



wwPDB EM Validation Summary Report ⓘ

Jun 9, 2024 – 08:27 PM EDT

PDB ID : 8FYD
EMDB ID : EMD-29565
Title : Cryo-EM structure of Cas1:Cas2-DEDDh:half-site integration complex bent CRISPR repeat conformation
Authors : Skopintsev, P.; Tuck, O.T.; Soczek, K.M.; Doudna, J.
Deposited on : 2023-01-25
Resolution : 3.90 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 ⓘ 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:

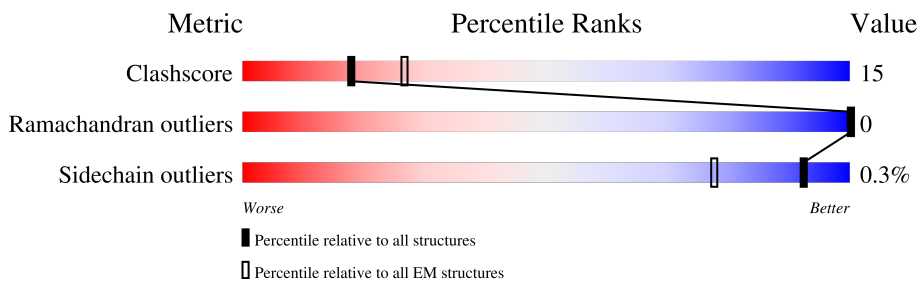
EMDB validation analysis : 0.0.1.dev92
MolProbity : 4.02b-467
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ : 1.9.13
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.36.2

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 3.90 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.



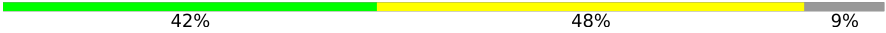

Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

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 $\leq 5\%$. 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	J	78	
2	A	289	
2	D	289	
3	B	316	
3	C	316	
3	E	316	
3	F	316	
4	G	64	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	H	33	 42% 48% 9%
6	I	44	 16% 14% 70%

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 13612 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 DNA chain called DNA (49-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
1	J	49	1005	476	190	291	48	0	0

- Molecule 2 is a protein called Cas2-DEDDh.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	A	93	746	473	129	138	6	0	0
2	D	93	746	473	129	138	6	0	0

- Molecule 3 is a protein called Cas1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	B	292	2295	1454	415	421	5	0	0
3	C	278	2163	1366	400	392	5	0	0
3	E	288	2261	1431	411	414	5	0	0
3	F	281	2188	1383	403	397	5	0	0

- Molecule 4 is a DNA chain called DNA (64-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
4	G	64	1330	631	245	390	64	1	0

- Molecule 5 is a DNA chain called DNA (30-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
5	H	30	613	293	112	179	29	0	0

- Molecule 6 is a DNA chain called DNA (13-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
6	I	13	265	127	47	78	13	0	0

LEU ARG
GLN ALA
ILE PHE
LYS LYS
TYR TYR
GLY GLY
GLU LEU
TYR TYR
LYS LYS
LEU LEU
ASN ASN
PHE PHE

LYS
CYS
LEU
ALA
ALA
LYS
LYS
LEU
ASN
PHE

● Molecule 3: Cas1

Chain B:  65% 28% 8%

MET
ALA
GLY
PRO
ILE
ILE
ALA
GLY
SER
SER
SER
GLU
LEU
PRO
THR
VAL
GLU
D20
I25
I25
H29
K31
I32
N33
R34
V35
D36
A42
V48
R49
I50
V57
L58
L59
L60
G61
G62
G63
T64
I66
D65
S67
H68
V71
E72
L73
L74
G78
T79
A80
L81

V82
W83
V89
R90
Y91
Y92
A93
S94
C95
A99
T113
S117
Y229
K230
V121
A122
R123
R124
M125
Y126
R129
F130
S136
K137
L138
T139
M140
L143
L144
E147
R150
V151
R152
R153
R156
S159
K160
I66
V164
K167
K168
R169
V170
Q184
A185
H190
A80
VAL

L196
H197
H198
G209
V213
H214
I222
Y223
D224
L228
Y229
K230
G248
I251
R256
L257
R260
L269
K270
V273
Q277
L278
L279
I288
E289
A290
E291
P292
L293
S294
L295
V296
D297
K298
K299
E300
K301
Y309
S310
F311
VAL
THR
SER
CYS
PRO

● Molecule 3: Cas1

Chain C:  59% 29% 12%

MET
A2
K9
S10
P16
E19
R21
I25
E28
H29
A30
K31
I32
K33
R34
V41
A42
E43
V48
P51
A52
A53
H54
I55
L59
L60
G61
T64
H68
R69
A70
V71
E72
D76
L81
W82
W83
W84
Q87
G88
W89
R90
Y91
Y92
A93

R103
F104
L105
L111
V112
T113
M114
R116
S117
R118
L119
R120
R123
A124
M125
Y126
Q127
M128
V135
S136
K137
L138
T139
M140
Q141
Q142
L143
R144
S145
H146
R150
Y155
L158
S159
K160
W166
K167
Y171
N172
P173
D174
D175
F176
A177
G178
P181
L182
H183
Q184

H190
V191
Y194
G195
L196
W201
L206
L210
W213
I222
Y223
D224
D227
L228
Y229
K230
I233
T234
Y235
P236
F239
R260
F263
I268
L269
L279
LEU
GLU
ILE
PRO
GLU
GLY
GLN
ILE
GLU
ALA
GLU
PRD
L74
P176
SER
LEU
SER
TRP
ASP
ASP
LYS

GLU
LYS
LEU
VAL
TYR
GLY
VAL
ASN
TYR
SER
GLU
VAL
THR
SER
CYS
PRO

● Molecule 3: Cas1

Chain E:  61% 30% 9%

MET
ALA
GLY
PRO
ILE
ILE
ALA
GLY
SER
SER
SER
GLU
LEU
PRO
ARG
VAL
VAL
D20
R21
K31
I32
N33
R34
V35
D36
S37
T40
V41
V48
R49
I50
I55
L58
L59
L60
T64
D65
I66
S67
H68
R69
E72
L73
L74
G75
D76
T77
G78
T79
A80
L81

Q87
G88
R90
R96
A97
A99
L105
W106
K107
Q108
A109
R110
L111
V112
T113
S117
R118
L119
Q127
N128
R129
PHE
PRO
THR
GLU
D134
V135
T139
M140
Q141
Q142
S145
R150
V151
R152
K154
L158
K161
R169
F176
N183
Q184
V191

A192
L193
Y194
H198
V201
L206
S207
P208
V213
H214
D218
R219
S220
F221
I222
Y223
D224
Y229
K230
A231
E232
I233
T234
Q253
R256
L257
R260
V264
R271
M272
V273
Q277
T278
L279
Q287
I288
E289
A290
S294
L295
V296
D297
D298
K299
V303

E311
VAL
THR
SER
CYS
PRO

4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	53545	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TALOS ARCTICA	Depositor
Voltage (kV)	200	Depositor
Electron dose ($e^-/\text{\AA}^2$)	50	Depositor
Minimum defocus (nm)	800	Depositor
Maximum defocus (nm)	2200	Depositor
Magnification	Not provided	
Image detector	GATAN K3 (6k x 4k)	Depositor
Maximum map value	0.562	Depositor
Minimum map value	-0.369	Depositor
Average map value	-0.001	Depositor
Map value standard deviation	0.018	Depositor
Recommended contour level	0.05	Depositor
Map size (Å)	267.6, 267.6, 267.6	wwPDB
Map dimensions	240, 240, 240	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.115, 1.115, 1.115	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	J	0.53	0/1128	0.86	0/1740
2	A	0.26	0/761	0.56	0/1029
2	D	0.26	0/761	0.55	0/1029
3	B	0.25	0/2338	0.51	0/3164
3	C	0.25	0/2201	0.55	0/2975
3	E	0.25	0/2301	0.51	0/3111
3	F	0.25	0/2226	0.55	0/3009
4	G	0.53	0/1491	0.88	0/2300
5	H	0.50	0/687	0.84	0/1059
6	I	0.52	0/296	0.93	0/454
All	All	0.34	0/14190	0.65	0/19870

There are no bond length outliers.

There are no bond angle outliers.

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	J	1005	0	550	45	0
2	A	746	0	740	21	0
2	D	746	0	740	27	0
3	B	2295	0	2325	67	0
3	C	2163	0	2212	77	0
3	E	2261	0	2295	69	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	F	2188	0	2240	79	0
4	G	1330	0	731	32	0
5	H	613	0	340	17	0
6	I	265	0	148	6	0
All	All	13612	0	12321	380	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 15.

The worst 5 of 380 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:D:17:GLY:O	2:D:20:THR:HG22	1.31	1.29
1:J:35:DA:H2''	1:J:36:DC:H5'	1.58	0.84
3:F:229:TYR:HE1	3:F:271:ARG:NH2	1.75	0.84
3:E:87:GLN:HE22	3:E:260:ARG:HG3	1.42	0.83
3:F:87:GLN:HE22	3:F:90:ARG:HD3	1.46	0.81

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	
2	A	91/289 (32%)	88 (97%)	3 (3%)	0	100	100
2	D	91/289 (32%)	91 (100%)	0	0	100	100
3	B	290/316 (92%)	279 (96%)	11 (4%)	0	100	100
3	C	276/316 (87%)	269 (98%)	7 (2%)	0	100	100
3	E	284/316 (90%)	268 (94%)	16 (6%)	0	100	100
3	F	279/316 (88%)	271 (97%)	8 (3%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	1311/1842 (71%)	1266 (97%)	45 (3%)	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	
2	A	80/245 (33%)	80 (100%)	0	100	100
2	D	80/245 (33%)	80 (100%)	0	100	100
3	B	238/258 (92%)	237 (100%)	1 (0%)	91	94
3	C	223/258 (86%)	222 (100%)	1 (0%)	91	94
3	E	234/258 (91%)	234 (100%)	0	100	100
3	F	226/258 (88%)	225 (100%)	1 (0%)	91	94
All	All	1081/1522 (71%)	1078 (100%)	3 (0%)	92	95

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	B	230	LYS
3	C	160	LYS
3	F	49	ARG

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

Mol	Chain	Res	Type
3	B	184	GLN
3	B	198	HIS
3	C	190	HIS
3	E	87	GLN
3	F	87	GLN

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.

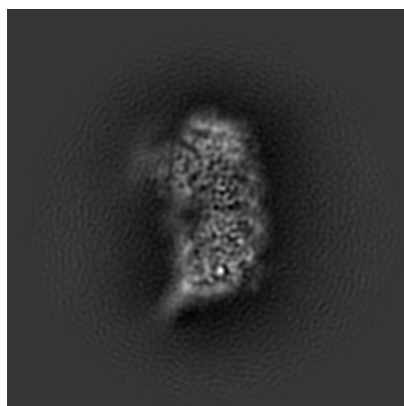
6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-29565. These allow visual inspection of the internal detail of the map and identification of artifacts.

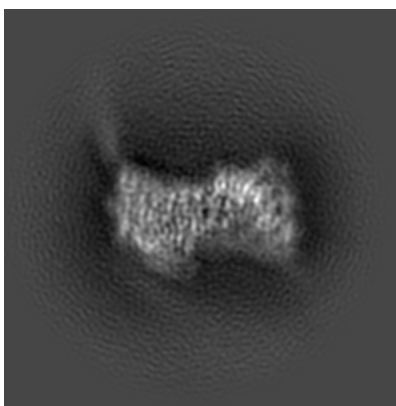
Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

6.1 Orthogonal projections [i](#)

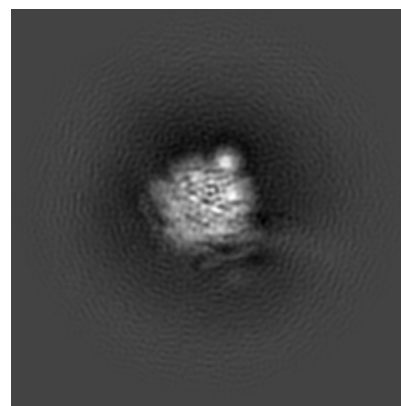
6.1.1 Primary map



X

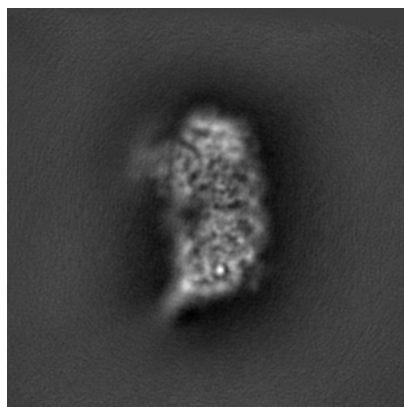


Y

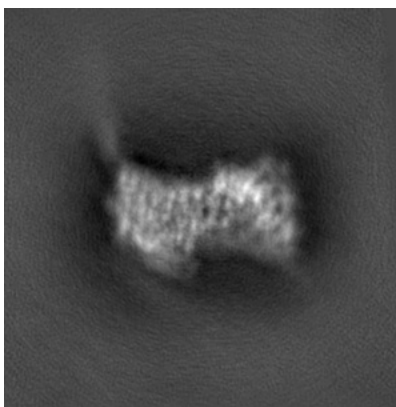


Z

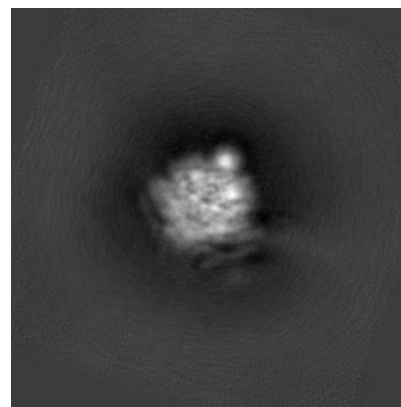
6.1.2 Raw map



X



Y

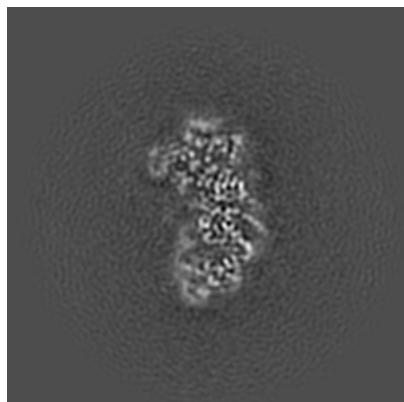


Z

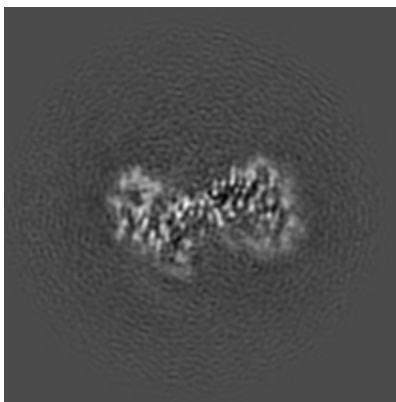
The images above show the map projected in three orthogonal directions.

6.2 Central slices [i](#)

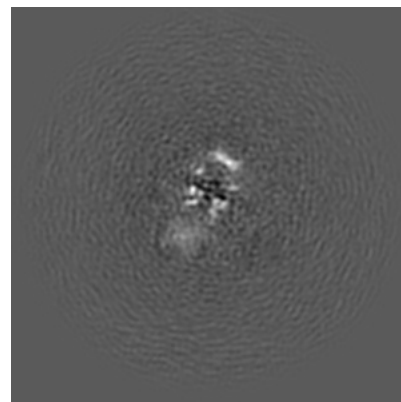
6.2.1 Primary map



X Index: 120

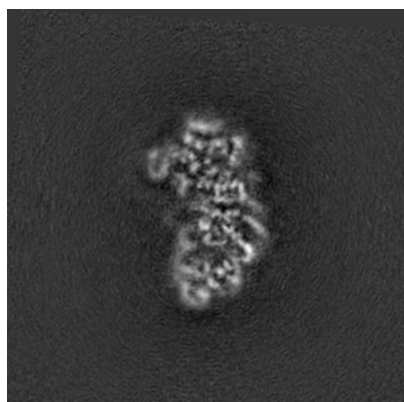


Y Index: 120

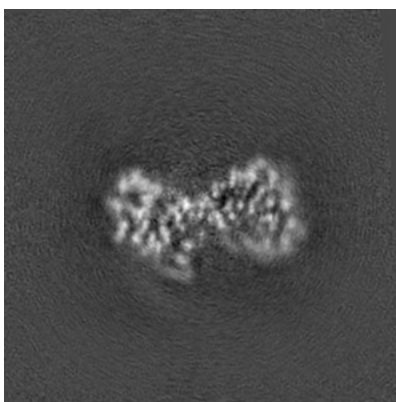


Z Index: 120

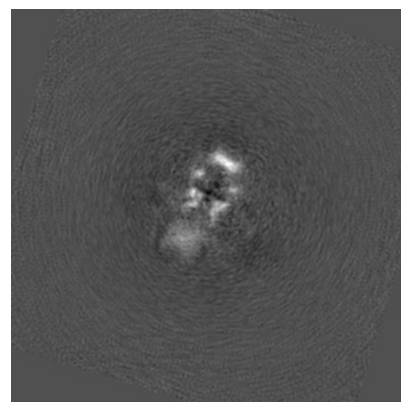
6.2.2 Raw map



X Index: 120



Y Index: 120

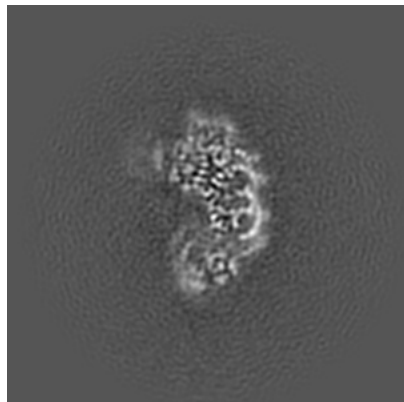


Z Index: 120

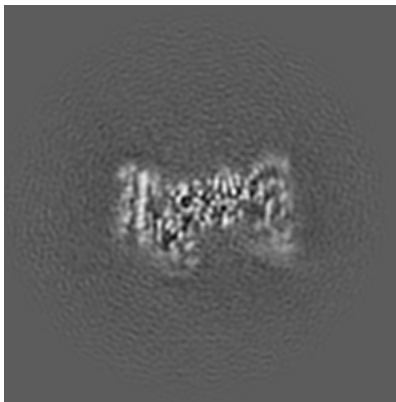
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices [i](#)

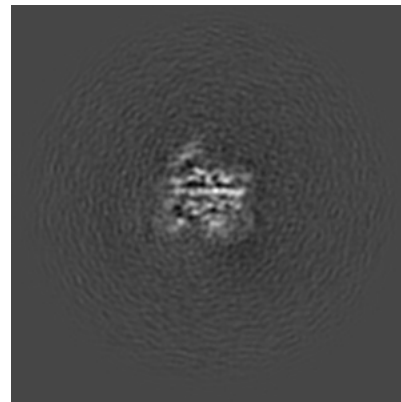
6.3.1 Primary map



X Index: 128

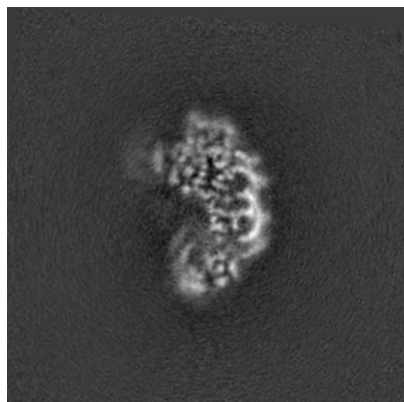


Y Index: 127

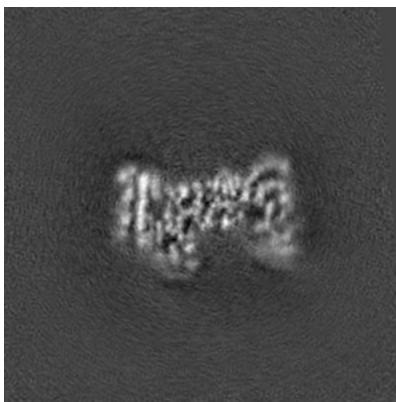


Z Index: 83

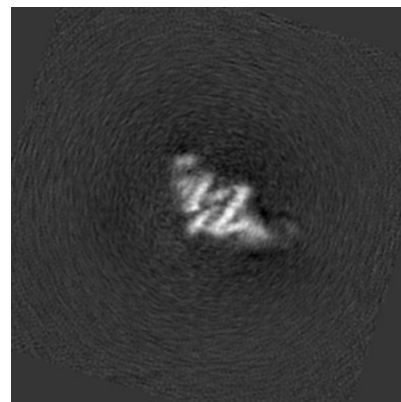
6.3.2 Raw map



X Index: 128



Y Index: 127

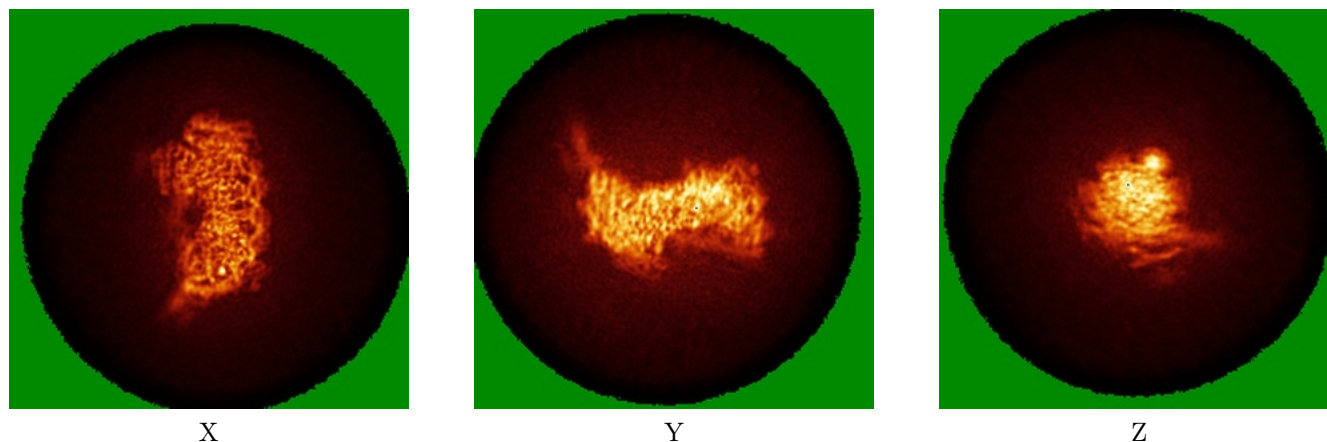


Z Index: 72

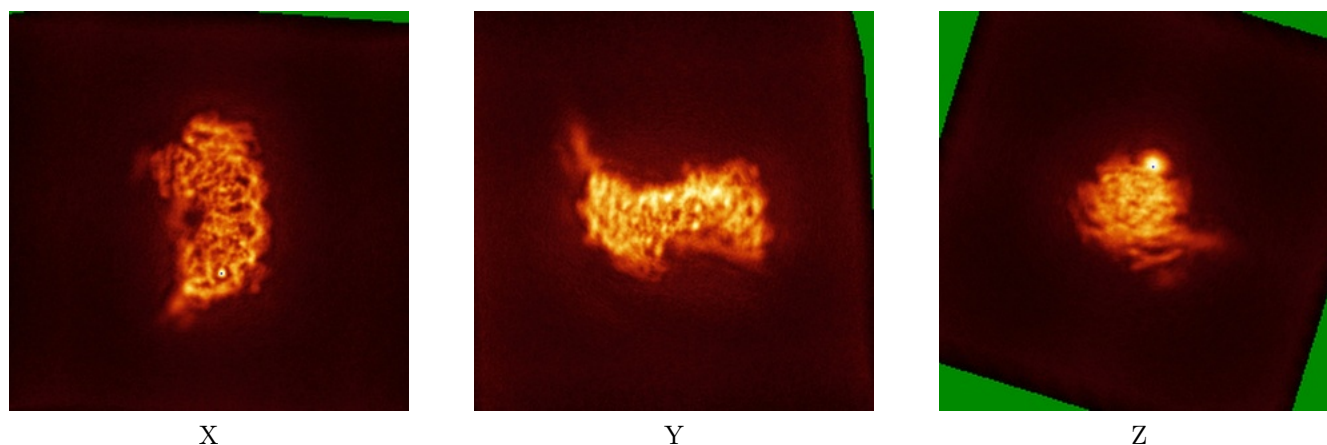
The images above show the largest variance slices of the map in three orthogonal directions.

6.4 Orthogonal standard-deviation projections (False-color) [i](#)

6.4.1 Primary map



6.4.2 Raw map



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



X



Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.05. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

6.5.2 Raw map



X



Y



Z

These images show the 3D surface of the raw map. The raw map's contour level was selected so that its surface encloses the same volume as the primary map does at its recommended contour level.

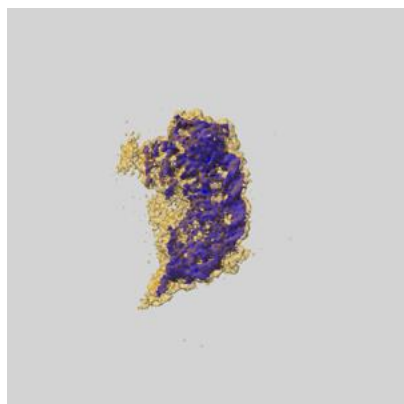
6.6 Mask visualisation [i](#)

This section shows the 3D surface view of the primary map at 50% transparency overlaid with the specified mask at 0% transparency

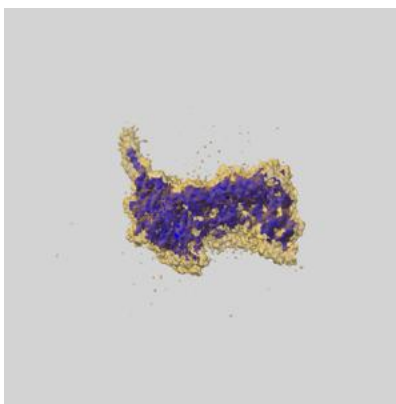
A mask typically either:

- Encompasses the whole structure
- Separates out a domain, a functional unit, a monomer or an area of interest from a larger structure

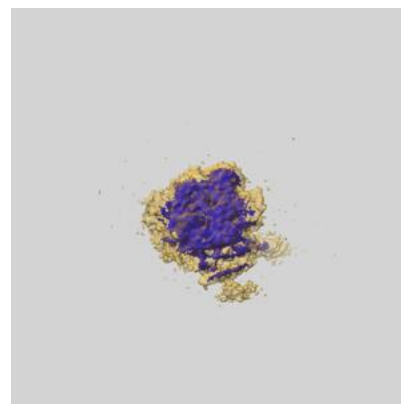
6.6.1 emd_29565_msk_1.map [i](#)



X



Y

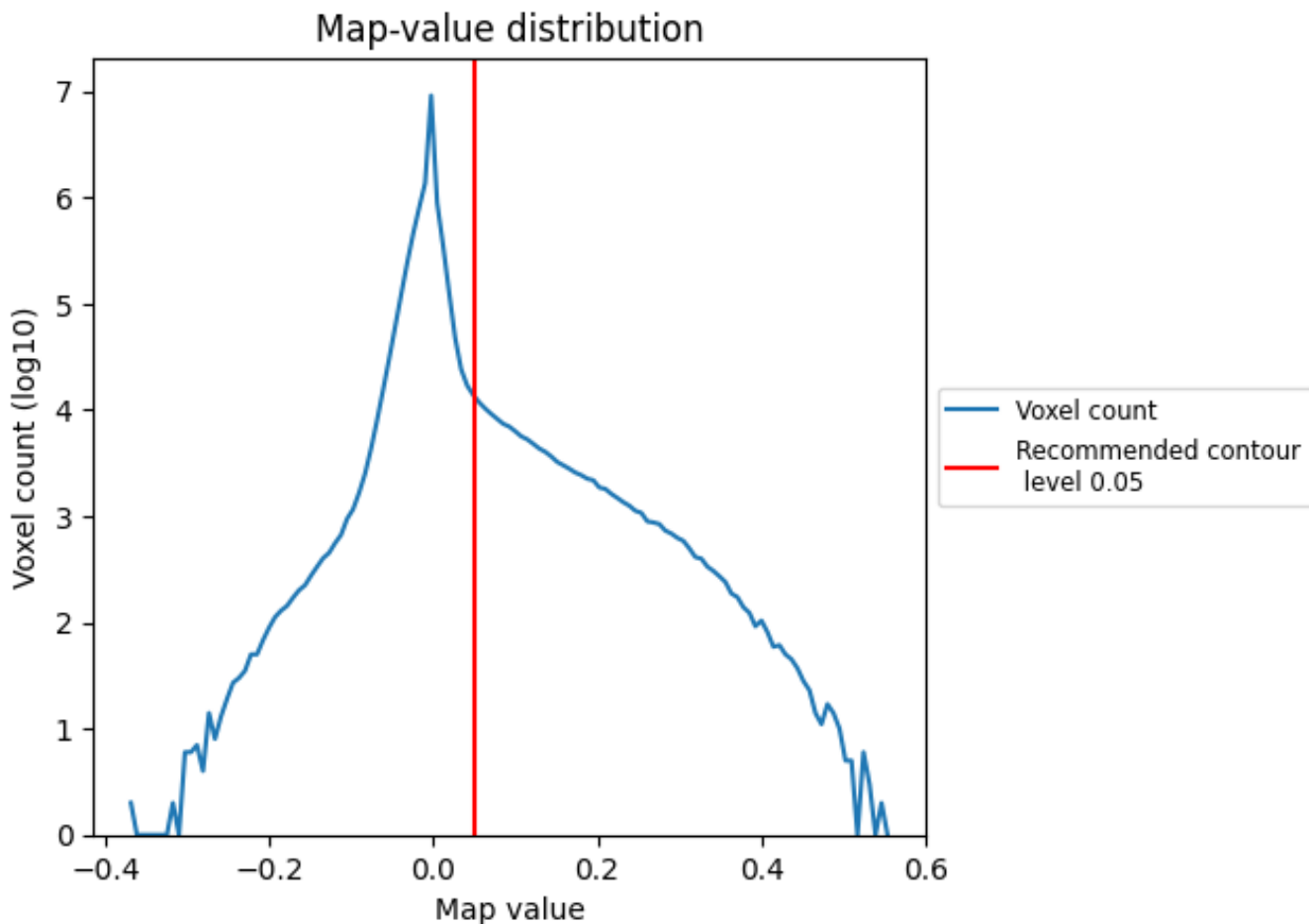


Z

7 Map analysis [i](#)

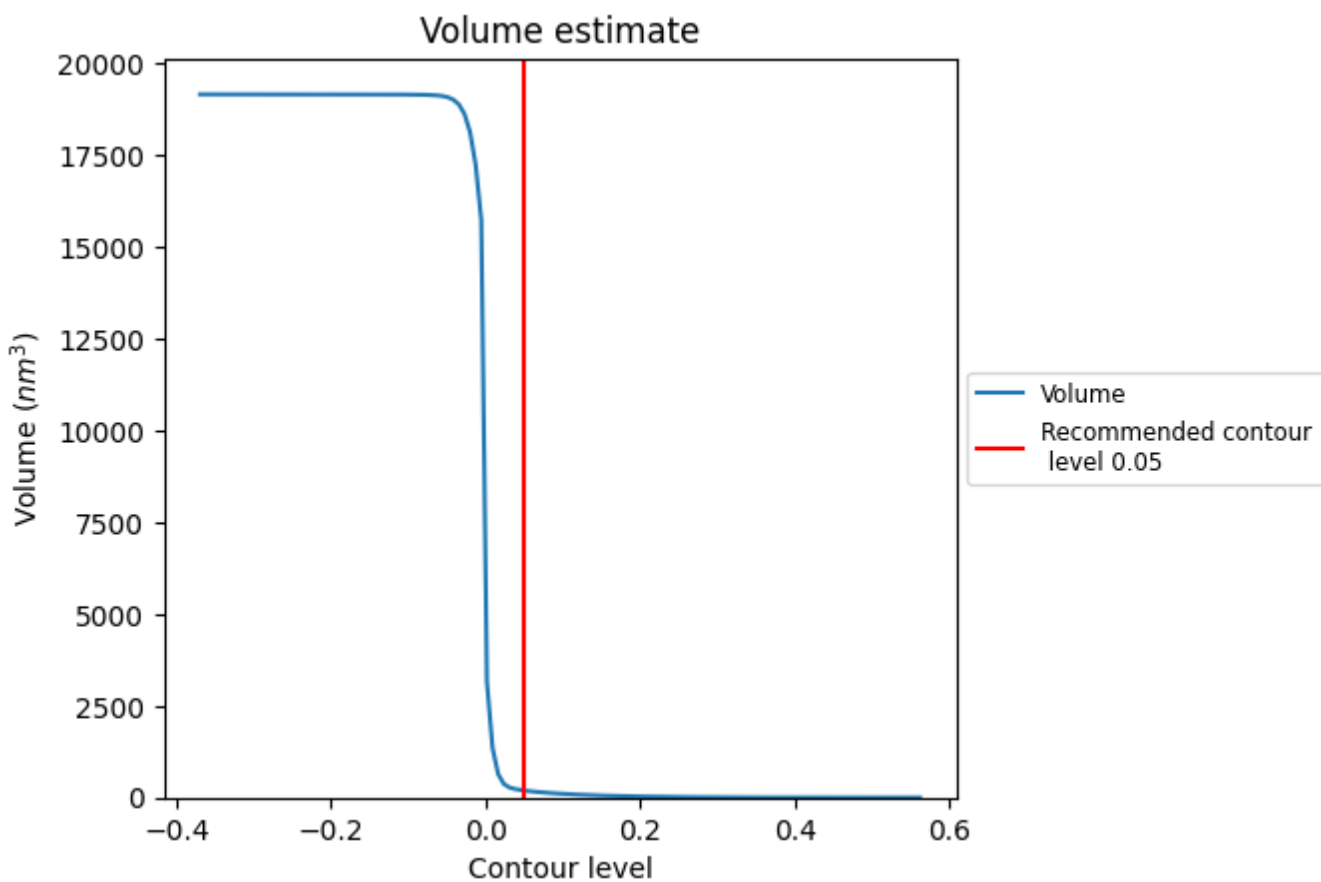
This section contains the results of statistical analysis of the map.

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.

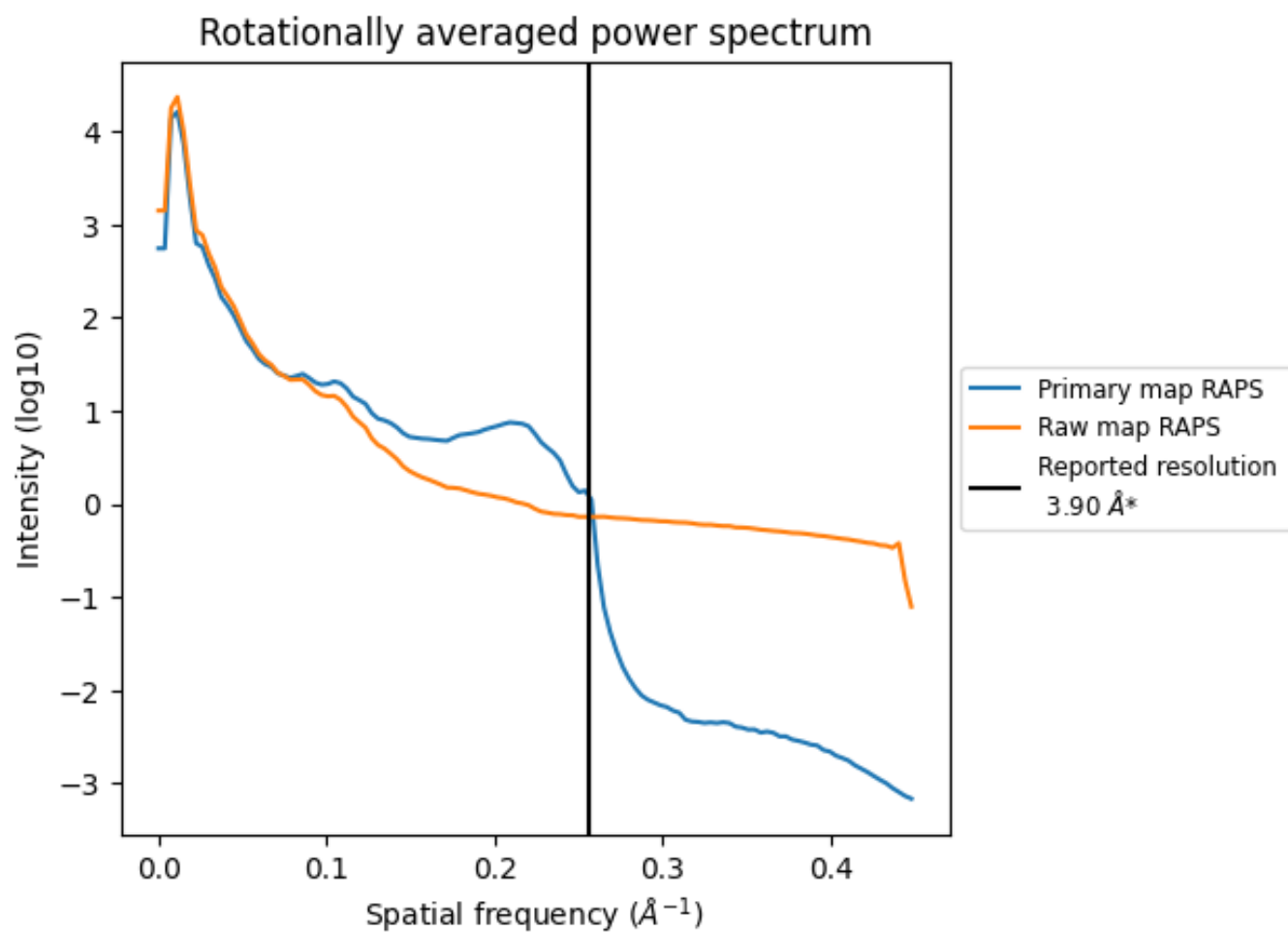
7.2 Volume estimate [\(i\)](#)



The volume at the recommended contour level is 192 nm³; this corresponds to an approximate mass of 173 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

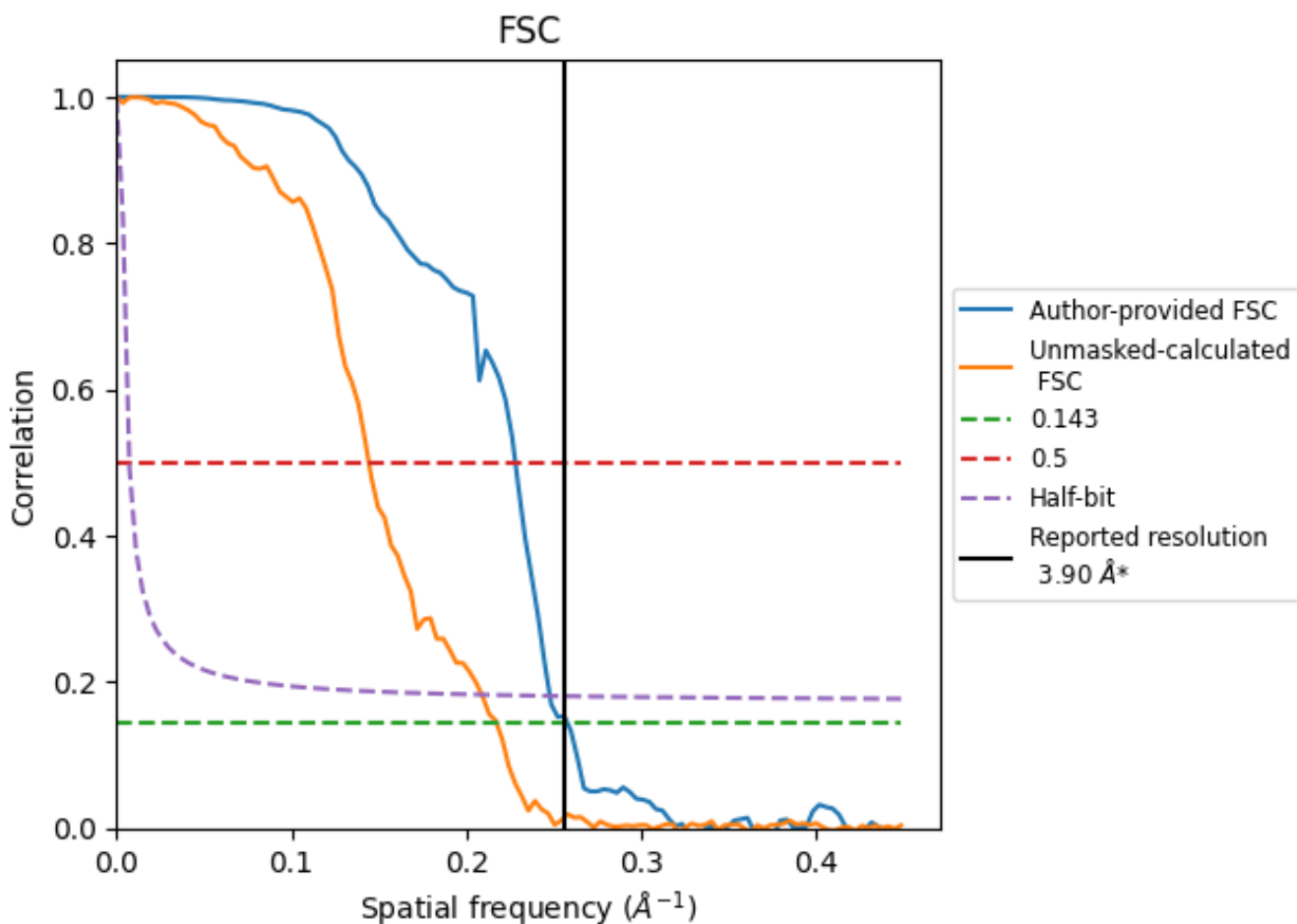


*Reported resolution corresponds to spatial frequency of 0.256 Å⁻¹

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 0.256 Å⁻¹

8.2 Resolution estimates [i](#)

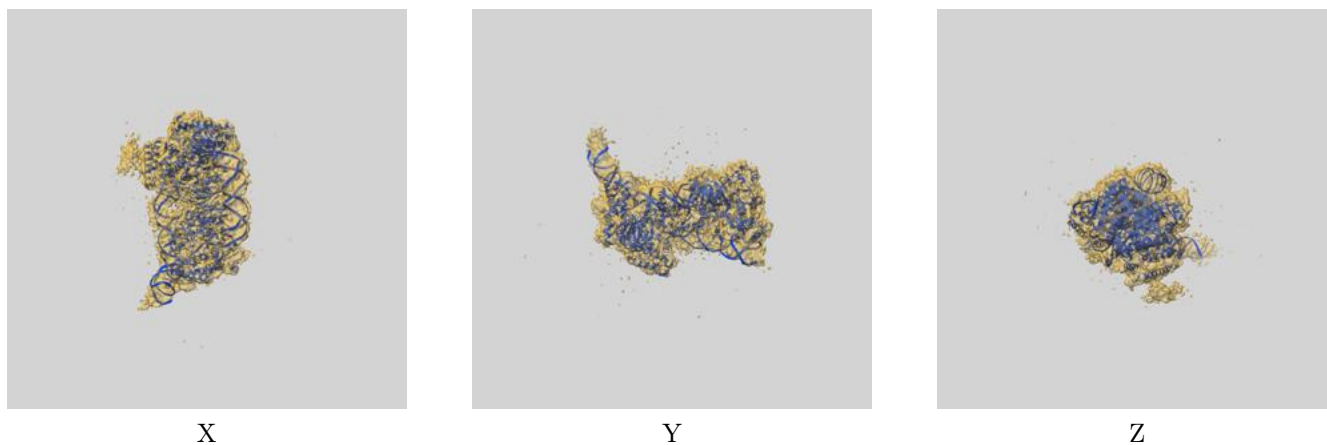
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.90	-	-
Author-provided FSC curve	3.88	4.39	4.04
Unmasked-calculated*	4.60	6.93	4.79

*Resolution estimate based on FSC curve calculated by comparison of deposited half-maps. The value from deposited half-maps intersecting FSC 0.143 CUT-OFF 4.60 differs from the reported value 3.9 by more than 10 %

9 Map-model fit [i](#)

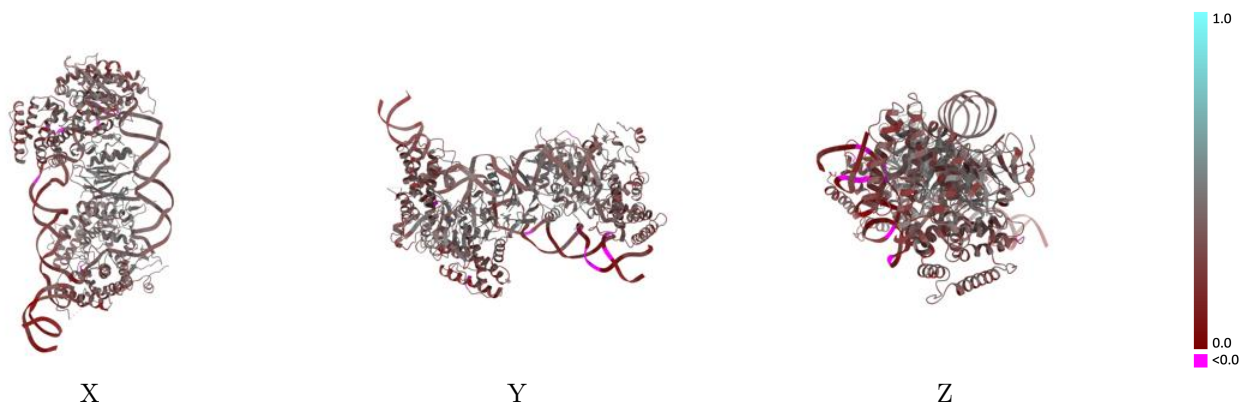
This section contains information regarding the fit between EMDB map EMD-29565 and PDB model 8FYD. Per-residue inclusion information can be found in section 3 on page 6.

9.1 Map-model overlay [i](#)



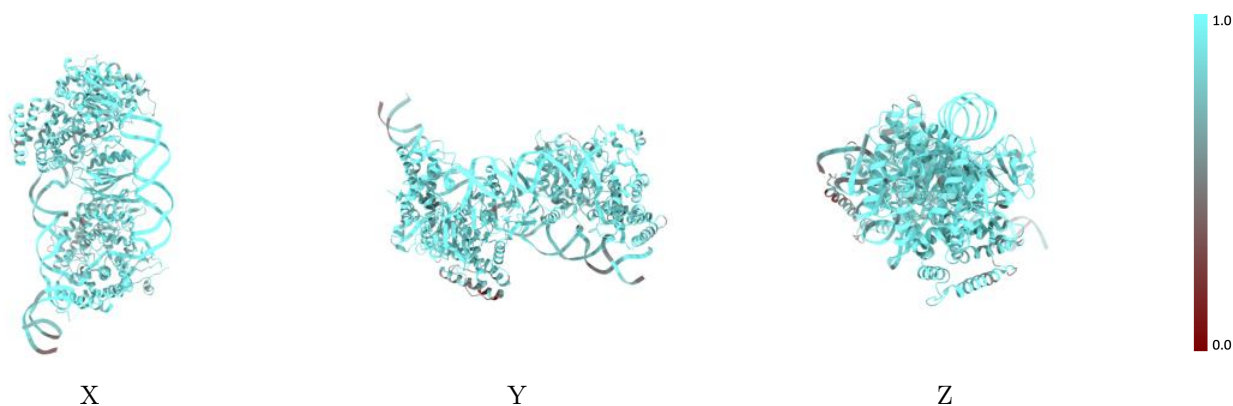
The images above show the 3D surface view of the map at the recommended contour level 0.05 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

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



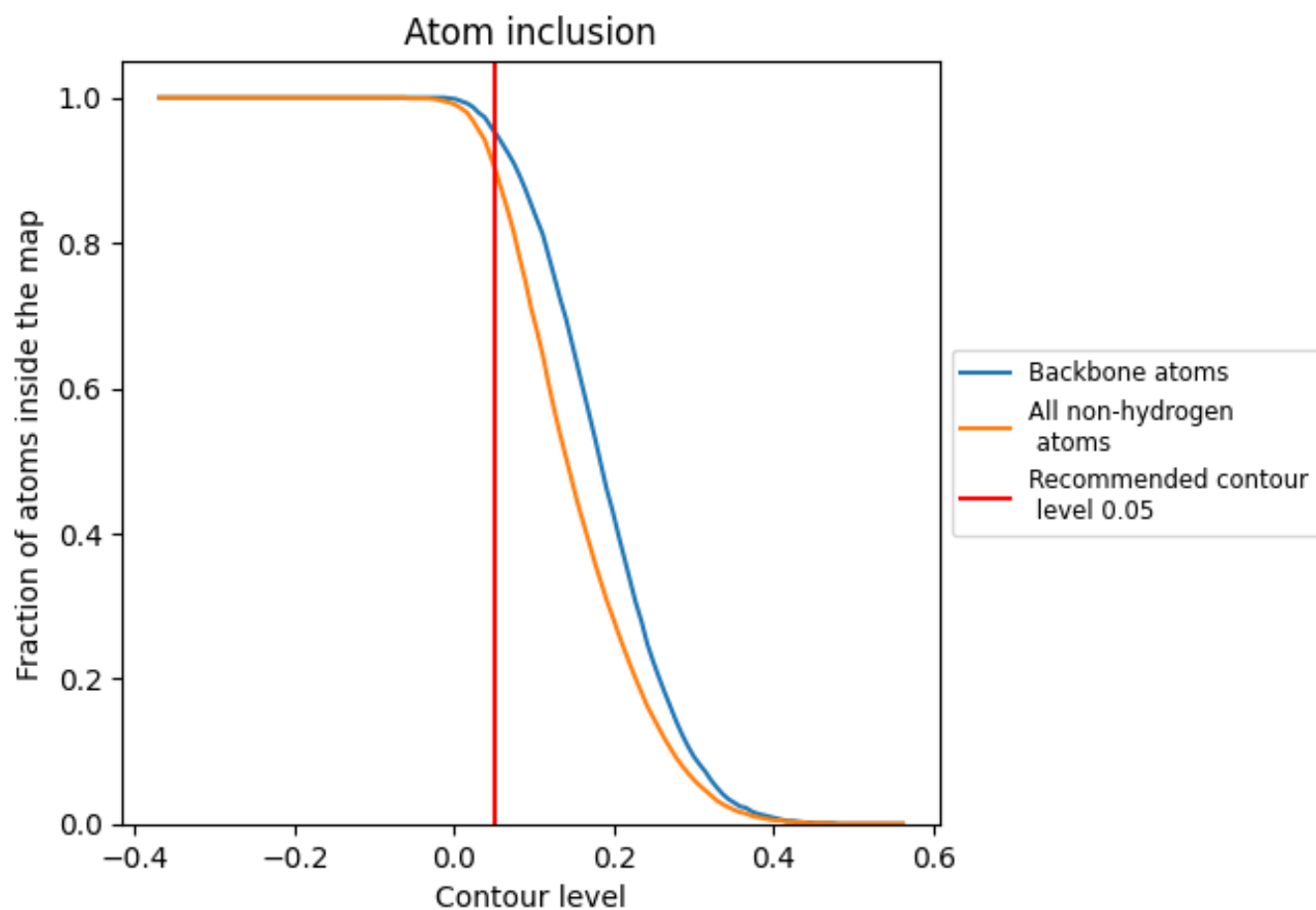
The images above show the model with each residue coloured according to its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions 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 (0.05).























9.4 Atom inclusion [i](#)



At the recommended contour level, 95% of all backbone atoms, 90% 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 (0.05) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	 0.9050	 0.3380
A	 0.9490	 0.4370
B	 0.9300	 0.3450
C	 0.9020	 0.3710
D	 0.9540	 0.4340
E	 0.9340	 0.3550
F	 0.8580	 0.3450
G	 0.8850	 0.2480
H	 0.9940	 0.3770
I	 0.7360	 0.1860
J	 0.8370	 0.1850

