

wwPDB EM Validation Summary Report (i)

Dec 11, 2022 - 07:41 pm GMT

PDB ID : 6T7D EMDB ID : EMD-10394 Title : Structure of human Sox11 transcription factor in complex with a nucleosome Authors Dodonova, S.O.; Zhu, F.; Dienemann, C.; Taipale, J.; Cramer, P. : Deposited on 2019-10-21 : 4.40 Å(reported) Resolution : Based on initial model 6FQ5 ·

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:

:	0.0.1. dev 43
:	4.02b-467
:	20191225.v01 (using entries in the PDB archive December 25th 2019)
:	1.9.9
:	Engh & Huber (2001)
:	Parkinson et al. (1996)
:	2.31.3
	::

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.

Clashscore

Ramachandran outliers

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	Percentile Ranks	Value
Clashscore		3
Ramachandran outliers		0
Worse	Better	
Percentile relativ	ve to all structures	
Percentile relativ	ve to all EM structures	
Metric	Whole archive	EM structures
	(# Entries)	(# Entries)

158937

154571

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.

4297

4023

Mol	Chain	Length	Quality of chain	
1	А	136	72%	28%
1	Е	136	71%	29%
2	В	103	6% 82%	18%
2	F	103	9%	• 16%
3	С	151	73%	27%
3	G	151	5%	27%
4	D	126	75%	25%
4	Н	126	- 75%	25%
5	Ι	151	72%	5% 23%

Continued on next page...



Continued from previous page...

Mol	Chain	Length	Quality of chain						
6	J	151	70%	8%	23%				
7	К	109	8% 71%		29%				



2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 8209 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.

Mol	Chain	Residues	Atoms				AltConf	Trace
1	Λ	98	Total	С	Ν	0	0	0
1	1 A		392	196	98	98	0	0
1	E	97	Total	С	Ν	0	0	0
1	Ľ	91	388	194	97	97	0	0

• Molecule 1 is a protein called Histone H3.2.

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
А	110	ALA	CYS	conflict	UNP Q71DI3
Е	110	ALA	CYS	conflict	UNP Q71DI3

• Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms				AltConf	Trace
9	В	84	Total	С	Ν	0	0	0
	2 В	04	336	168	84	84	0	0
0	Б	87	Total	С	Ν	0	0	0
	2 F	01	349	174	87	88	U	0

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

Mol	Chain	Residues	Atoms				AltConf	Trace
3	С	110	Total 440			O 110	0	0
3	G	110	Total 440	C 220	N 110	O 110	0	0

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
С	-21	MET	-	initiating methionine	UNP P04908
С	-20	GLY	-	expression tag	UNP P04908
С	-19	SER	-	expression tag	UNP P04908

Continued on next page...



Chain	Residue	Modelled	Actual	Comment	Reference
С	-18	SER	-	expression tag	UNP P04908
С	-17	HIS	-	expression tag	UNP P04908
С	-16	HIS	-	expression tag	UNP P04908
С	-15	HIS	-	expression tag	UNP P04908
С	-14	HIS	-	expression tag	UNP P04908
С	-13	HIS	-	expression tag	UNP P04908
С	-12	HIS	-	expression tag	UNP P04908
С	-11	GLU	-	expression tag	UNP P04908
С	-10	ASN	-	expression tag	UNP P04908
С	-9	LEU	-	expression tag	UNP P04908
С	-8	TYR	-	expression tag	UNP P04908
С	-7	PHE	-	expression tag	UNP P04908
С	-6	GLN	-	expression tag	UNP P04908
С	-5	SER	-	expression tag	UNP P04908
С	-4	ASN	-	expression tag	UNP P04908
С	-3	ALA	-	expression tag	UNP P04908
С	-2	PRO	-	expression tag	UNP P04908
С	-1	TRP	-	expression tag	UNP P04908
G	-21	MET	-	initiating methionine	UNP P04908
G	-20	GLY	-	expression tag	UNP P04908
G	-19	SER	-	expression tag	UNP P04908
G	-18	SER	-	expression tag	UNP P04908
G	-17	HIS	-	expression tag	UNP P04908
G	-16	HIS	-	expression tag	UNP P04908
G	-15	HIS	-	expression tag	UNP P04908
G	-14	HIS	-	expression tag	UNP P04908
G	-13	HIS	-	expression tag	UNP P04908
G	-12	HIS	-	expression tag	UNP P04908
G	-11	GLU	-	expression tag	UNP P04908
G	-10	ASN	-	expression tag	UNP P04908
G	-9	LEU	-	expression tag	UNP P04908
G	-8	TYR	-	expression tag	UNP P04908
G	-7	PHE	-	expression tag	UNP P04908
G	-6	GLN	-	expression tag	UNP P04908
G	-5	SER	-	expression tag	UNP P04908
G	-4	ASN	-	expression tag	UNP P04908
G	-3	ALA	-	expression tag	UNP P04908
G	-2	PRO	-	expression tag	UNP P04908
G	-1	TRP	-	expression tag	UNP P04908

Continued from previous page...

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



Mol	Chain	Residues	Atoms				AltConf	Trace
4	Л	95	Total	С	Ν	0	0	0
4	4 D		380	190	95	95	0	
4	Н	05	Total	С	Ν	0	0	0
4	11	95	380	190	95	95	0	

• Molecule 5 is a DNA chain called DNA (151-MER).

Mol	Chain	Residues	Atoms				AltConf	Trace	
5	Ι	117	Total 2376	C 1146	N 378	0 735	Р 117	0	0

• Molecule 6 is a DNA chain called DNA (151-MER).

Mol	Chain	Residues	Atoms				AltConf	Trace	
6	J	117	Total 2421	C 1150	N 485	O 669	Р 117	0	0

• Molecule 7 is a protein called Transcription factor SOX-11.

Mol	Chain	Residues	Atoms			AltConf	Trace	
7	K	77	Total 307	C 154	N 77	O 76	0	0

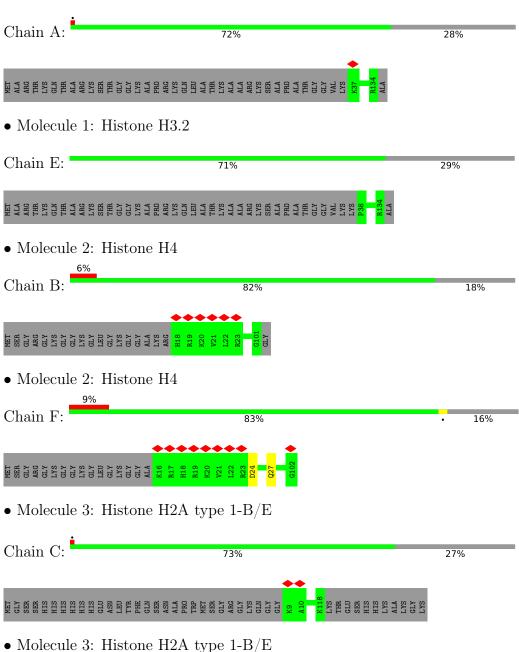
There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
K	30	SER	-	expression tag	UNP P35716
K	31	ASN	-	expression tag	UNP P35716
K	32	ALA	-	expression tag	UNP P35716



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

5%		
Chain G:	73%	27%

MET GLY SER SER HIS HIS HIS HIS HIS HIS ASN CLEU ZVR	ALLA SER ALA ALA ALA ALA ALA CLY CLY CLY CLY CLY CLY CLY CLY CLY CLY	A14 G106 A113 A113 A113 A113 A115 CLU CLV SER H1S ALA ALA ALA CLV CLY CLY
• Molecule 4: Hist	one H2B type 1-K	
<u>.</u>	01	
Chain D:	75%	25%
	•	
MET PRO GLU GLU PRO ALA ALA PRO PRO LYS LYS	sur LYS LYS LYS ALA ALA ALA ALA CLYS LYS LYS LYS LYS LYS LYS LYS LYS LYS	
• Molecule 4: Hist	one H2B type 1_K	
• Molecule 4. Inst	one HZD type I-K	
Chain H:	75%	25%
	•••	_
MET PRO GLU GLU PRO ALA ALA ALA ALA CLYS CLYS	SEL LYS ALA ALA ALA ALA ALA ALA ALA ALA ALA AL	LYS
		-
• Molecule 5: DNA	A (151-MER)	
Chain I:	72%	5% 23%
DA DT DT C1 C2 C2 C6 C6 C6 T54 T54		000 010 011 011 011 011
• Molecule 6: DNA	A (151-MER)	
	- (101	
Chain J:	70%	8% 23%
		u u
D D D D D D D D D D D D D D D D D D D		A11 T11 T11 T11 T11 T11 T11 T11
• Molecule 7: Tran	nscription factor SOX-11	
8%	-	
Chain K:	71%	29%
SER ASN ALA ALA ALA LEU ASP CLU SER ASP PRO ASP TRP TRP	ALA ALA ALA ALA ALA C47 C47 C47 C47 C47 C47 C47 C47 C47 C47	ASP PRO SER ALA ALA PRO SER CLN SER PRO PRO



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	40402	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 $(e^-/\text{\AA}^2)$	1.125	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	3500	Depositor
Magnification	130000	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	0.266	Depositor
Minimum map value	-0.146	Depositor
Average map value	0.004	Depositor
Map value standard deviation	0.020	Depositor
Recommended contour level	0.045	Depositor
Map size (Å)	147.0, 147.0, 147.0	wwPDB
Map dimensions	140, 140, 140	wwPDB
Map angles ($^{\circ}$)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.05, 1.05, 1.05	Depositor



5 Model quality (i)

5.1 Standard geometry (i)

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	Bo	ond angles
	Unam	RMSZ	# Z > 5	RMSZ	# Z > 5
1	А	0.28	0/391	0.49	0/487
1	Е	0.26	0/387	0.48	0/482
2	В	0.29	0/335	0.52	0/417
2	F	0.28	0/348	0.52	0/432
3	С	0.27	0/439	0.52	0/547
3	G	0.27	0/439	0.52	0/547
4	D	0.28	0/379	0.50	0/472
4	Н	0.27	0/379	0.50	0/472
5	Ι	0.66	0/2649	1.14	0/4086
6	J	0.68	0/2731	1.01	1/4213~(0.0%)
7	К	0.29	0/305	0.49	0/377
All	All	0.55	0/8782	0.93	1/12532~(0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Ζ	$Observed(^{o})$	$Ideal(^{o})$
6	J	53	DA	O4'-C1'-C2'	6.21	110.87	105.90

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	А	392	0	99	0	0

Continued on next page...



Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	Е	388	0	99	0	0
2	В	336	0	98	0	0
2	F	349	0	103	1	0
3	С	440	0	122	0	0
3	G	440	0	122	0	0
4	D	380	0	102	0	0
4	Н	380	0	102	0	0
5	Ι	2376	0	1341	17	0
6	J	2421	0	1308	19	0
7	Κ	307	0	74	0	0
All	All	8209	0	3570	23	0

Continued from previous page...

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 3.

The worst 5 of 23 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)
5:I:2:DC:C2	6:J:145:DG:N1	2.20	1.08
5:I:2:DC:O2	6:J:145:DG:C2	2.13	1.02
5:I:2:DC:C2	6:J:145:DG:C2	2.48	1.01
5:I:2:DC:N3	6:J:145:DG:N1	2.11	0.97
5:I:2:DC:N3	6:J:145:DG:C6	2.35	0.94

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	Perce	ntiles
1	А	96/136~(71%)	96 (100%)	0	0	100	100
1	Е	95/136~(70%)	93~(98%)	2(2%)	0	100	100

Continued on next page...



Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
2	В	82/103~(80%)	79~(96%)	3~(4%)	0	100 100
2	F	85/103~(82%)	81~(95%)	4(5%)	0	100 100
3	\mathbf{C}	108/151~(72%)	105~(97%)	3~(3%)	0	100 100
3	G	108/151~(72%)	106 (98%)	2(2%)	0	100 100
4	D	93/126~(74%)	92~(99%)	1 (1%)	0	100 100
4	Н	93/126~(74%)	92~(99%)	1 (1%)	0	100 100
7	Κ	73/109~(67%)	72~(99%)	1 (1%)	0	100 100
All	All	833/1141 (73%)	816 (98%)	17~(2%)	0	100 100

Continued from previous page...

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains (i)

There are no protein residues with a non-rotameric sidechain to report in this entry.

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)

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)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
7	K	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	K	116:TYR	С	117:LYS	Ν	3.68



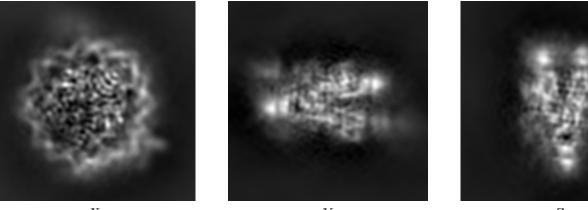
6 Map visualisation (i)

This section contains visualisations of the EMDB entry EMD-10394. 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

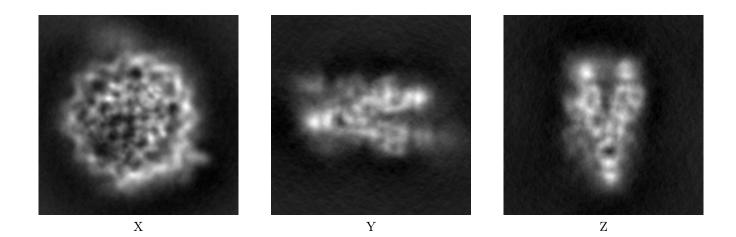


Х



Ζ

6.1.2 Raw map

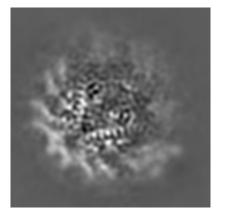


The images above show the map projected in three orthogonal directions.

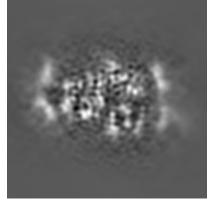


6.2 Central slices (i)

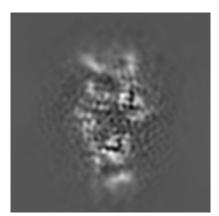
6.2.1 Primary map



X Index: 70

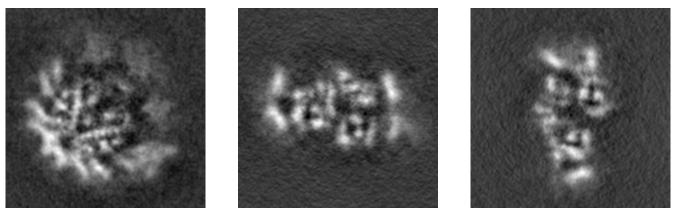


Y Index: 70



Z Index: 70

6.2.2 Raw map



X Index: 70

Y Index: 70

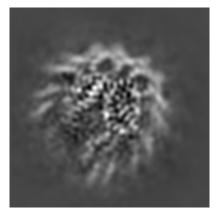
Z Index: 70

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

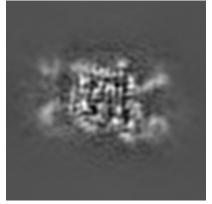


6.3 Largest variance slices (i)

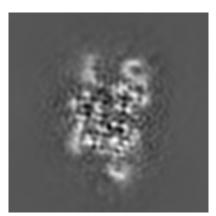
6.3.1 Primary map



X Index: 82

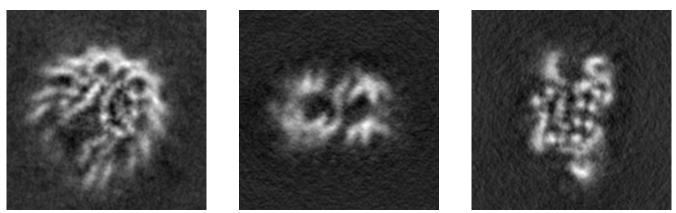


Y Index: 78



Z Index: 81

6.3.2 Raw map



X Index: 82

Y Index: 98

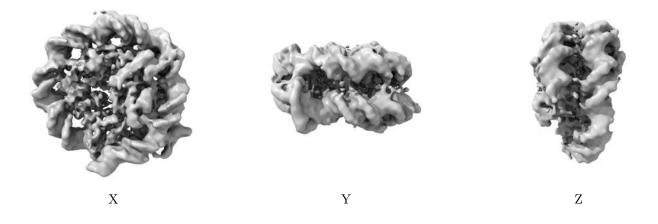


The images above show the largest variance slices of the map in three orthogonal directions.



6.4 Orthogonal surface views (i)

6.4.1 Primary map



The images above show the 3D surface view of the map at the recommended contour level 0.045. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

6.4.2 Raw map



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.

6.5 Mask visualisation (i)

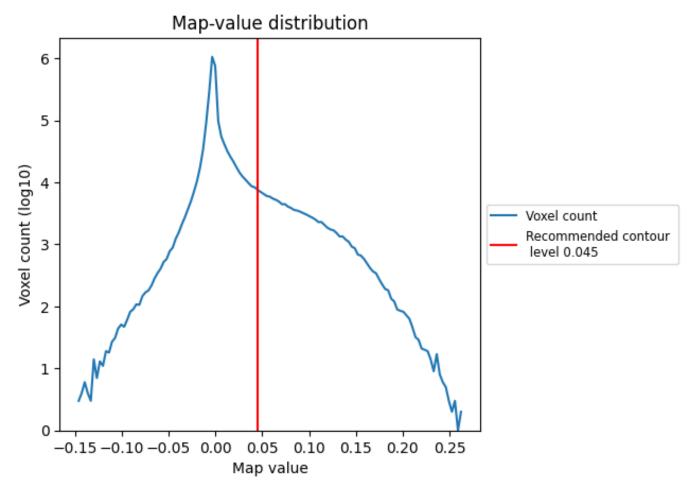
This section was not generated. No masks/segmentation were deposited.



7 Map analysis (i)

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

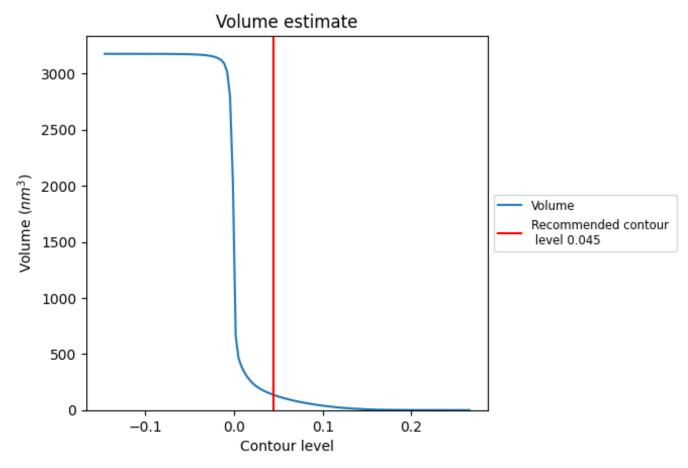
7.1 Map-value distribution (i)



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)

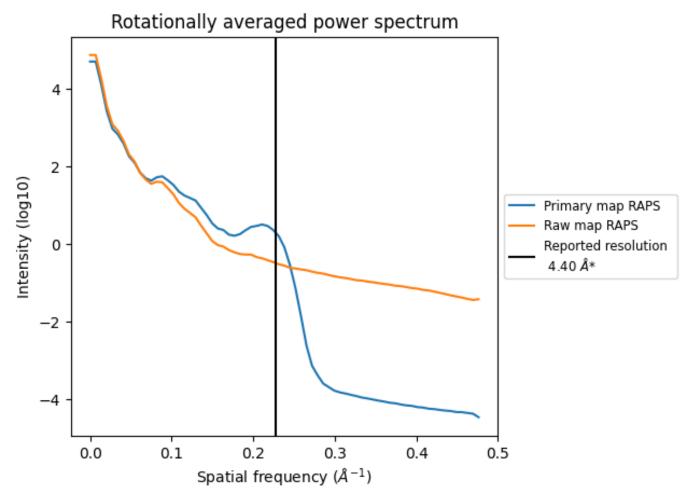


The volume at the recommended contour level is 135 nm^3 ; this corresponds to an approximate mass of 122 kDa.

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)



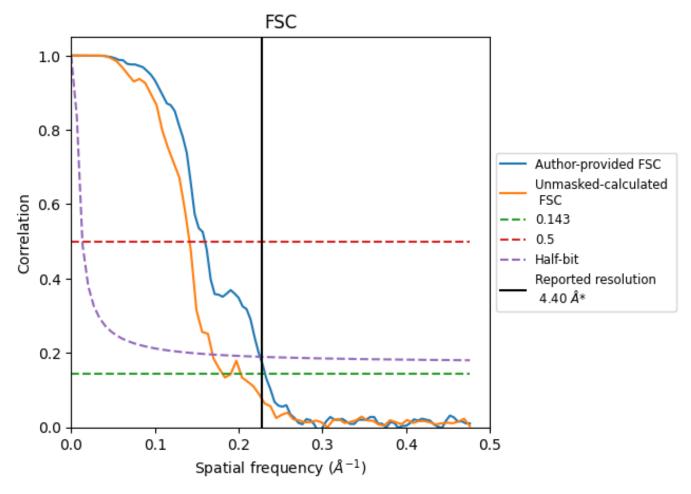
*Reported resolution corresponds to spatial frequency of 0.227 ${\rm \AA^{-1}}$



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 0.227 ${\rm \AA^{-1}}$



8.2 Resolution estimates (i)

Resolution estimate (Å)	Estimation criterion (FSC cut-off)			
Resolution estimate (A)	0.143	0.5	Half-bit	
Reported by author	4.40	-	-	
Author-provided FSC curve	4.32	6.26	4.42	
Unmasked-calculated*	5.53	7.08	5.92	

*Resolution estimate based on FSC curve calculated by comparison of deposited half-maps. The value from deposited half-maps intersecting FSC 0.143 CUT-OFF 5.53 differs from the reported value 4.4 by more than 10 %



9 Map-model fit (i)

This section contains information regarding the fit between EMDB map EMD-10394 and PDB model 6T7D. Per-residue inclusion information can be found in section 3 on page 7.

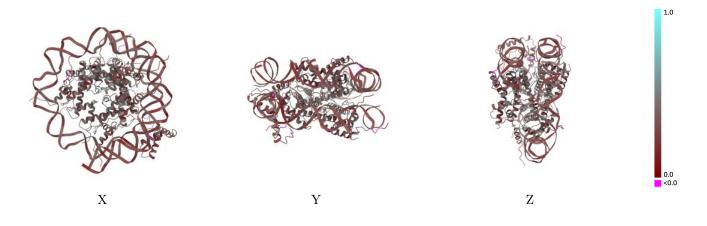
9.1 Map-model overlay (i)



The images above show the 3D surface view of the map at the recommended contour level 0.045 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.

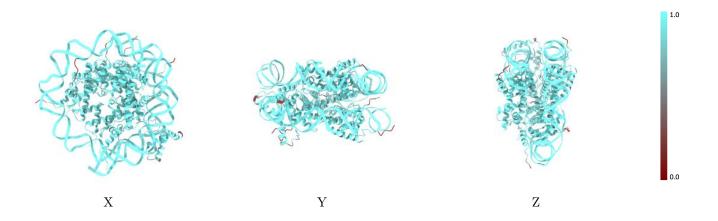


9.2 Q-score mapped to coordinate model (i)



The images above show the model with each residue coloured according 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.

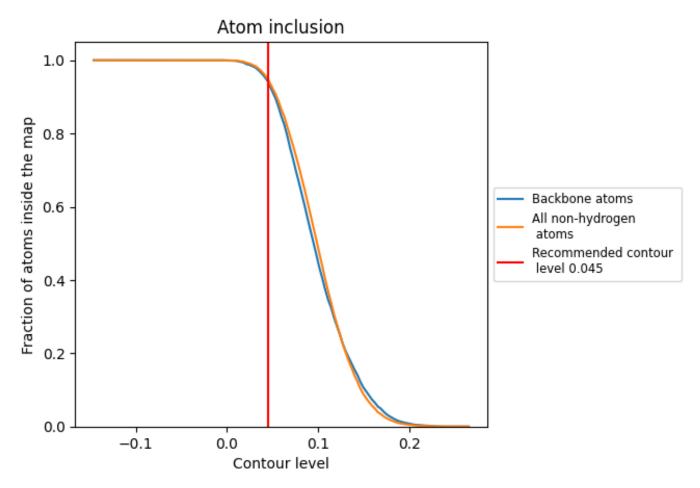
9.3 Atom inclusion mapped to coordinate model (i)



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 (0.045).



9.4 Atom inclusion (i)



At the recommended contour level, 94% of all backbone atoms, 95% of all non-hydrogen atoms, are inside the map.



Map-model fit summary (i) 9.5

The table lists the average atom inclusion at the recommended contour level (0.045) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score	
All	0.9490	0.3230	1.0
A	0.9592	0.4030	
В	0.9107	0.3710	
С	0.9500	0.3820	
D	0.9711	0.3720	
E	0.9768	0.3780	
F	0.8739	0.3680	
G	0.9091	0.3410	
Н	0.9500	0.3580	
Ι	0.9625	0.2920	0.0 <0.0
J	0.9632	0.2970	
K	0.8371	0.2790	

