



wwPDB EM Validation Summary Report ⓘ

May 18, 2024 – 07:05 pm BST

PDB ID : 6T9I
EMDB ID : EMD-10412
Title : cryo-EM structure of transcription coactivator SAGA
Authors : Wang, H.; Cheung, A.; Cramer, P.
Deposited on : 2019-10-28
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

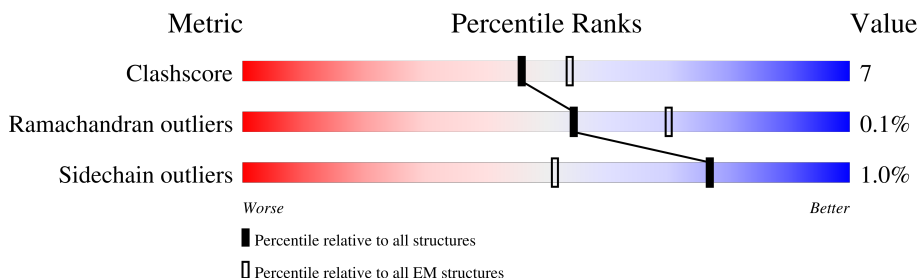
1 Overall quality at a glance

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	B	604	
2	C	337	
3	D	798	
4	E	516	
5	F	157	
6	G	206	
7	H	488	
8	I	539	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
9	K	1332	
10	Q	657	
11	T	3744	
12	U	155	

2 Entry composition [i](#)

There are 12 unique types of molecules in this entry. The entry contains 48821 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 Transcription factor SPT20.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	B	224	1891	1207	318	358	8	0	0

- Molecule 2 is a protein called Protein SPT3.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	C	214	1774	1130	319	318	7	0	0

- Molecule 3 is a protein called Transcription initiation factor TFIID subunit 5.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	D	566	4474	2830	772	854	18	0	0

- Molecule 4 is a protein called Transcription initiation factor TFIID subunit 6.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	E	433	3441	2207	572	657	5	0	0

- Molecule 5 is a protein called Transcription initiation factor TFIID subunit 9.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
5	F	111	902	581	158	159	4	0	0

- Molecule 6 is a protein called Transcription initiation factor TFIID subunit 10.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
6	G	140	1108	689	199	218	2	0	0

- Molecule 7 is a protein called Transcriptional coactivator HFI1/ADA1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
7	H	227	1803	1121	314	359	9	0	0

- Molecule 8 is a protein called Transcription initiation factor TFIID subunit 12.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
8	I	173	1351	836	243	268	4	0	0

- Molecule 9 is a protein called Transcriptional activator SPT7.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
9	K	279	2258	1427	392	428	11	0	0

- Molecule 10 is a protein called SAGA-associated factor 73.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
10	Q	84	675	426	119	127	3	0	0

- Molecule 11 is a protein called Transcription-associated protein 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
11	T	3513	28729	18596	4776	5237	120	0	0

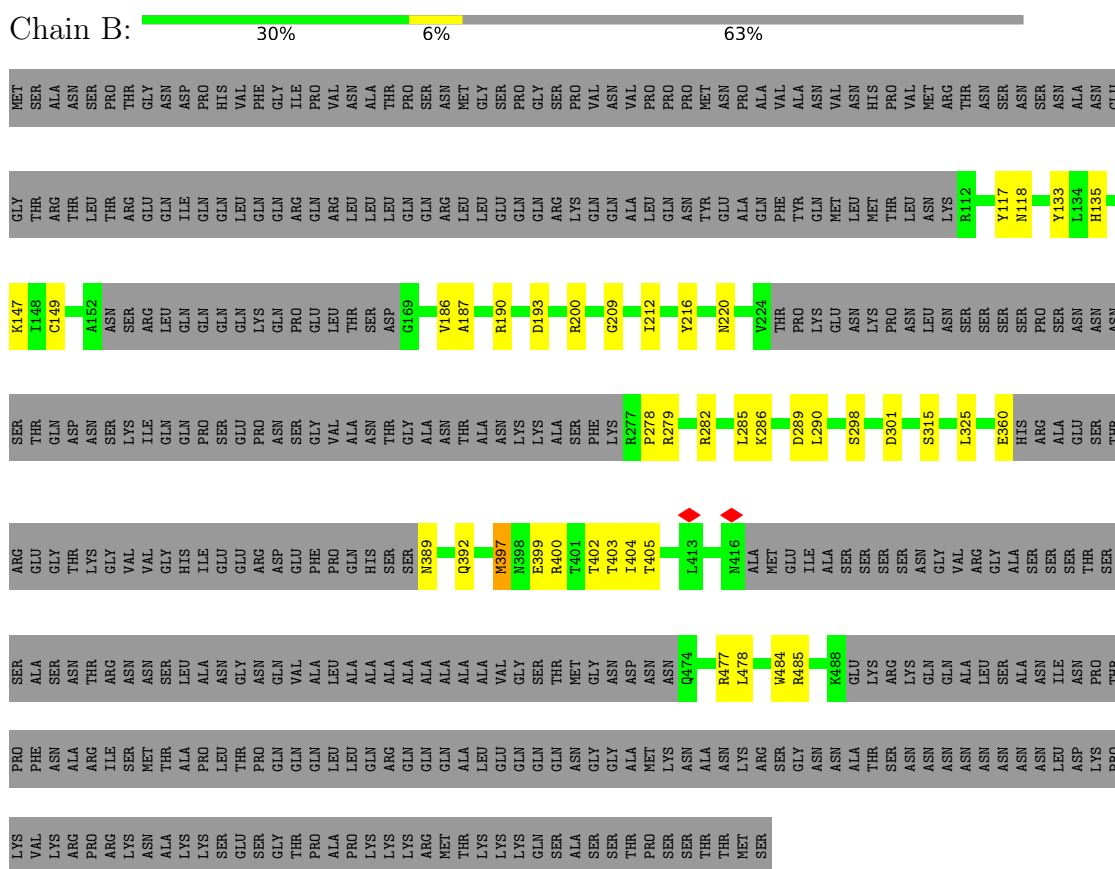
- Molecule 12 is a protein called unassigned sequence.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
12	U	83	415	249	83	83	0	0

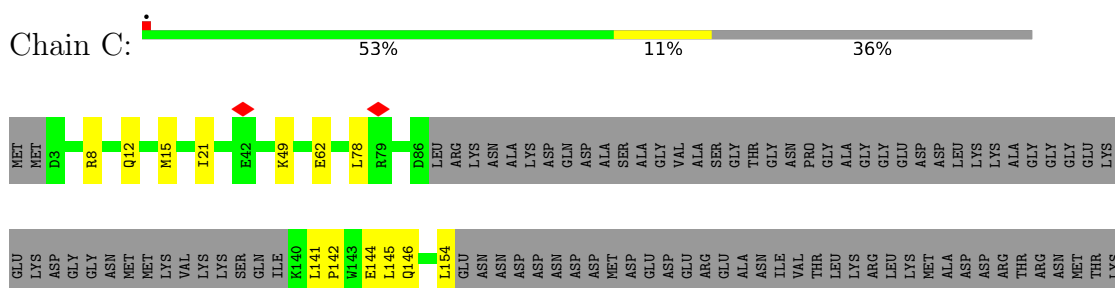
3 Residue-property plots [i](#)

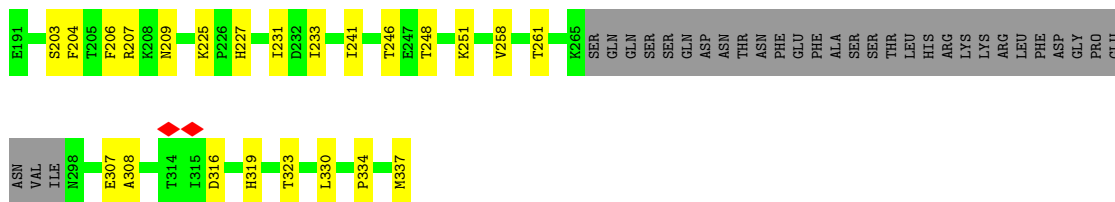
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and atom inclusion 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: Transcription factor SPT20

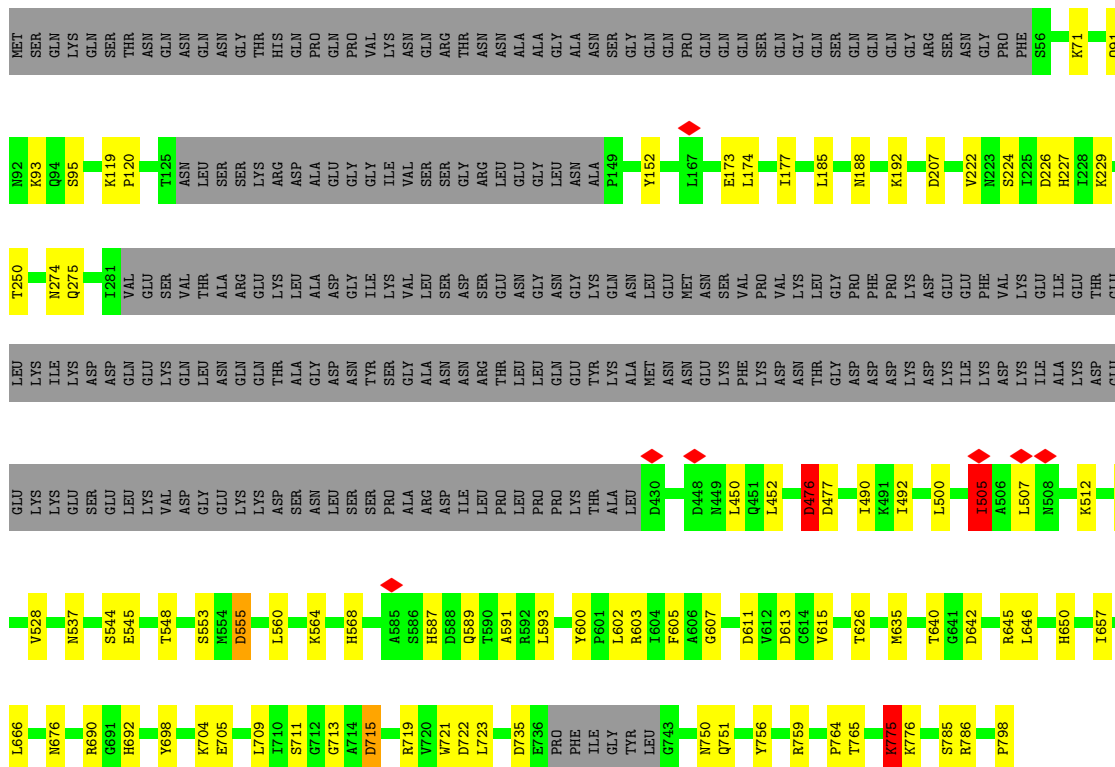


- Molecule 2: Protein SPT3

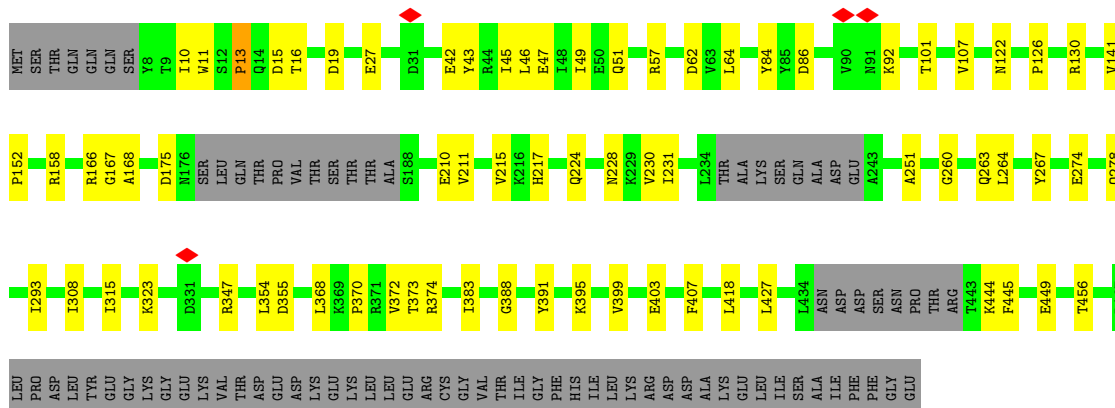




• Molecule 3: Transcription initiation factor TFIID subunit 5




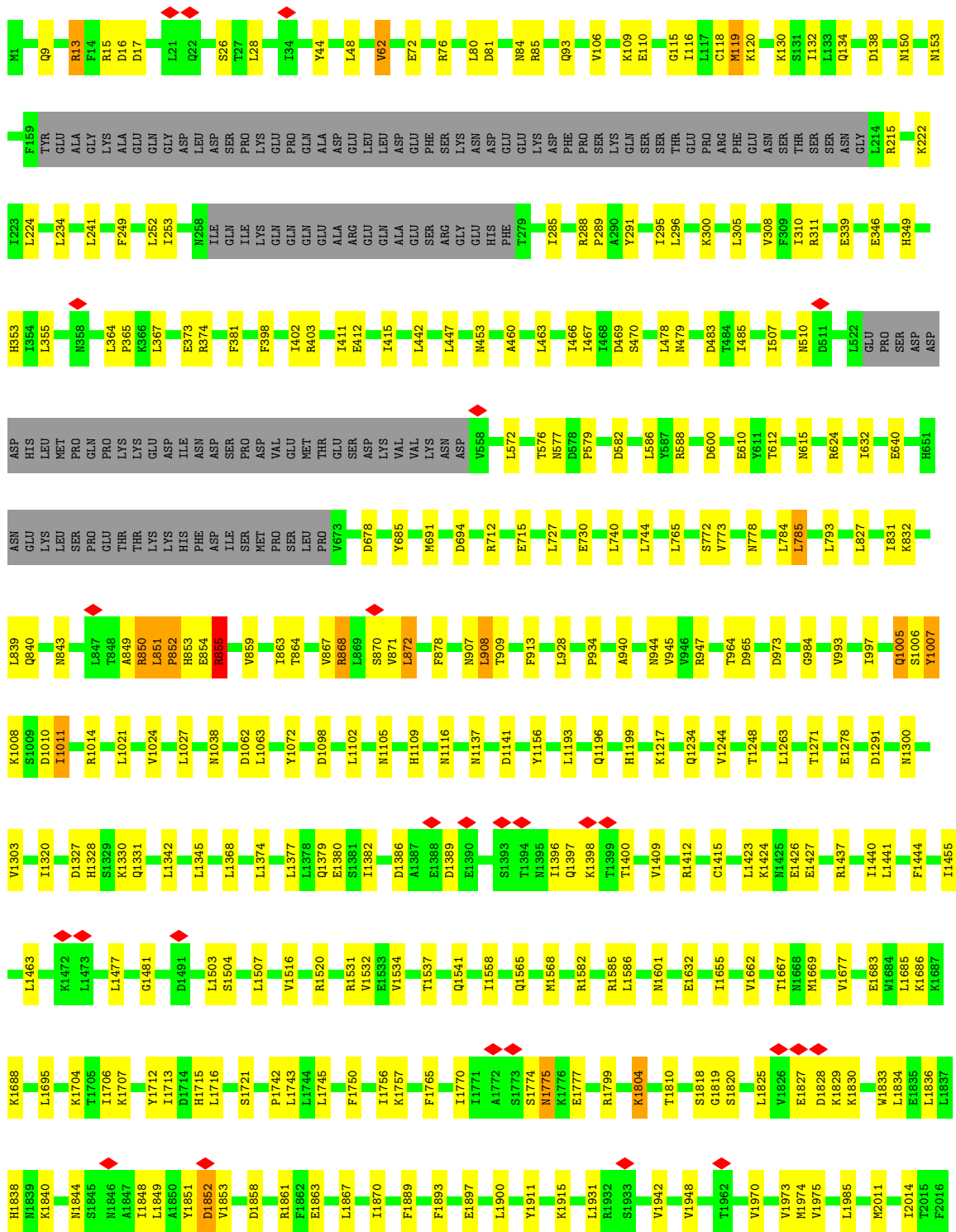
• Molecule 4: Transcription initiation factor TFIID subunit 6



SER
VAL
ASN
PRO
TYR
ASN
GLY
ARG
ILE
ASN

• Molecule 11: Transcription-associated protein 1

Chain T:  79% 14% 6%



4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	27602	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 ($e^-/\text{\AA}^2$)	42.45	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	0.250	Depositor
Minimum map value	-0.106	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.004	Depositor
Recommended contour level	0.015	Depositor
Map size (Å)	419.99997, 419.99997, 419.99997	wwPDB
Map dimensions	400, 400, 400	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.05, 1.05, 1.05	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	B	0.29	0/1931	0.53	0/2607
2	C	0.26	0/1810	0.48	0/2435
3	D	0.31	0/4582	0.58	3/6211 (0.0%)
4	E	0.29	0/3504	0.56	1/4759 (0.0%)
5	F	0.30	0/922	0.54	0/1249
6	G	0.29	0/1123	0.51	0/1518
7	H	0.29	0/1826	0.55	0/2456
8	I	0.33	0/1373	0.58	1/1863 (0.1%)
9	K	0.26	0/2288	0.54	1/3080 (0.0%)
10	Q	0.29	0/692	0.53	0/937
11	T	0.38	0/29358	0.60	12/39779 (0.0%)
12	U	0.28	0/412	0.46	0/572
All	All	0.35	0/49821	0.58	18/67466 (0.0%)

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
3	D	0	1

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
11	T	1291	ASP	CB-CG-OD2	6.86	124.47	118.30
11	T	855	ARG	NE-CZ-NH1	6.79	123.69	120.30
3	D	477	ASP	CB-CG-OD1	6.78	124.40	118.30
11	T	2504	LEU	CA-CB-CG	6.39	130.01	115.30
11	T	785	LEU	CA-CB-CG	6.34	129.88	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	D	476	ASP	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	B	1891	0	1840	34	0
2	C	1774	0	1790	23	0
3	D	4474	0	4377	60	0
4	E	3441	0	3493	52	0
5	F	902	0	919	22	0
6	G	1108	0	1110	20	0
7	H	1803	0	1800	37	0
8	I	1351	0	1351	72	0
9	K	2258	0	2308	28	0
10	Q	675	0	666	15	0
11	T	28729	0	29144	368	0
12	U	415	0	415	3	0
All	All	48821	0	49213	642	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 7.

The worst 5 of 642 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)
8:I:418:ARG:HG3	8:I:421:ARG:NE	1.08	1.39
8:I:418:ARG:NE	8:I:421:ARG:HH11	1.24	1.32
11:T:850:ARG:CB	11:T:854:GLU:OE2	1.80	1.29
8:I:418:ARG:CD	8:I:421:ARG:NH1	1.98	1.26
11:T:850:ARG:HB2	11:T:854:GLU:OE2	1.10	1.25

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	B	214/604 (35%)	202 (94%)	12 (6%)	0	100	100
2	C	206/337 (61%)	199 (97%)	7 (3%)	0	100	100
3	D	558/798 (70%)	515 (92%)	42 (8%)	1 (0%)	47	79
4	E	425/516 (82%)	393 (92%)	31 (7%)	1 (0%)	47	79
5	F	107/157 (68%)	104 (97%)	3 (3%)	0	100	100
6	G	138/206 (67%)	127 (92%)	11 (8%)	0	100	100
7	H	221/488 (45%)	204 (92%)	17 (8%)	0	100	100
8	I	171/539 (32%)	157 (92%)	14 (8%)	0	100	100
9	K	271/1332 (20%)	253 (93%)	18 (7%)	0	100	100
10	Q	82/657 (12%)	75 (92%)	7 (8%)	0	100	100
11	T	3489/3744 (93%)	3209 (92%)	278 (8%)	2 (0%)	51	84
12	U	77/155 (50%)	73 (95%)	4 (5%)	0	100	100
All	All	5959/9533 (62%)	5511 (92%)	444 (8%)	4 (0%)	54	84

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
11	T	851	LEU
3	D	505	ILE
11	T	852	PRO
4	E	13	PRO

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	B	212/536 (40%)	211 (100%)	1 (0%)	88	93
2	C	198/302 (66%)	196 (99%)	2 (1%)	76	86
3	D	501/701 (72%)	494 (99%)	7 (1%)	67	81
4	E	391/464 (84%)	390 (100%)	1 (0%)	92	95
5	F	95/131 (72%)	95 (100%)	0	100	100
6	G	122/177 (69%)	121 (99%)	1 (1%)	81	89
7	H	206/434 (48%)	206 (100%)	0	100	100
8	I	150/483 (31%)	142 (95%)	8 (5%)	22	52
9	K	257/1225 (21%)	257 (100%)	0	100	100
10	Q	77/579 (13%)	77 (100%)	0	100	100
11	T	3238/3452 (94%)	3201 (99%)	37 (1%)	73	84
All	All	5447/8484 (64%)	5390 (99%)	57 (1%)	77	86

5 of 57 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
11	T	311	ARG
11	T	3307	TYR
11	T	908	LEU
11	T	2276	THR
11	T	1851	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 86 such sidechains are listed below:

Mol	Chain	Res	Type
11	T	1892	ASN
11	T	2906	GLN
11	T	2203	GLN
11	T	2521	ASN
11	T	3064	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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
12	U	2

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	U	29:ALA	C	101:ALA	N	57.64
1	U	9:ALA	C	11:ALA	N	3.53

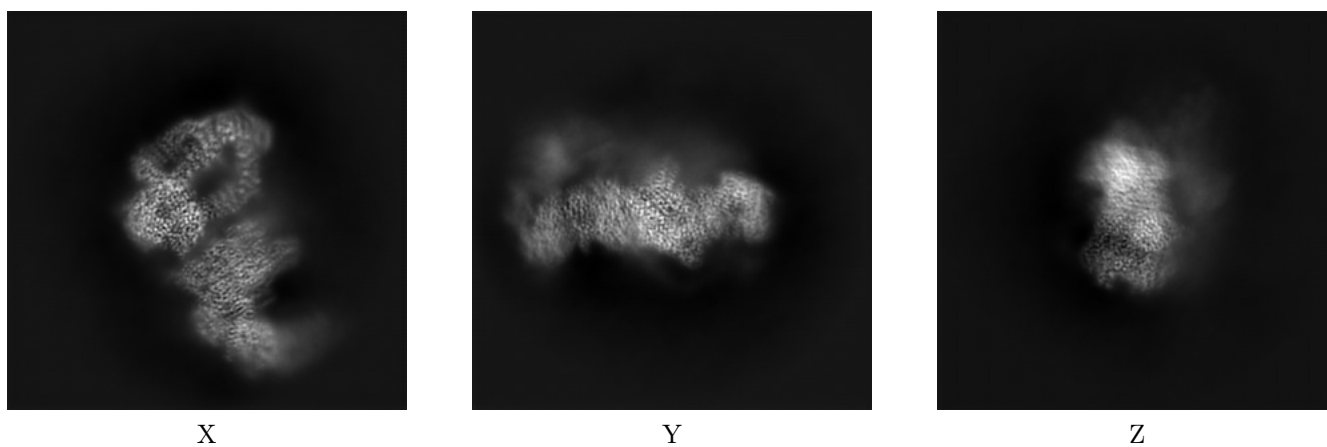
6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-10412. 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](#)

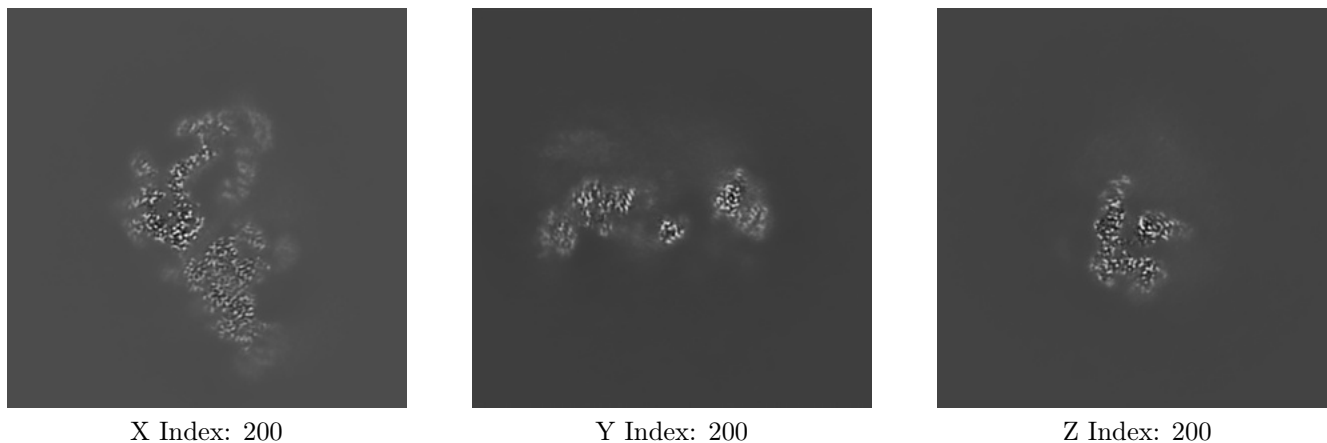
6.1.1 Primary map



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

6.2 Central slices [i](#)

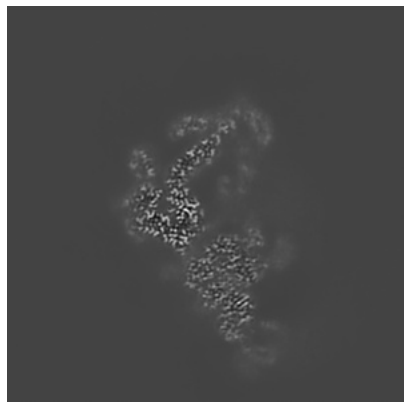
6.2.1 Primary map



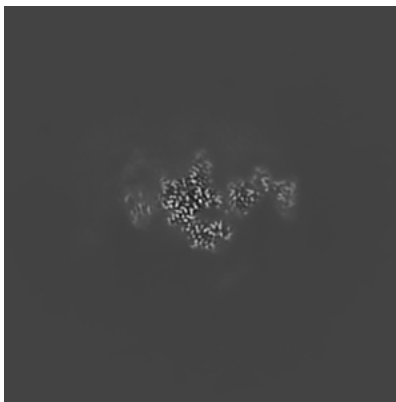
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices [i](#)

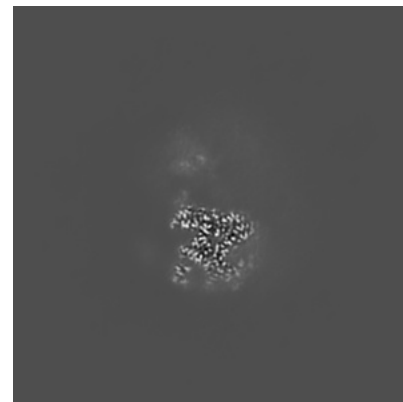
6.3.1 Primary map



X Index: 204



Y Index: 178

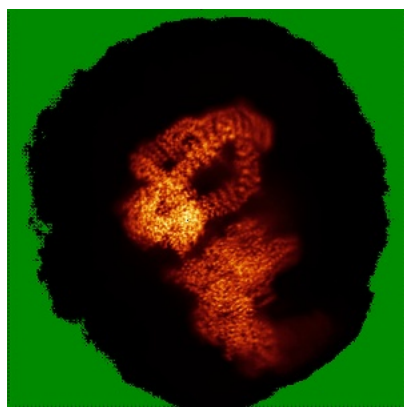


Z Index: 188

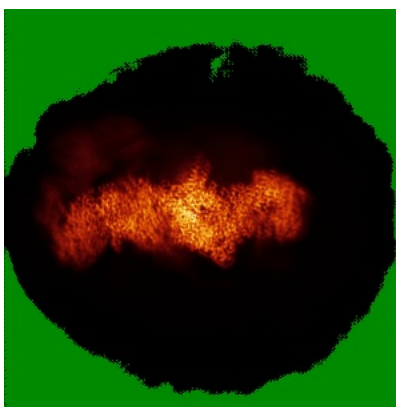
The images above show the largest variance slices of the map in three orthogonal directions.

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

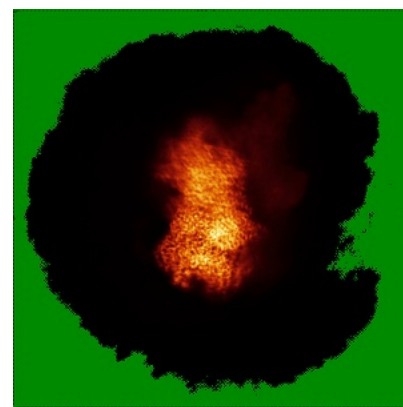
6.4.1 Primary map



X



Y



Z

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

6.5 Orthogonal surface views [i](#)

6.5.1 Primary map



The images above show the 3D surface view of the map at the recommended contour level 0.015. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

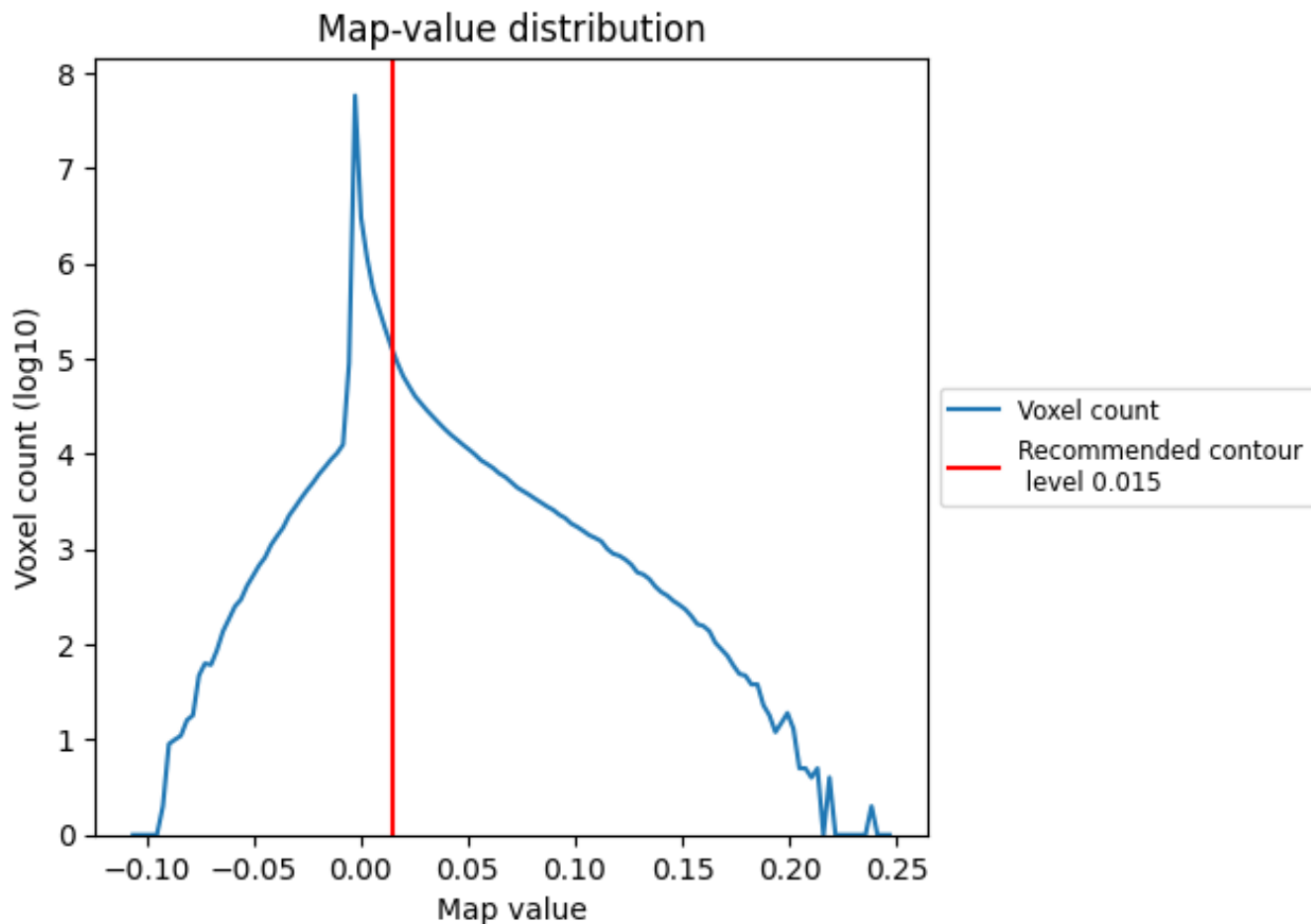
6.6 Mask visualisation [i](#)

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

7 Map analysis [i](#)

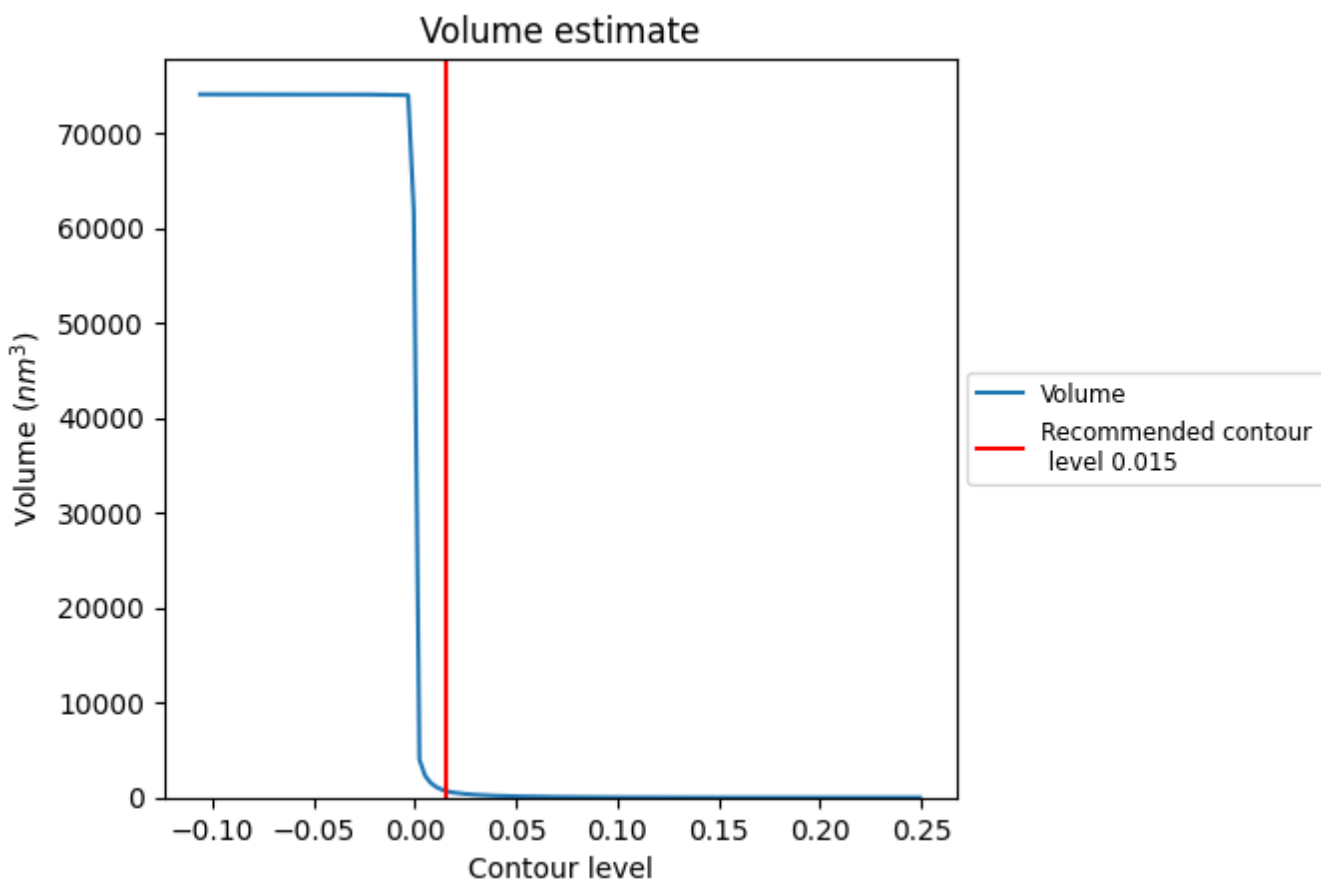
This section contains the results of statistical analysis of the map.

7.1 Map-value distribution [i](#)



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

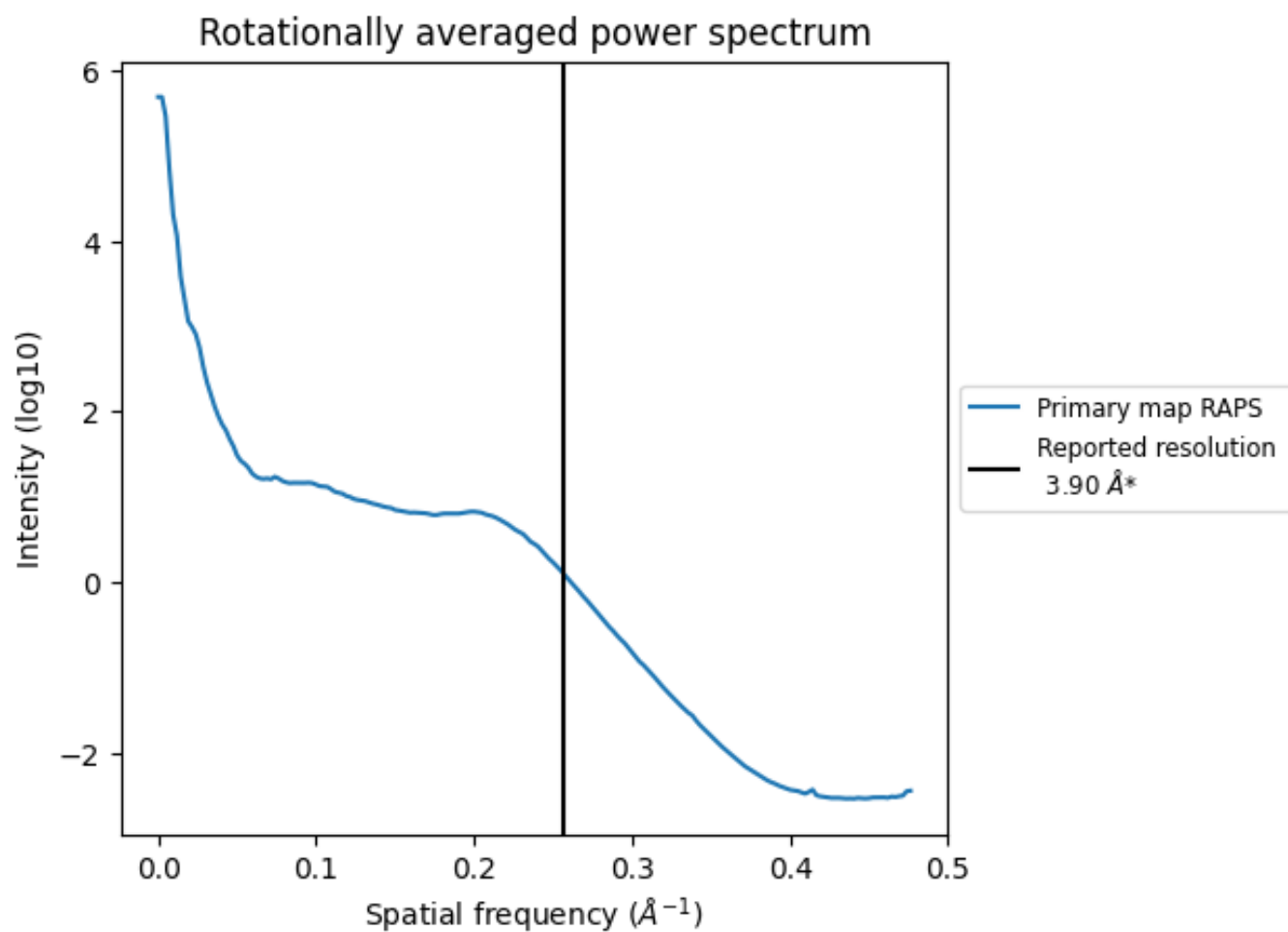
7.2 Volume estimate [\(i\)](#)



The volume at the recommended contour level is 724 nm^3 ; this corresponds to an approximate mass of 654 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](#)

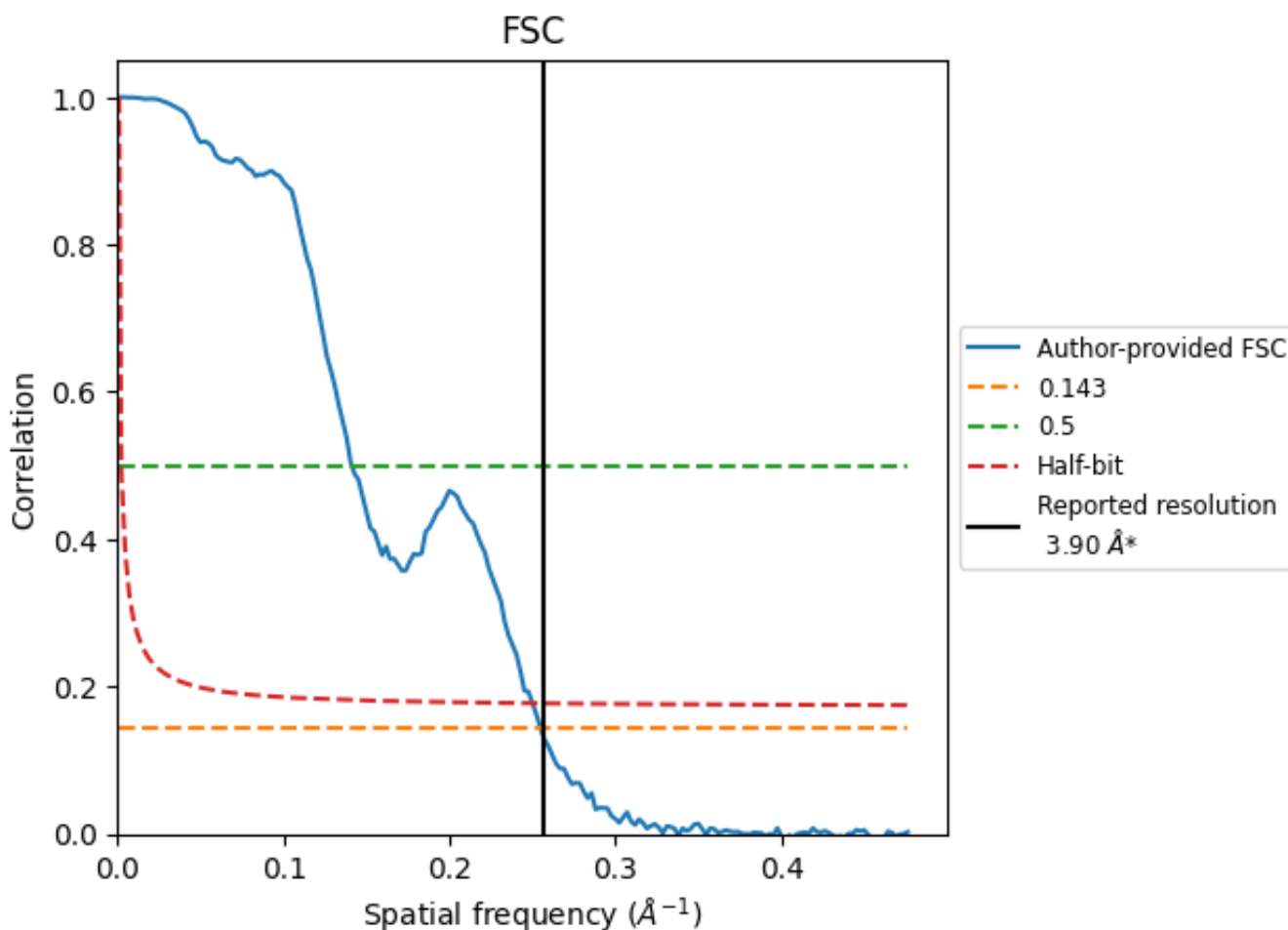


*Reported resolution corresponds to spatial frequency of 0.256 Å⁻¹

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](#)

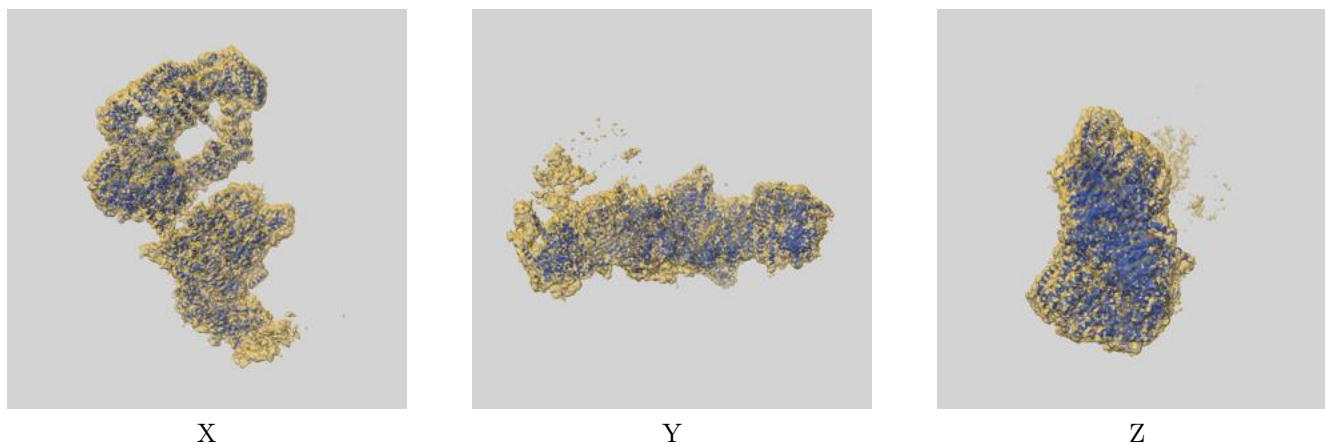
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.90	-	-
Author-provided FSC curve	3.92	7.08	4.00
Unmasked-calculated*	-	-	-

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

9 Map-model fit [i](#)

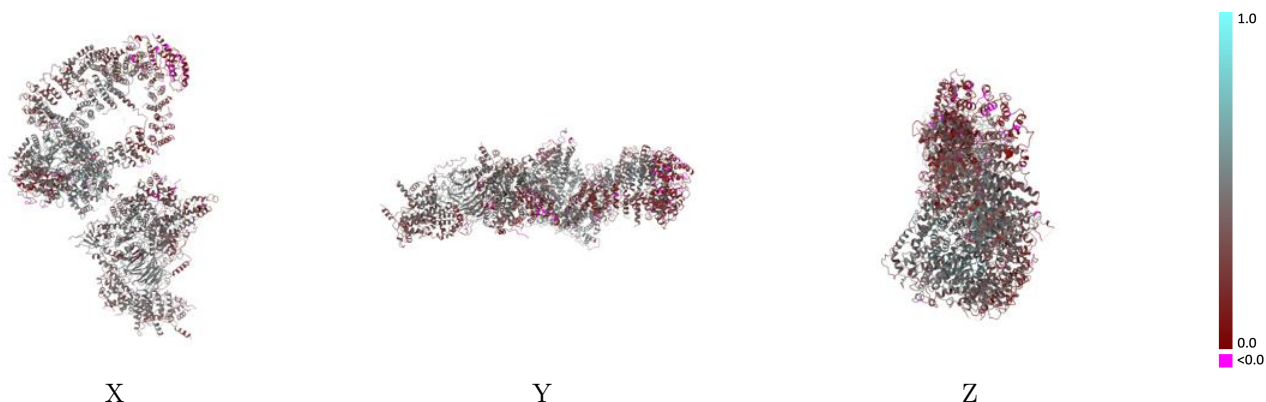
This section contains information regarding the fit between EMDB map EMD-10412 and PDB model 6T9I. Per-residue inclusion information can be found in section 3 on page 6.

9.1 Map-model overlay [i](#)



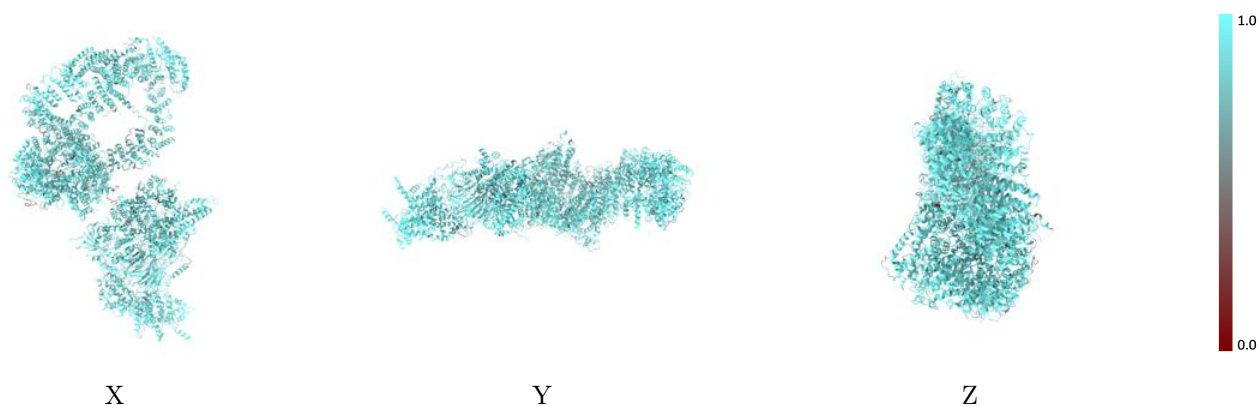
The images above show the 3D surface view of the map at the recommended contour level 0.015 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

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



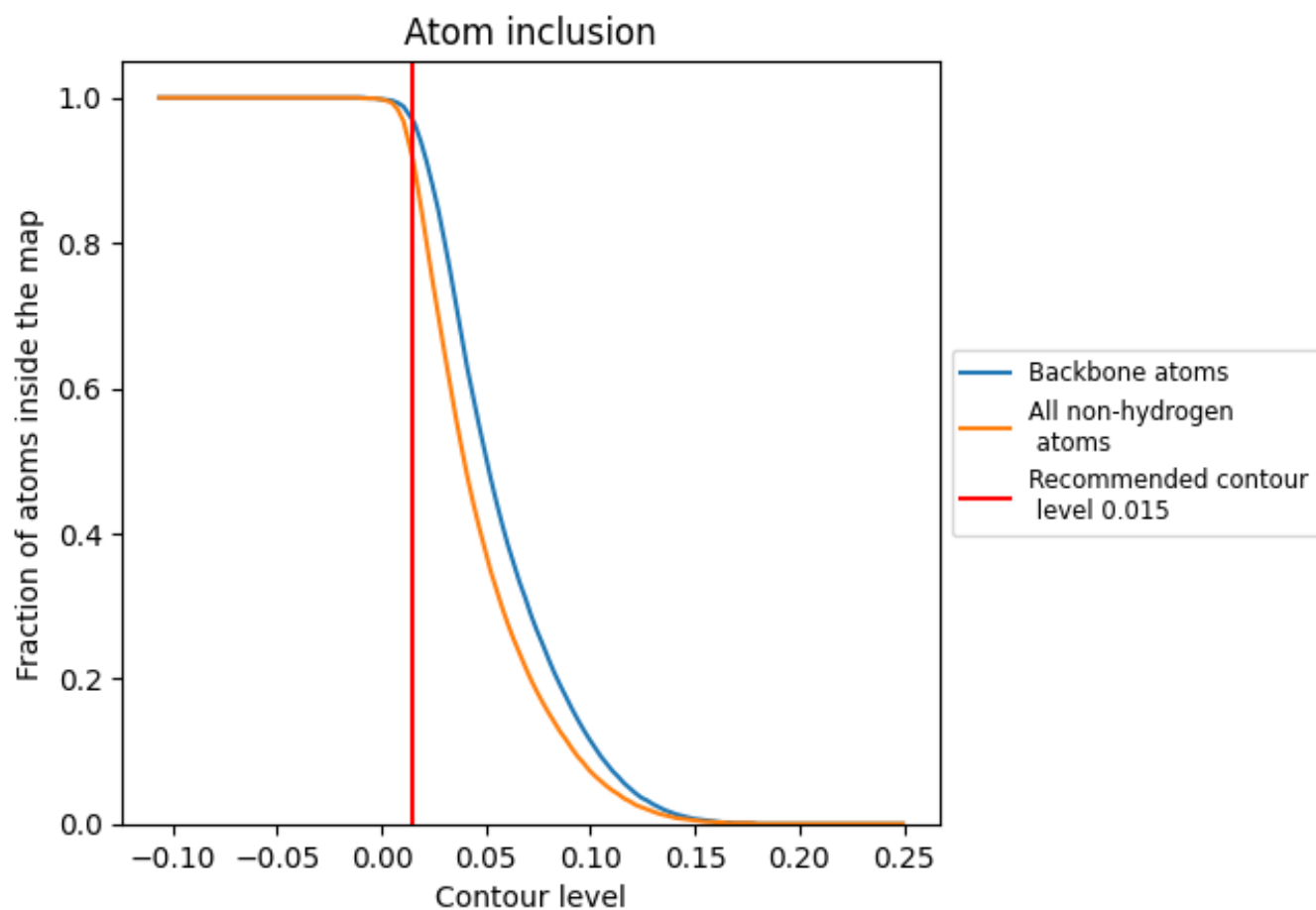
The images above show the model with each residue coloured according to its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions 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 (0.015).



























9.4 Atom inclusion [i](#)



At the recommended contour level, 97% of all backbone atoms, 92% 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 (0.015) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	 0.9160	 0.3800
B	 0.9140	 0.4350
C	 0.8820	 0.2750
D	 0.9250	 0.4160
E	 0.9230	 0.3980
F	 0.9260	 0.4540
G	 0.9610	 0.3520
H	 0.9080	 0.4100
I	 0.8770	 0.4240
K	 0.9340	 0.3490
Q	 0.9480	 0.4250
T	 0.9130	 0.3710
U	 0.9980	 0.3980

