wwPDB EM Validation Summary Report

Dec 19, 2022 – 09:44 am GMT

PDB ID	:	7NHX
EMDB ID	:	EMD-12342
Title	:	1918 H1N1 Viral influenza polymerase heterotrimer - full transcriptase (Class1)
Authors	:	Keown, J.R.; Carrique, L.; Fodor, E.; Grimes, J.M.
Deposited on	:	2021-02-11
Resolution	:	3.23 Å(reported)
Based on initial model	:	6RR7

This is a wwPDB EM Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references ⓘ) were used in the production of this report:

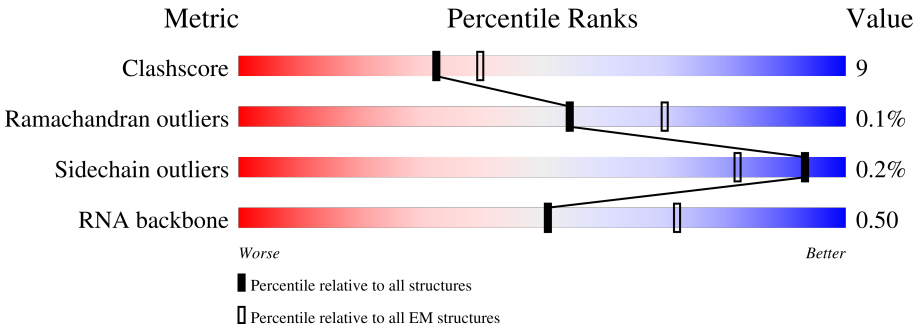
EMDB validation analysis	:	0.0.1.dev43
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.9
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.3

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.23 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
RNA backbone	4643	859

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	716	84% 16%
2	B	757	77% 20%
3	C	905	17% 55% 19% 25%
4	E	17	24% 24% 53%
5	F	15	53% 40% 7%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 34661 atoms, of which 17149 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 Polymerase acidic protein.

Mol	Chain	Residues	Atoms						AltConf	Trace
			Total	C	H	N	O	S		
1	A	716	11524	3671	5726	985	1102	40	0	0

- Molecule 2 is a protein called RNA-directed RNA polymerase catalytic subunit.

Mol	Chain	Residues	Atoms						AltConf	Trace
			Total	C	H	N	O	S		
2	B	729	11619	3664	5790	1020	1101	44	0	0

- Molecule 3 is a protein called Polymerase basic protein 2,Immunoglobulin G-binding protein A.

Mol	Chain	Residues	Atoms						AltConf	Trace
			Total	C	H	N	O	S		
3	C	676	10772	3386	5382	971	994	39	0	0

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	760	GLU	-	linker	UNP Q3HM41
C	761	ASN	-	linker	UNP Q3HM41
C	762	LEU	-	linker	UNP Q3HM41
C	763	TYR	-	linker	UNP Q3HM41
C	764	PHE	-	linker	UNP Q3HM41
C	765	GLN	-	linker	UNP Q3HM41
C	766	GLY	-	linker	UNP Q3HM41
C	767	GLU	-	linker	UNP Q3HM41
C	768	LEU	-	linker	UNP Q3HM41
C	769	LYS	-	linker	UNP Q3HM41
C	770	THR	-	linker	UNP Q3HM41
C	771	ALA	-	linker	UNP Q3HM41
C	772	ALA	-	linker	UNP Q3HM41
C	773	LEU	-	linker	UNP Q3HM41

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	774	ALA	-	linker	UNP Q3HM41
C	775	GLN	-	linker	UNP Q3HM41
C	776	HIS	-	linker	UNP Q3HM41
C	777	ASP	-	linker	UNP Q3HM41
C	778	GLU	-	linker	UNP Q3HM41
C	779	ALA	-	linker	UNP Q3HM41
C	780	VAL	-	linker	UNP Q3HM41
C	781	ASP	-	linker	UNP Q3HM41
C	782	ASN	-	linker	UNP Q3HM41
C	783	LYS	-	linker	UNP Q3HM41
C	797	HIS	ASN	conflict	UNP P38507
C	798	LEU	MET	conflict	UNP P38507
C	808	ALA	GLY	conflict	UNP P38507
C	832	ASP	GLU	conflict	UNP P38507
C	833	ALA	SER	conflict	UNP P38507
C	838	VAL	ALA	conflict	UNP P38507
C	866	ALA	GLY	conflict	UNP P38507
C	873	ALA	ASP	conflict	UNP P38507
C	890	GLY	ASP	conflict	UNP P38507
C	896	VAL	ALA	conflict	UNP P38507
C	898	ALA	-	expression tag	UNP P38507
C	899	ASN	-	expression tag	UNP P38507
C	900	SER	-	expression tag	UNP P38507
C	901	ALA	-	expression tag	UNP P38507
C	902	GLY	-	expression tag	UNP P38507
C	903	LYS	-	expression tag	UNP P38507
C	904	SER	-	expression tag	UNP P38507
C	905	THR	-	expression tag	UNP P38507

- Molecule 4 is a RNA chain called RNA (5'-R(P*GP*GP*CP*CP*UP*GP*CP*U)-3').

Mol	Chain	Residues	Atoms						AltConf	Trace
4	E	8	Total	C	H	N	O	P	0	0
			255	75	86	28	58	8		

- Molecule 5 is a RNA chain called RNA (5'-R(P*AP*GP*UP*AP*GP*AP*AP*AP*CP*AP *AP*GP*GP*CP*C)-3').

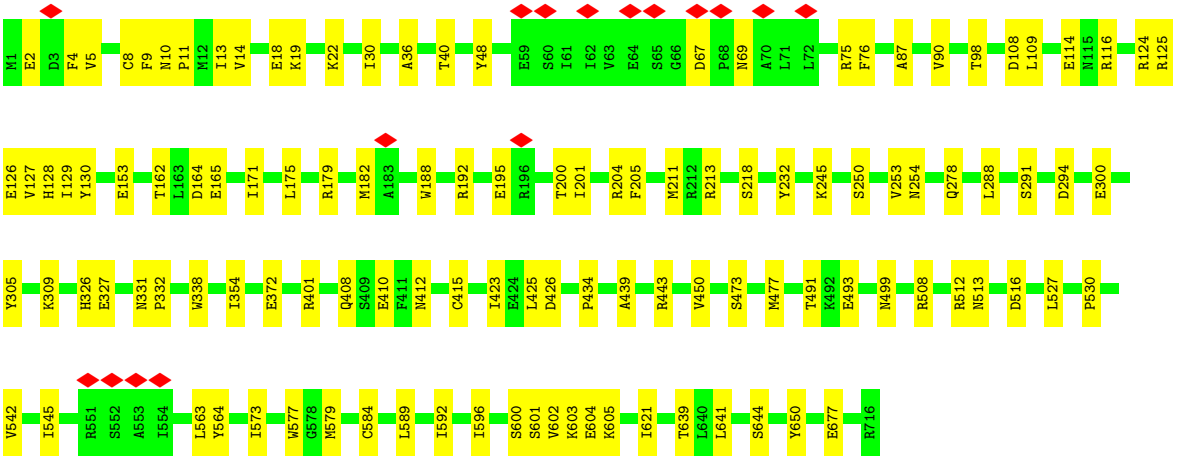
Mol	Chain	Residues	Atoms						AltConf	Trace
5	F	15	Total	C	H	N	O	P	0	0
			491	146	165	66	99	15		

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.

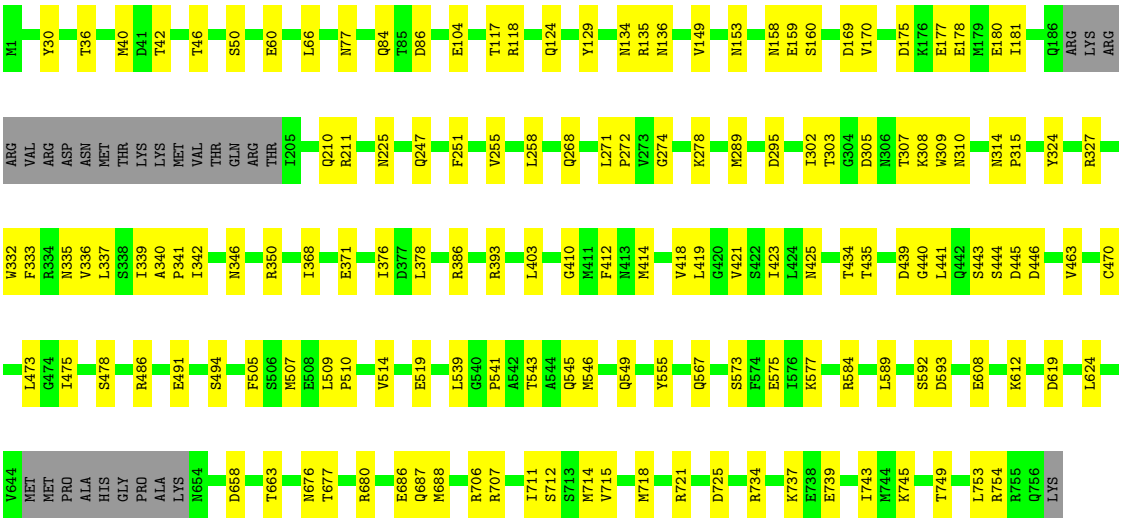
- Molecule 1: Polymerase acidic protein

Chain A:



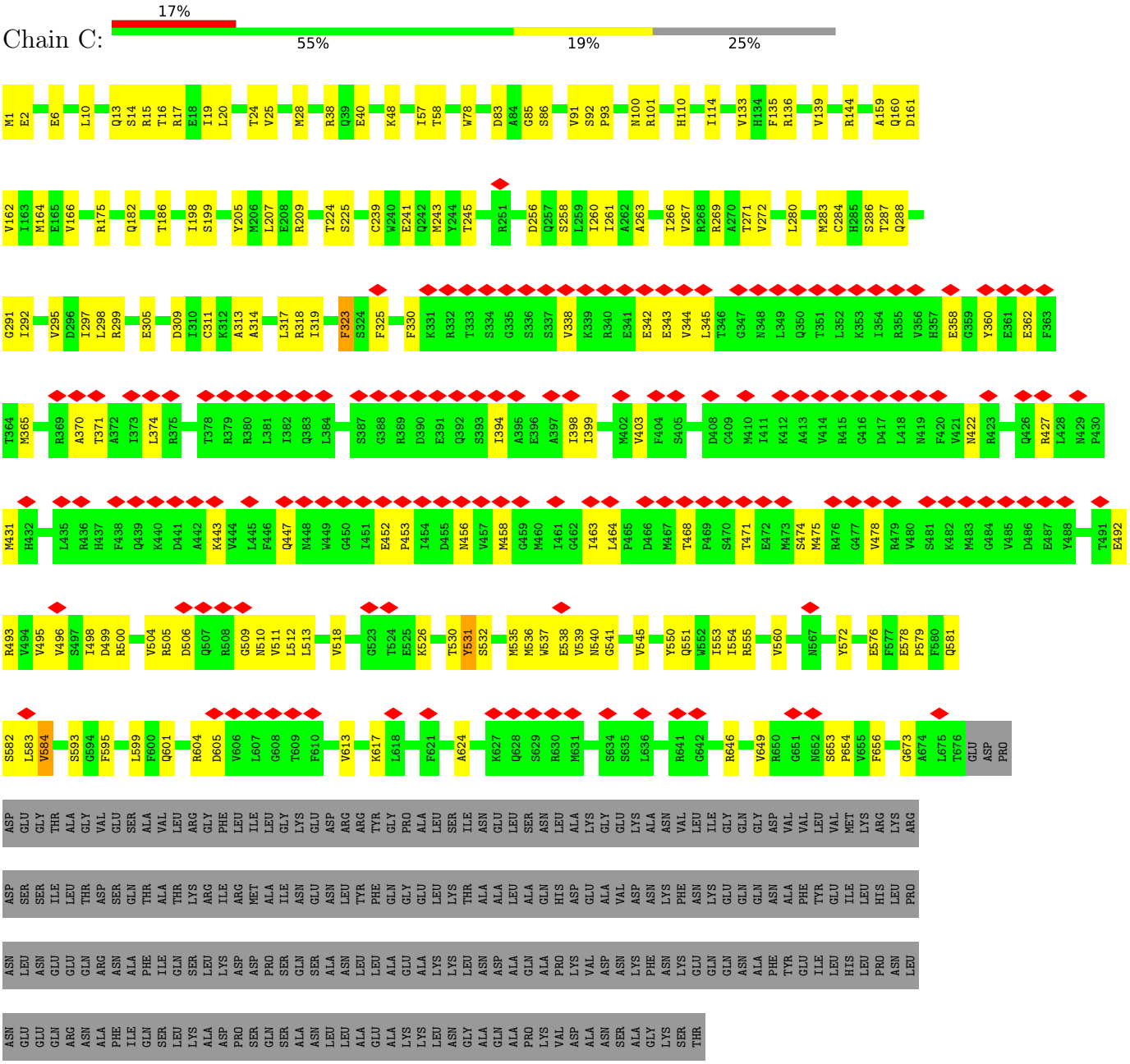
- Molecule 2: RNA-directed RNA polymerase catalytic subunit

Chain B:



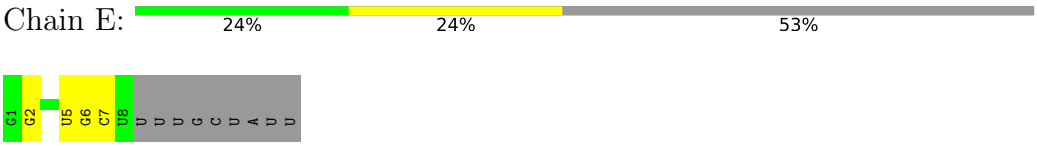
• Molecule 3: Polymerase basic protein 2,Immunoglobulin G-binding protein A

Chain C:



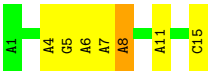
• Molecule 4: RNA (5'-R(P*GP*GP*CP*CP*UP*GP*CP*U)-3')

Chain E:



• Molecule 5: RNA (5'-R(P*AP*GP*UP*AP*GP*AP*AP*AP*CP*AP*AP*GP*GP*CP*C)-3')

Chain F:


A1
A4
G5
A6
A7
A8
A11
C15

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	58204	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	TFS KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	61.5	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	165000	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	0.375	Depositor
Minimum map value	-0.145	Depositor
Average map value	0.002	Depositor
Map value standard deviation	0.018	Depositor
Recommended contour level	0.0485	Depositor
Map size (Å)	211.2, 211.2, 211.2	wwPDB
Map dimensions	256, 256, 256	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	0.825, 0.825, 0.825	Depositor

5 Model quality

5.1 Standard geometry

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	0.26	0/5921	0.52	0/7979
2	B	0.26	0/5941	0.55	0/8011
3	C	0.26	0/5484	0.58	0/7398
4	E	0.16	0/187	0.71	0/289
5	F	0.18	0/366	0.72	0/569
All	All	0.26	0/17899	0.56	0/24246

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
3	C	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	C	422	ASN	Sidechain

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	5798	5726	5732	89	0
2	B	5829	5790	5803	114	0
3	C	5390	5382	5495	152	0
4	E	169	86	87	5	0
5	F	326	165	165	3	0
All	All	17512	17149	17282	322	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 322 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:C:136:ARG:NH1	3:C:536:MET:SD	2.32	1.03
1:A:245:LYS:NZ	2:B:84:GLN:OE1	2.00	0.94
3:C:286:SER:O	3:C:532:SER:OG	1.87	0.93
3:C:505:ARG:NH2	3:C:509:GLY:O	2.02	0.91
1:A:48:TYR:OH	1:A:162:THR:OG1	1.91	0.89

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	714/716 (100%)	678 (95%)	36 (5%)	0	100	100
2	B	723/757 (96%)	700 (97%)	23 (3%)	0	100	100
3	C	674/905 (74%)	631 (94%)	40 (6%)	3 (0%)	34	68
All	All	2111/2378 (89%)	2009 (95%)	99 (5%)	3 (0%)	54	83

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	323	PHE
3	C	584	VAL
3	C	531	TYR

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	640/640 (100%)	639 (100%)	1 (0%)	93 97
2	B	644/669 (96%)	642 (100%)	2 (0%)	92 96
3	C	598/788 (76%)	597 (100%)	1 (0%)	93 97
All	All	1882/2097 (90%)	1878 (100%)	4 (0%)	93 97

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	125	ARG
2	B	734	ARG
2	B	737	LYS
3	C	427	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	510	HIS
2	B	425	ASN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
4	E	7/17 (41%)	1 (14%)	0
5	F	14/15 (93%)	4 (28%)	0
All	All	21/32 (65%)	5 (23%)	0

All (5) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
4	E	2	G
5	F	6	A
5	F	8	A
5	F	11	A
5	F	15	C

There are no RNA pucker outliers to report.

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)

There are no chain breaks in this entry.

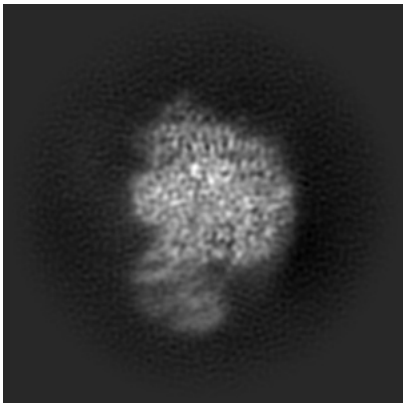
6 Map visualisation

This section contains visualisations of the EMDB entry EMD-12342. These allow visual inspection of the internal detail of the map and identification of artifacts.

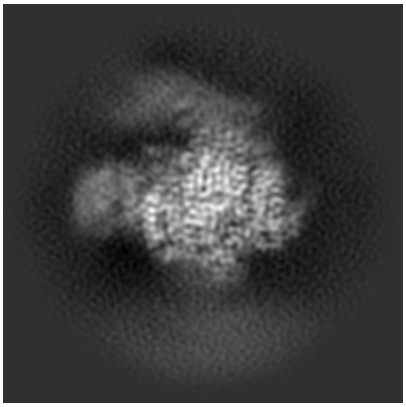
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections

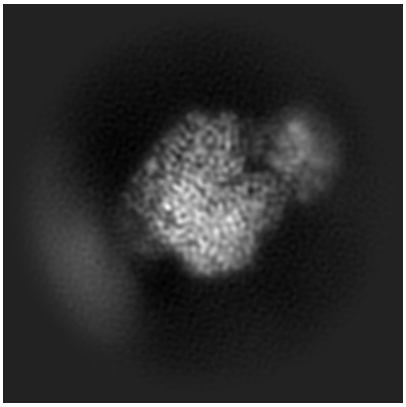
6.1.1 Primary map



X



Y

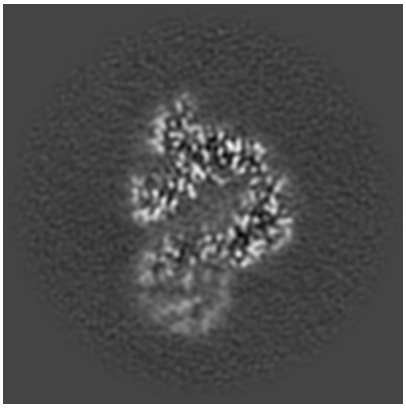


Z

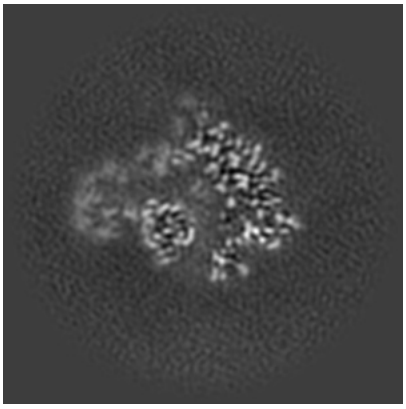
The images above show the map projected in three orthogonal directions.

6.2 Central slices

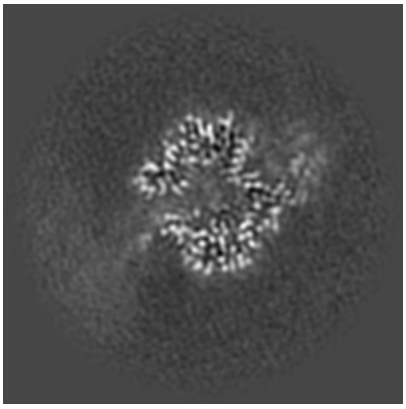
6.2.1 Primary map



X Index: 128



Y Index: 128

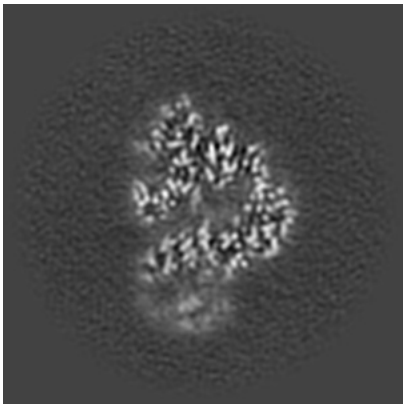


Z Index: 128

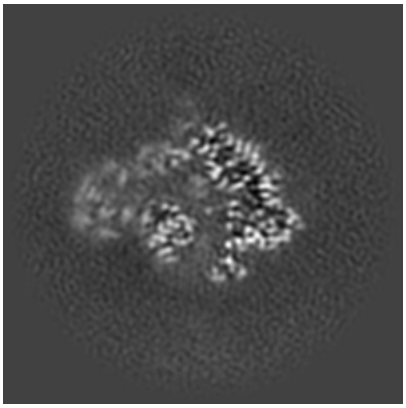
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

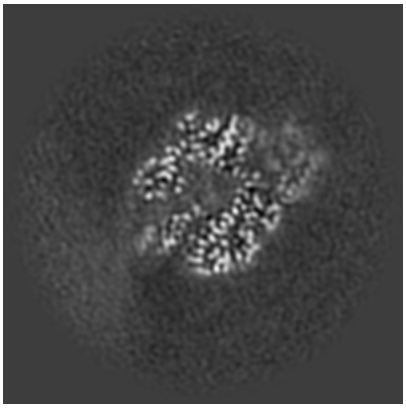
6.3.1 Primary map



X Index: 125



Y Index: 126

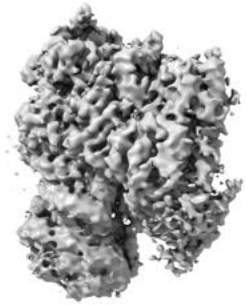


Z Index: 131

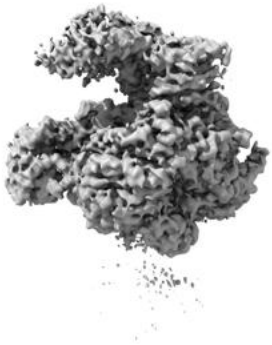
The images above show the largest variance slices of the map in three orthogonal directions.

6.4 Orthogonal surface views (i)

6.4.1 Primary map



X



Y

Z

The images above show the 3D surface view of the map at the recommended contour level 0.0485. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

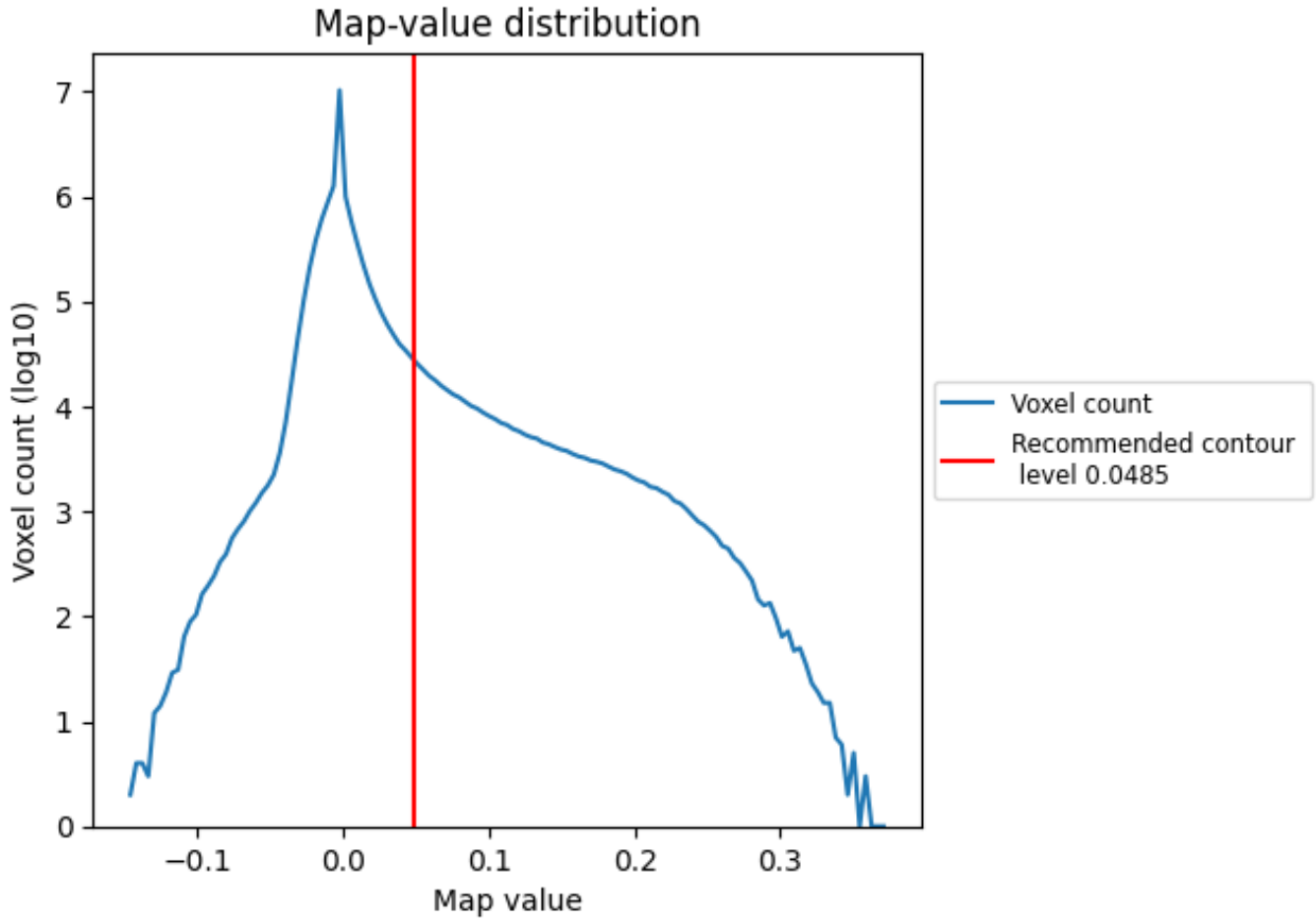
6.5 Mask visualisation (i)

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

7 Map analysis

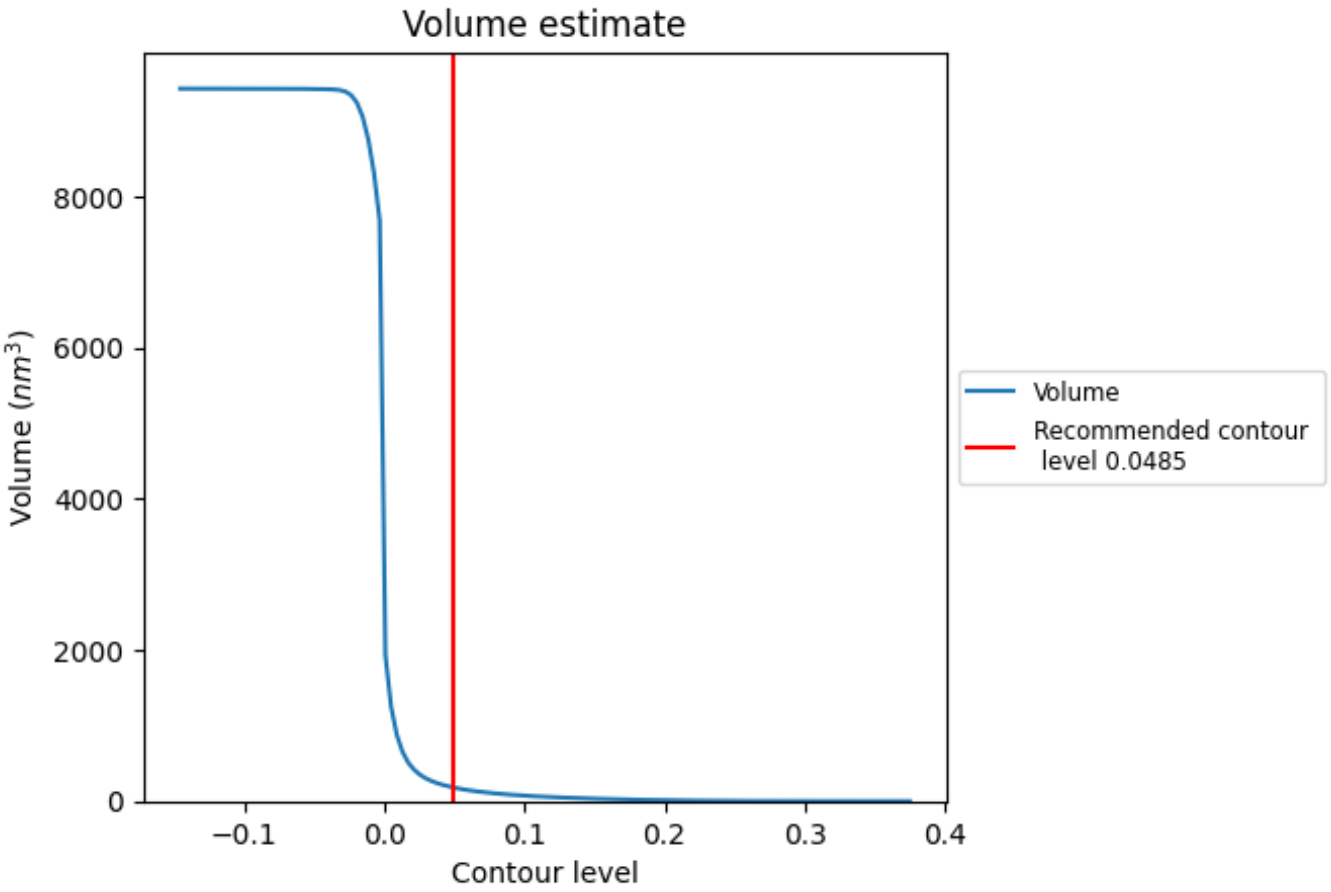
This section contains the results of statistical analysis of the map.

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.

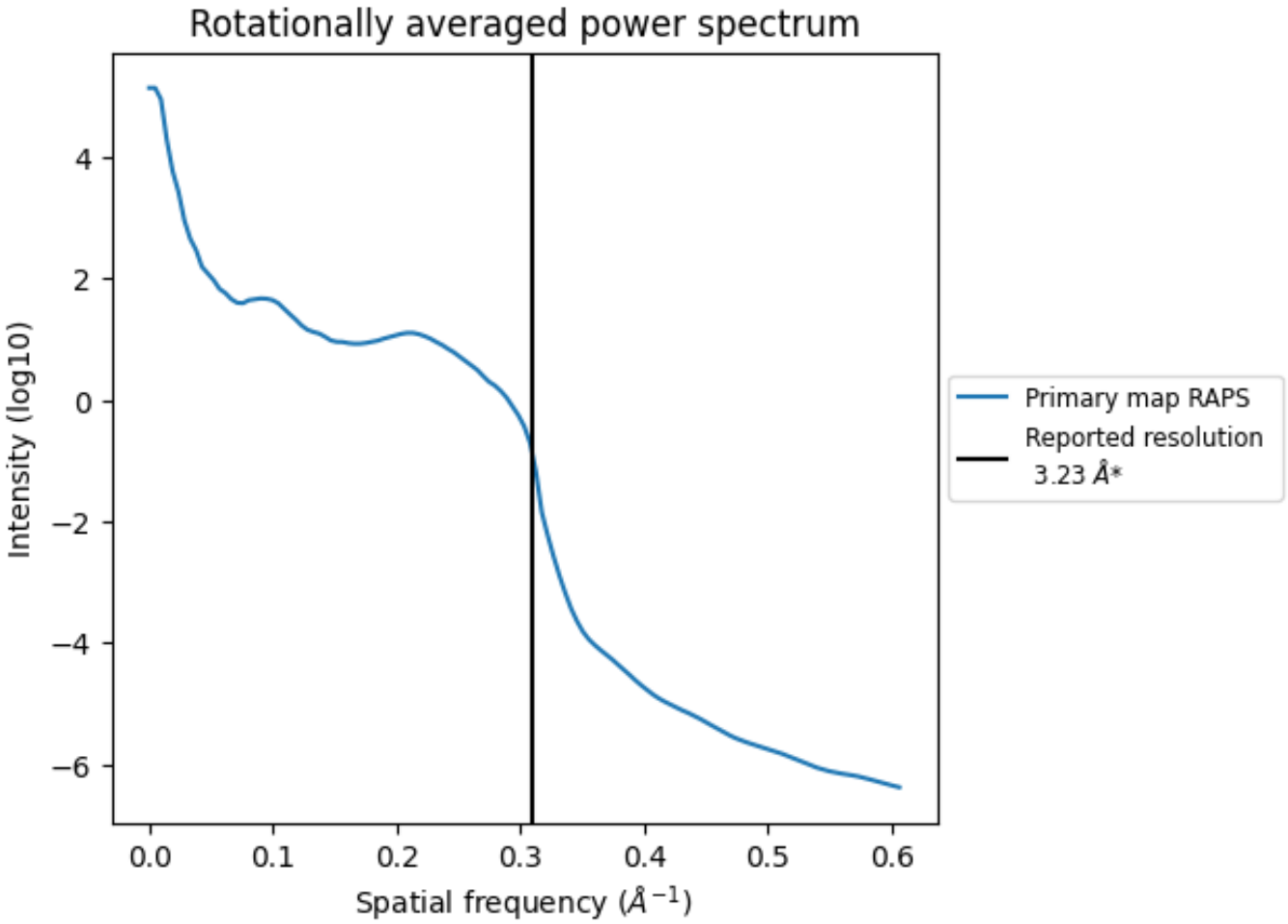
7.2 Volume estimate (i)



The volume at the recommended contour level is 184 nm^3; this corresponds to an approximate mass of 166 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)

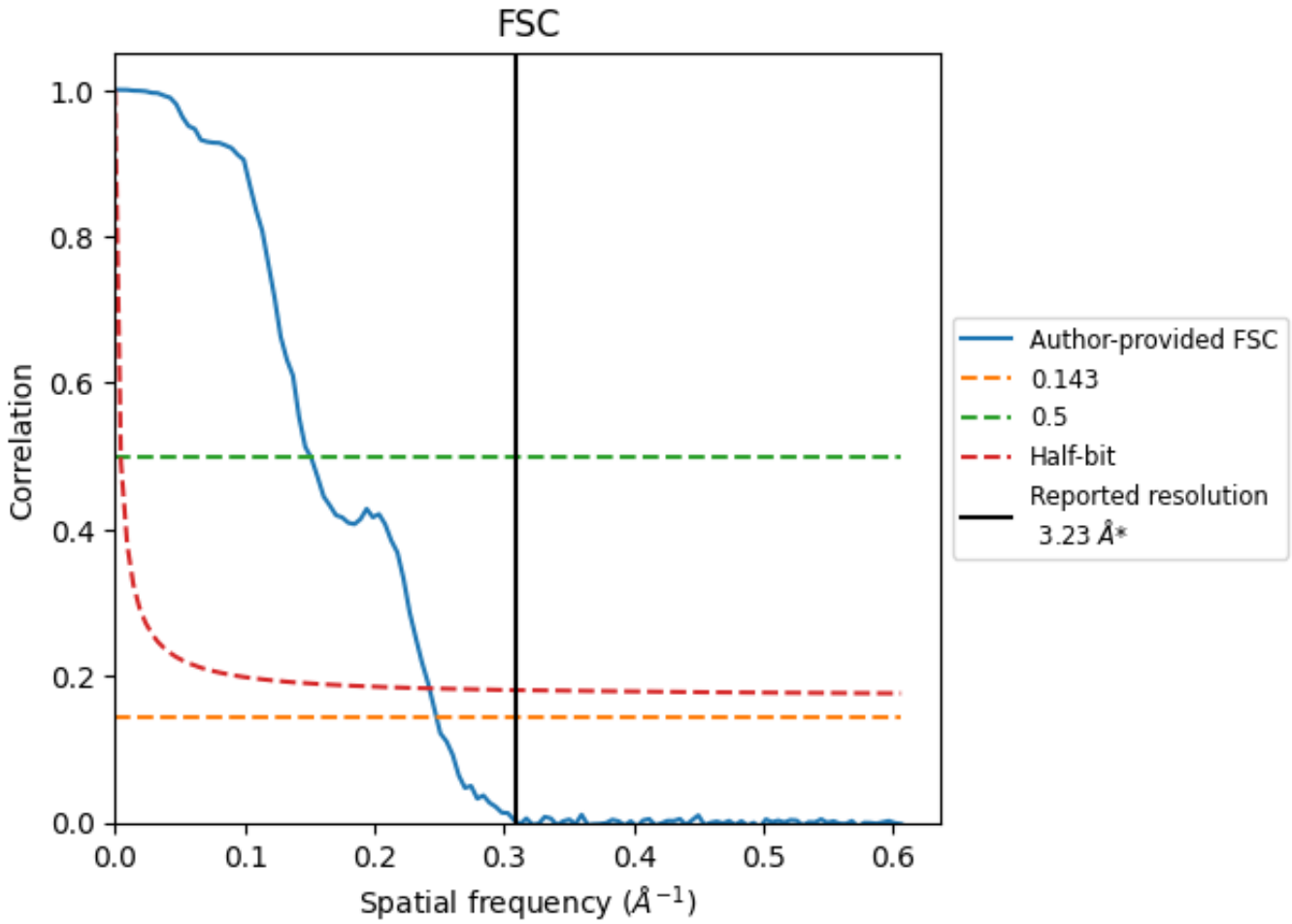


*Reported resolution corresponds to spatial frequency of 0.310 Å^{-1}

8 Fourier-Shell correlation

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



*Reported resolution corresponds to spatial frequency of 0.310 Å^{-1}

8.2 Resolution estimates (i)

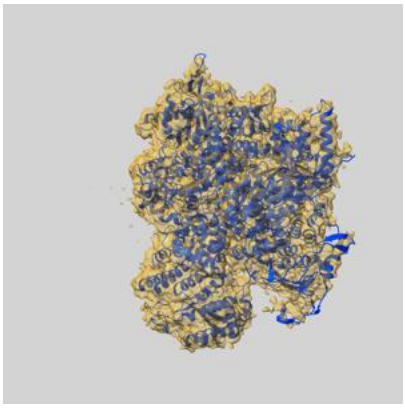
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.23	-	-
Author-provided FSC curve	4.03	6.64	4.13
Unmasked-calculated*	-	-	-

*Resolution estimate based on FSC curve calculated by comparison of deposited half-maps. The value from author-provided FSC intersecting FSC 0.143 CUT-OFF 4.03 differs from the reported value 3.23 by more than 10 %

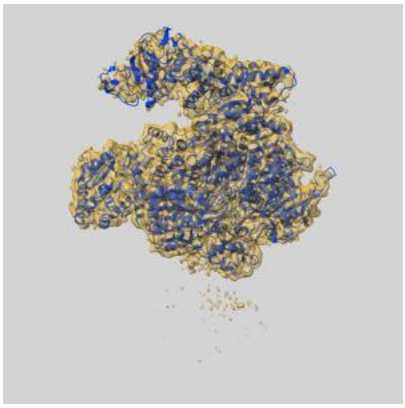
9 Map-model fit

This section contains information regarding the fit between EMDB map EMD-12342 and PDB model 7NHX. Per-residue inclusion information can be found in section 3 on page 5.

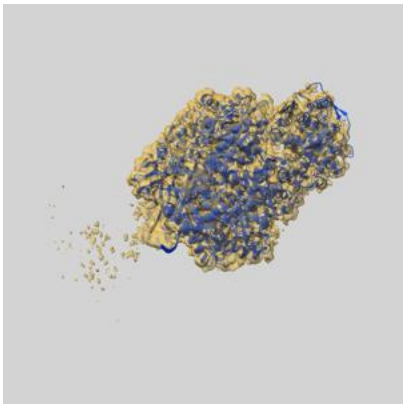
9.1 Map-model overlay



X



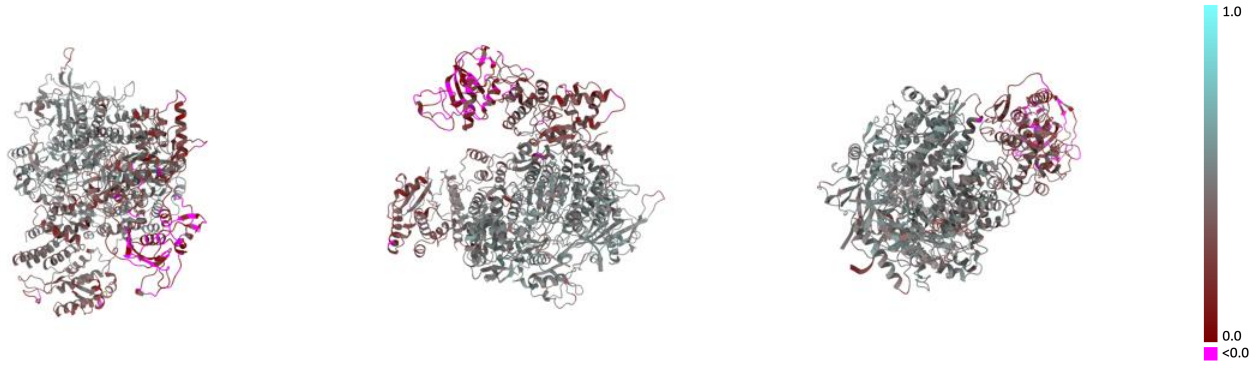
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.0485 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.

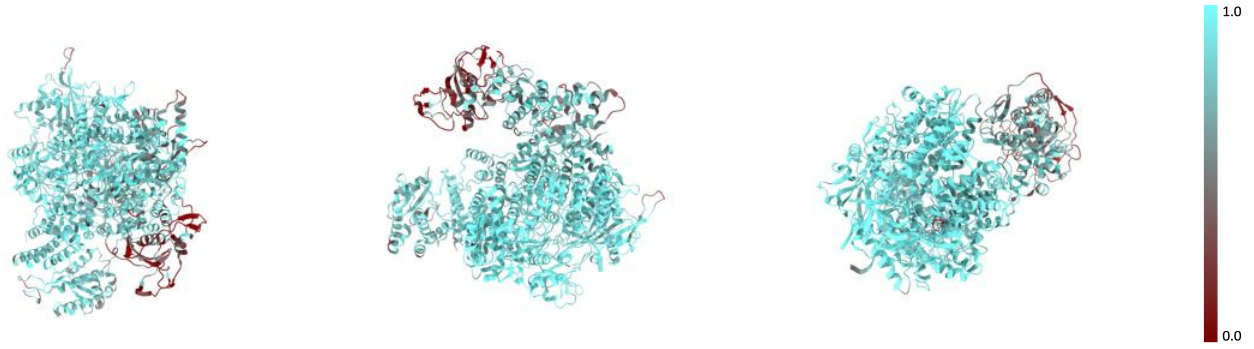
9.2 Q-score mapped to coordinate model (i)



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.

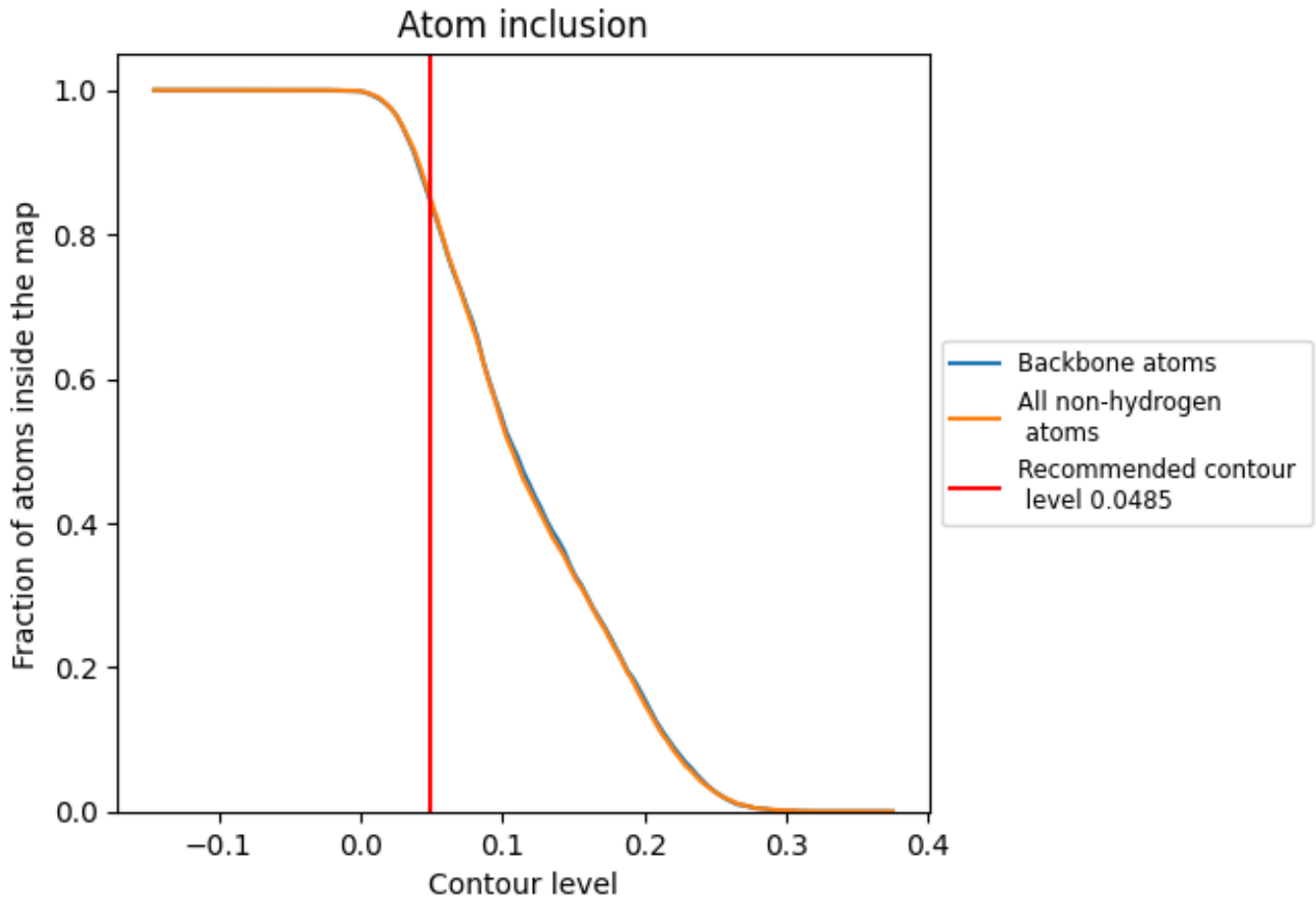
9.3 Atom inclusion mapped to coordinate model (i)



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 (0.0485).

9.4 Atom inclusion (i)



At the recommended contour level, 85% of all backbone atoms, 85% 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 (0.0485) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.8531	0.4060
A	0.9042	0.4360
B	0.9439	0.4800
C	0.6883	0.2900
E	0.8225	0.3090
F	0.9877	0.5060