wwPDB EM Validation Summary Report (i)

Nov 14, 2022 – 07:59 AM EST

PDB ID	:	7K79
EMDB ID	:	EMD-22697
Title	:	CBF3
Authors	:	Ruifang, G.; Yawen, B.
Deposited on	:	2020-09-22
Resolution	:	4.00 Å(reported)

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

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the (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 (i)) were used in the production of this report:

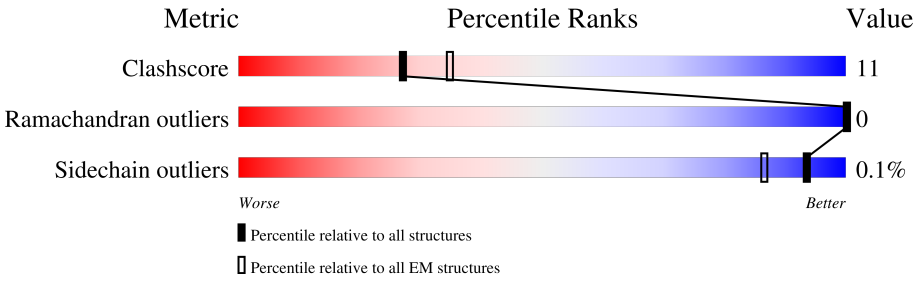
EMDB validation analysis	:	0.0.1.dev43
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.9
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 4.00 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



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

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the EM map (all-atom inclusion < 40%). The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	K	519	28% 50% 27% • 21%
2	L	620	10% 61% 22% • 15%
2	O	620	12% 61% 21% • 17%
3	N	194	21% 53% 24% • 22%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 13272 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 Centromere DNA-binding protein complex CBF3 subunit C.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	K	408	3413	2226	576	599	12	0	0

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
K	0	MET	-	expression tag	UNP P35203
K	1	GLY	-	expression tag	UNP P35203
K	479	GLY	-	expression tag	UNP P35203
K	480	SER	-	expression tag	UNP P35203
K	481	ARG	-	expression tag	UNP P35203
K	482	SER	-	expression tag	UNP P35203
K	483	GLY	-	expression tag	UNP P35203
K	484	SER	-	expression tag	UNP P35203
K	485	GLU	-	expression tag	UNP P35203
K	486	ASN	-	expression tag	UNP P35203
K	487	LEU	-	expression tag	UNP P35203
K	488	TYR	-	expression tag	UNP P35203
K	489	PHE	-	expression tag	UNP P35203
K	490	GLN	-	expression tag	UNP P35203
K	491	GLY	-	expression tag	UNP P35203
K	492	SER	-	expression tag	UNP P35203
K	493	LYS	-	expression tag	UNP P35203
K	494	ARG	-	expression tag	UNP P35203
K	495	ARG	-	expression tag	UNP P35203
K	496	TRP	-	expression tag	UNP P35203
K	497	LYS	-	expression tag	UNP P35203
K	498	LYS	-	expression tag	UNP P35203
K	499	ASN	-	expression tag	UNP P35203
K	500	PHE	-	expression tag	UNP P35203
K	501	ILE	-	expression tag	UNP P35203
K	502	ALA	-	expression tag	UNP P35203
K	503	VAL	-	expression tag	UNP P35203
K	504	SER	-	expression tag	UNP P35203

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
K	505	ALA	-	expression tag	UNP P35203
K	506	ALA	-	expression tag	UNP P35203
K	507	ASN	-	expression tag	UNP P35203
K	508	ARG	-	expression tag	UNP P35203
K	509	PHE	-	expression tag	UNP P35203
K	510	LYS	-	expression tag	UNP P35203
K	511	LYS	-	expression tag	UNP P35203
K	512	ILE	-	expression tag	UNP P35203
K	513	SER	-	expression tag	UNP P35203
K	514	SER	-	expression tag	UNP P35203
K	515	SER	-	expression tag	UNP P35203
K	516	GLY	-	expression tag	UNP P35203
K	517	ALA	-	expression tag	UNP P35203
K	518	LEU	-	expression tag	UNP P35203

- Molecule 2 is a protein called Centromere DNA-binding protein complex CBF3 subunit B.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	L	525	4355	2837	698	799	21	0	0
			Total	C	N	O	S		
2	O	517	4284	2789	690	784	21	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
L	609	GLY	-	expression tag	UNP P40969
L	610	SER	-	expression tag	UNP P40969
L	611	GLY	-	expression tag	UNP P40969
L	612	GLY	-	expression tag	UNP P40969
L	613	SER	-	expression tag	UNP P40969
L	614	GLY	-	expression tag	UNP P40969
L	615	GLU	-	expression tag	UNP P40969
L	616	ASN	-	expression tag	UNP P40969
L	617	LEU	-	expression tag	UNP P40969
L	618	TYR	-	expression tag	UNP P40969
L	619	PHE	-	expression tag	UNP P40969
L	620	GLN	-	expression tag	UNP P40969
O	609	GLY	-	expression tag	UNP P40969
O	610	SER	-	expression tag	UNP P40969
O	611	GLY	-	expression tag	UNP P40969
O	612	GLY	-	expression tag	UNP P40969

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
O	613	SER	-	expression tag	UNP P40969
O	614	GLY	-	expression tag	UNP P40969
O	615	GLU	-	expression tag	UNP P40969
O	616	ASN	-	expression tag	UNP P40969
O	617	LEU	-	expression tag	UNP P40969
O	618	TYR	-	expression tag	UNP P40969
O	619	PHE	-	expression tag	UNP P40969
O	620	GLN	-	expression tag	UNP P40969

- Molecule 3 is a protein called Suppressor of kinetochore protein 1.

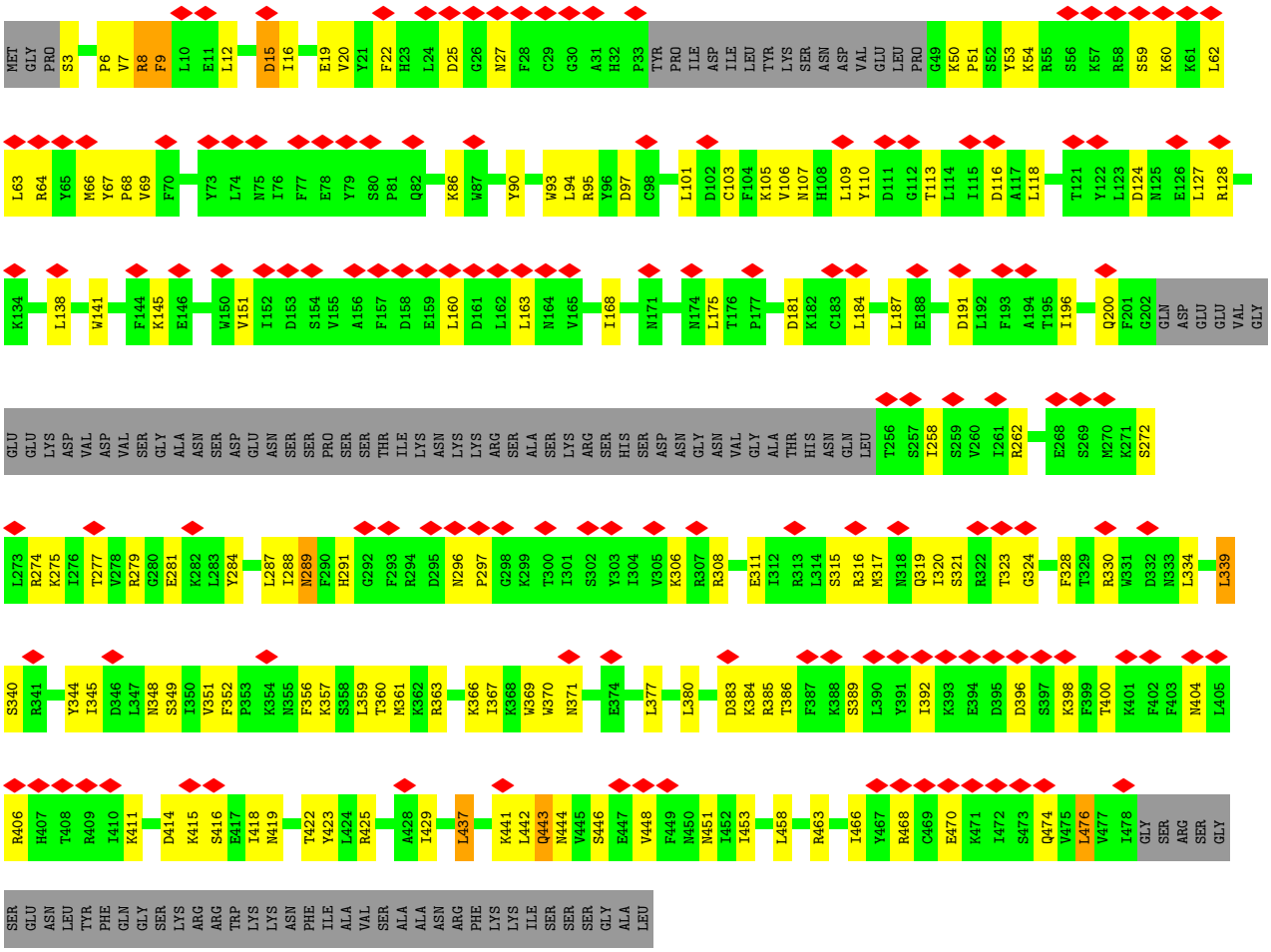
Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	N	151	1220	764	211	241	4	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and atom inclusion in map density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red diamond above a residue indicates a poor fit to the EM map for this residue (all-atom inclusion < 40%). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

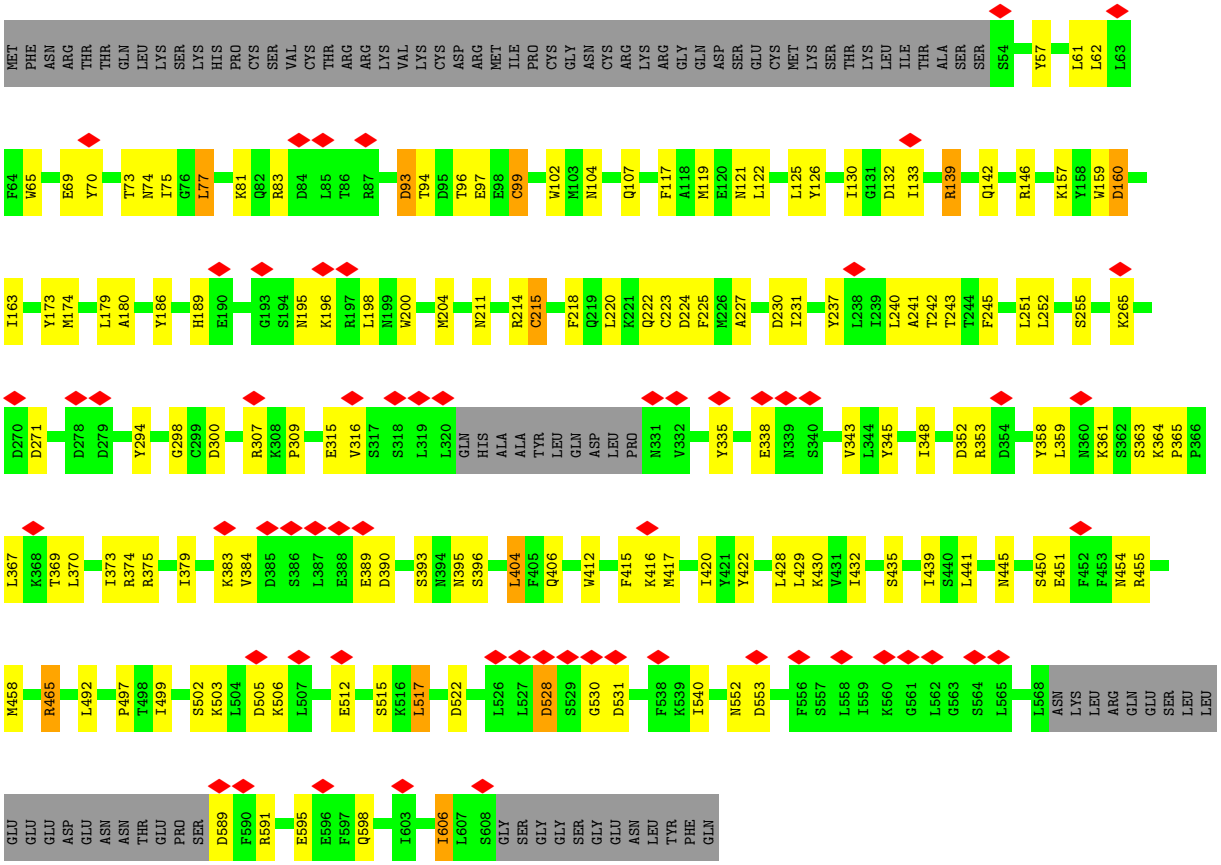
• Molecule 1: Centromere DNA-binding protein complex CBF3 subunit C

Chain K: 28% 50% 27% • 21%



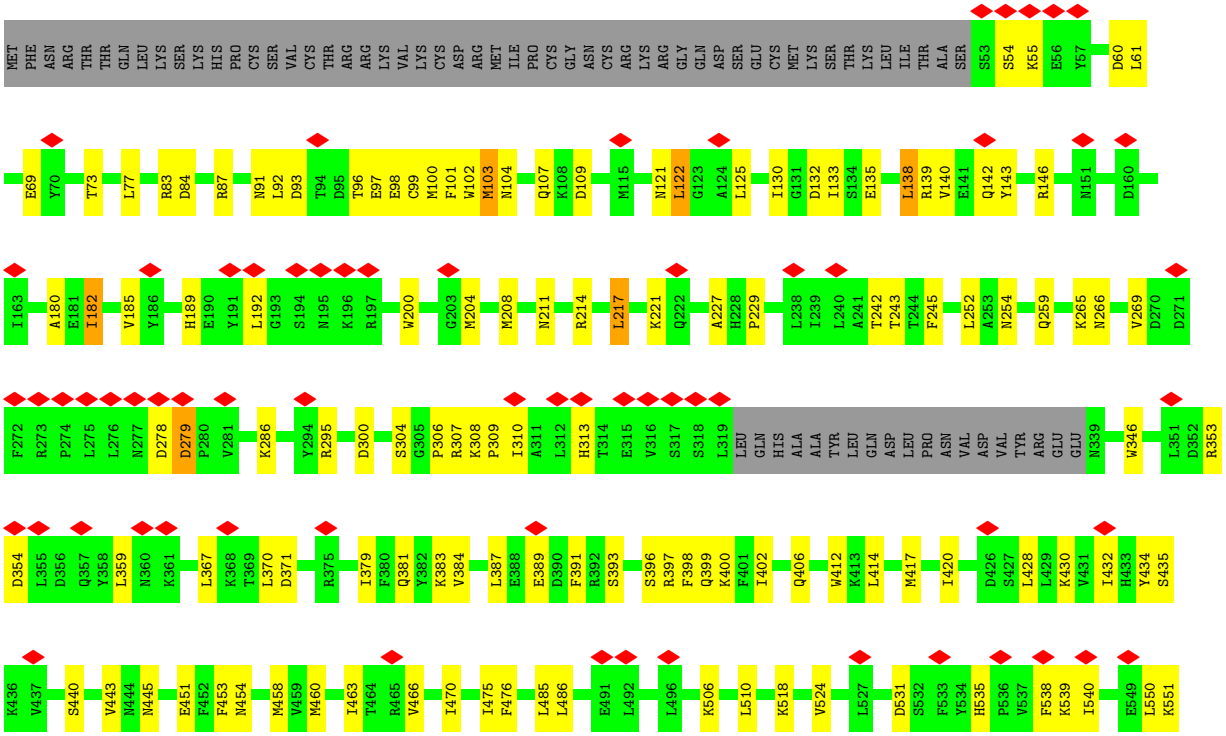
• Molecule 2: Centromere DNA-binding protein complex CBF3 subunit B

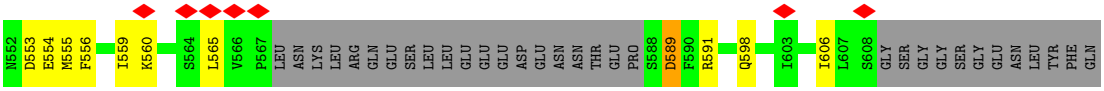
Chain L: 10% 61% 22% • 15%



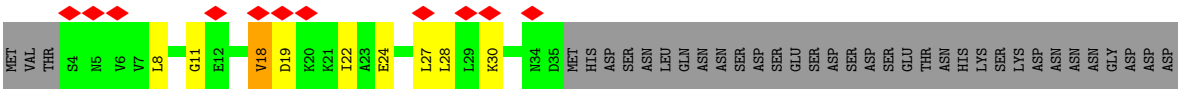
• Molecule 2: Centromere DNA-binding protein complex CBF3 subunit B

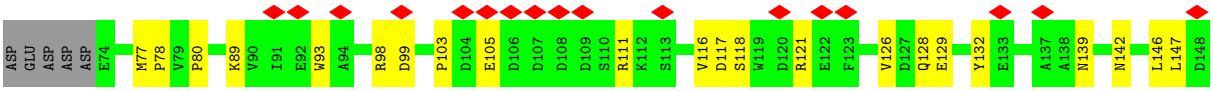
Chain O:





• Molecule 3: Suppressor of kinetochore protein 1





4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	115666	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	NONE	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	71	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	0.057	Depositor
Minimum map value	-0.020	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.002	Depositor
Recommended contour level	0.02	Depositor
Map size (Å)	296.8, 296.8, 296.8	wwPDB
Map dimensions	280, 280, 280	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.06, 1.06, 1.06	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	K	0.45	0/3489	0.99	20/4709 (0.4%)
2	L	0.49	2/4461 (0.0%)	0.97	19/6045 (0.3%)
2	O	0.47	0/4389	0.97	21/5946 (0.4%)
3	N	0.47	1/1241 (0.1%)	0.93	3/1680 (0.2%)
All	All	0.47	3/13580 (0.0%)	0.97	63/18380 (0.3%)

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
1	K	0	1
3	N	0	1
All	All	0	2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	L	99	CYS	CB-SG	-5.37	1.73	1.81
3	N	151	CYS	CB-SG	-5.15	1.73	1.81
2	L	215	CYS	CB-SG	-5.06	1.73	1.81

The worst 5 of 63 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	L	352	ASP	CB-CG-OD1	8.66	126.10	118.30
1	K	414	ASP	CB-CG-OD1	8.60	126.04	118.30
2	L	122	LEU	CA-CB-CG	8.44	134.71	115.30
2	O	60	ASP	CB-CG-OD1	8.06	125.55	118.30
1	K	116	ASP	CB-CG-OD1	7.99	125.49	118.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	K	9	PHE	Peptide
3	N	186	ARG	Peptide

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	K	3413	0	3462	88	0
2	L	4355	0	4318	92	0
2	O	4284	0	4246	89	0
3	N	1220	0	1183	42	0
All	All	13272	0	13209	286	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 11.

The worst 5 of 286 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:L:412:TRP:HE1	2:L:435:SER:HG	1.31	0.79
2:O:430:LYS:O	2:O:434:TYR:HB2	1.81	0.79
3:N:191:ALA:O	3:N:192:GLU:HG2	1.89	0.72
2:L:57:TYR:O	2:L:61:LEU:HB2	1.90	0.72
2:O:391:PHE:HA	2:O:397:ARG:HH12	1.55	0.71

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	K	402/519 (78%)	336 (84%)	66 (16%)	0	100	100
2	L	519/620 (84%)	484 (93%)	35 (7%)	0	100	100
2	O	511/620 (82%)	477 (93%)	34 (7%)	0	100	100
3	N	147/194 (76%)	130 (88%)	17 (12%)	0	100	100
All	All	1579/1953 (81%)	1427 (90%)	152 (10%)	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	K	378/481 (79%)	376 (100%)	2 (0%)	88	93
2	L	489/577 (85%)	489 (100%)	0	100	100
2	O	480/577 (83%)	480 (100%)	0	100	100
3	N	135/179 (75%)	135 (100%)	0	100	100
All	All	1482/1814 (82%)	1480 (100%)	2 (0%)	93	97

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

Mol	Chain	Res	Type
1	K	107	ASN
1	K	289	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	K	451	ASN
1	K	474	GLN
2	L	222	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	O	142	GLN
2	O	395	ASN

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)

There are no chain breaks in this entry.

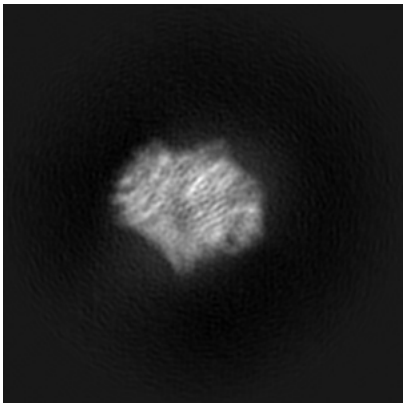
6 Map visualisation (i)

This section contains visualisations of the EMDB entry EMD-22697. These allow visual inspection of the internal detail of the map and identification of artifacts.

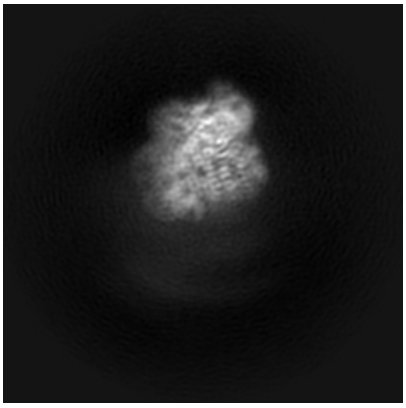
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections (i)

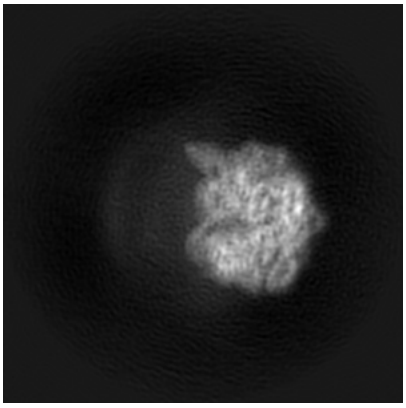
6.1.1 Primary map



X



Y

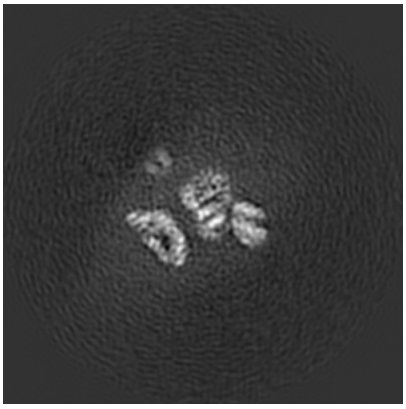


Z

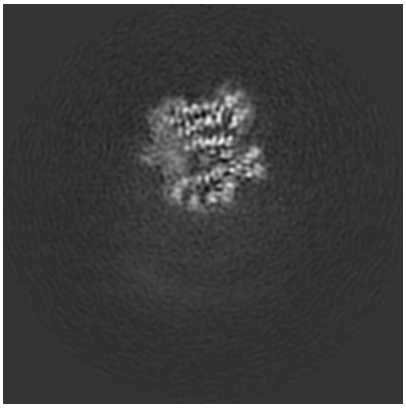
The images above show the map projected in three orthogonal directions.

6.2 Central slices (i)

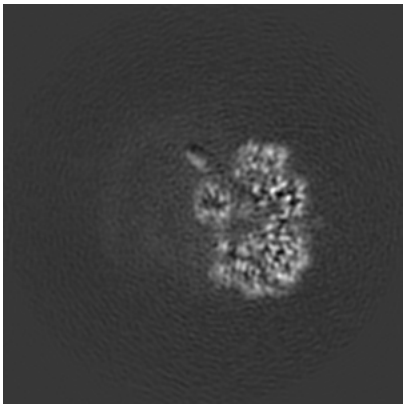
6.2.1 Primary map



X Index: 140



Y Index: 140

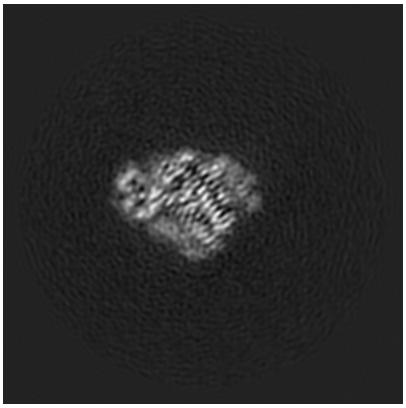


Z Index: 140

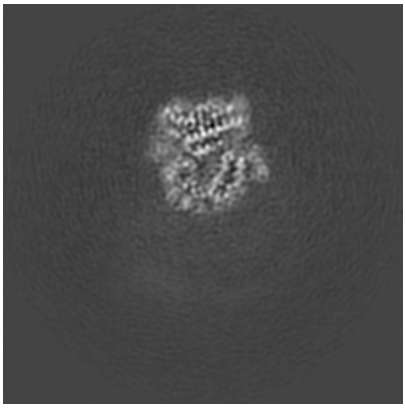
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

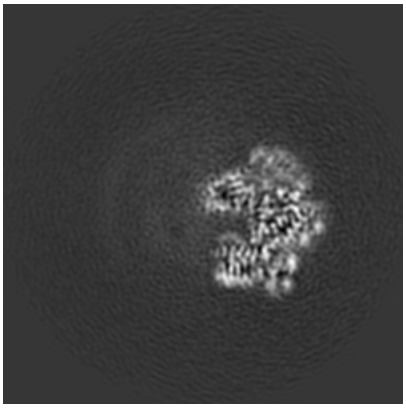
6.3.1 Primary map



X Index: 198



Y Index: 149

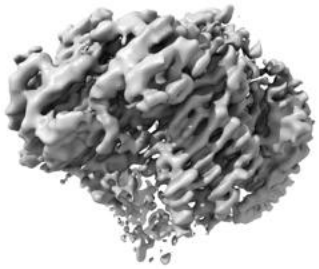


Z Index: 155

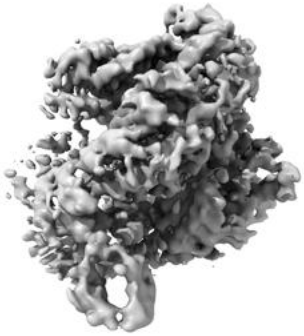
The images above show the largest variance slices of the map in three orthogonal directions.

6.4 Orthogonal surface views (i)

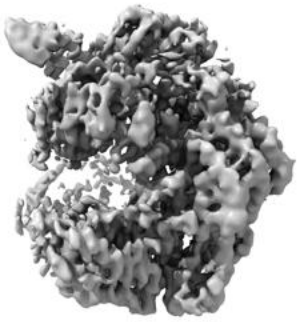
6.4.1 Primary map



X



Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.02. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

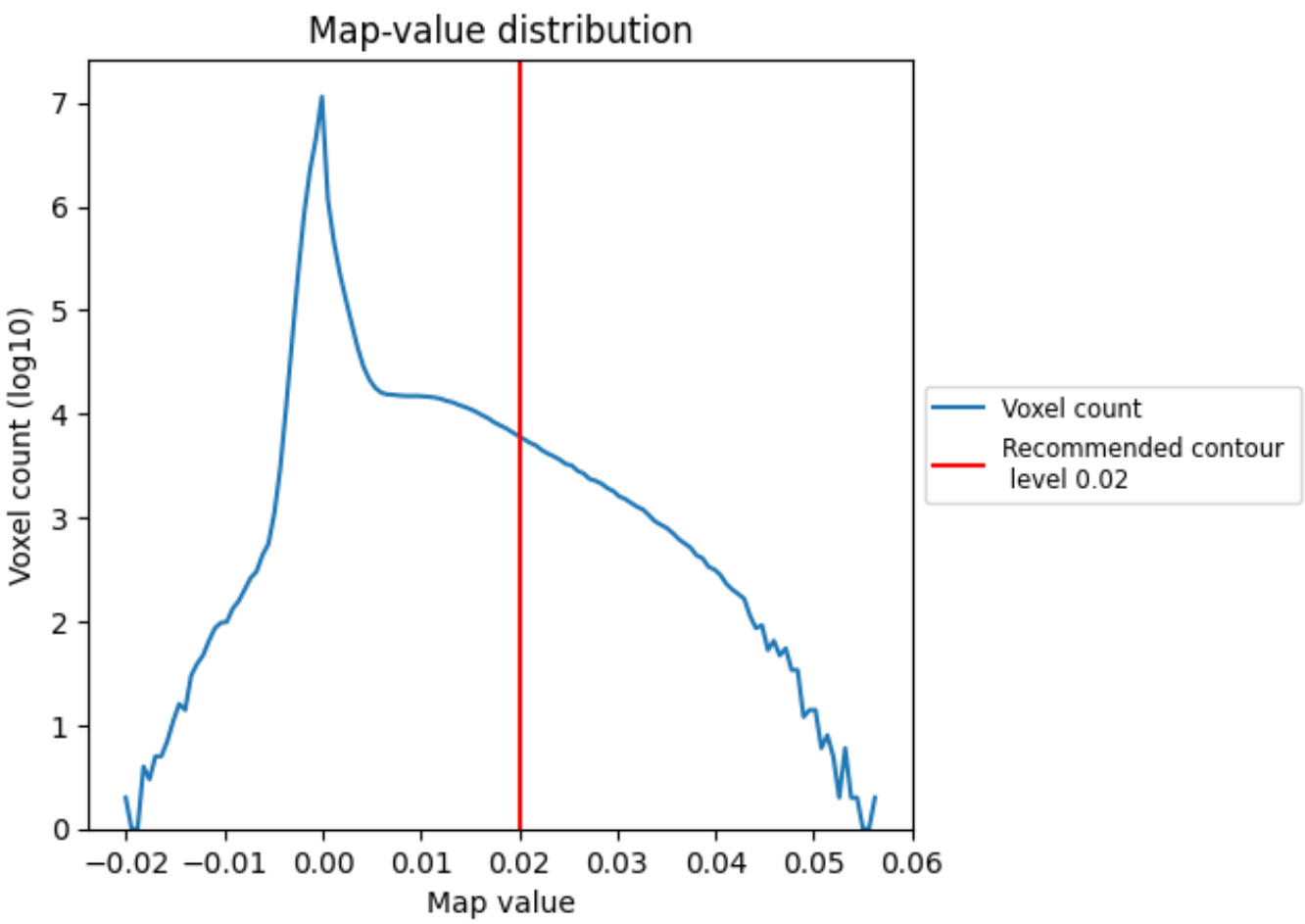
6.5 Mask visualisation (i)

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

7 Map analysis (i)

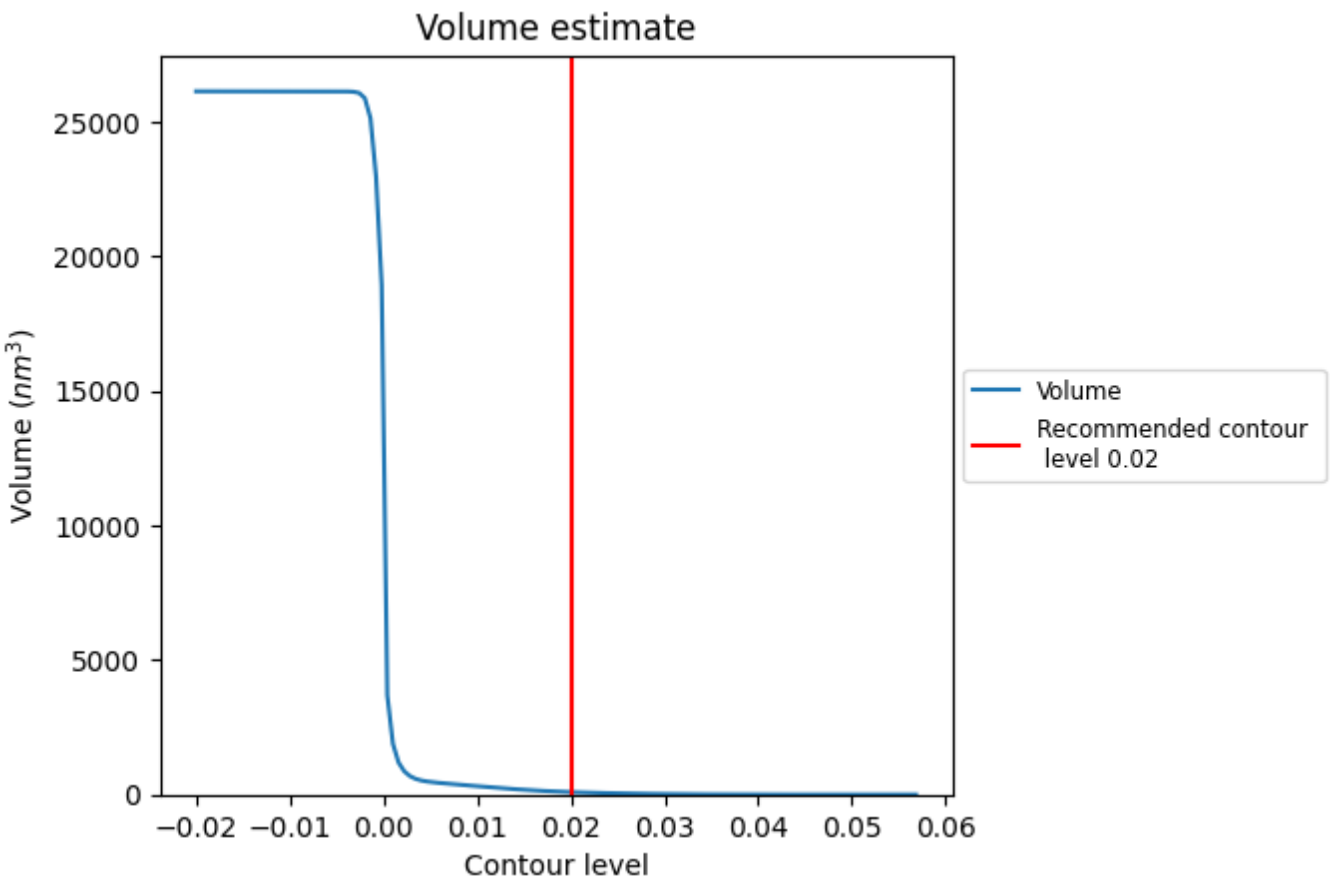
This section contains the results of statistical analysis of the map.

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.

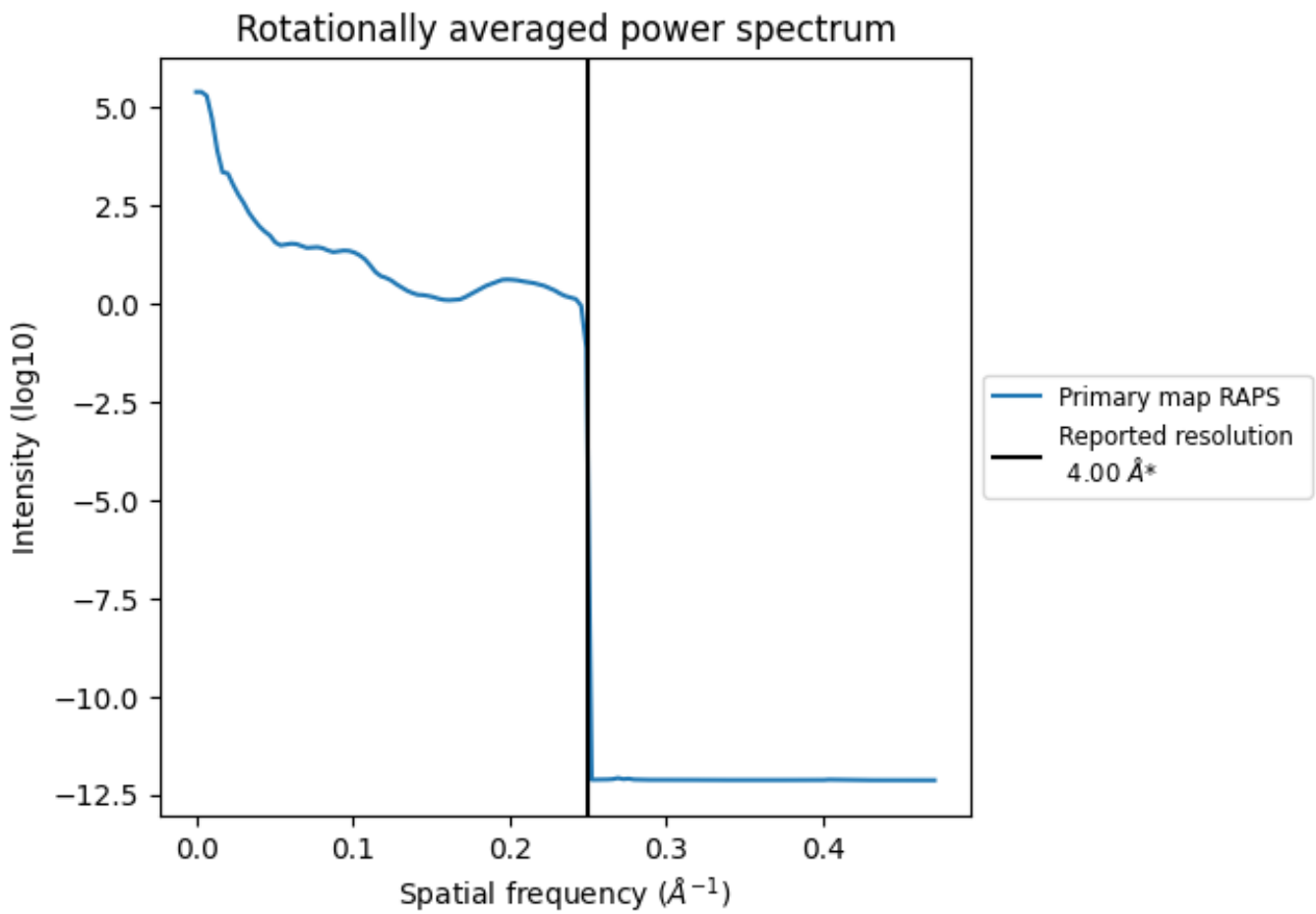
7.2 Volume estimate (i)



The volume at the recommended contour level is 91 nm^3; this corresponds to an approximate mass of 82 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 0.250 Å^{-1}

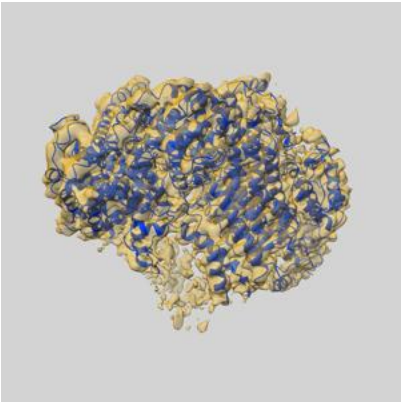
8 Fourier-Shell correlation (i)

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

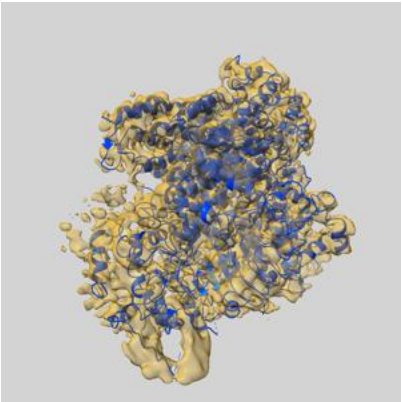
9 Map-model fit (i)

This section contains information regarding the fit between EMDB map EMD-22697 and PDB model 7K79. Per-residue inclusion information can be found in section 3 on page 6.

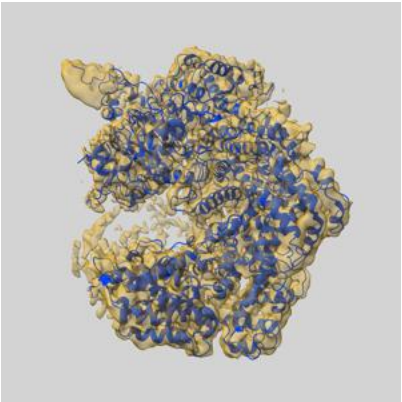
9.1 Map-model overlay (i)



X



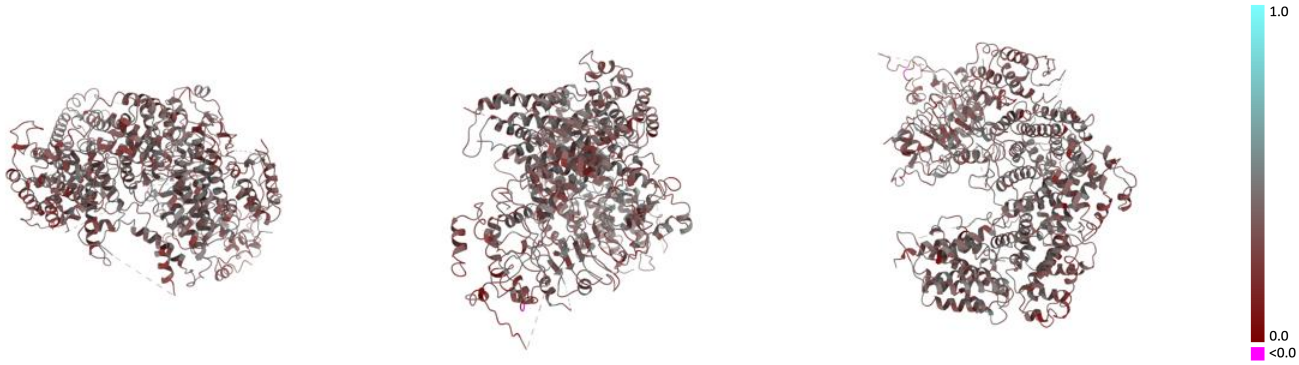
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.02 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.

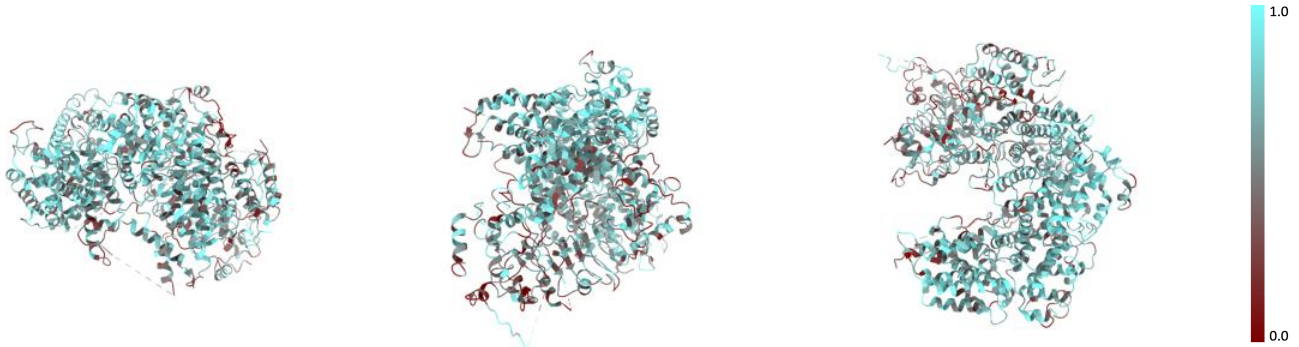
9.2 Q-score mapped to coordinate model (i)



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.

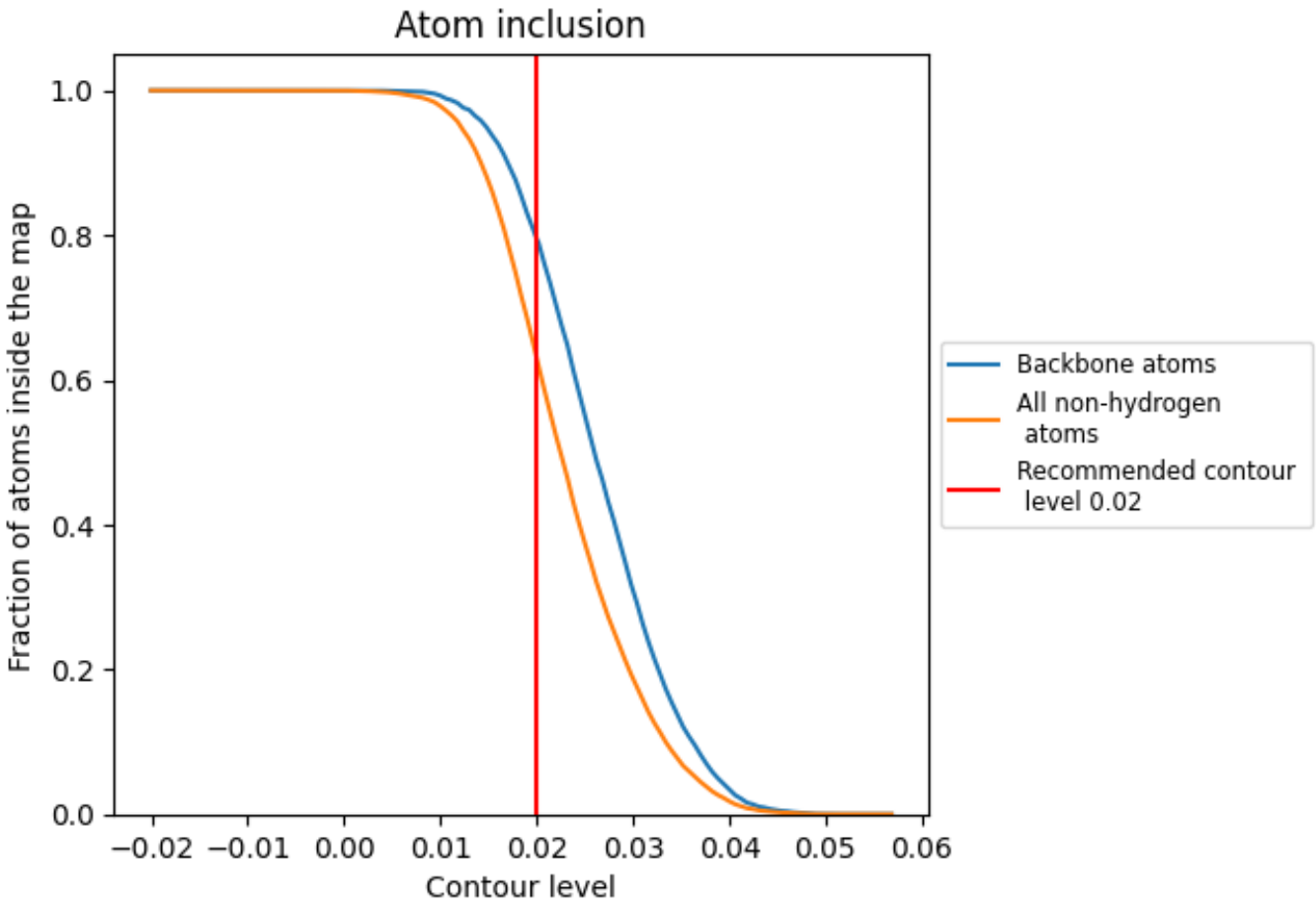
9.3 Atom inclusion mapped to coordinate model (i)



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 (0.02).

9.4 Atom inclusion (i)



At the recommended contour level, 80% of all backbone atoms, 63% of all non-hydrogen atoms, are inside the map.

9.5 Map-model fit summary (i)

The table lists the average atom inclusion at the recommended contour level (0.02) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.6302	0.3610
K	0.5091	0.3430
L	0.6968	0.3840
N	0.5638	0.3330
O	0.6775	0.3610