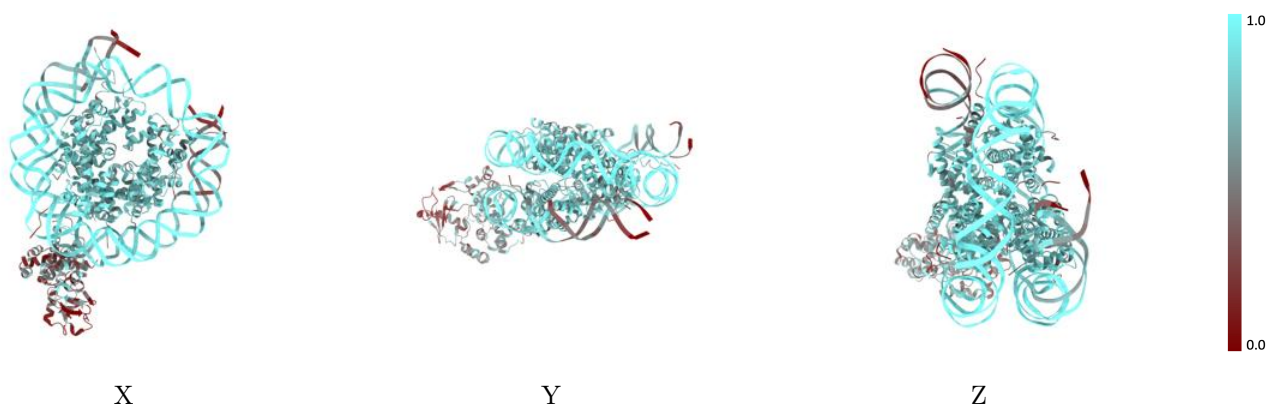
The images above show the 3D surface view of the map at the recommended contour level 8.0 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

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



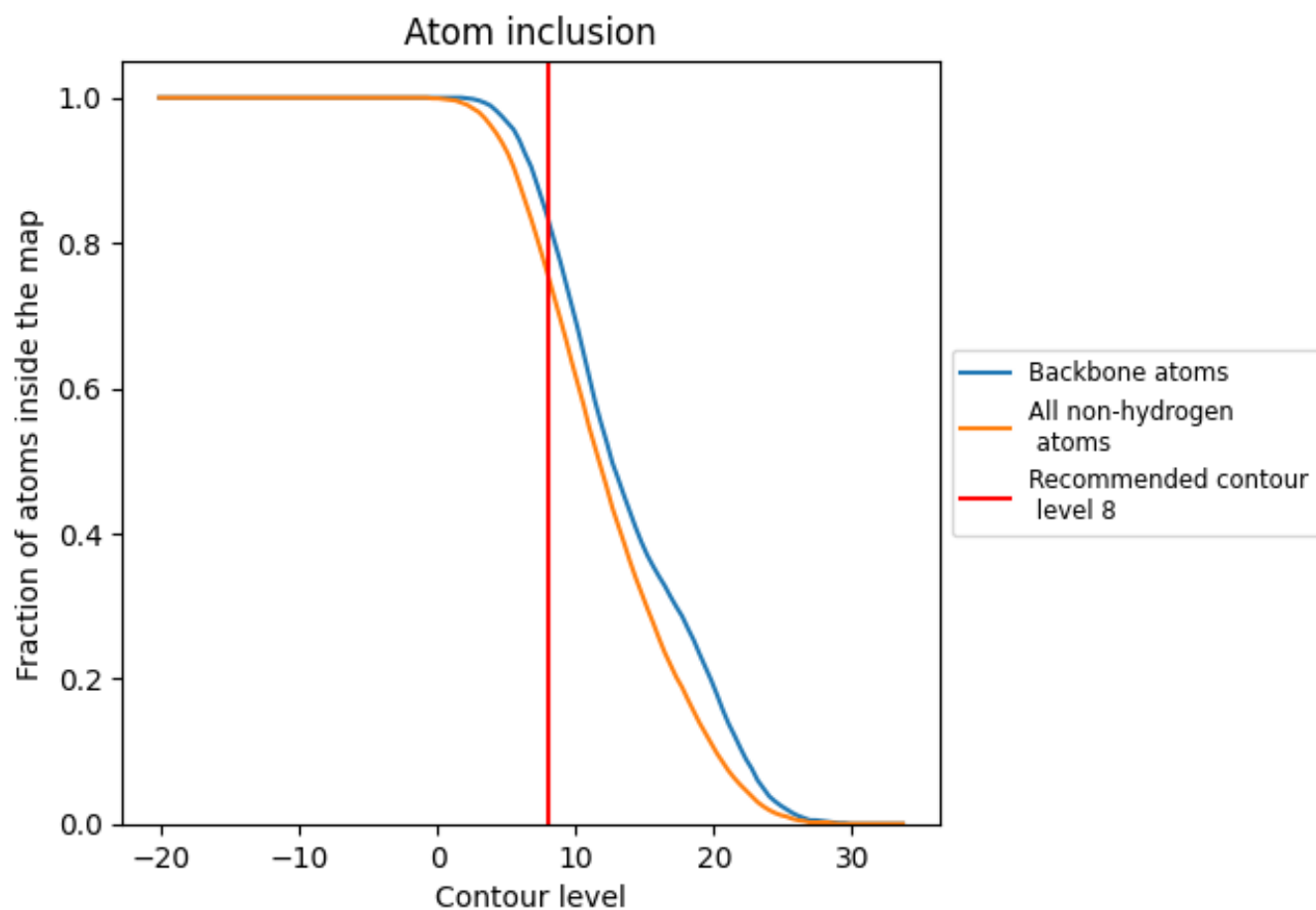
The images above show the model with each residue coloured according to its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions 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 (8).

## 9.4 Atom inclusion [i](#)



























At the recommended contour level, 83% of all backbone atoms, 76% 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 (8) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	 0.7560	 0.3590
A	 0.8060	 0.4500
B	 0.8100	 0.4690
C	 0.7700	 0.4430
D	 0.7970	 0.4410
E	 0.7850	 0.4460
F	 0.8210	 0.4700
G	 0.7970	 0.4550
H	 0.8230	 0.4520
I	 0.8420	 0.3390
J	 0.8480	 0.3430
K	 0.4260	 0.1710

