W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB EM Validation Summary Report (i)

May 17, 2025 – 09:06 PM EDT

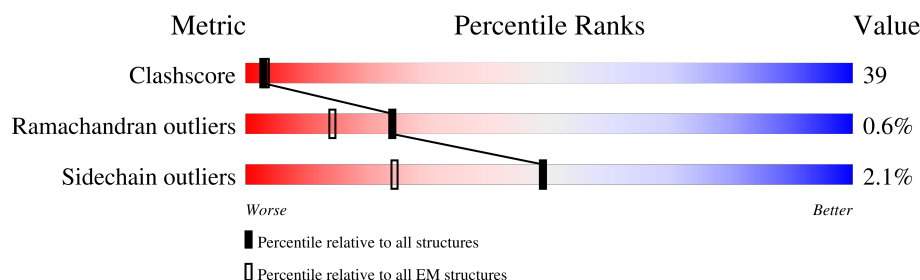|		
PDB ID	:	7MI9 / pdb_00007mi9
EMDB ID	:	EMD-23843
Title	:	Full integration complex of Cas1/Cas2 from Cas4-containing system
Authors	:	Hu, C.Y.; Ke, A.K.
Deposited on	:	2021-04-16
Resolution	:	3.89 Å(reported)

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 3.89 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	Percentile Ranks	Value
Clashscore		39
Ramachandran outliers		0.6%
Sidechain outliers		2.1%

Worse   *Better*

▌ Percentile relative to all structures

▯ Percentile relative to all EM structures

Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	210492	15764
Ramachandran outliers	207382	16835
Sidechain outliers	206894	16415

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	559	10% 28% 33% • 38%
1	B	559	19% 24% 34% • 40%
1	C	559	21% 25% 35% • 39%
1	D	559	38% 28% 33% 39%
2	E	95	• 37% 61% •
2	F	95	• 35% 65%
3	G	80	14% 31% 69%
4	H	72	12% 19% 81%

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	I	12	58% ... 17% ... 83%
6	J	6	33% ... 33% ... 67%

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 15924 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 CRISPR-associated exonuclease Cas4/endonuclease Cas1 fusion.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	348	Total	C	N	O	S	0	0
			2772	1750	514	492	16		
1	B	338	Total	C	N	O	S	0	0
			2696	1702	498	480	16		
1	C	339	Total	C	N	O	S	0	0
			2707	1708	502	481	16		
1	D	339	Total	C	N	O	S	0	0
			2696	1702	499	479	16		

- Molecule 2 is a protein called CRISPR-associated endoribonuclease Cas2.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	E	95	Total	C	N	O	S	0	0
			786	504	145	132	5		
2	F	95	Total	C	N	O	S	0	0
			786	504	145	132	5		

- Molecule 3 is a DNA chain called DNA (80-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
3	G	80	Total	C	N	O	P	0	0
			1639	777	300	482	80		

- Molecule 4 is a DNA chain called DNA (72-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
4	H	72	Total	C	N	O	P	0	0
			1468	699	252	445	72		

- Molecule 5 is a DNA chain called DNA (5'-D(P*CP*GP*GP*AP*AP*AP*AP*GP*AP*G P*CP*C)-3').

Mol	Chain	Residues	Atoms					AltConf	Trace
5	I	12	Total 250	C 117	N 54	O 67	P 12	0	0

- Molecule 6 is a DNA chain called DNA (5'-D(P*GP*AP*AP*GP*CP*C)-3').

Mol	Chain	Residues	Atoms					AltConf	Trace
6	J	6	Total 124	C 58	N 26	O 34	P 6	0	0

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.
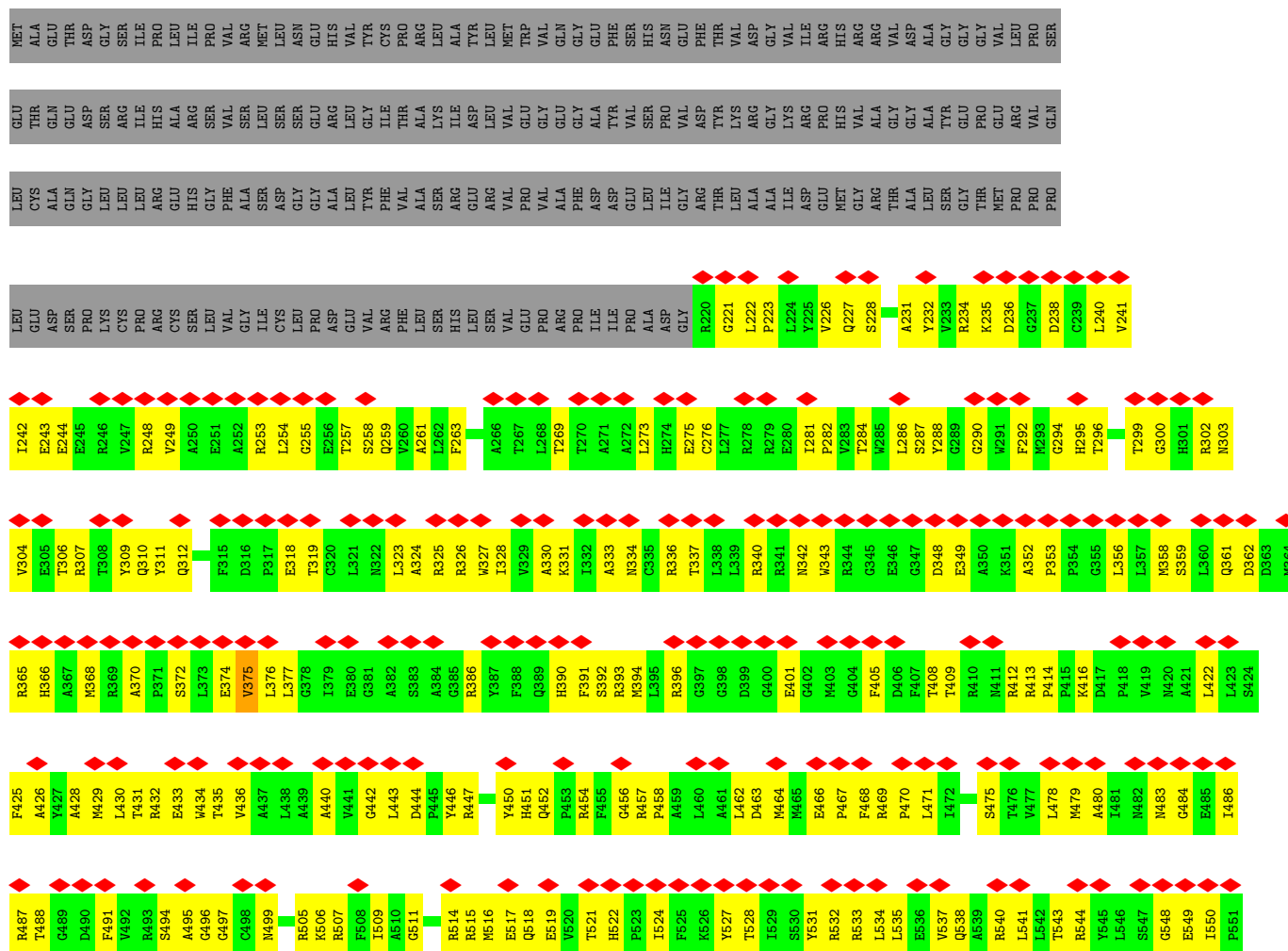
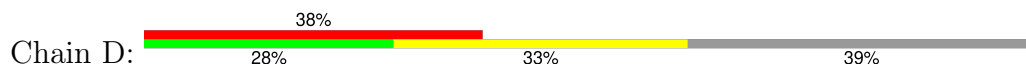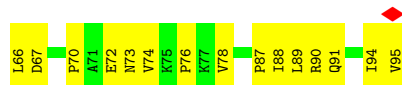• Molecule 1: CRISPR-associated exonuclease Cas4/endonuclease Cas1 fusion

• Molecule 1: CRISPR-associated exonuclease Cas4/endonuclease Cas1 fusion
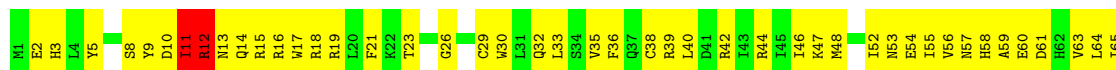
• Molecule 1: CRISPR-associated exonuclease Cas4/endonuclease Cas1 fusion

Chain C:

● Molecule 1: CRISPR-associated exonuclease Cas4/endonuclease Cas1 fusion
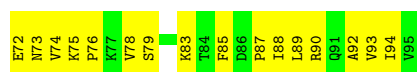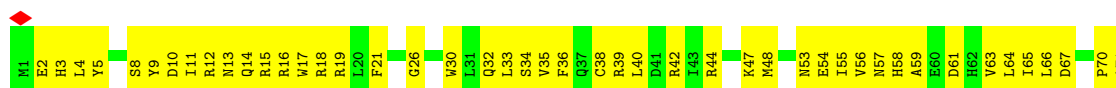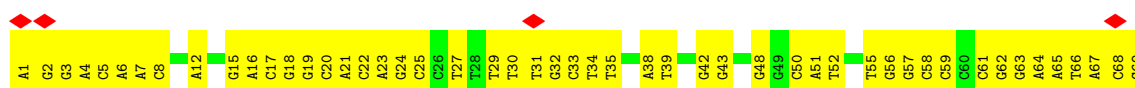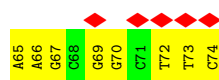

Chain D:

A552 Y553 P554 N555 F556 V557 T558 ARG

- Molecule 2: CRISPR-associated endoribonuclease Cas2

Chain E:

37% 61%

M1 E2 H3 L4 Y5 S8 Y9 D10 I11 R12 N13 Q14 R15 R16 W17 R18 R19 L20 F21 K22 T23 G26 C29 W30 L31 Q32 L33 S34 V35 F36 Q37 C38 R39 L40 D41 R42 I43 R44 I45 I46 K47 M48 I52 N53 E54 I55 V56 N57 H58 A59 E60 D61 H62 V63 L64 I65

L66 D67 P70 A71 E72 N73 V74 K75 P76 K77 V78 P87 I88 L89 R90 Q91 I94 V95

- Molecule 2: CRISPR-associated endoribonuclease Cas2

Chain F:

35% 65%

M1 E2 H3 L4 Y5 S8 Y9 D10 I11 R12 N13 Q14 R15 R16 W17 R18 R19 L20 F21 G26 W30 L31 Q32 L33 S34 V35 F36 Q37 C38 R39 L40 D41 R42 I43 R44 K47 M48 N53 E54 I55 V56 N57 H58 A59 E60 D61 H62 V63 L64 I65 L66 D67 P70 A71

E72 N73 V74 K75 P76 K77 V78 S79 K83 T84 F85 D86 P87 I88 L89 R90 Q91 A92 V93 I94 V95

- Molecule 3: DNA (80-MER)

Chain G:

14% 31% 69%

A1 G2 G3 A4 C5 A6 A7 C8 A12 G15 A16 C17 G18 G19 C20 A21 C22 A23 G24 C25 C26 T27 T28 T29 T30 T31 G32 C33 T34 T35 A38 T39 G42 G43 G48 G49 C50 A51 T52 T55 G56 G57 C58 C59 C60 C61 G62 G63 A64 A65 T66 A67 C68 G69

G70 C71 T72 C73 T74 T75 T76 T77 C78 C79 G80

- Molecule 4: DNA (72-MER)

Chain H:

12% 19% 81%

C3 T4 G5 T6 G7 C8 C9 G10 T11 C12 C13 G14 T15 A16 A17 C18 G19 T20 T21 G22 T23 C24 G25 A26 T27 T28 T29 T30 T31 G32 T33 A34 C37 C38 G39 G40 G41 G42 C43 C44 G47 A48 T49 G50 C51 C52 C53 C54 G55 G56 C57 C58 T59 C60 A61 I62 T63 G64

A65 A66 G67 C68 G69 G70 C71 T72 T73 C74

- Molecule 5: DNA (5'-D(P*CP*GP*GP*AP*AP*AP*AP*GP*AP*GP*CP*C)-3')

Chain I:

58% 17% 83%

• Molecule 6: DNA (5'-D(P*GP*AP*AP*GP*CP*C)-3')

Chain J:
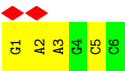
4 Experimental information ⓘ

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	80000	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	TFS TALOS	Depositor
Voltage (kV)	200	Depositor
Electron dose $(e^-/\text{Å}^2)$	50	Depositor
Minimum defocus (nm)	1500	Depositor
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K3 BIOQUANTUM (6k x 4k)	Depositor
Maximum map value	0.129	Depositor
Minimum map value	-0.054	Depositor
Average map value	-0.000	Depositor
Map value standard deviation	0.002	Depositor
Recommended contour level	0.014	Depositor
Map size (Å)	346.896, 346.896, 346.896	wwPDB
Map dimensions	264, 264, 264	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.314, 1.314, 1.314	Depositor

5 Model quality ⓘ

5.1 Standard geometry ⓘ

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$|Z|$ >5	RMSZ	#$|Z|$ >5
1	A	0.39	0/2838	0.62	0/3841
1	B	0.37	0/2759	0.67	4/3733 (0.1%)
1	C	0.38	0/2770	0.60	0/3747
1	D	0.18	0/2759	0.48	0/3733
2	E	0.32	0/802	0.52	0/1080
2	F	0.26	0/802	0.47	0/1080
3	G	0.28	0/1837	0.48	0/2833
4	H	0.29	0/1640	0.54	0/2529
5	I	0.26	0/282	0.40	0/433
6	J	0.21	0/139	0.37	0/212
All	All	0.32	0/16628	0.56	4/23221 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	524	ILE	N-CA-C	-6.43	107.60	113.71
1	B	412	ARG	CA-C-N	-5.29	108.89	121.80
1	B	412	ARG	C-N-CA	-5.29	108.89	121.80
1	B	388	PHE	N-CA-C	-5.10	106.74	113.17

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2772	0	2758	243	0
1	B	2696	0	2676	300	0
1	C	2707	0	2689	234	0
1	D	2696	0	2673	176	0
2	E	786	0	813	77	0
2	F	786	0	813	74	0
3	G	1639	0	899	102	0
4	H	1468	0	815	130	0
5	I	250	0	133	13	0
6	J	124	0	67	4	0
All	All	15924	0	14336	1165	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 39.

The worst 5 of 1165 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)
1:B:391:PHE:CE2	1:B:423:LEU:HD11	1.54	1.42
1:B:388:PHE:CE1	1:B:410:ARG:HB2	1.61	1.35
1:B:291:TRP:CB	4:H:26:DA:H1'	1.59	1.32
1:B:334:ASN:O	1:B:337:THR:HG22	1.29	1.32
1:B:344:ARG:O	1:B:346:GLU:HG3	1.33	1.26

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	346/559 (62%)	298 (86%)	47 (14%)	1 (0%)	37	70
1	B	336/559 (60%)	269 (80%)	64 (19%)	3 (1%)	14	48

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	337/559 (60%)	301 (89%)	34 (10%)	2 (1%)	22	57
1	D	337/559 (60%)	310 (92%)	26 (8%)	1 (0%)	37	70
2	E	93/95 (98%)	75 (81%)	16 (17%)	2 (2%)	5	32
2	F	93/95 (98%)	77 (83%)	16 (17%)	0	100	100
All	All	1542/2426 (64%)	1330 (86%)	203 (13%)	9 (1%)	24	57

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	345	GLY
1	D	375	VAL
1	A	340	ARG
1	B	343	TRP
2	E	11	ILE

5.3.2   Protein sidechains ⓘ

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	287/462 (62%)	281 (98%)	6 (2%)	48	67
1	B	279/462 (60%)	268 (96%)	11 (4%)	27	51
1	C	280/462 (61%)	272 (97%)	8 (3%)	37	59
1	D	278/462 (60%)	278 (100%)	0	100	100
2	E	86/86 (100%)	84 (98%)	2 (2%)	45	64
2	F	86/86 (100%)	86 (100%)	0	100	100
All	All	1296/2020 (64%)	1269 (98%)	27 (2%)	49	67

5 of 27 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	388	PHE
1	C	331	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	359	SER
1	B	491	PHE
1	C	332	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	518	GLN
1	D	227	GLN
2	F	32	GLN
1	D	334	ASN
1	B	538	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6   Map visualisation ⓘ

This section contains visualisations of the EMDB entry EMD-23843. These allow visual inspection of the internal detail of the map and identification of artifacts.

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

6.1   Orthogonal projections ⓘ

6.1.1   Primary map

| X | Y | Z |
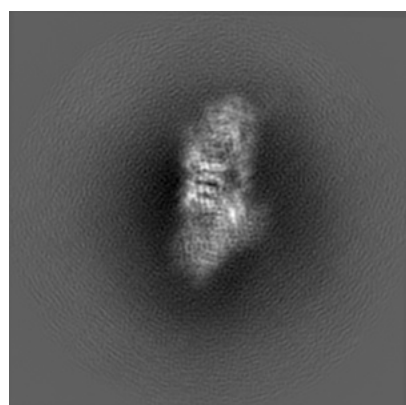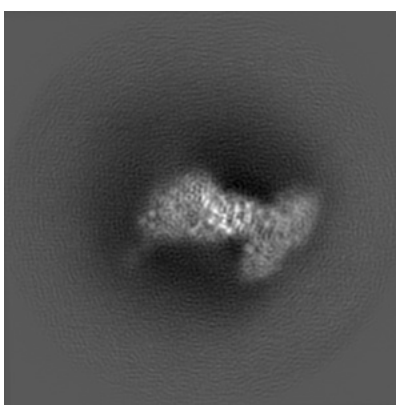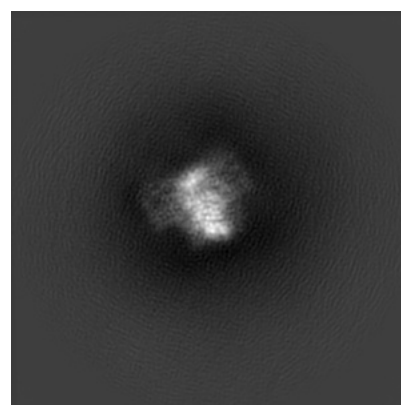
The images above show the map projected in three orthogonal directions.

6.2   Central slices ⓘ
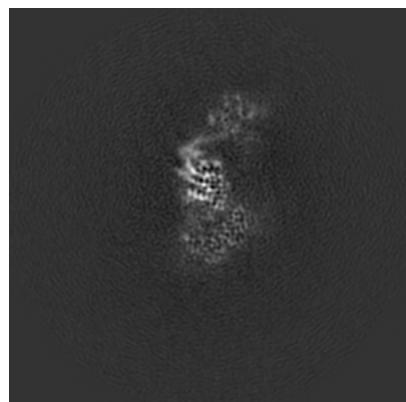
6.2.1   Primary map
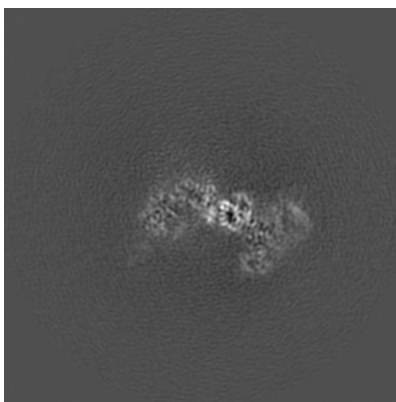
| X Index: 132 | Y Index: 132 | Z Index: 132 |
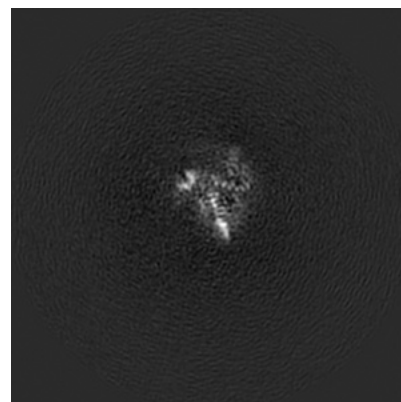
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices ⓘ
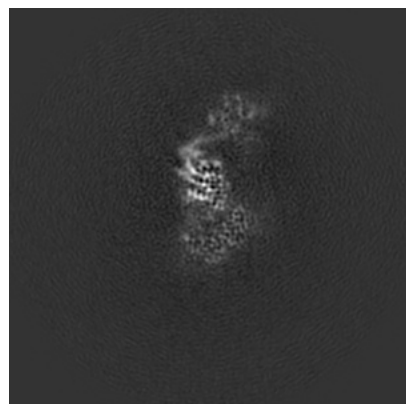
6.3.1 Primary map
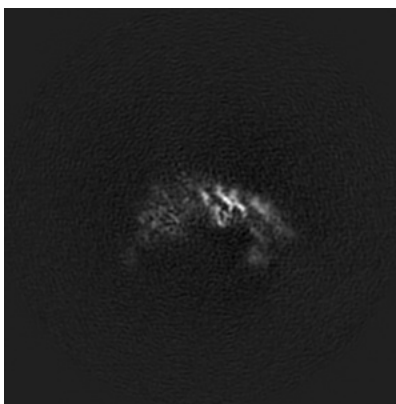
| X Index: 132 | Y Index: 121 | Z Index: 145 |
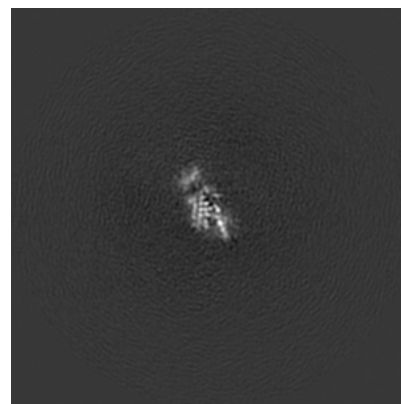
The images above show the largest variance slices of the map in three orthogonal directions.

6.4 Orthogonal standard-deviation projections (False-color) ⓘ

6.4.1 Primary map

| X | Y | Z |
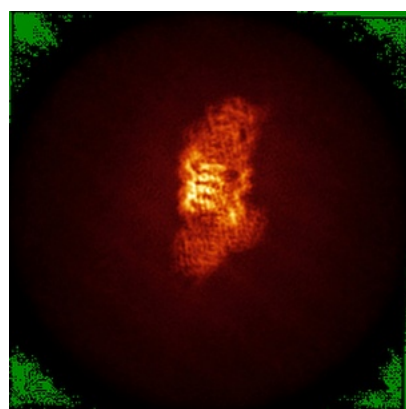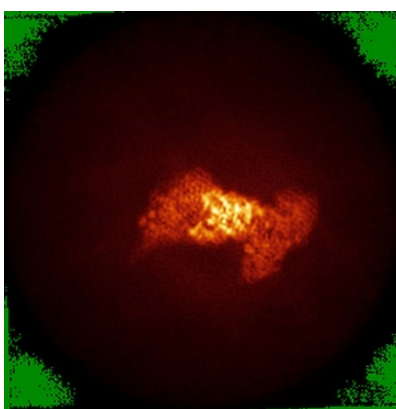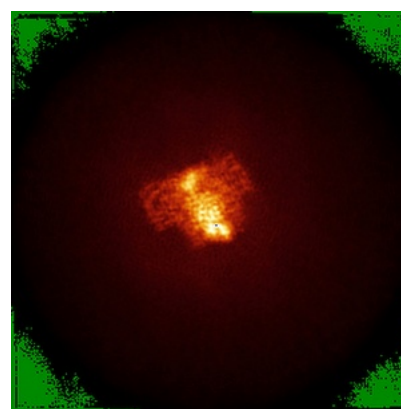
The images above show the map standard deviation projections with false color in three orthogonal directions. Minimum values are shown in green, max in blue, and dark to light orange shades represent small to large values respectively.

6.5   Orthogonal surface views (i)

6.5.1   Primary map

X   Y   Z

The images above show the 3D surface view of the map at the recommended contour level 0.014. 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 ⓘ

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

7.1 Map-value distribution ⓘ
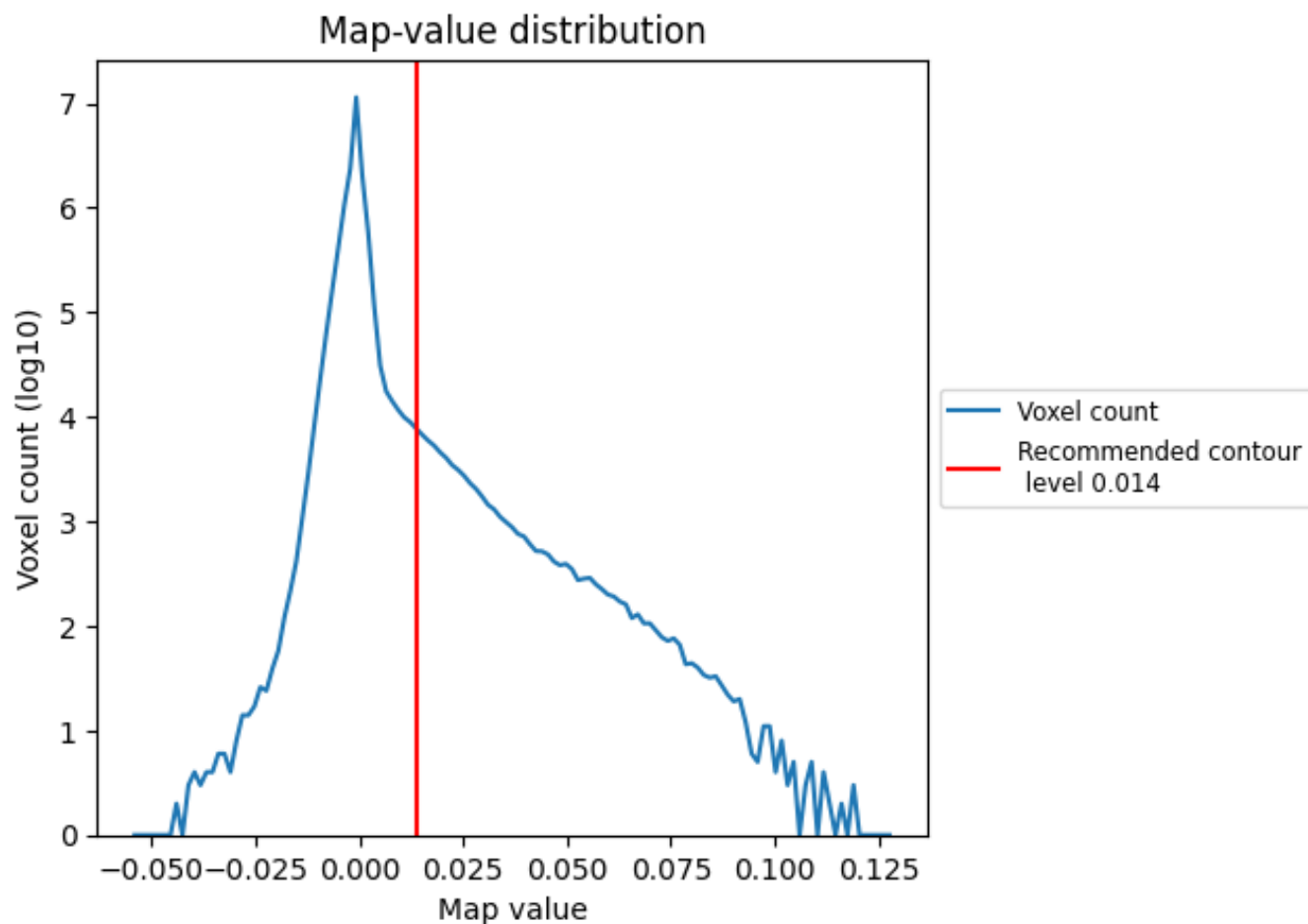
The map-value distribution is plotted in 128 intervals along the x-axis. The y-axis is logarithmic. A spike in this graph at zero usually indicates that the volume has been masked.

7.2 Volume estimate ⓘ
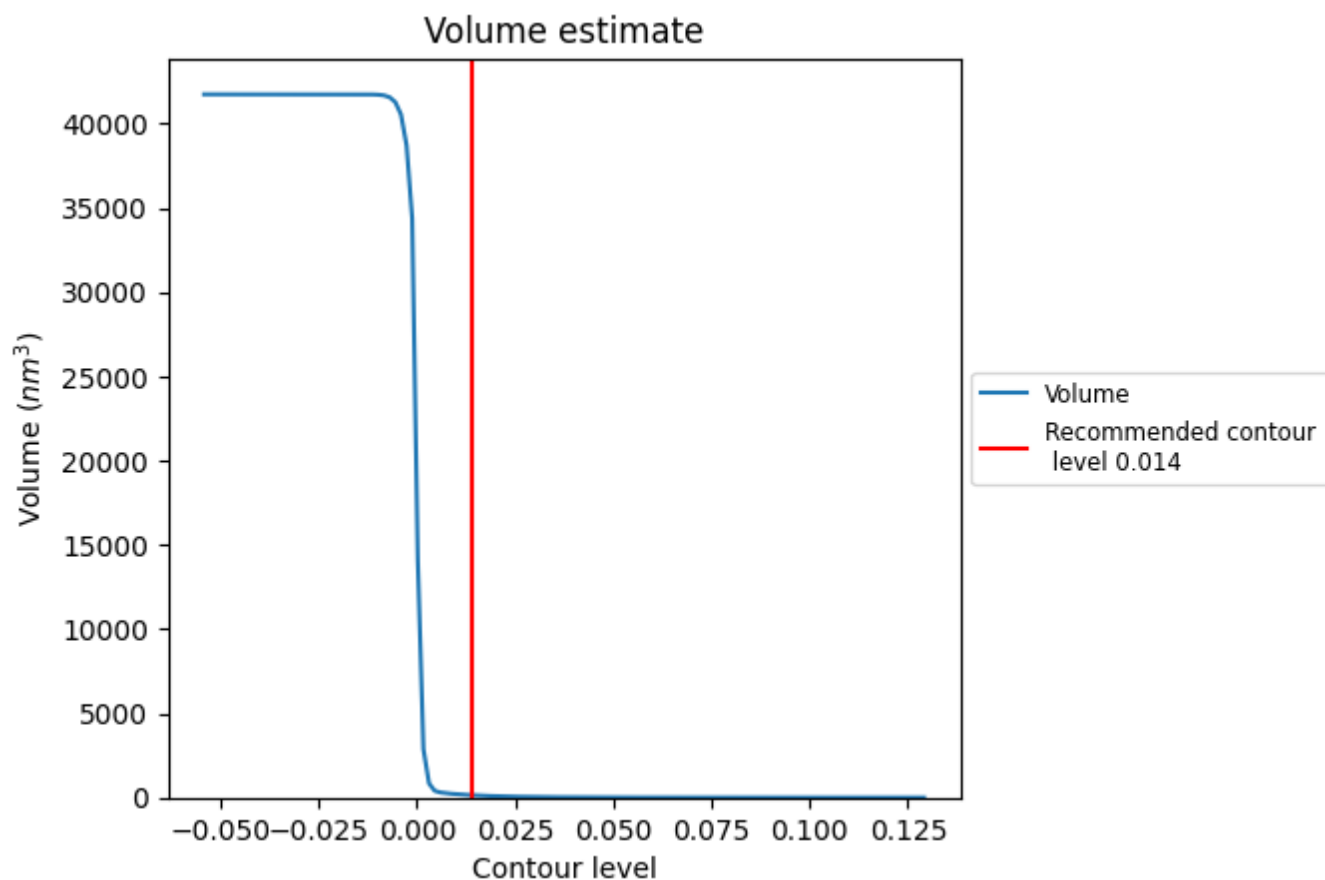
The volume at the recommended contour level is 141 nm^3; this corresponds to an approximate mass of 127 kDa.

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   ⓘ
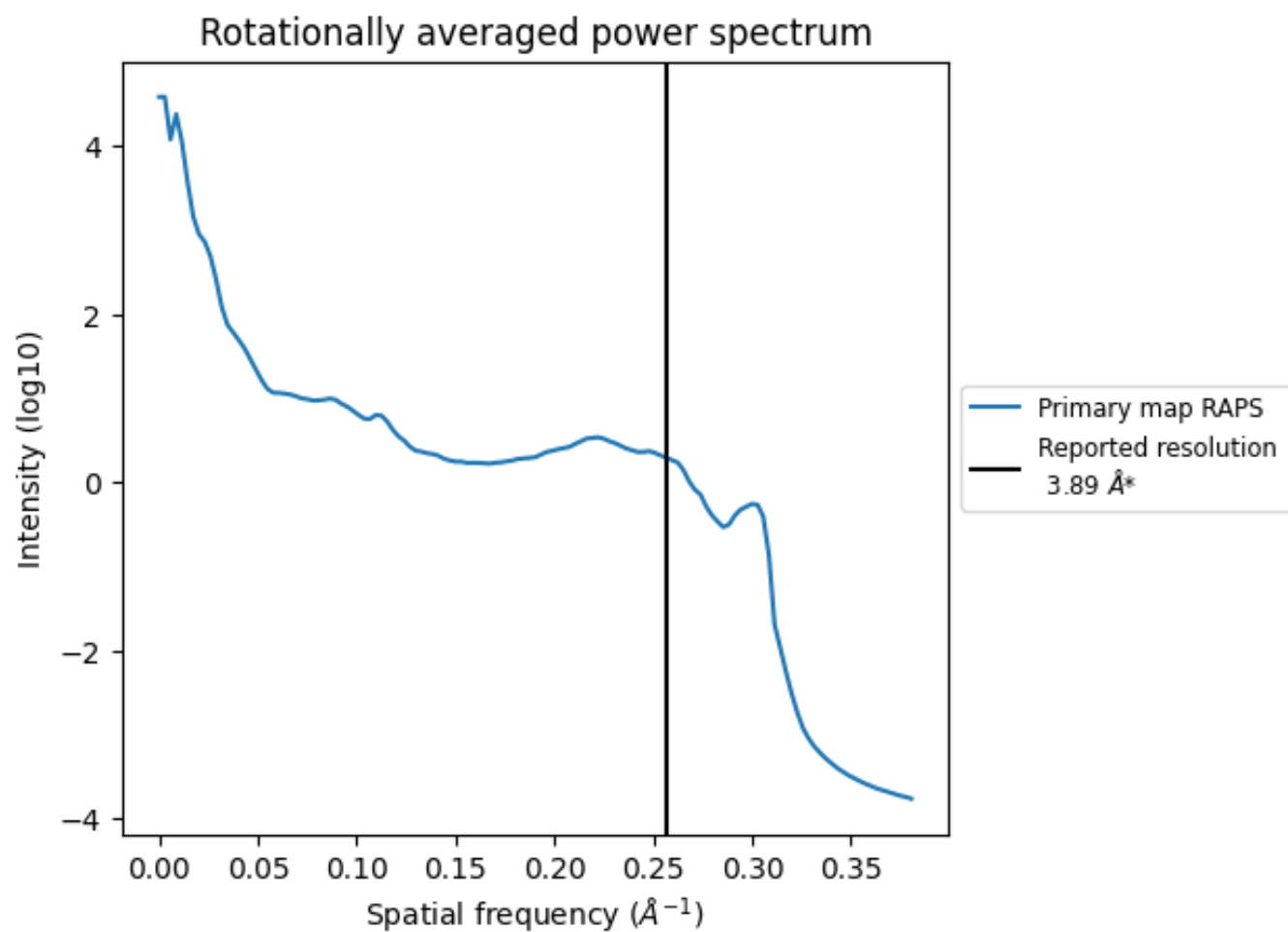
*Reported resolution corresponds to spatial frequency of 0.257 Å^{-1}

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-23843 and PDB model 7MI9. Per-residue inclusion information can be found in section 3 on page 6.

9.1 Map-model overlay (i)
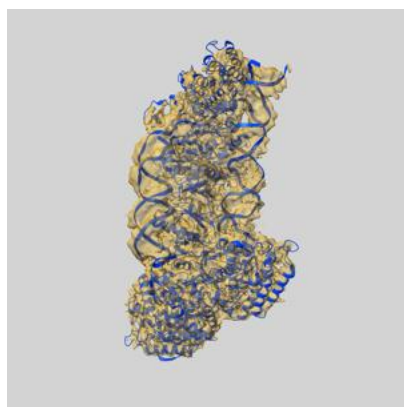
X	Y	Z
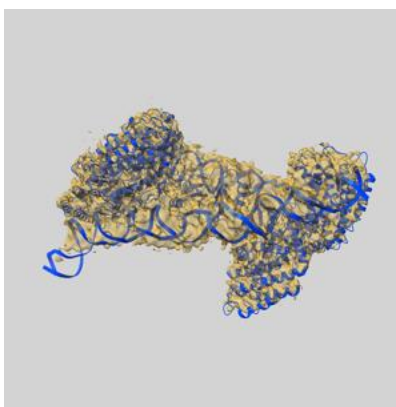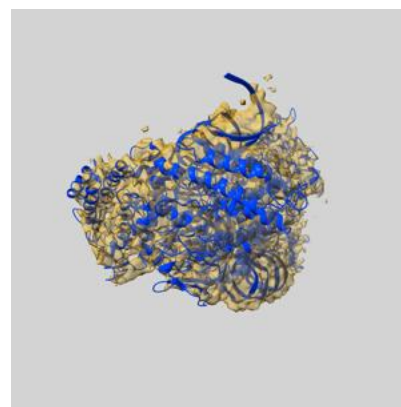
The images above show the 3D surface view of the map at the recommended contour level 0.014 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

9.2   Q-score mapped to coordinate model ⓘ

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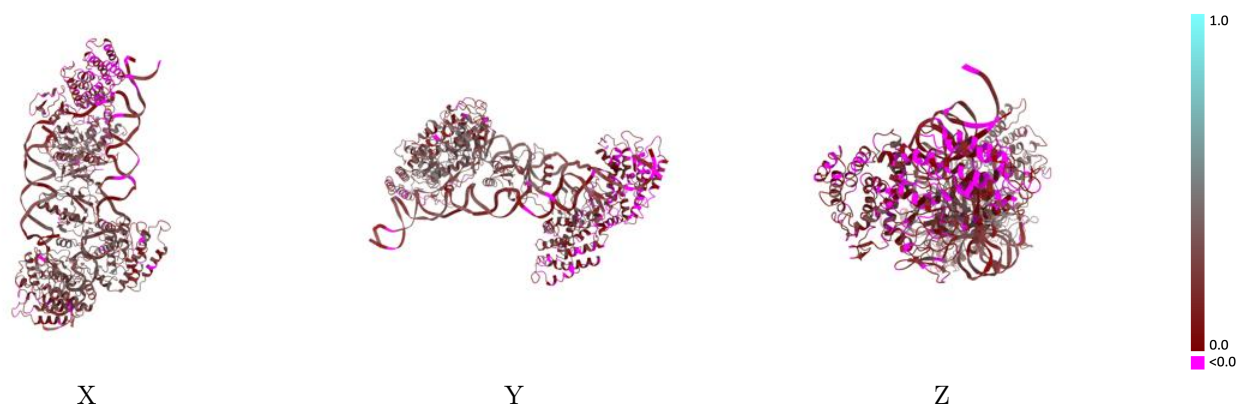
The images above show the model with each residue coloured according its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions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

X   Y   Z
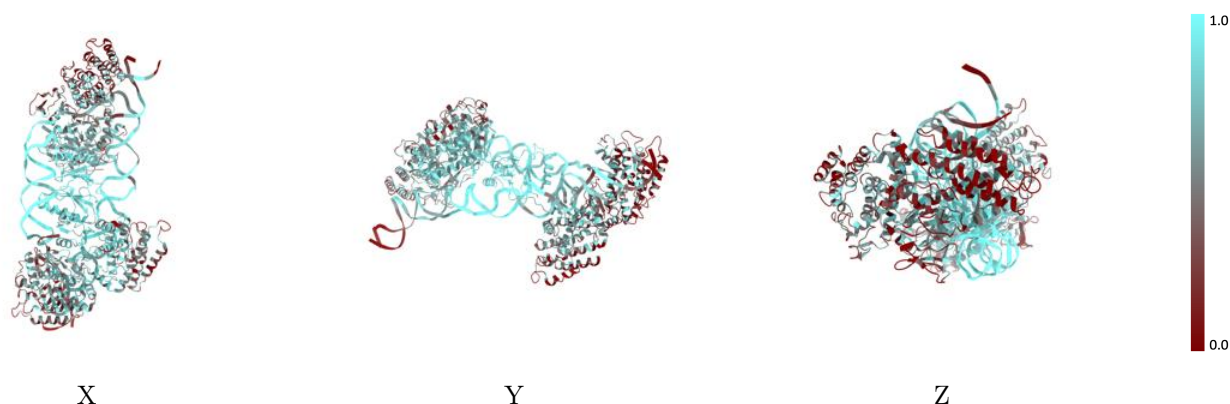
The images above show the model with each residue coloured according to its atom inclusion. This shows to what extent they are inside the map at the recommended contour level (0.014).

9.4　Atom inclusion ⓘ
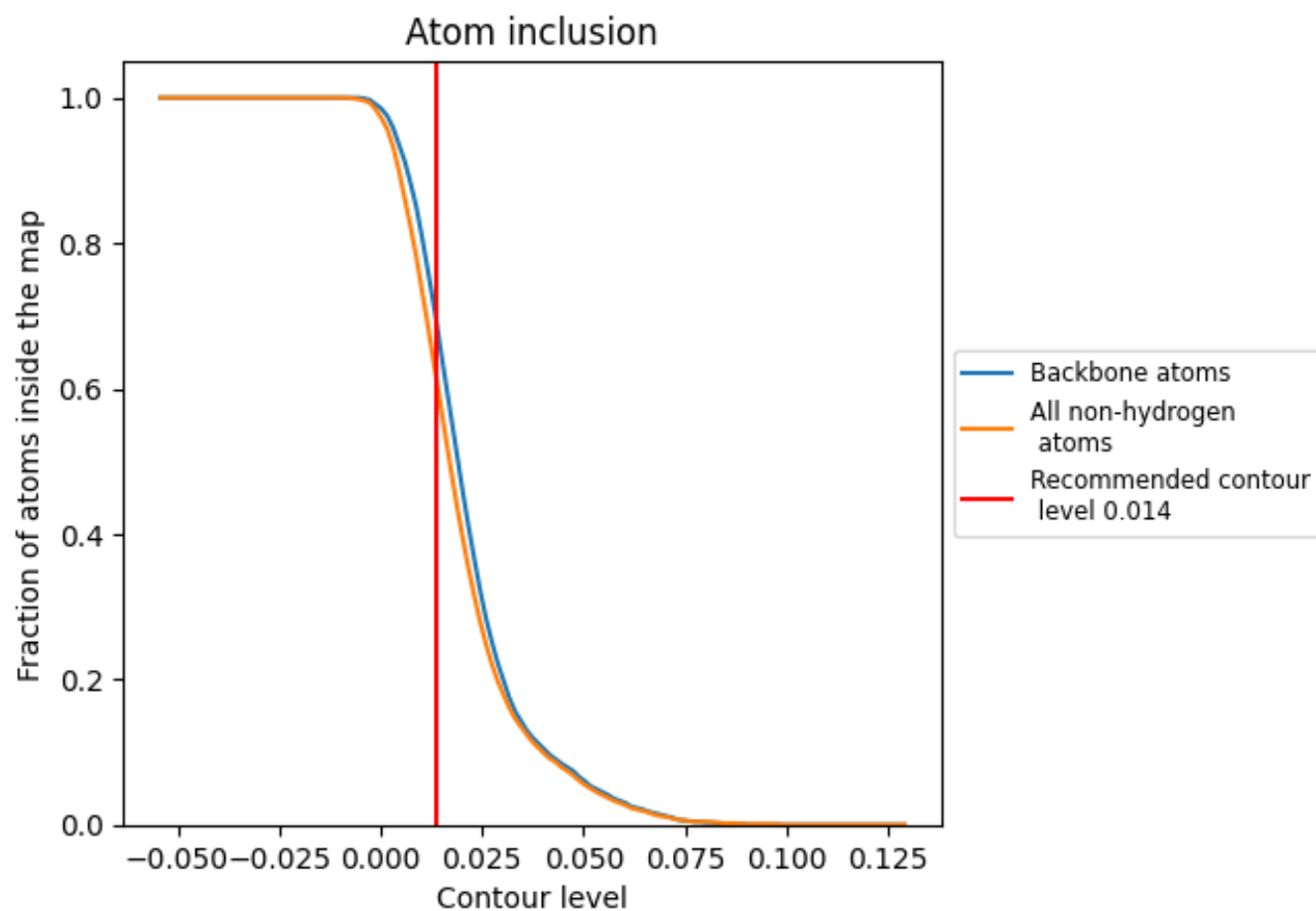
At the recommended contour level, 69% of all backbone atoms, 60% of all non-hydrogen atoms, are inside the map.

9.5 Map-model fit summary ⓘ

The table lists the average atom inclusion at the recommended contour level (0.014) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.6050	0.2010
A	0.6850	0.3050
B	0.5510	0.2510
C	0.5220	0.1610
D	0.3320	0.0510
E	0.9320	0.2870
F	0.9240	0.2950
G	0.7330	0.1850
H	0.7760	0.1960
I	0.2680	0.1570
J	0.4920	0.0940

1.0

0.0
<0.0