



## wwPDB EM Validation Summary Report ⓘ

Nov 10, 2024 – 12:17 am GMT

PDB ID : 7NK2  
EMDB ID : EMD-12429  
Title : 1918 H1N1 Viral influenza polymerase heterotrimer with Nb8202 core  
Authors : Keown, J.R.; Carrique, L.; Fodor, E.; Grimes, J.M.  
Deposited on : 2021-02-17  
Resolution : 4.84 Å(reported)  
Based on initial model : 7HNA

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  
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

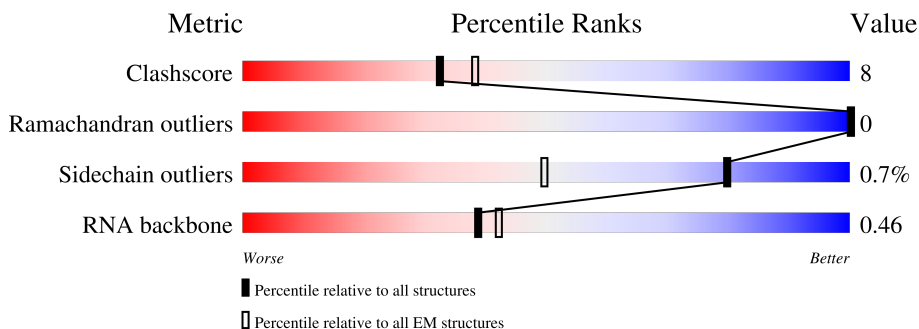
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 4.84 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
RNA backbone	6643	2191

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	716	<div> <div>31%</div> <div>77%</div> <div>23%</div> </div>
2	B	757	<div> <div>19%</div> <div>80%</div> <div>17%</div> <div>.</div> </div>
3	C	905	<div> <div>16%</div> <div>21%</div> <div>6%</div> <div>72%</div> </div>
4	E	17	<div> <div>18%</div> <div>24%</div> <div>24%</div> <div>53%</div> </div>
5	D	15	<div> <div>60%</div> <div>33%</div> <div>7%</div> </div>
6	F	134	<div> <div>61%</div> <div>8%</div> <div>.</div> <div>30%</div> </div>

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 28776 atoms, of which 13868 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
1	A	716	Total	C	H	N	O	S	0	0
			11524	3671	5726	985	1102	40		

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

Mol	Chain	Residues	Atoms						AltConf	Trace
2	B	735	Total	C	H	N	O	S	0	0
			11704	3691	5833	1028	1106	46		

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

Mol	Chain	Residues	Atoms						AltConf	Trace
3	C	249	Total	C	H	N	O	S	0	0
			4075	1271	2058	361	369	16		

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	D	15	Total	C	H	N	O	P	0	0
			491	146	165	66	99	15		

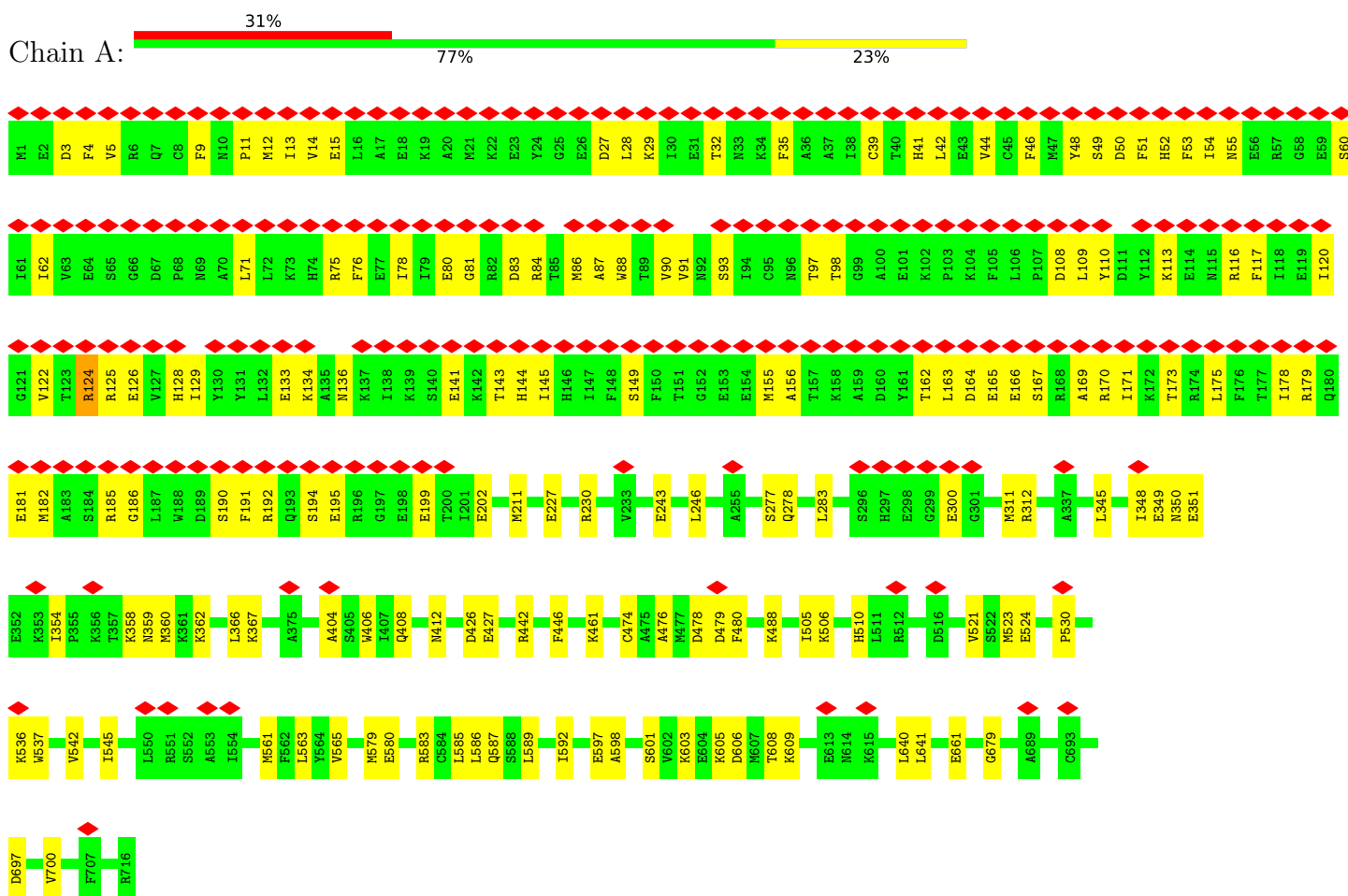
- Molecule 6 is a protein called Nanobody8202.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	F	94	Total	C	N	O	S	0	0
			727	457	127	138	5		

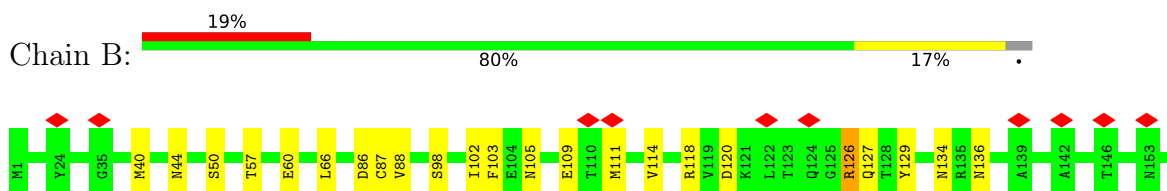
### 3 Residue-property plots

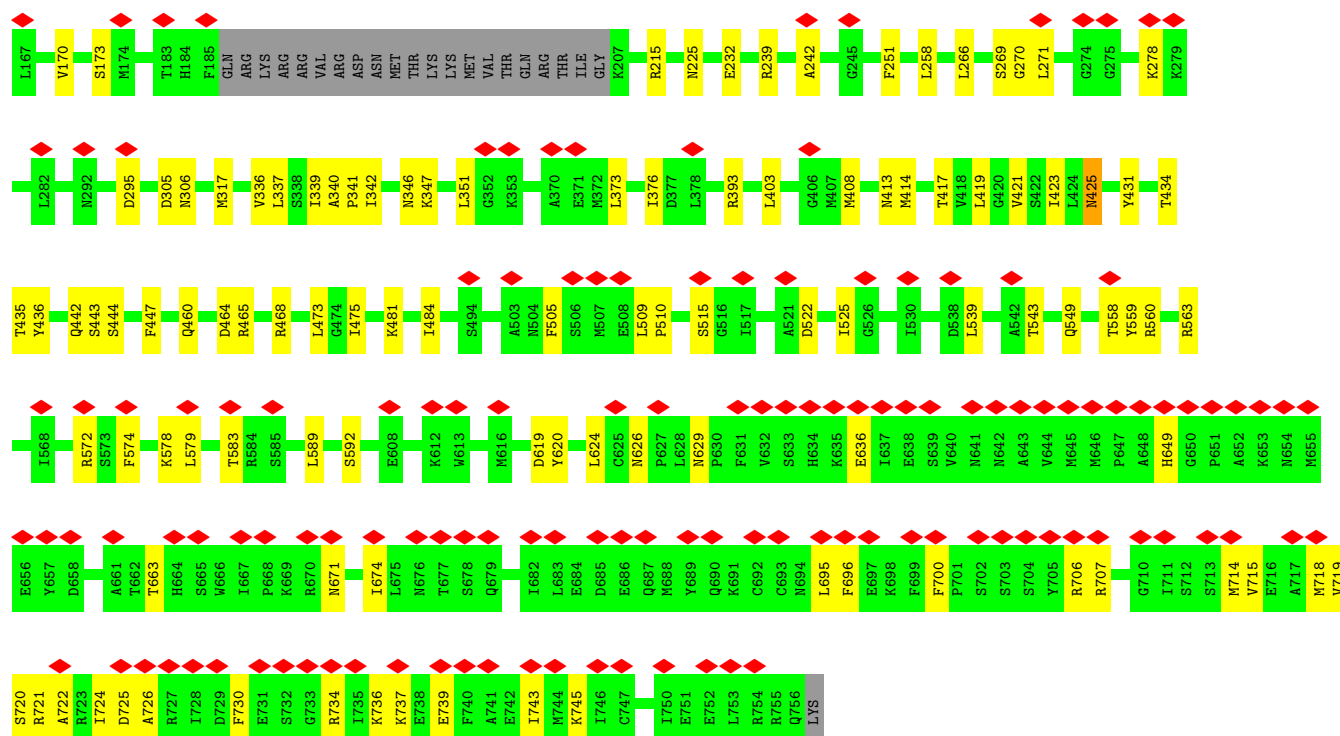
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and atom inclusion 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



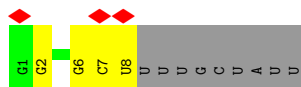
- Molecule 2: RNA-directed RNA polymerase catalytic subunit





LYS  
LYS  
LEU  
ASN  
GLY  
ALA  
GLN  
ALA  
PRO  
LYS  
VAL  
ASP  
ALA  
ASN  
SER  
ALA  
GLY  
LYS  
SER  
THR

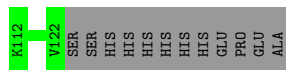
- Chain E:  18% 24% 24% 34%



- Chain D:  60% 33% 7%



- Chain F:  61% 8% 30%





## 4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	198338	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	TFS GLACIOS	Depositor
Voltage (kV)	200	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	69.5	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	FEI FALCON III (4k x 4k)	Depositor
Maximum map value	13.565	Depositor
Minimum map value	-8.125	Depositor
Average map value	0.027	Depositor
Map value standard deviation	0.342	Depositor
Recommended contour level	1.81	Depositor
Map size (Å)	280.0, 280.0, 280.0	wwPDB
Map dimensions	140, 140, 140	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	2.0, 2.0, 2.0	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		Bond angles	
		RMSZ	# Z  >5	RMSZ	# Z  >5
1	A	0.26	0/5921	0.52	1/7979 (0.0%)
2	B	0.25	0/5987	0.49	0/8075
3	C	0.26	0/2057	0.51	0/2773
4	E	0.18	0/187	0.75	0/289
5	D	0.20	0/366	0.71	0/569
6	F	0.83	0/738	1.25	8/989 (0.8%)
All	All	0.31	0/15256	0.58	9/20674 (0.0%)

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	113	LYS	CD-CE-NZ	11.77	138.77	111.70
6	F	72	ARG	NE-CZ-NH2	-9.02	115.79	120.30
6	F	72	ARG	NE-CZ-NH1	8.87	124.73	120.30
6	F	34	MET	CG-SD-CE	-6.46	89.87	100.20
6	F	61	MET	CG-SD-CE	-6.01	90.58	100.20

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	5798	5726	5732	125	0
2	B	5871	5833	5847	95	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	C	2017	2058	2058	47	0
4	E	169	86	87	3	0
5	D	326	165	165	2	0
6	F	727	0	704	2	0
All	All	14908	13868	14593	245	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 8.

The worst 5 of 245 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)
2:B:522:ASP:OD2	2:B:559:TYR:OH	1.91	0.87
1:A:124:ARG:NH1	1:A:191:PHE:O	2.08	0.86
5:D:5:G:OP2	5:D:5:G:N2	2.10	0.82
1:A:88:TRP:CZ3	3:C:174:ALA:HB2	2.18	0.79
1:A:300:GLU:O	1:A:488:LYS:NZ	2.17	0.76

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone

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%)	677 (95%)	37 (5%)	0	100	100
2	B	731/757 (97%)	709 (97%)	22 (3%)	0	100	100
3	C	247/905 (27%)	233 (94%)	14 (6%)	0	100	100
6	F	86/134 (64%)	83 (96%)	3 (4%)	0	100	100
All	All	1778/2512 (71%)	1702 (96%)	76 (4%)	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	640/640 (100%)	638 (100%)	2 (0%)	91	92
2	B	648/669 (97%)	644 (99%)	4 (1%)	84	88
3	C	223/788 (28%)	223 (100%)	0	100	100
6	F	74/108 (68%)	69 (93%)	5 (7%)	13	34
All	All	1585/2205 (72%)	1574 (99%)	11 (1%)	80	87

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
6	F	13	GLN
6	F	65	LYS
6	F	82	GLN
6	F	76	LYS
2	B	734	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
4	E	7/17 (41%)	1 (14%)	0
5	D	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	D	6	A
5	D	8	A
5	D	11	A

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
5	D	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 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.

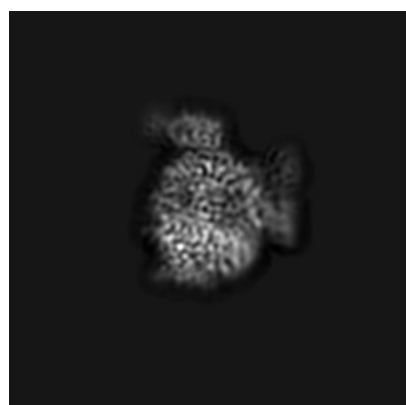
## 6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-12429. These allow visual inspection of the internal detail of the map and identification of artifacts.

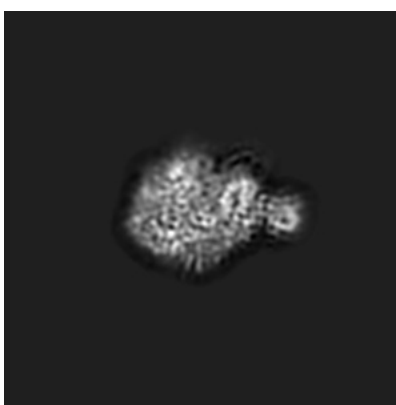
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

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

#### 6.1.1 Primary map



X



Y

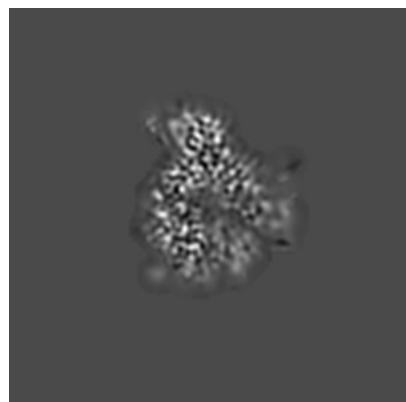


Z

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

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

#### 6.2.1 Primary map



X Index: 70



Y Index: 70

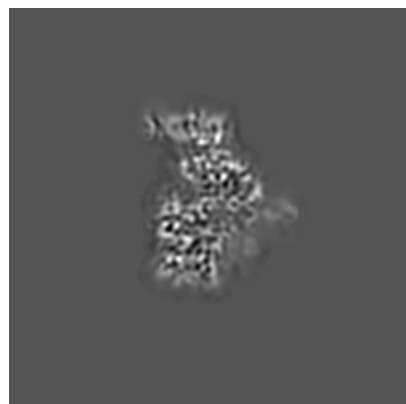


Z Index: 70

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: 65



Y Index: 69



Z Index: 60

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

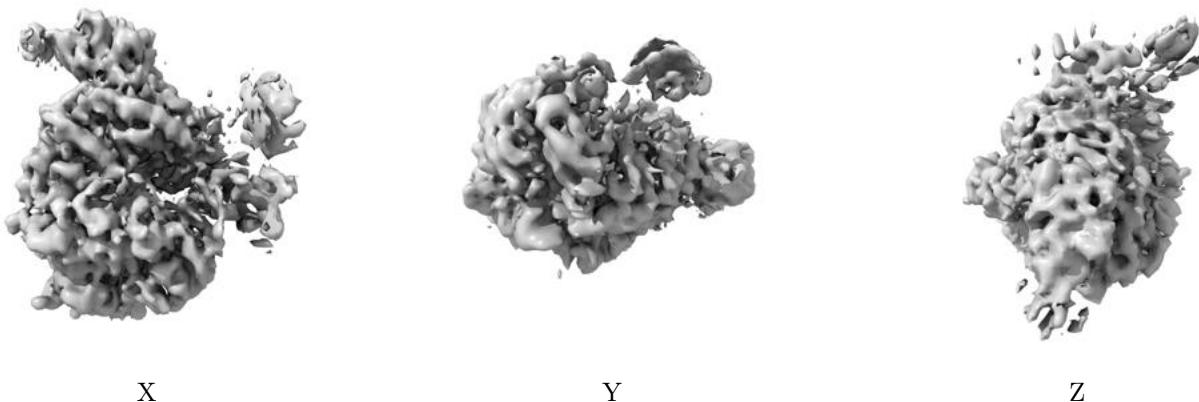


Z

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



The images above show the 3D surface view of the map at the recommended contour level 1.81. 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](#)

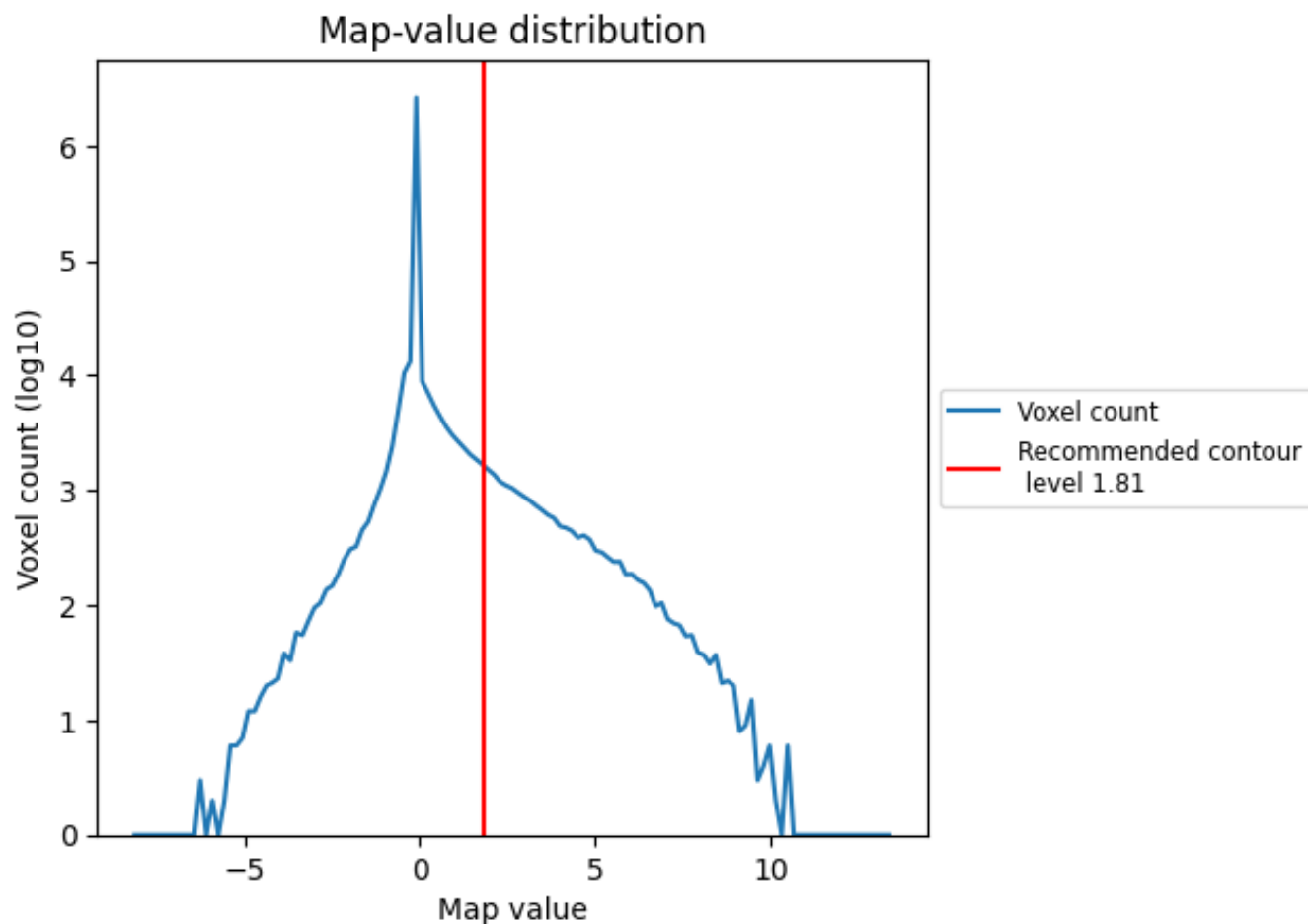
This section was not generated. No masks/segmentation were deposited.



## 7 Map analysis [i](#)

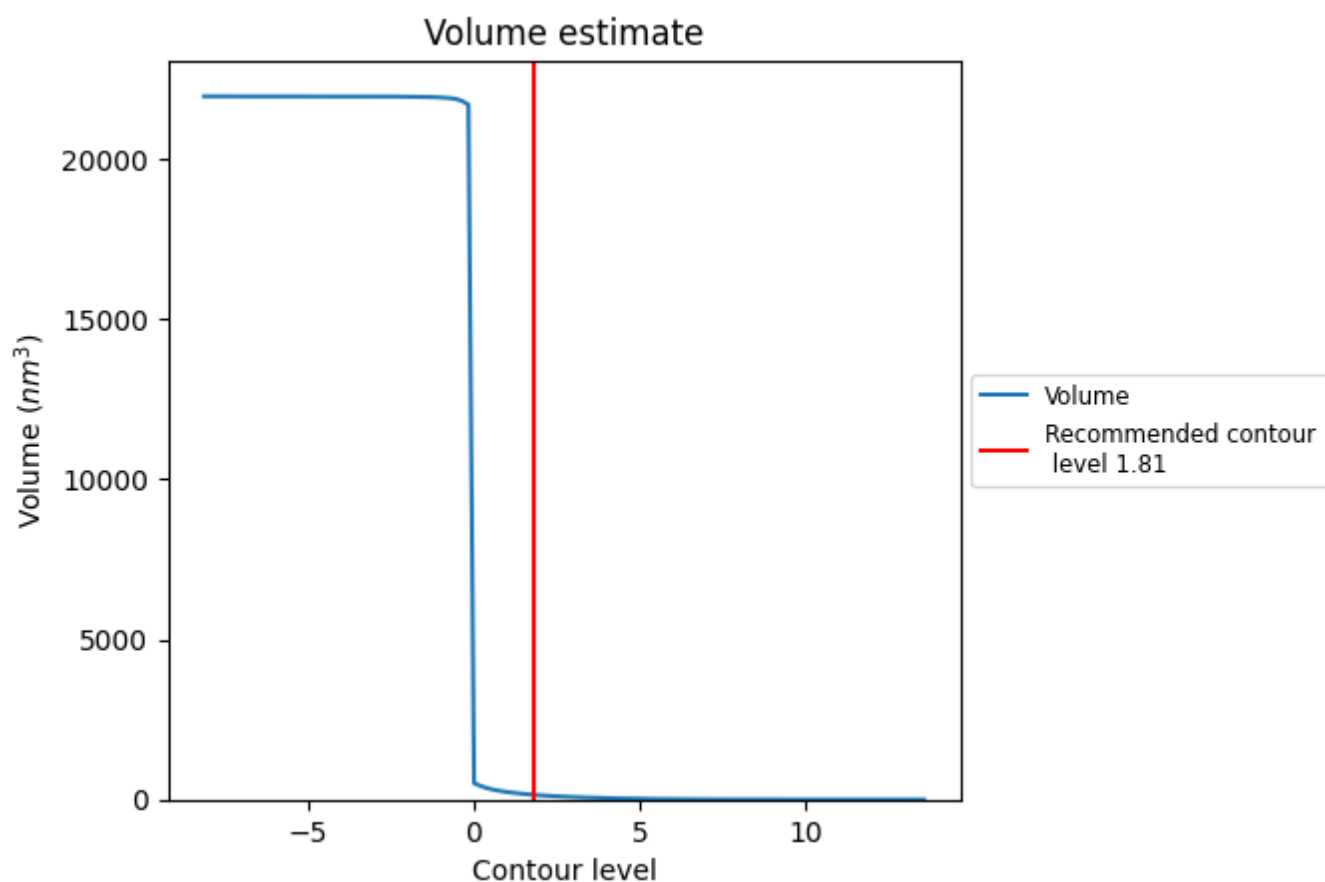
This section contains the results of statistical analysis of the map.

### 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.

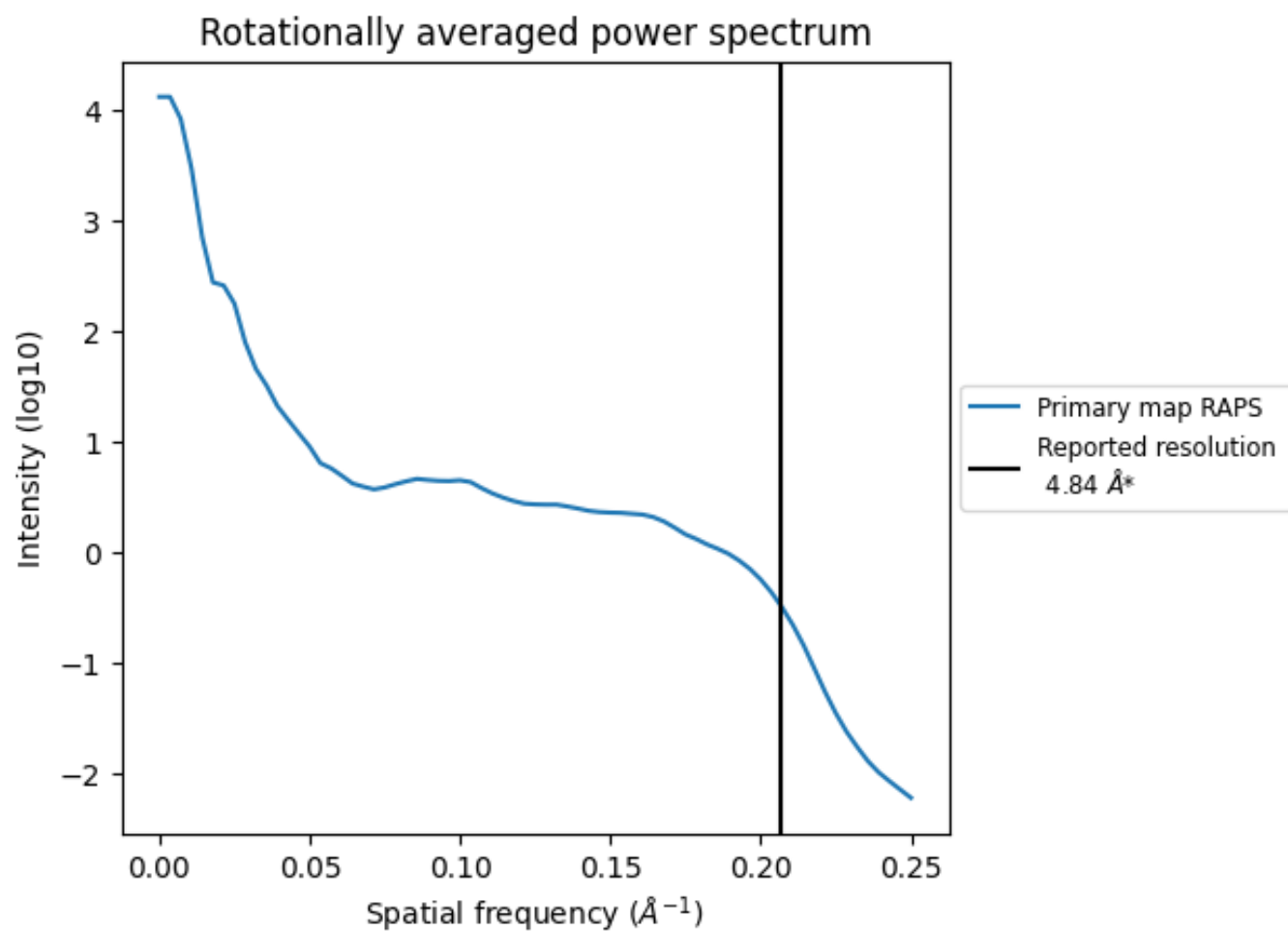
## 7.2 Volume estimate [i](#)



The volume at the recommended contour level is 148 nm<sup>3</sup>; this corresponds to an approximate mass of 133 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 ⓘ



\*Reported resolution corresponds to spatial frequency of 0.207 Å<sup>-1</sup>

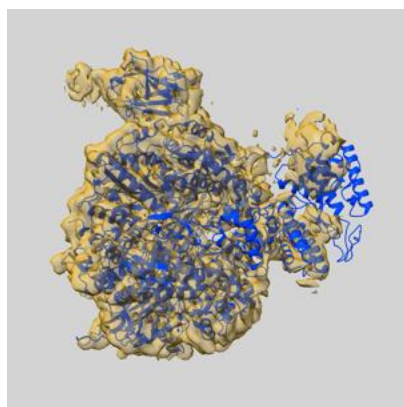
## 8 Fourier-Shell correlation

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

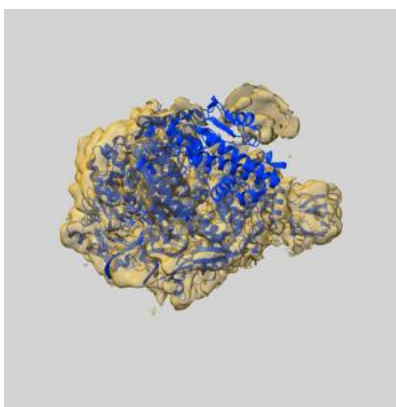
## 9 Map-model fit [i](#)

This section contains information regarding the fit between EMDB map EMD-12429 and PDB model 7NK2. Per-residue inclusion information can be found in [section 3](#) on [page 6](#).

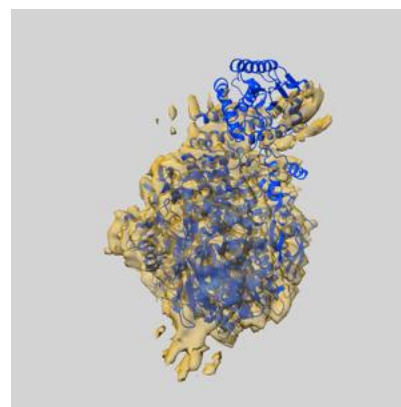
### 9.1 Map-model overlay [i](#)



X



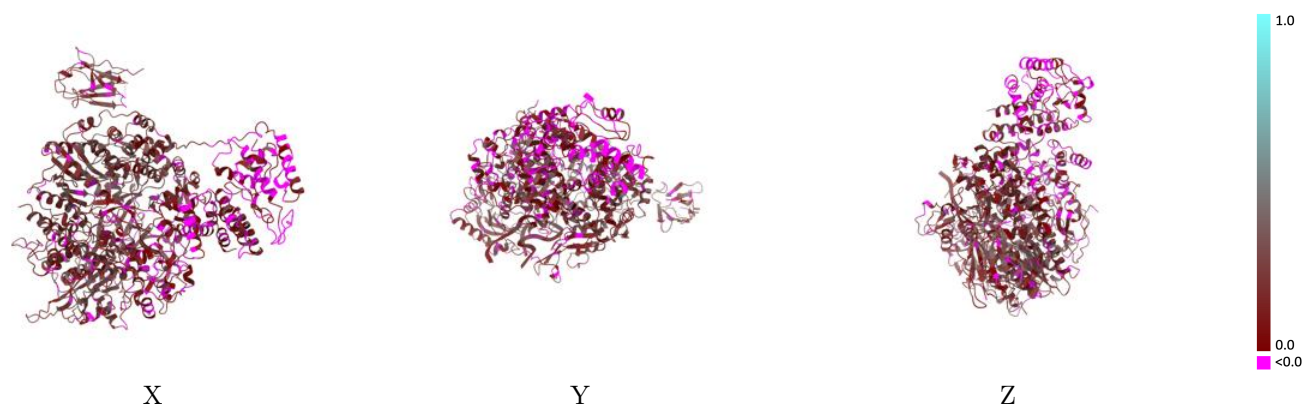
Y



Z

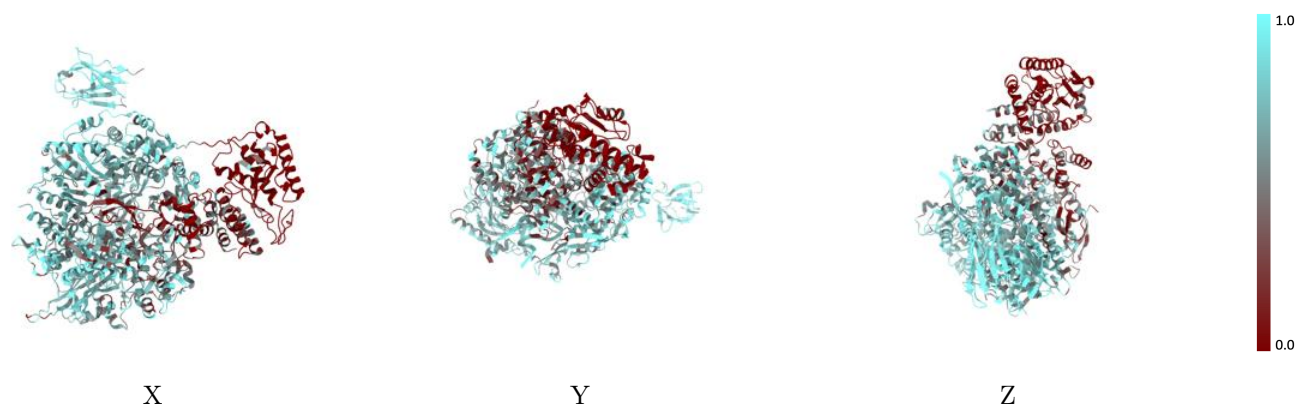
The images above show the 3D surface view of the map at the recommended contour level 1.81 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

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



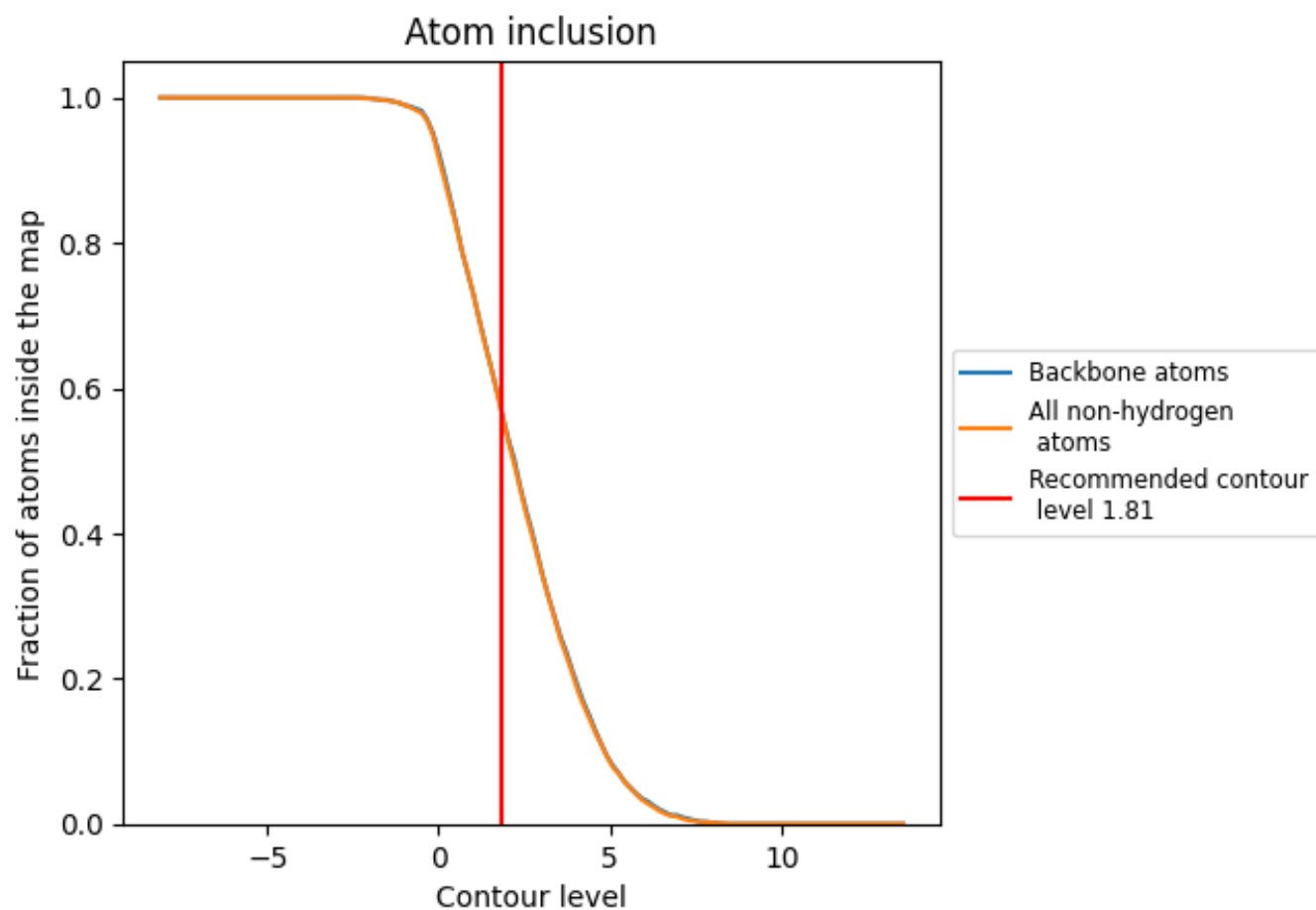
The images above show the model with each residue coloured according to its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions 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 (1.81).

## 9.4 Atom inclusion [i](#)



At the recommended contour level, 58% of all backbone atoms, 57% 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 (1.81) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	<div></div> 0.5730	<div></div> 0.1450
A	<div></div> 0.5420	<div></div> 0.1330
B	<div></div> 0.6360	<div></div> 0.1720
C	<div></div> 0.3610	<div></div> 0.0660
D	<div></div> 0.8280	<div></div> 0.2190
E	<div></div> 0.5980	<div></div> 0.2040
F	<div></div> 0.8390	<div></div> 0.1960

1.0

0.0

<0.0