wwPDB EM Validation Summary Report (i)

Nov 8, 2022 – 03:55 PM EST

PDB ID	:	6MUS
EMDB ID	:	EMD-9254
Title	:	Cryo-EM structure of larger Csm-crRNA-target RNA ternary complex in type III-A CRISPR-Cas system
Authors	:	Jia, N.; Wang, C.; Eng, E.T.
Deposited on	:	2018-10-23
Resolution	:	3.60 Å(reported)

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

We welcome your comments at *validation@mail.wwpdb.org*
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:

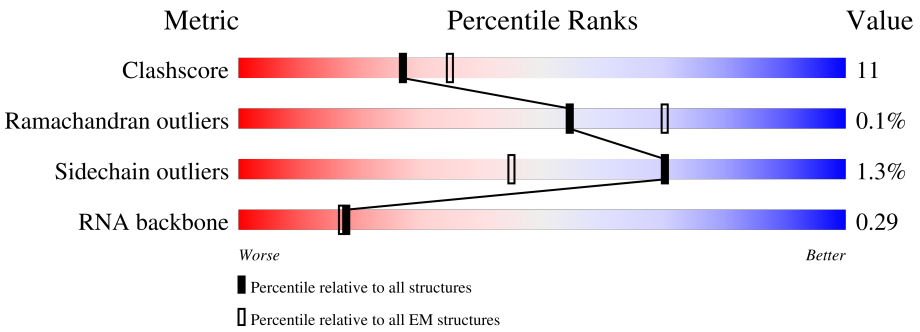
EMDB validation analysis	:	0.0.1.dev43
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.9
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

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 3.60 Å.

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



Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	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	791	82% 15% • •
2	B	187	13% 67% 9% 25%
2	J	187	13% 57% 8% 35%
3	C	291	71% 23% • 5%
3	D	291	70% 27% • •
3	K	291	71% 22% • 5%
4	E	289	85% 12% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	G	38	5%, 26%, 29%, 32%, 13%
6	H	40	12%, 25%, 18%, 8%, 38%
7	F	403	50%, 74%, 19%, 6%

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 20864 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 Uncharacterized protein Csm1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	772	6155	3944	1073	1119	19	0	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-13	MET	-	initiating methionine	UNP B6YWB8
A	-12	GLY	-	expression tag	UNP B6YWB8
A	-11	SER	-	expression tag	UNP B6YWB8
A	-10	SER	-	expression tag	UNP B6YWB8
A	-9	HIS	-	expression tag	UNP B6YWB8
A	-8	HIS	-	expression tag	UNP B6YWB8
A	-7	HIS	-	expression tag	UNP B6YWB8
A	-6	HIS	-	expression tag	UNP B6YWB8
A	-5	HIS	-	expression tag	UNP B6YWB8
A	-4	HIS	-	expression tag	UNP B6YWB8
A	-3	SER	-	expression tag	UNP B6YWB8
A	-2	GLN	-	expression tag	UNP B6YWB8
A	-1	ASP	-	expression tag	UNP B6YWB8
A	0	PRO	-	expression tag	UNP B6YWB8
A	14	ALA	HIS	conflict	UNP B6YWB8
A	15	ASN	ASP	conflict	UNP B6YWB8

- Molecule 2 is a protein called Uncharacterized protein Csm2.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	B	141	1040	662	184	190	4	0	0
2	J	121	907	582	159	162	4	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	SER	-	expression tag	UNP B6YWB9
J	0	SER	-	expression tag	UNP B6YWB9

- Molecule 3 is a protein called Uncharacterized protein Csm3.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	C	277	2210	1404	388	412	6	0	0
			Total	C	N	O	S		
3	D	287	2284	1447	401	429	7	0	0
			Total	C	N	O	S		
3	K	275	2198	1396	385	410	7	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	0	SER	-	expression tag	UNP B6YWC0
C	36	ALA	ASP	engineered mutation	UNP B6YWC0
D	0	SER	-	expression tag	UNP B6YWC0
D	36	ALA	ASP	engineered mutation	UNP B6YWC0
K	0	SER	-	expression tag	UNP B6YWC0
K	36	ALA	ASP	engineered mutation	UNP B6YWC0

- Molecule 4 is a protein called Uncharacterized protein Csm4.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	E	279	2207	1436	373	394	4	0	0

- Molecule 5 is a RNA chain called RNA (33-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
5	G	33	717	319	136	230	32	0	0

- Molecule 6 is a RNA chain called RNA (25-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
6	H	25	524	235	92	172	25	0	0

- Molecule 7 is a protein called Uncharacterized protein Csm5.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
7	F	377	2618	1653	469	487	9	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	398	HIS	-	expression tag	UNP B6YWC2
F	399	HIS	-	expression tag	UNP B6YWC2
F	400	HIS	-	expression tag	UNP B6YWC2
F	401	HIS	-	expression tag	UNP B6YWC2
F	402	HIS	-	expression tag	UNP B6YWC2
F	403	HIS	-	expression tag	UNP B6YWC2

- Molecule 8 is ZINC ION (three-letter code: ZN) (formula: Zn).

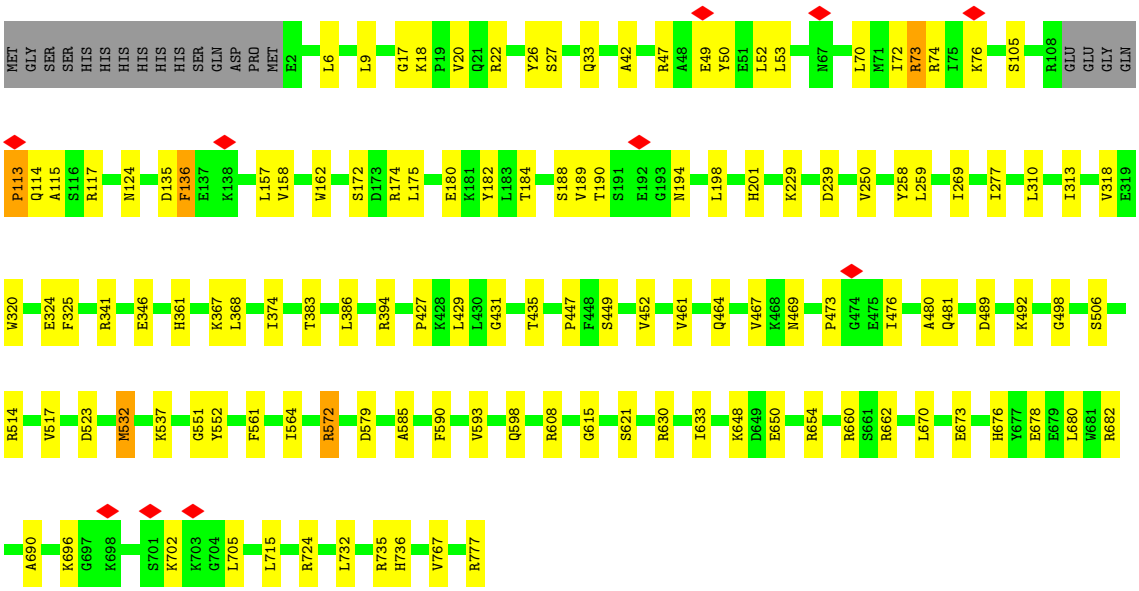
Mol	Chain	Residues	Atoms		AltConf
			Total	Zn	
8	A	1	1	1	0
8	C	1	1	1	0
8	D	1	1	1	0
8	K	1	1	1	0

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.

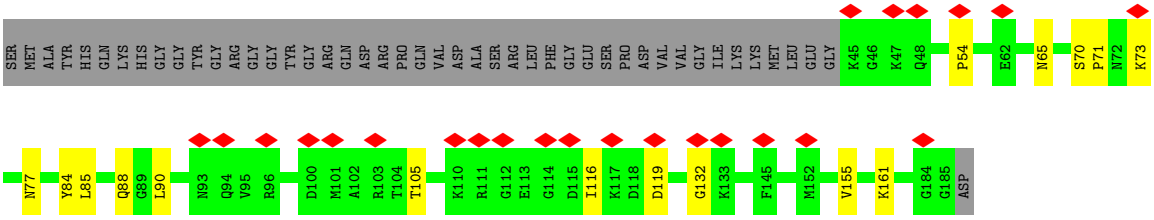
- Molecule 1: Uncharacterized protein Csm1

Chain A:



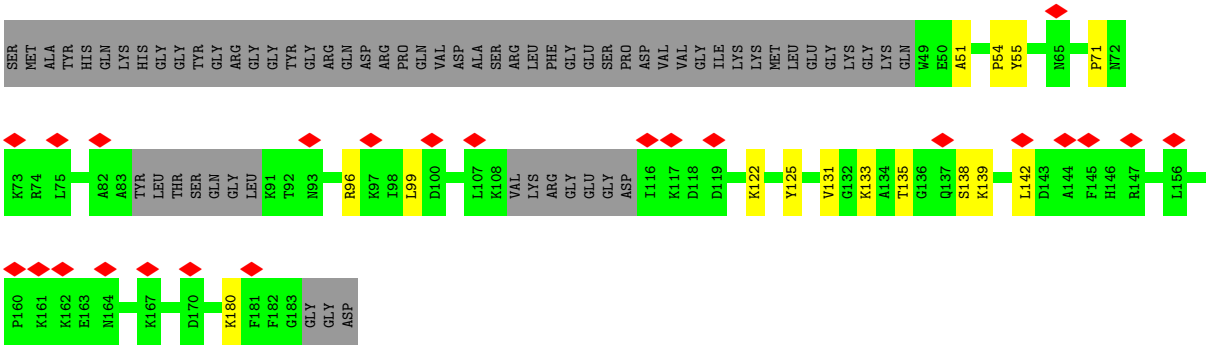
- Molecule 2: Uncharacterized protein Csm2

Chain B:



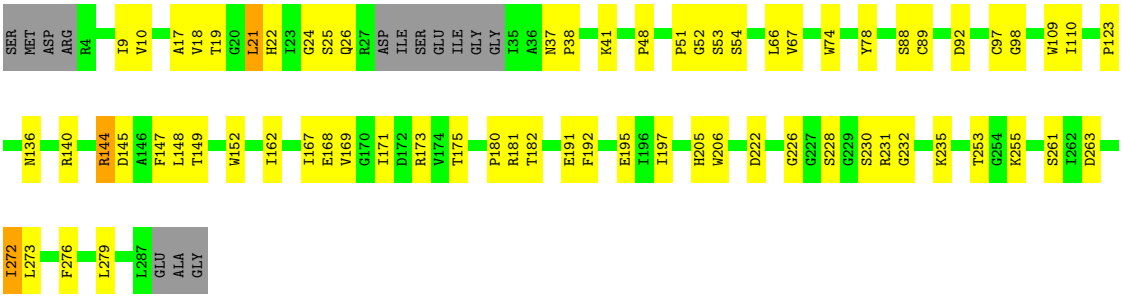
- Molecule 2: Uncharacterized protein Csm2

Chain J:



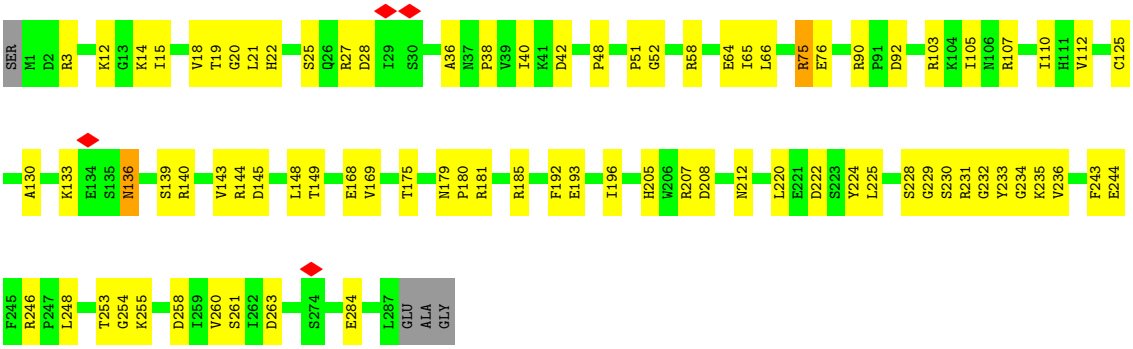
• Molecule 3: Uncharacterized protein Csm3

Chain C:



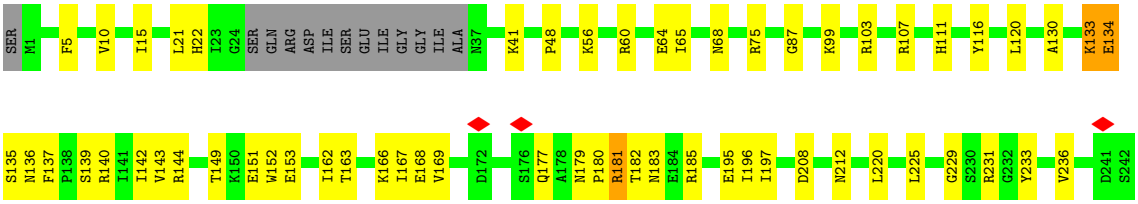
• Molecule 3: Uncharacterized protein Csm3

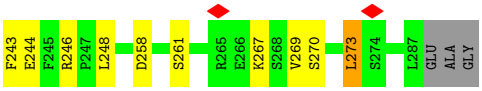
Chain D:



• Molecule 3: Uncharacterized protein Csm3

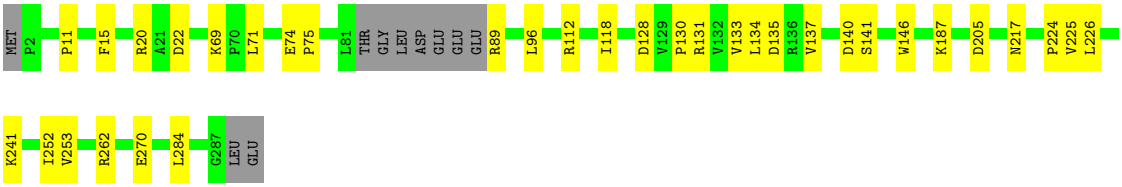
Chain K:





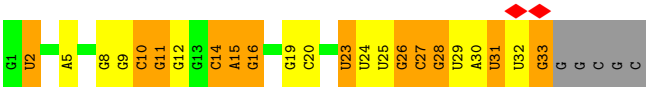
• Molecule 4: Uncharacterized protein Csm4

Chain E:



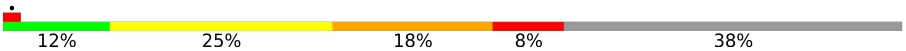
• Molecule 5: RNA (33-MER)

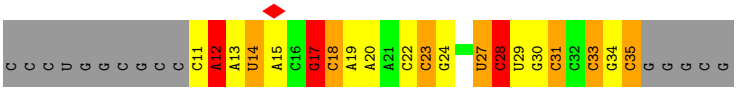
Chain G:



• Molecule 6: RNA (25-MER)

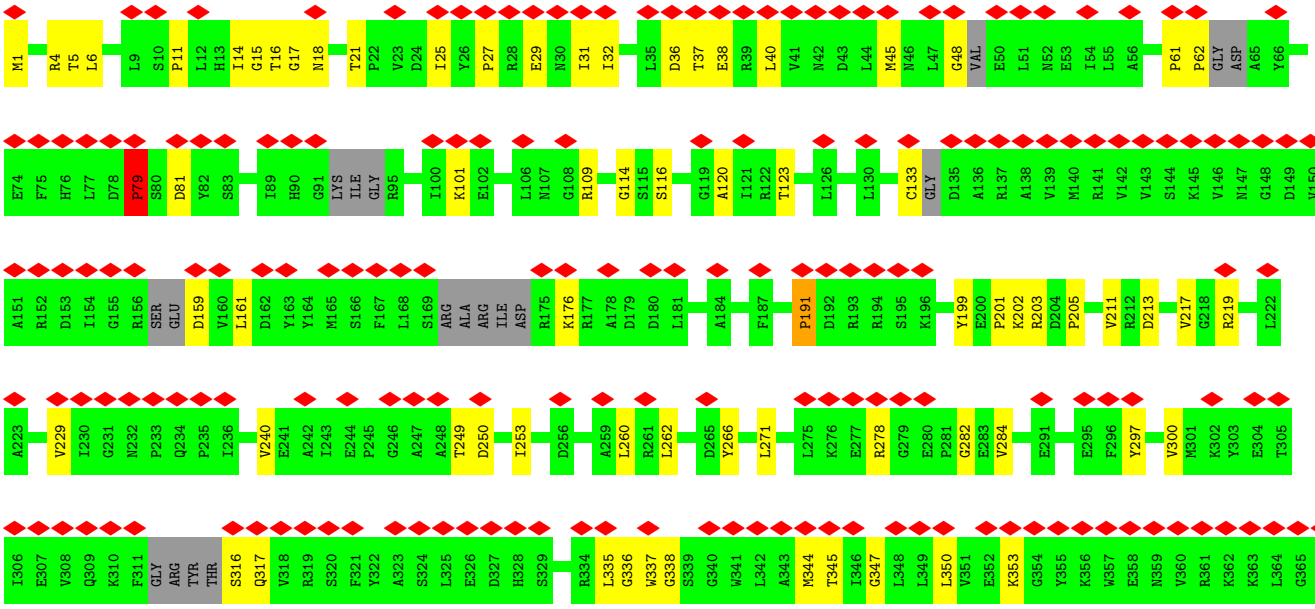
Chain H:

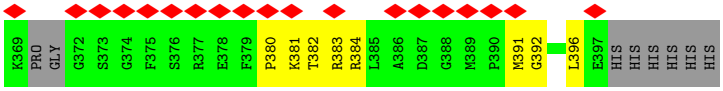




• Molecule 7: Uncharacterized protein Csm5

Chain F:




K369
PRO
GLY
G372
S373
G374
F375
S376
R377
E378
F379
P380
K381
T382
R383
R384
L385
A386
D387
G388
M389
P390
M391
G392
L396
E397
HIS
HIS
HIS
HIS
HIS
HIS

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	30431	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 ($e^-/Å^2$)	1.35	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	0.232	Depositor
Minimum map value	-0.152	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.005	Depositor
Recommended contour level	0.02	Depositor
Map size (Å)	278.784, 278.784, 278.784	wwPDB
Map dimensions	256, 256, 256	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.089, 1.089, 1.089	Depositor

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	0.42	0/6295	0.58	1/8496 (0.0%)
2	B	0.29	0/1055	0.57	2/1421 (0.1%)
2	J	0.28	0/919	0.58	2/1236 (0.2%)
3	C	0.54	0/2259	0.60	0/3046
3	D	0.50	0/2334	0.65	0/3147
3	K	0.41	0/2247	0.57	0/3029
4	E	0.50	0/2267	0.60	0/3070
5	G	0.89	2/804 (0.2%)	1.08	4/1256 (0.3%)
6	H	0.64	3/583 (0.5%)	1.18	5/904 (0.6%)
7	F	0.38	0/2658	0.63	4/3614 (0.1%)
All	All	0.47	5/21421 (0.0%)	0.65	18/29219 (0.1%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	G	33	G	C1'-N9	-6.86	1.37	1.46
6	H	12	A	C1'-N9	-5.84	1.38	1.46
6	H	17	G	C1'-N9	-5.71	1.38	1.46
5	G	32	U	C1'-N1	5.23	1.56	1.48
6	H	14	U	C1'-N1	5.22	1.56	1.48

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
6	H	28	C	N1-C2-O2	8.20	123.82	118.90
5	G	20	C	N1-C2-O2	7.99	123.69	118.90
5	G	20	C	N3-C2-O2	-7.29	116.80	121.90
5	G	20	C	C2-N1-C1'	6.98	126.48	118.80
6	H	33	C	C5-C6-N1	6.71	124.36	121.00

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	6155	0	6145	74	0
2	B	1040	0	994	7	0
2	J	907	0	872	11	0
3	C	2210	0	2181	83	0
3	D	2284	0	2252	125	0
3	K	2198	0	2168	74	0
4	E	2207	0	2221	23	0
5	G	717	0	358	59	0
6	H	524	0	273	21	0
7	F	2618	0	2284	91	0
8	A	1	0	0	0	0
8	C	1	0	0	0	0
8	D	1	0	0	0	0
8	K	1	0	0	0	0
All	All	20864	0	19748	443	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 443 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:66:LEU:CD2	3:D:248:LEU:HD11	1.57	1.32
5:G:26:G:O6	7:F:344:MET:HG2	1.28	1.30
3:C:25:SER:O	3:C:26:GLN:HG3	1.34	1.24
3:D:27:ARG:NE	3:D:36:ALA:HA	1.49	1.24
5:G:26:G:N2	7:F:335:LEU:HA	1.53	1.21

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	768/791 (97%)	712 (93%)	56 (7%)	0	100	100
2	B	139/187 (74%)	129 (93%)	9 (6%)	1 (1%)	22	61
2	J	115/187 (62%)	106 (92%)	9 (8%)	0	100	100
3	C	273/291 (94%)	250 (92%)	23 (8%)	0	100	100
3	D	285/291 (98%)	257 (90%)	28 (10%)	0	100	100
3	K	271/291 (93%)	244 (90%)	27 (10%)	0	100	100
4	E	275/289 (95%)	256 (93%)	19 (7%)	0	100	100
7	F	359/403 (89%)	327 (91%)	30 (8%)	2 (1%)	25	64
All	All	2485/2730 (91%)	2281 (92%)	201 (8%)	3 (0%)	54	83

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
7	F	79	PRO
7	F	191	PRO
2	B	70	SER

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	643/664 (97%)	634 (99%)	9 (1%)	67	85
2	B	95/154 (62%)	95 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	J	84/154 (54%)	84 (100%)	0	100	100
3	C	237/247 (96%)	234 (99%)	3 (1%)	69	86
3	D	245/247 (99%)	240 (98%)	5 (2%)	55	79
3	K	236/247 (96%)	230 (98%)	6 (2%)	47	75
4	E	231/240 (96%)	230 (100%)	1 (0%)	91	97
7	F	213/346 (62%)	212 (100%)	1 (0%)	88	95
All	All	1984/2299 (86%)	1959 (99%)	25 (1%)	70	86

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	136	ASN
4	E	89	ARG
7	F	18	ASN
3	D	236	VAL
3	K	75	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	179	ASN
7	F	18	ASN
3	D	26	GLN
7	F	263	ASN
3	K	136	ASN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
5	G	32/38 (84%)	12 (37%)	4 (12%)
6	H	24/40 (60%)	14 (58%)	0
All	All	56/78 (71%)	26 (46%)	4 (7%)

5 of 26 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
5	G	2	U
5	G	9	G

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
5	G	10	C
5	G	11	G
5	G	14	C

All (4) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
5	G	14	C
5	G	15	A
5	G	26	G
5	G	27	C

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)

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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.

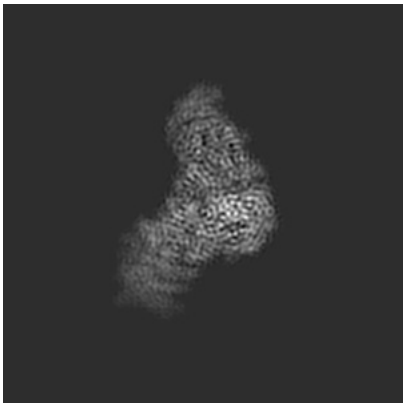
6 Map visualisation (i)

This section contains visualisations of the EMDB entry EMD-9254. 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

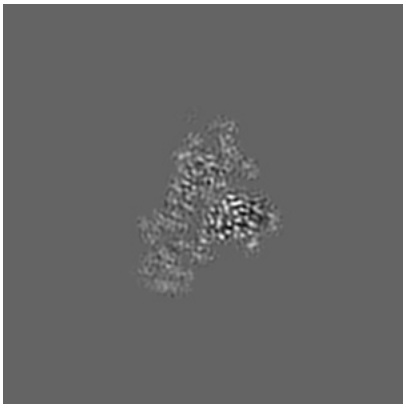
Y

Z

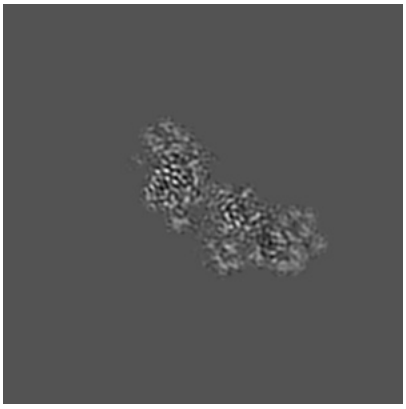
The images above show the map projected in three orthogonal directions.

6.2 Central slices (i)

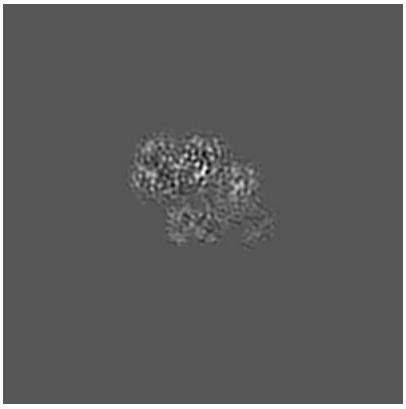
6.2.1 Primary map



X Index: 128



Y Index: 128

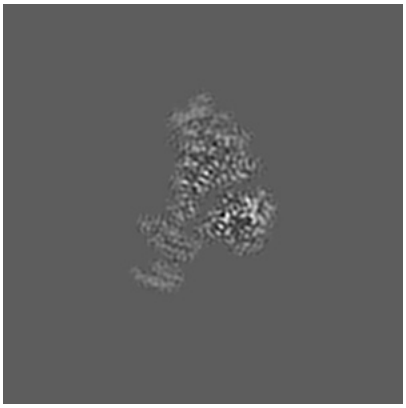


Z Index: 128

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: 121

Y Index: 144

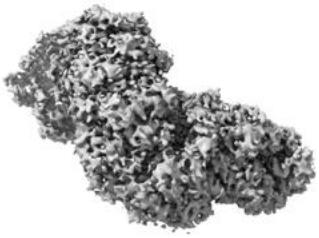
Z Index: 118

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.

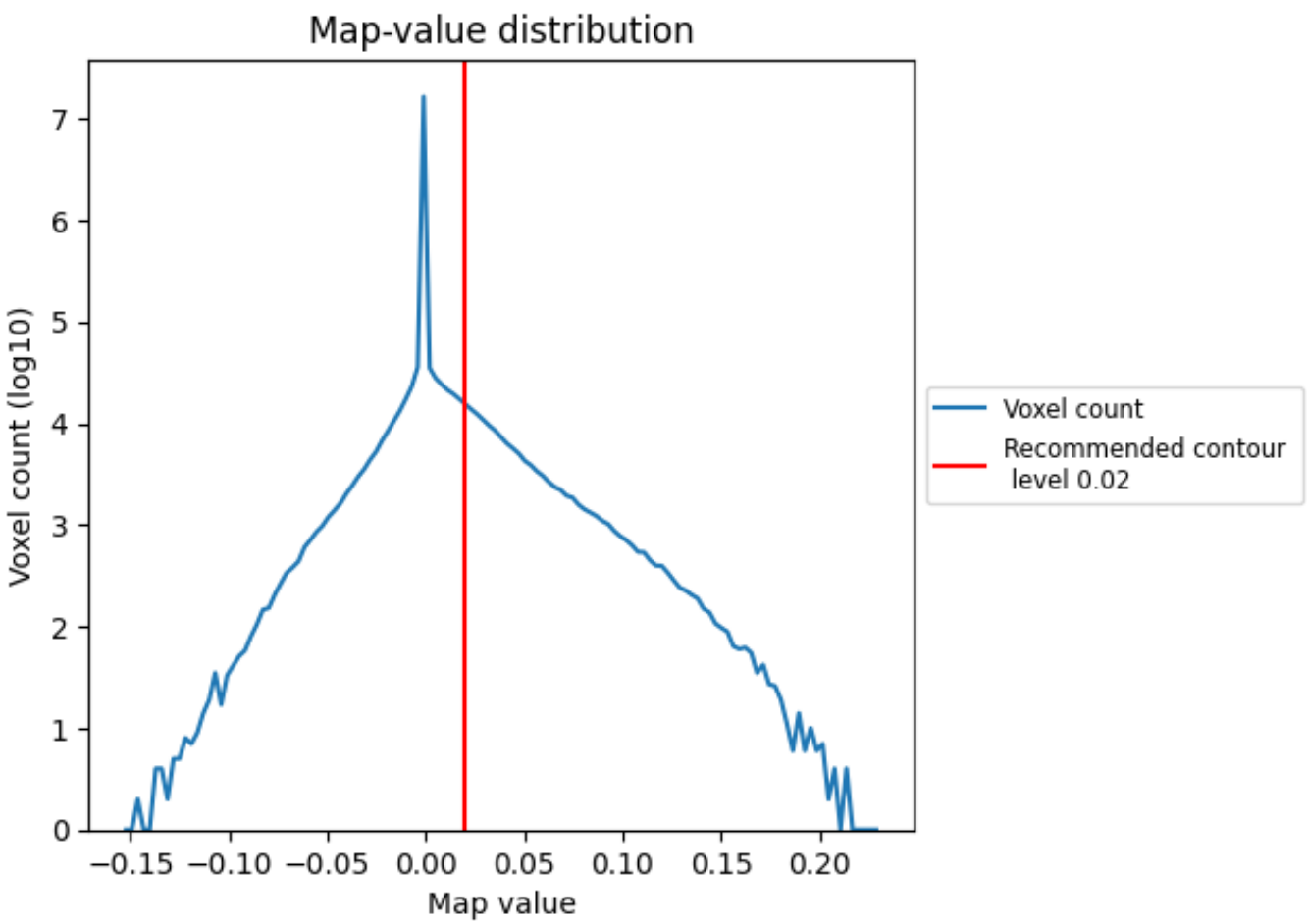
6.5 Mask visualisation (i)

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

7 Map analysis (i)

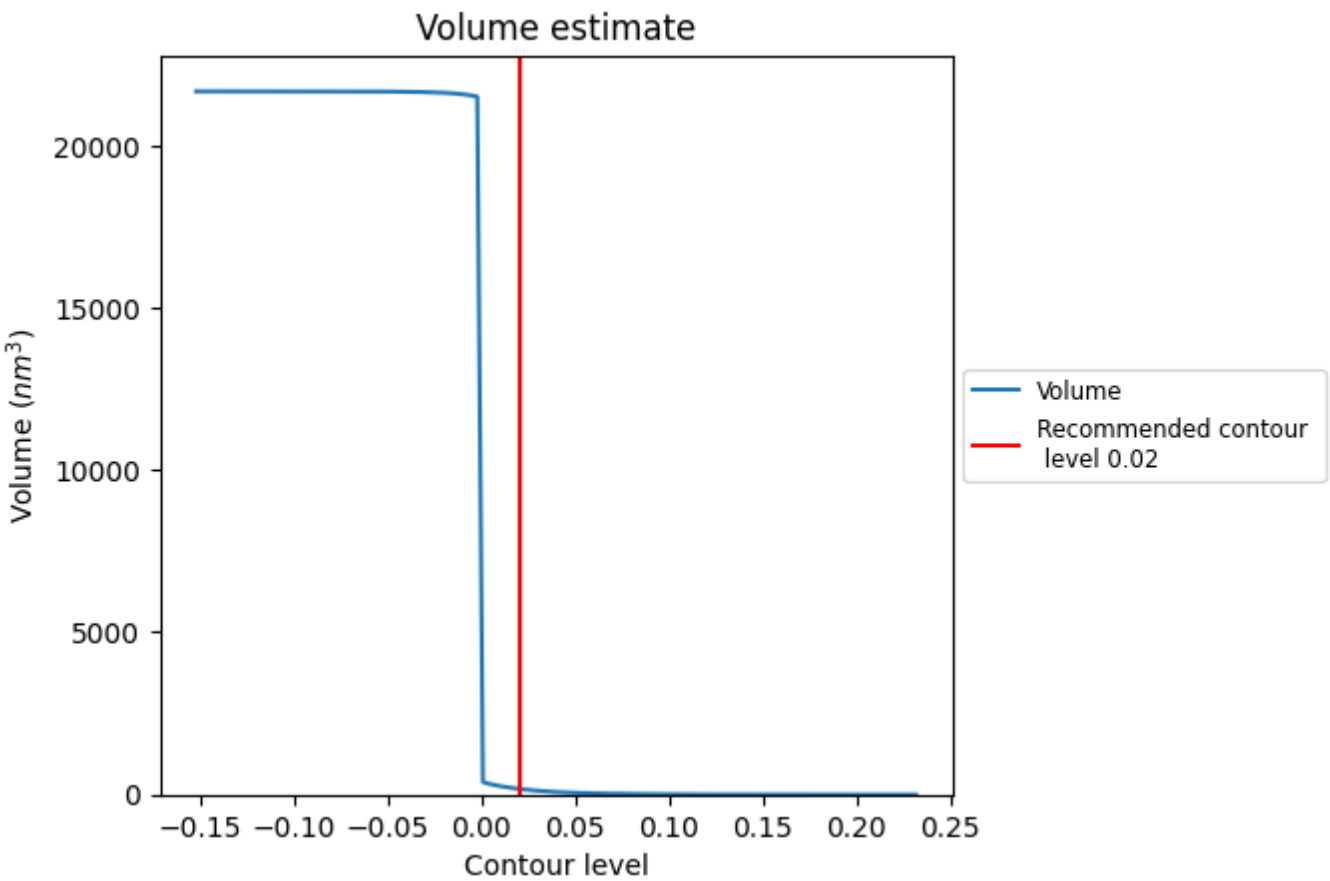
This section contains the results of statistical analysis of the map.

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.

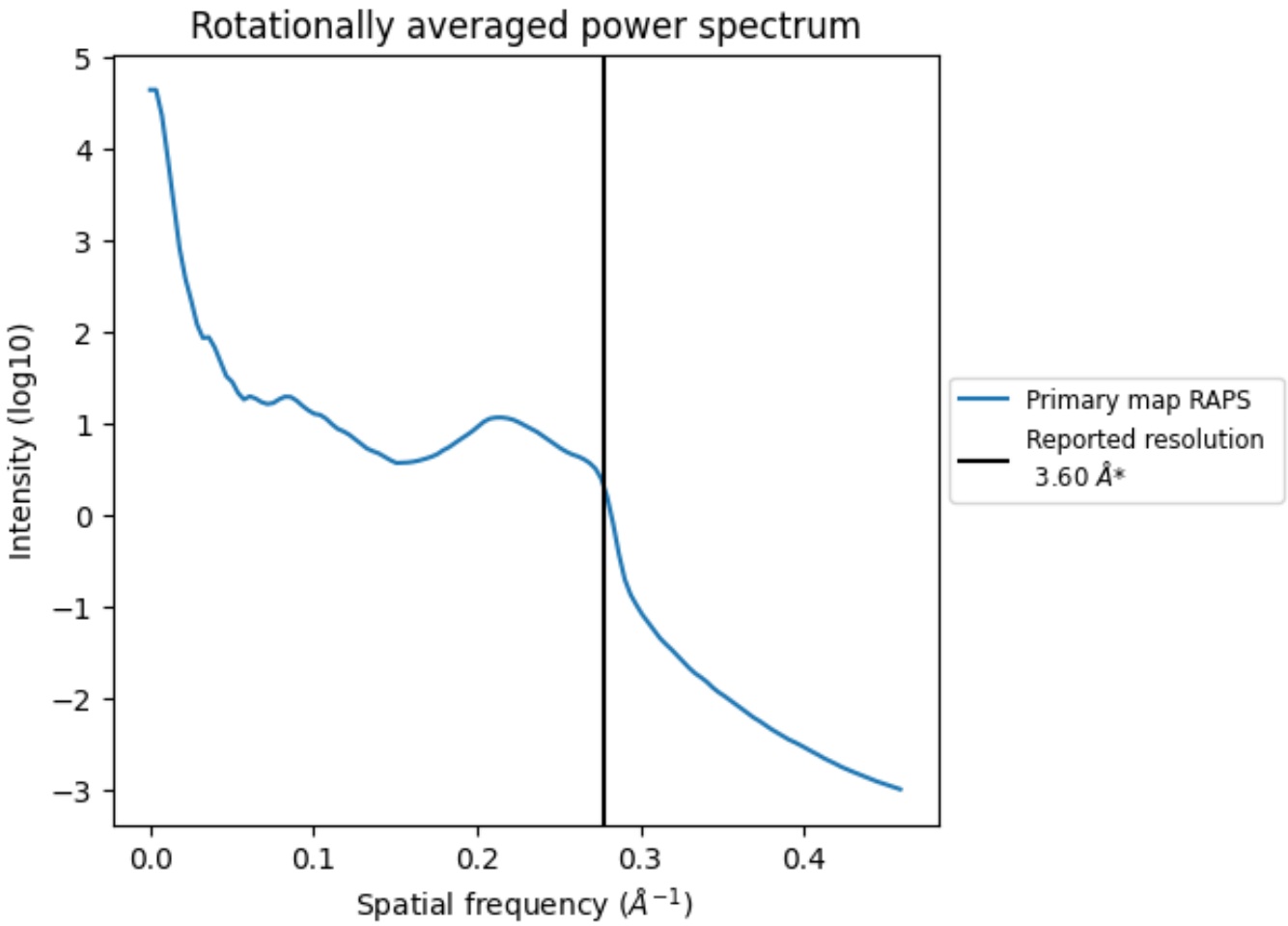
7.2 Volume estimate (i)



The volume at the recommended contour level is 176 nm^3; this corresponds to an approximate mass of 159 kDa.

The volume estimate graph shows how the enclosed volume varies with the contour level. The recommended contour level is shown as a vertical line and the intersection between the line and the curve gives the volume of the enclosed surface at the given level.

7.3 Rotationally averaged power spectrum (i)

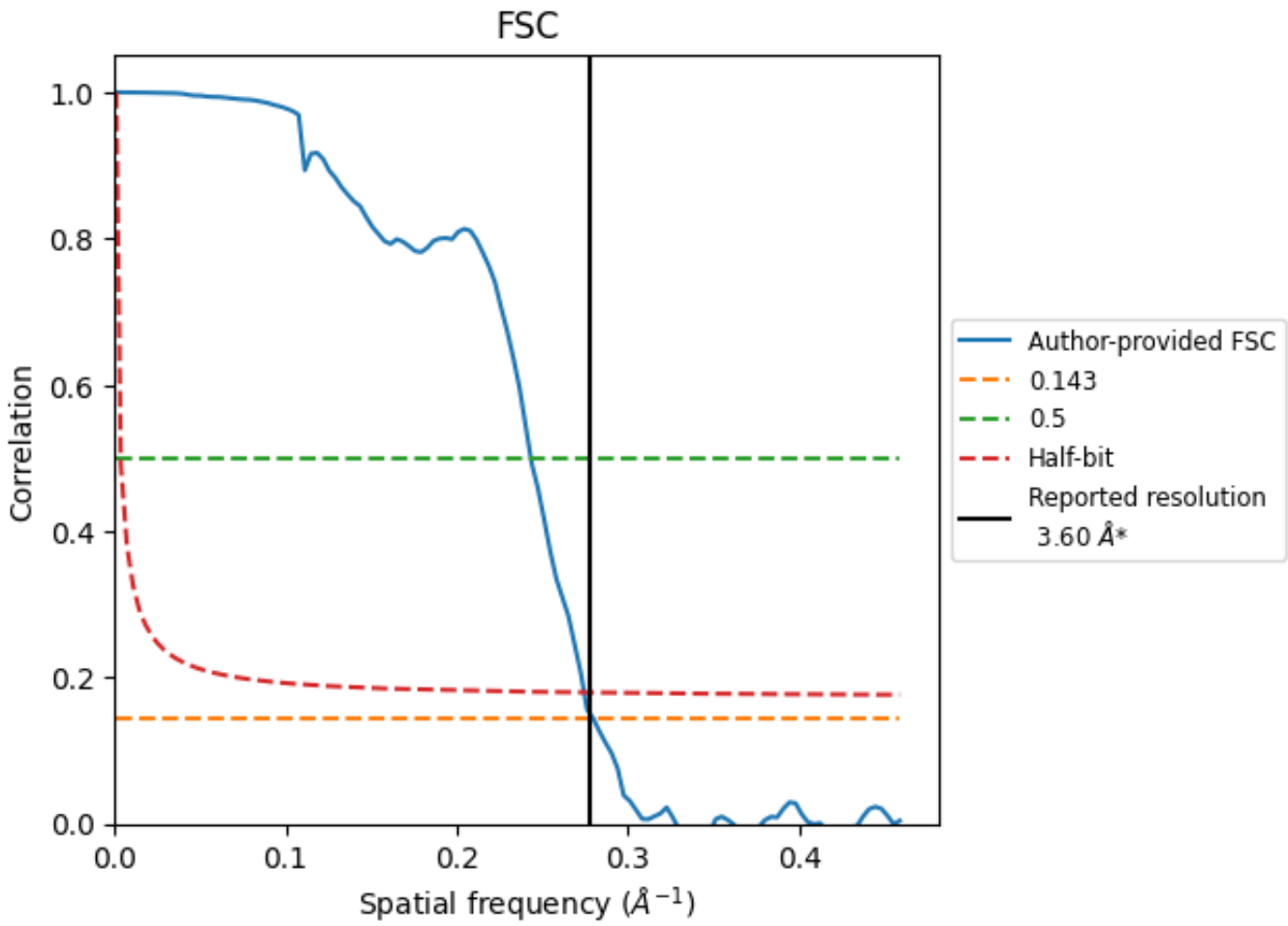


*Reported resolution corresponds to spatial frequency of 0.278 Å^{-1}

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 0.278 Å^{-1}

8.2 Resolution estimates (i)

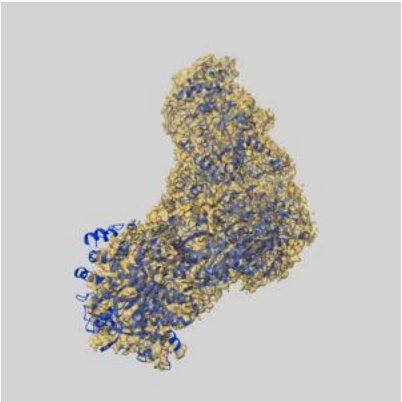
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.60	-	-
Author-provided FSC curve	3.57	4.11	3.64
Unmasked-calculated*	-	-	-

*Resolution estimate based on FSC curve calculated by comparison of deposited half-maps.

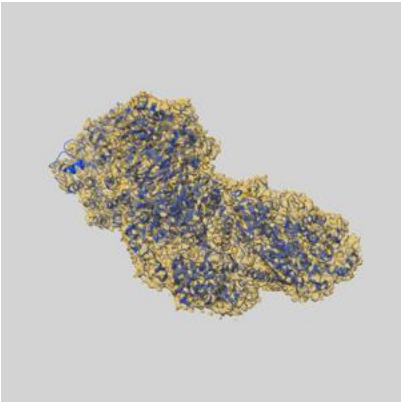
9 Map-model fit (i)

This section contains information regarding the fit between EMDB map EMD-9254 and PDB model 6MUS. Per-residue inclusion information can be found in section 3 on page 7.

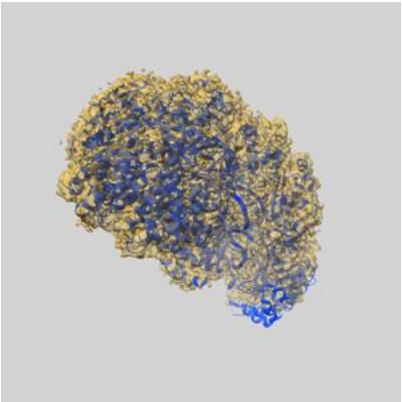
9.1 Map-model overlay (i)



X



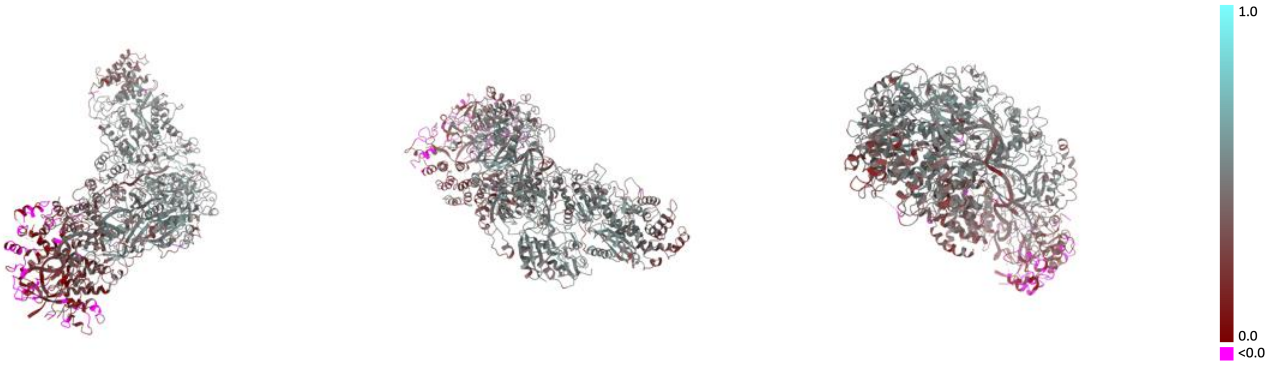
Y



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.

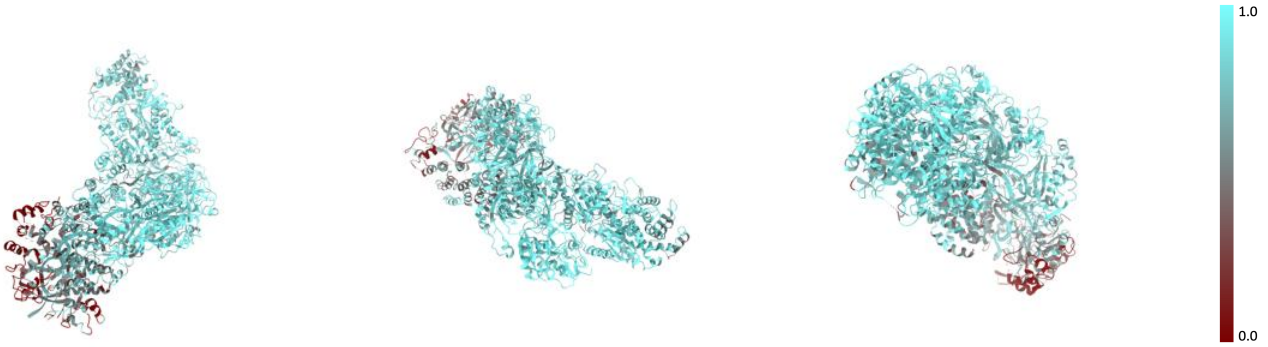
9.2 Q-score mapped to coordinate model (i)



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.

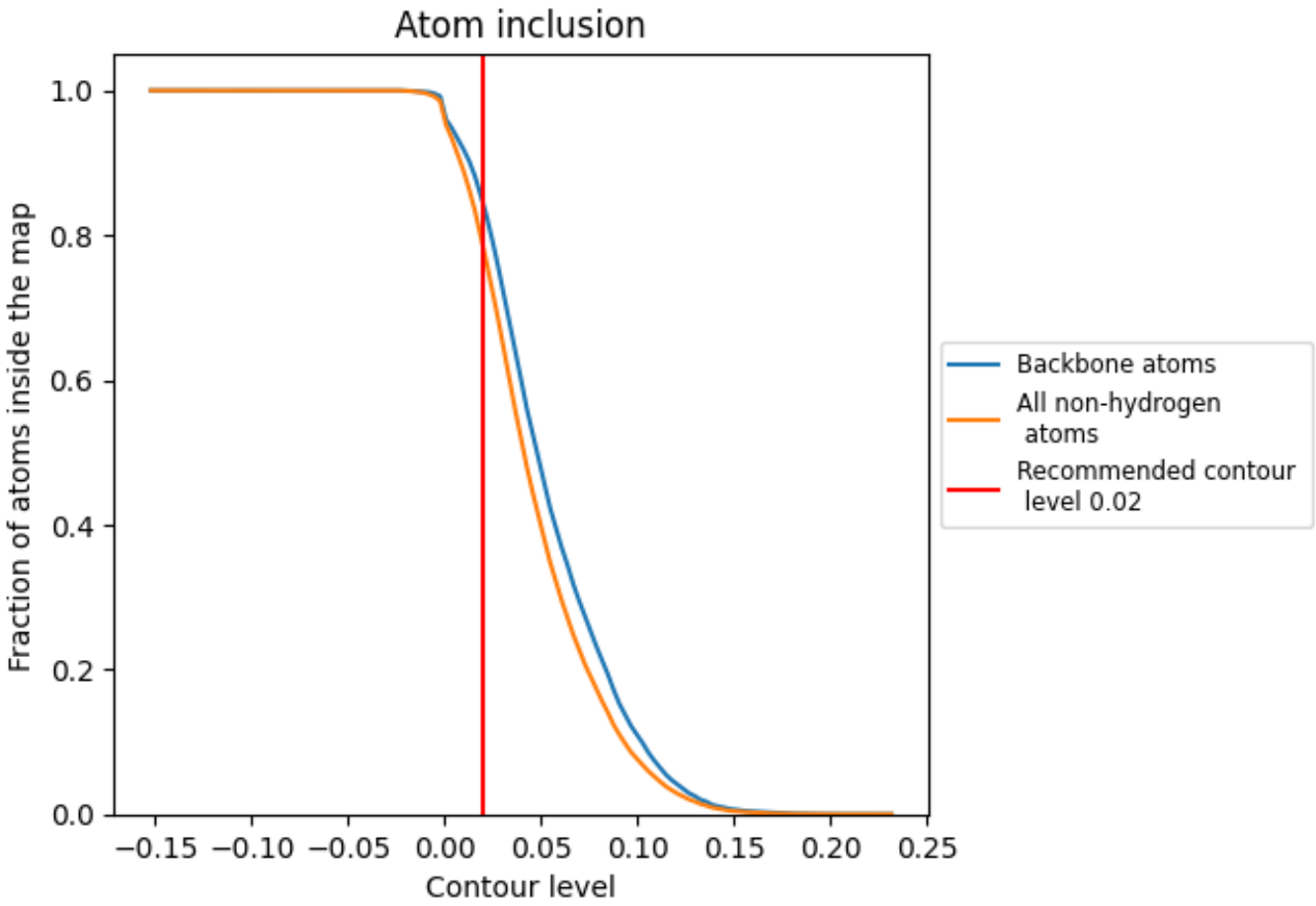
9.3 Atom inclusion mapped to coordinate model (i)



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, 84% of all backbone atoms, 78% 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	0.7842	0.4020
A	0.8544	0.4420
B	0.6468	0.3070
C	0.8937	0.4850
D	0.8904	0.4850
E	0.9269	0.5030
F	0.3893	0.1410
G	0.8605	0.4510
H	0.6813	0.3020
J	0.5886	0.2520
K	0.8450	0.4430