wwPDB EM Validation Summary Report

Nov 19, 2022 – 06:32 pm GMT

PDB ID	:	5OF4
EMDB ID	:	EMD-3802
Title	:	The cryo-EM structure of human TFIIH
Authors	:	Greber, B.J.; Nguyen, T.H.D.; Fang, J.; Afonine, P.V.; Adams, P.D.; Nogales, E.
Deposited on	:	2017-07-10
Resolution	:	4.40 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:

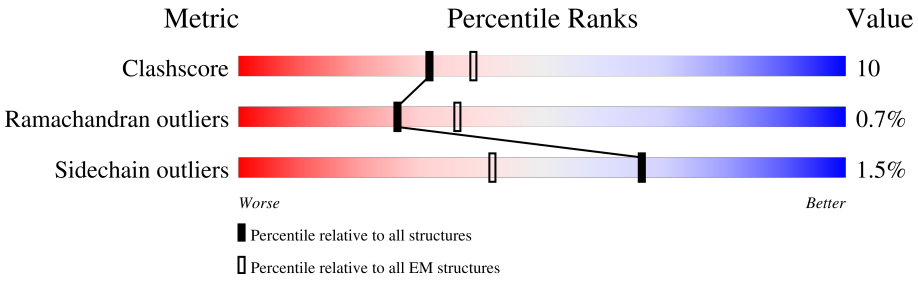
EMDB validation analysis	:	0.0.1.dev43
Mogul	:	1.8.4, CSD as541be (2020)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.9
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 4.40 Å.

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



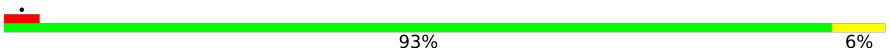
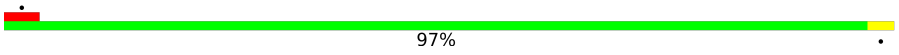
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

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

Mol	Chain	Length	Quality of chain
1	A	553	17% 61% 27% • 10%
2	B	760	13% 63% 29% • 5%
3	D	85	33% 76% 18% • 5%
4	E	395	5% 33% 13% • 53%
5	F	308	7% 54% 12% • 33%
6	G	71	25% 65% 25% • 7%
7	H	124	6% 99% •
8	Z	270	25% 97% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
9	Y	232	
10	X	78	

2 Entry composition (i)

There are 11 unique types of molecules in this entry. The entry contains 17200 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 TFIIH basal transcription factor complex helicase XPB subunit,XPB,TFIIH basal transcription factor complex helicase XPB subunit.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	499	Total	C	N	O	S	0	0
			3887	2469	676	718	24		

- Molecule 2 is a protein called TFIIH basal transcription factor complex helicase XPD subunit.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	724	Total	C	N	O	S	0	0
			5671	3626	992	1025	28		

- Molecule 3 is a protein called General transcription factor IIH subunit 4,p52,General transcription factor IIH subunit 4.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	D	81	Total	C	N	O	S	0	0
			643	412	118	112	1		

- Molecule 4 is a protein called General transcription factor IIH subunit 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	E	184	Total	C	N	O	S	0	0
			1451	918	251	272	10		

- Molecule 5 is a protein called General transcription factor IIH subunit 3.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	F	205	Total	C	N	O	S	0	0
			1498	954	257	276	11		

- Molecule 6 is a protein called General transcription factor IIH subunit 5.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	G	66	Total	C	N	O	S	0	0
			522	336	83	100	3		

- Molecule 7 is a protein called MAT1.

Mol	Chain	Residues	Atoms				AltConf	Trace
7	H	124	Total	C	N	O	0	0
			620	372	124	124		

- Molecule 8 is a protein called Unassigned secondary structure elements..

Mol	Chain	Residues	Atoms				AltConf	Trace
8	Z	270	Total	C	N	O	0	0
			1350	810	270	270		

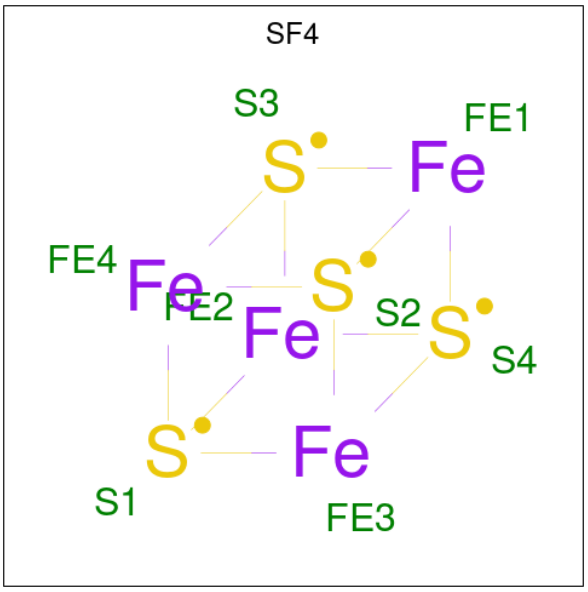
- Molecule 9 is a protein called Unassigned secondary structure elements (p52 region).

Mol	Chain	Residues	Atoms				AltConf	Trace
9	Y	232	Total	C	N	O	0	0
			1160	696	232	232		

- Molecule 10 is a protein called Unassigned secondary structure elements (XPB NTE region).

Mol	Chain	Residues	Atoms				AltConf	Trace
10	X	78	Total	C	N	O	0	0
			390	234	78	78		

- Molecule 11 is IRON/SULFUR CLUSTER (three-letter code: SF4) (formula: Fe_4S_4).

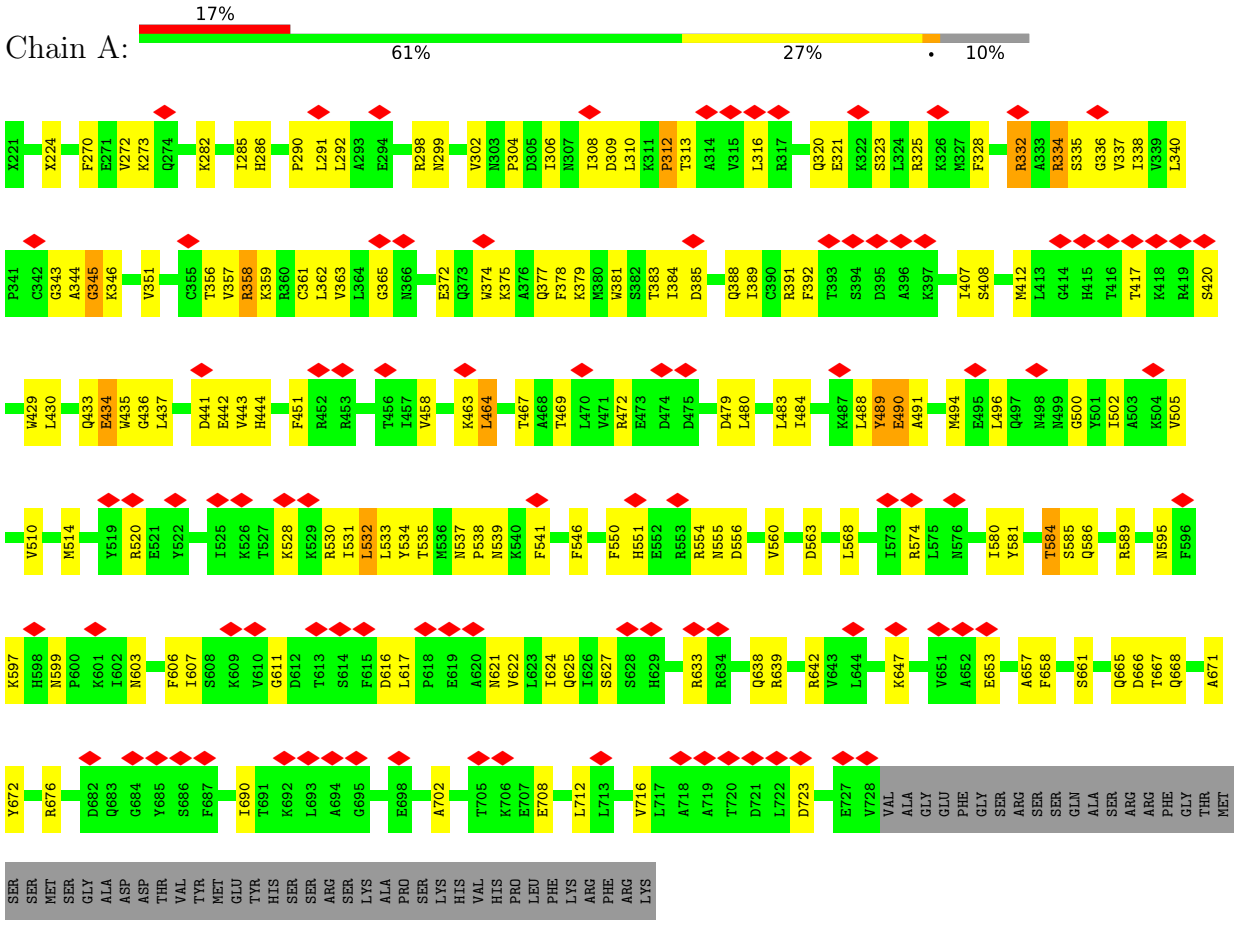


Mol	Chain	Residues	Atoms			AltConf
			Total	Fe	S	
11	B	1	8	4	4	0

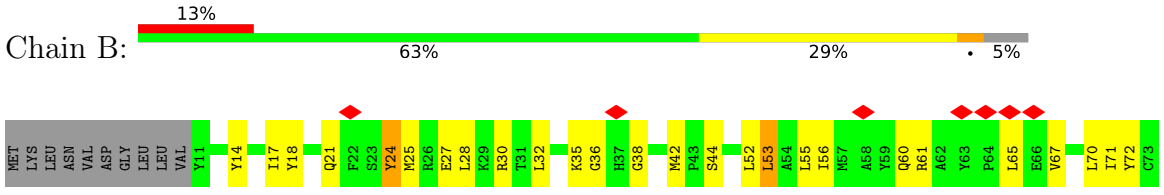
3 Residue-property plots (i)

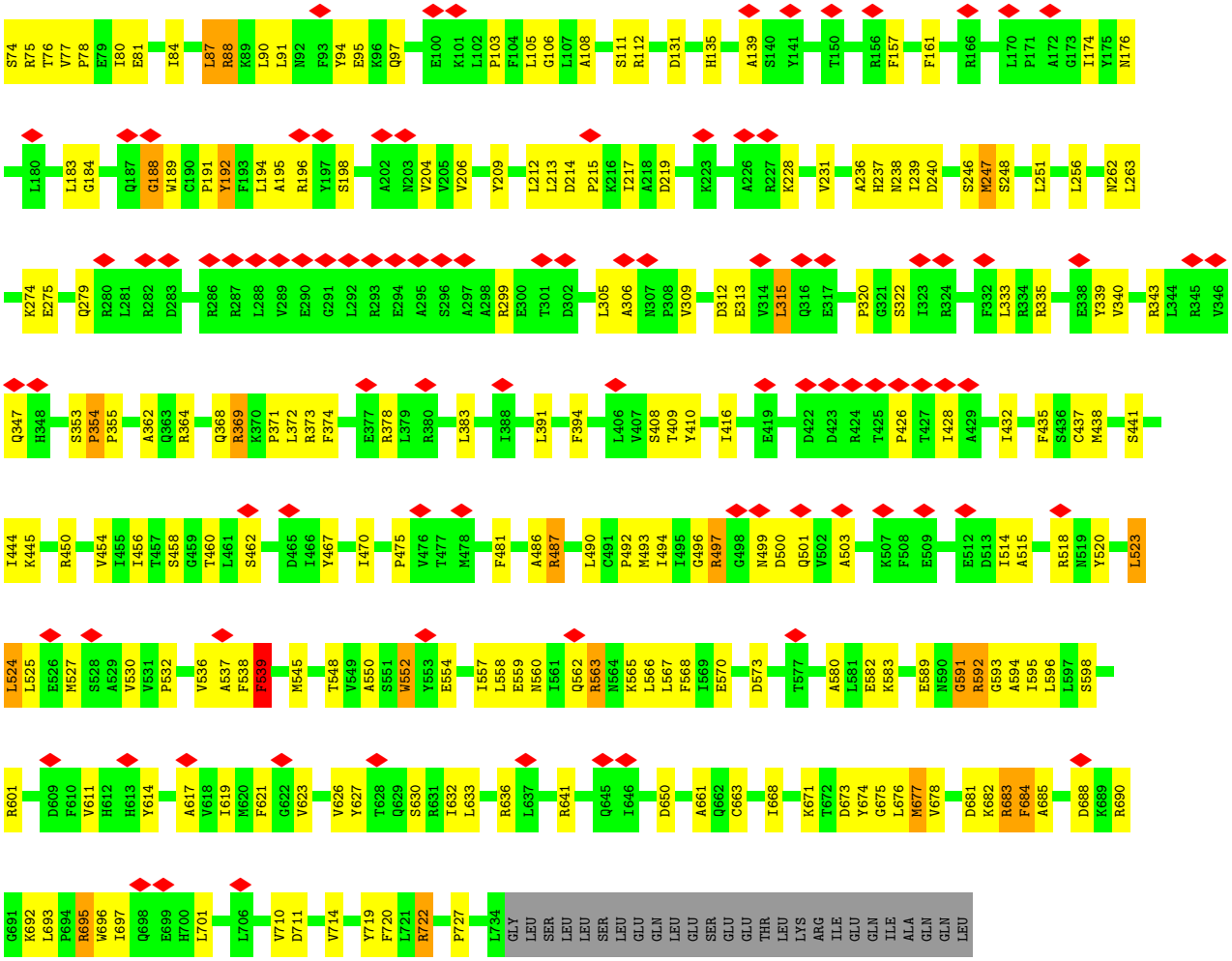
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and atom inclusion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

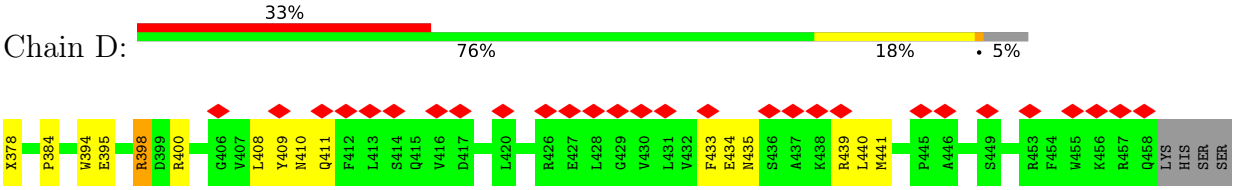


• Molecule 2: TFIIH basal transcription factor complex helicase XPD subunit

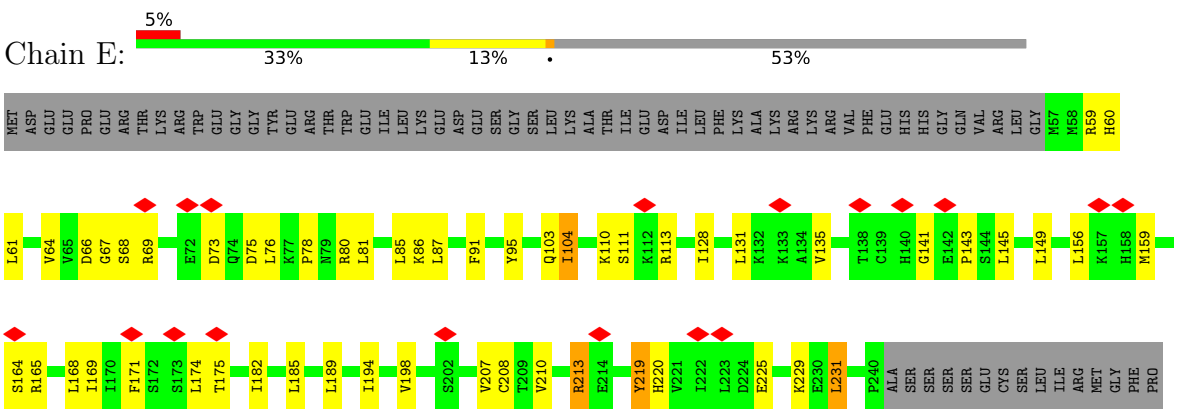


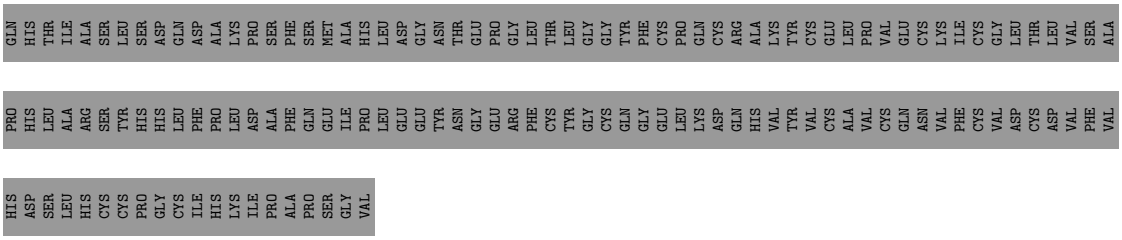


• Molecule 3: General transcription factor IIH subunit 4,p52,General transcription factor IIH subunit 4

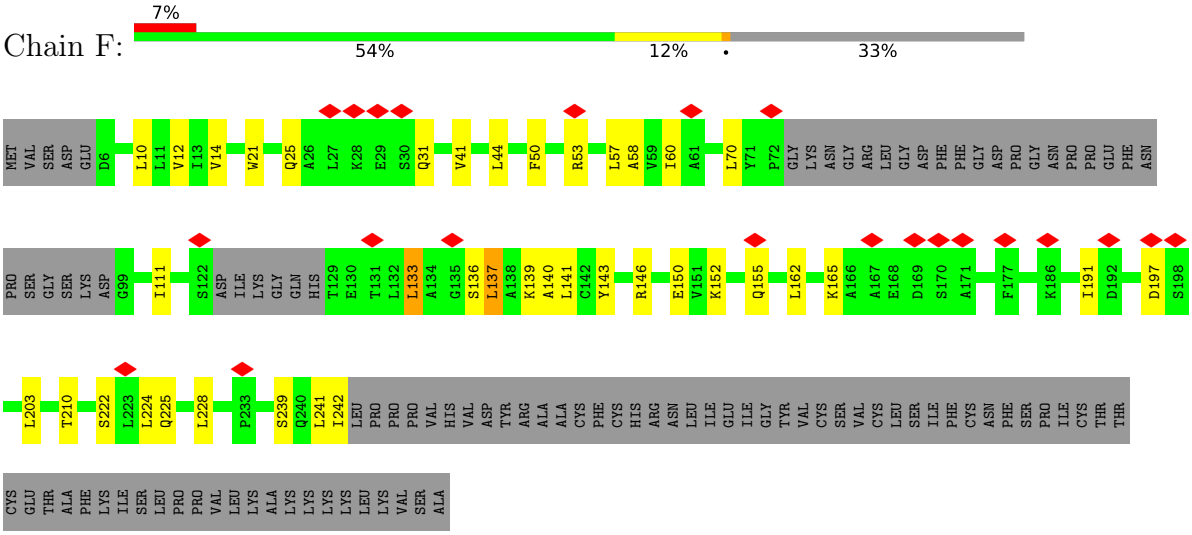


• Molecule 4: General transcription factor IIH subunit 2

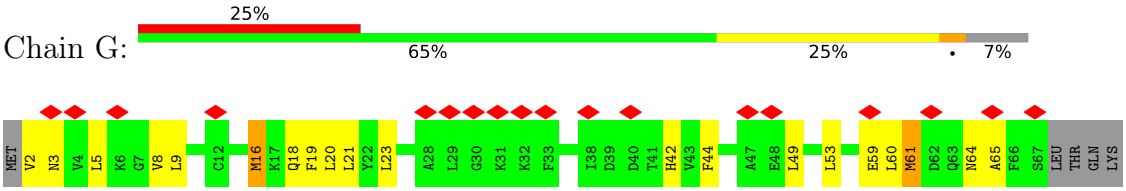




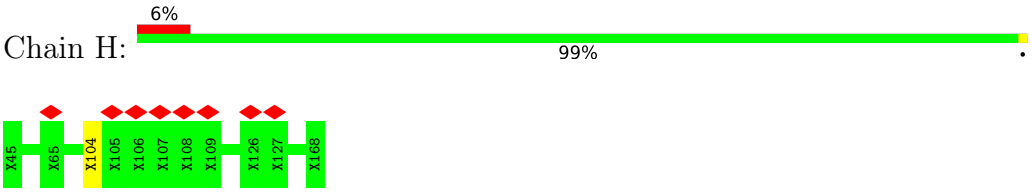
• Molecule 5: General transcription factor IIH subunit 3



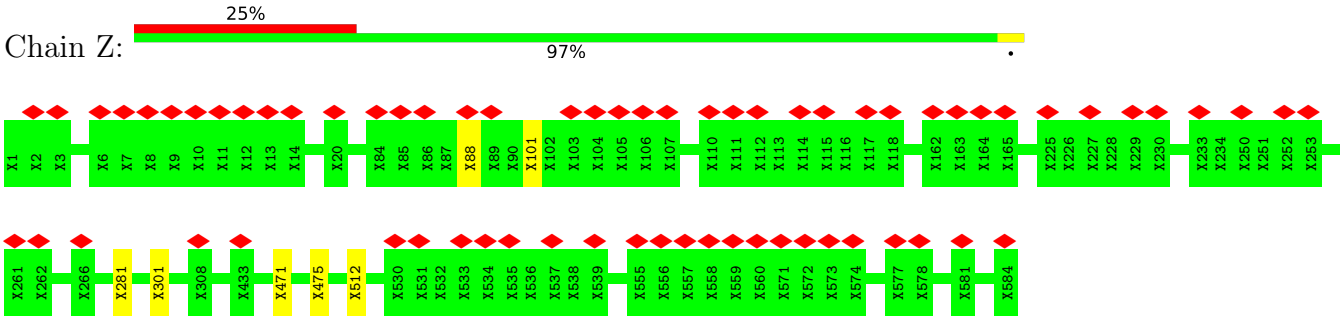
• Molecule 6: General transcription factor IIH subunit 5



• Molecule 7: MAT1

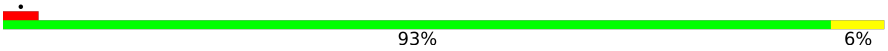


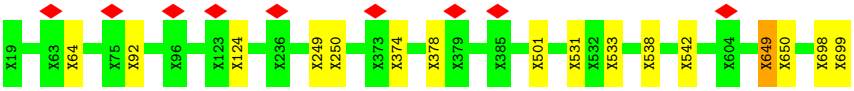
• Molecule 8: Unassigned secondary structure elements.



• Molecule 9: Unassigned secondary structure elements (p52 region)

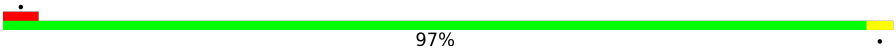
Chain Y:

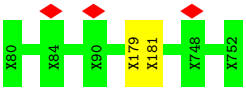




• Molecule 10: Unassigned secondary structure elements (XPB NTE region)

Chain X:





4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	122900	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION; Correction in RELION based on values determined in CTFFIND4.	Depositor
Microscope	FEI TITAN	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	40	Depositor
Minimum defocus (nm)	2000	Depositor
Maximum defocus (nm)	4500	Depositor
Magnification	37879	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	0.129	Depositor
Minimum map value	-0.075	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.004	Depositor
Recommended contour level	0.03	Depositor
Map size (Å)	337.92, 337.92, 337.92	wwPDB
Map dimensions	256, 256, 256	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.32, 1.32, 1.32	Depositor

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: SF4

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	1.00	7/3779 (0.2%)	1.07	17/5102 (0.3%)
2	B	1.08	11/5790 (0.2%)	1.22	44/7853 (0.6%)
3	D	0.92	1/631 (0.2%)	0.96	2/855 (0.2%)
4	E	1.13	3/1478 (0.2%)	1.24	10/2001 (0.5%)
5	F	1.02	1/1518 (0.1%)	1.11	12/2061 (0.6%)
6	G	0.72	0/528	0.86	2/713 (0.3%)
All	All	1.04	23/13724 (0.2%)	1.15	87/18585 (0.5%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	6
2	B	0	9
3	D	0	1
6	G	0	1
7	H	0	1
8	Z	0	2
9	Y	0	3
All	All	0	23

The worst 5 of 23 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	489	TYR	C-O	-9.39	1.05	1.23
3	D	394	TRP	CE3-CZ3	9.33	1.54	1.38
4	E	91	PHE	CG-CD1	7.93	1.50	1.38
2	B	591	GLY	C-O	7.24	1.35	1.23
2	B	539	PHE	N-CA	-7.20	1.31	1.46

The worst 5 of 87 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	E	207	VAL	CG1-CB-CG2	10.98	128.47	110.90
2	B	196	ARG	NE-CZ-NH1	10.96	125.78	120.30
1	A	639	ARG	NE-CZ-NH2	-9.64	115.48	120.30
2	B	684	PHE	CB-CG-CD1	9.34	127.34	120.80
1	A	332	ARG	NE-CZ-NH1	8.44	124.52	120.30

There are no chirality outliers.

5 of 23 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	272	VAL	Peptide
1	A	312	PRO	Peptide
1	A	328	PHE	Peptide
1	A	412	MET	Peptide
1	A	434	GLU	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	A	3887	0	3750	94	0
2	B	5671	0	5575	140	0
3	D	643	0	622	14	0
4	E	1451	0	1492	32	0
5	F	1498	0	1414	18	0
6	G	522	0	528	14	0
7	H	620	0	131	0	0
8	Z	1350	0	315	4	0
9	Y	1160	0	279	10	0
10	X	390	0	96	1	0
11	B	8	0	0	0	0
All	All	17200	0	14202	301	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 10.

The worst 5 of 301 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:B:106:GLY:HA2	2:B:204:VAL:O	1.71	0.91
1:A:671:ALA:HB2	6:G:65:ALA:HB1	1.56	0.88
1:A:224:UNK:O	1:A:292:LEU:HA	1.74	0.87
1:A:429:TRP:O	1:A:433:GLN:HB2	1.75	0.85
3:D:408:LEU:HB3	6:G:5:LEU:HD12	1.59	0.84

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	461/553 (83%)	402 (87%)	57 (12%)	2 (0%)	34	72
2	B	722/760 (95%)	614 (85%)	101 (14%)	7 (1%)	15	54
3	D	75/85 (88%)	71 (95%)	3 (4%)	1 (1%)	12	48
4	E	182/395 (46%)	171 (94%)	11 (6%)	0	100	100
5	F	199/308 (65%)	182 (92%)	15 (8%)	2 (1%)	15	54
6	G	64/71 (90%)	56 (88%)	8 (12%)	0	100	100
All	All	1703/2172 (78%)	1496 (88%)	195 (12%)	12 (1%)	26	62

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	313	THR
2	B	428	ILE
2	B	592	ARG
2	B	14	TYR
2	B	685	ALA

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	400/454 (88%)	395 (99%)	5 (1%)	69 82
2	B	586/664 (88%)	576 (98%)	10 (2%)	60 78
3	D	64/73 (88%)	64 (100%)	0	100 100
4	E	169/352 (48%)	165 (98%)	4 (2%)	49 69
5	F	144/272 (53%)	142 (99%)	2 (1%)	67 81
6	G	59/64 (92%)	59 (100%)	0	100 100
All	All	1422/1879 (76%)	1401 (98%)	21 (2%)	66 80

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	722	ARG
4	E	174	LEU
5	F	133	LEU
4	E	194	ILE
4	E	164	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
5	F	176	ASN
4	E	235	HIS
1	A	665	GLN
4	E	60	HIS
1	A	638	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)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
11	SF4	B	1000	2	0,12,12	-	-	-		

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
11	SF4	B	1000	2	-	-	0/6/5/5

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)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
9	Y	18
8	Z	15
10	X	6
1	A	1

The worst 5 of 40 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	Z	181:UNK	C	201:UNK	N	103.83
1	Z	165:UNK	C	171:UNK	N	91.40
1	Z	20:UNK	C	82:UNK	N	60.92
1	Y	320:UNK	C	373:UNK	N	57.49
1	Z	487:UNK	C	501:UNK	N	56.59

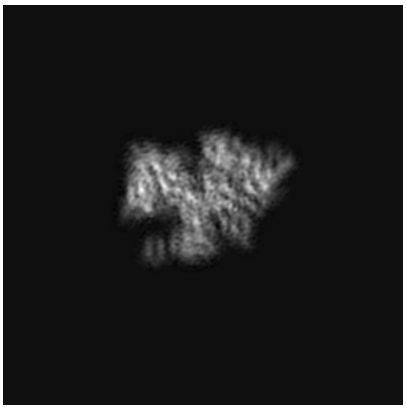
6 Map visualisation

This section contains visualisations of the EMDB entry EMD-3802. These allow visual inspection of the internal detail of the map and identification of artifacts.

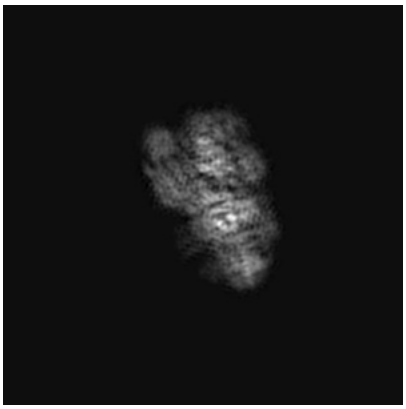
Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

6.1 Orthogonal projections

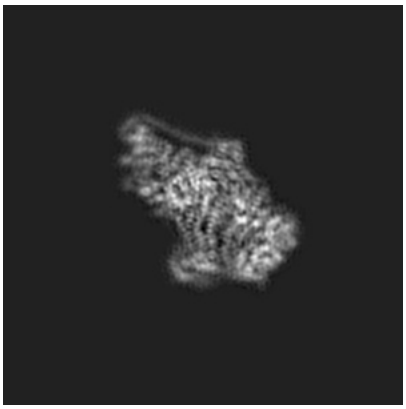
6.1.1 Primary map



X

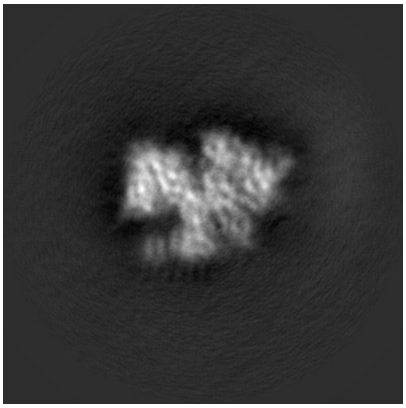


Y

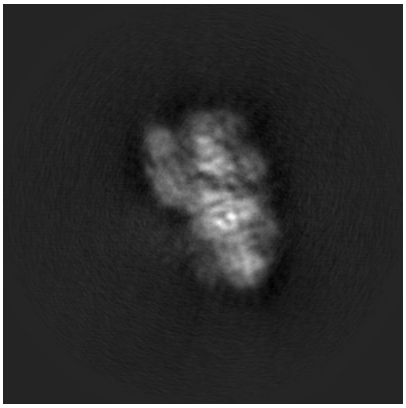


Z

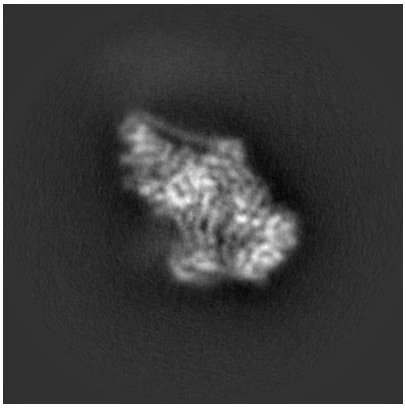
6.1.2 Raw map



X



Y

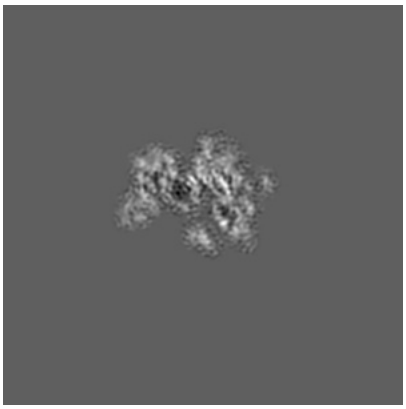


Z

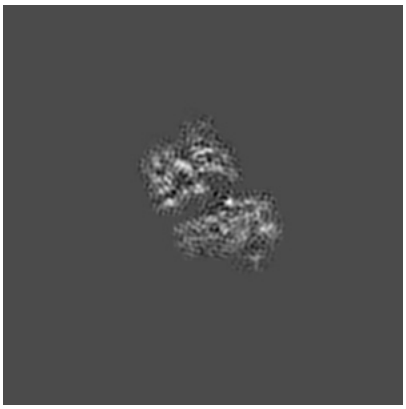
The images above show the map projected in three orthogonal directions.

6.2 Central slices (i)

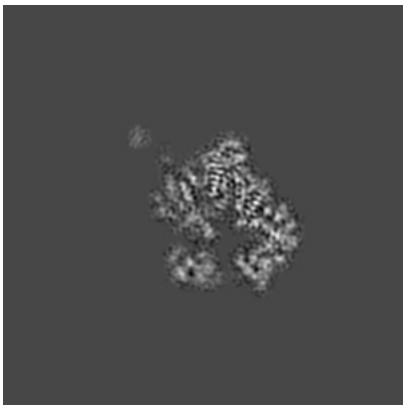
6.2.1 Primary map



X Index: 128

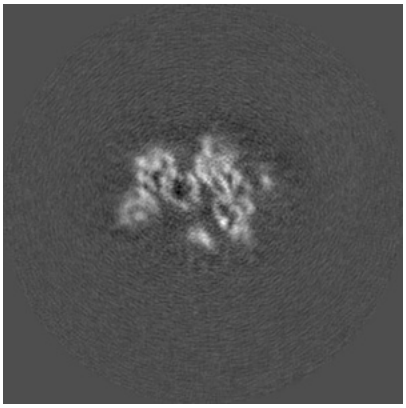


Y Index: 128

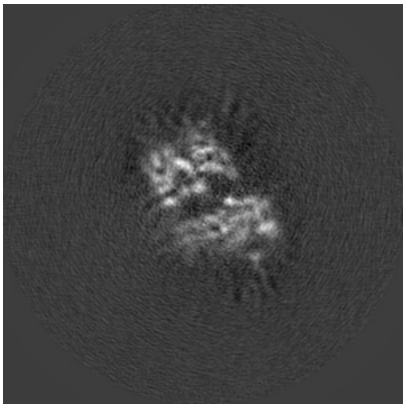


Z Index: 128

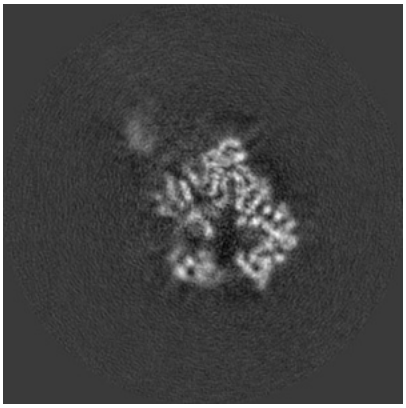
6.2.2 Raw map



X Index: 128



Y Index: 128

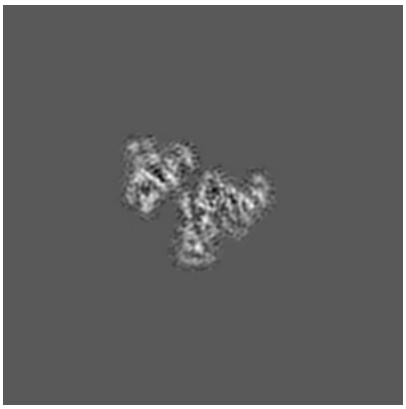


Z Index: 128

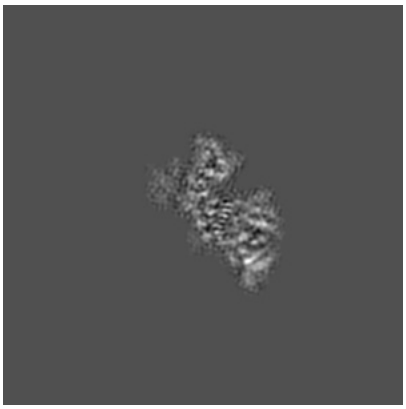
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

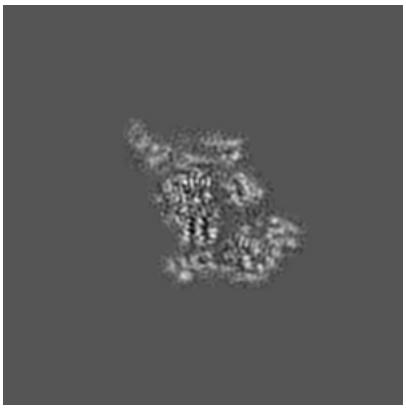
6.3.1 Primary map



X Index: 151

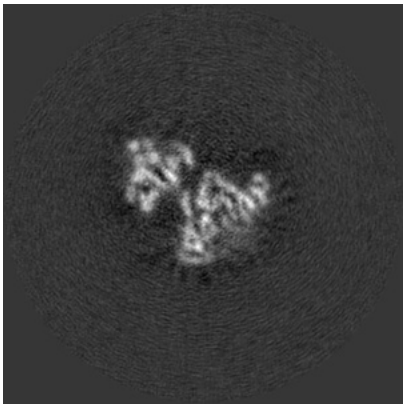


Y Index: 138

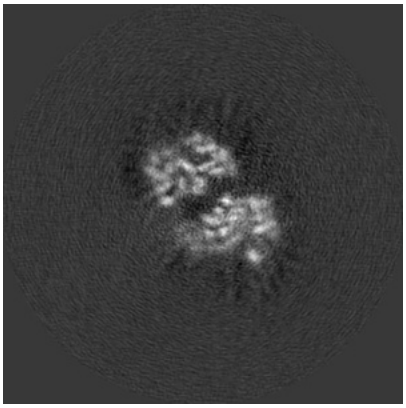


Z Index: 138

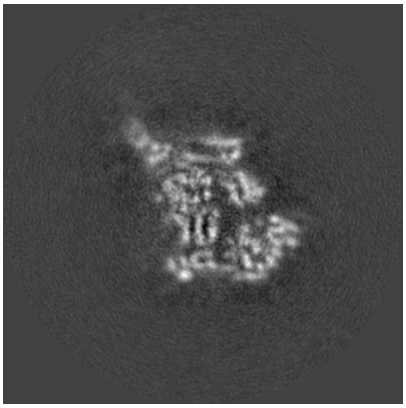
6.3.2 Raw map



X Index: 151



Y Index: 130



Z Index: 138

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 0.03. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

6.4.2 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.

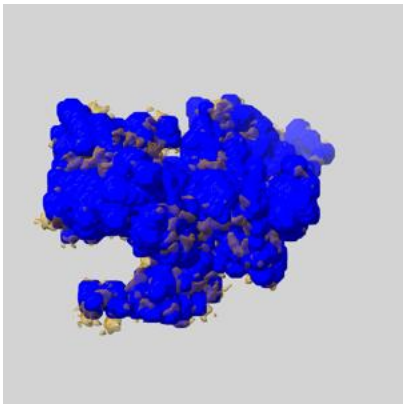
6.5 Mask visualisation (i)

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

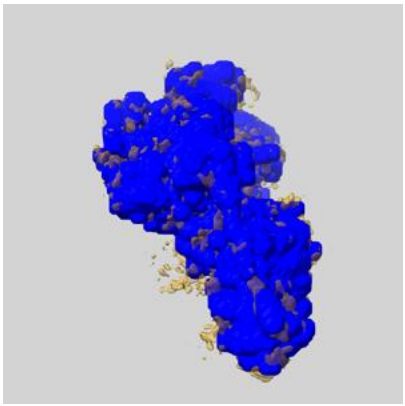
A mask typically either:

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

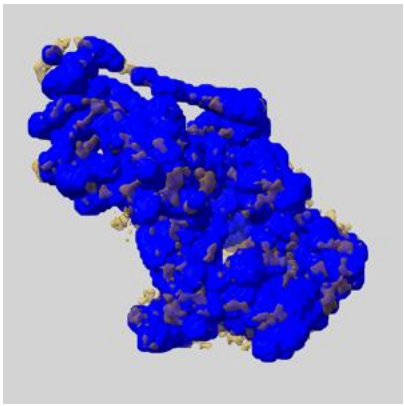
6.5.1 emd_3802_msk_1.map (i)



X



Y

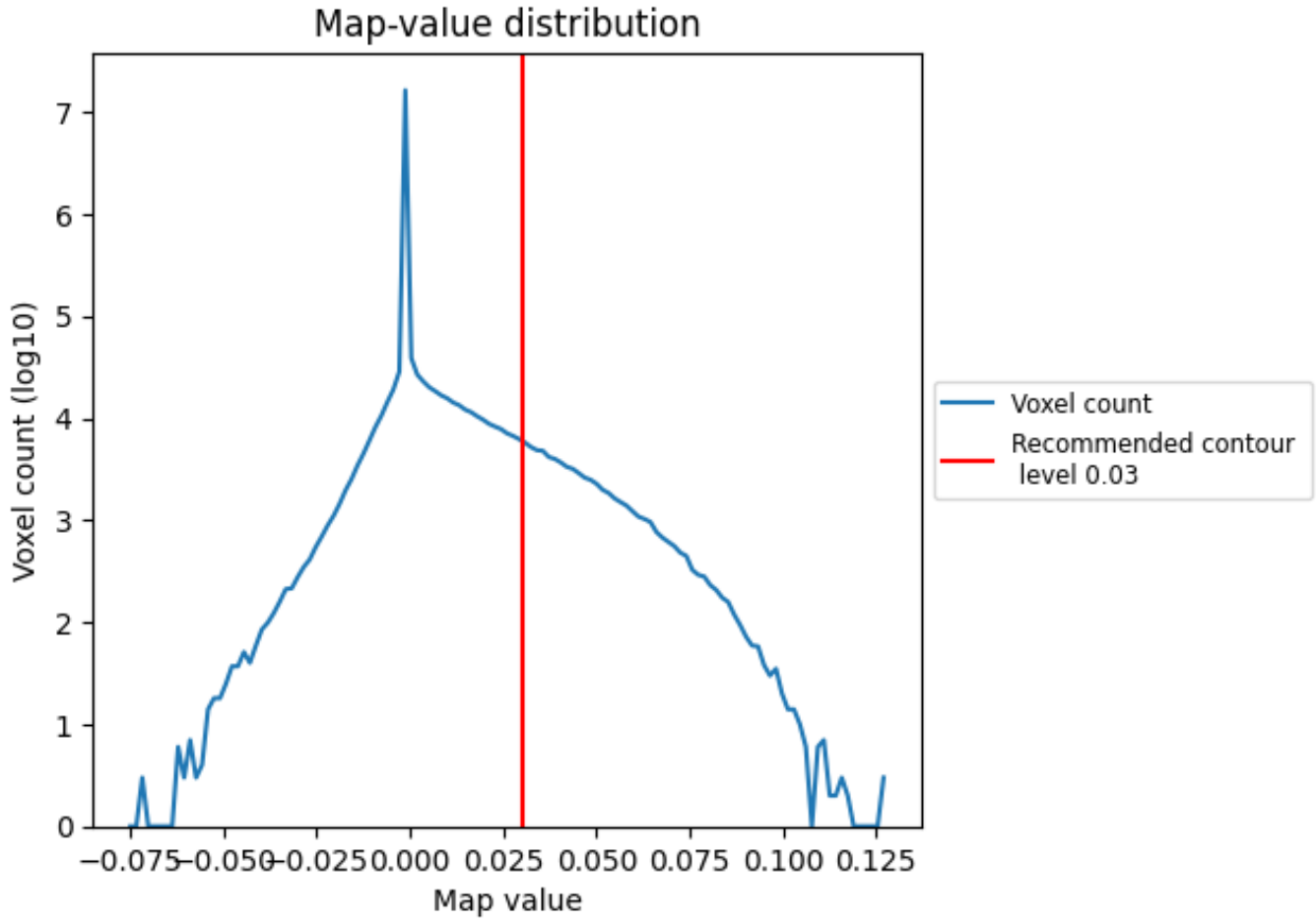


Z

7 Map analysis

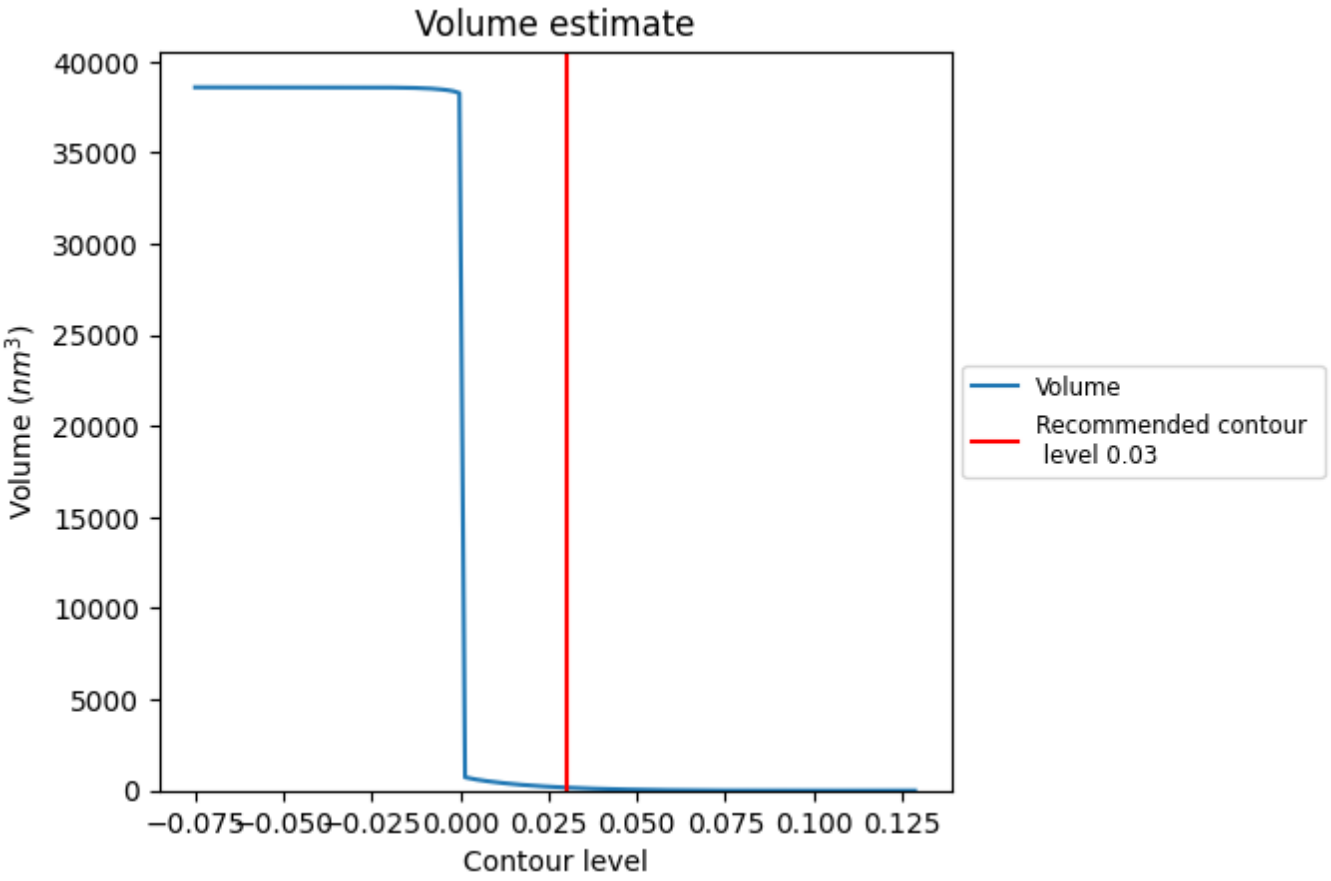
This section contains the results of statistical analysis of the map.

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.

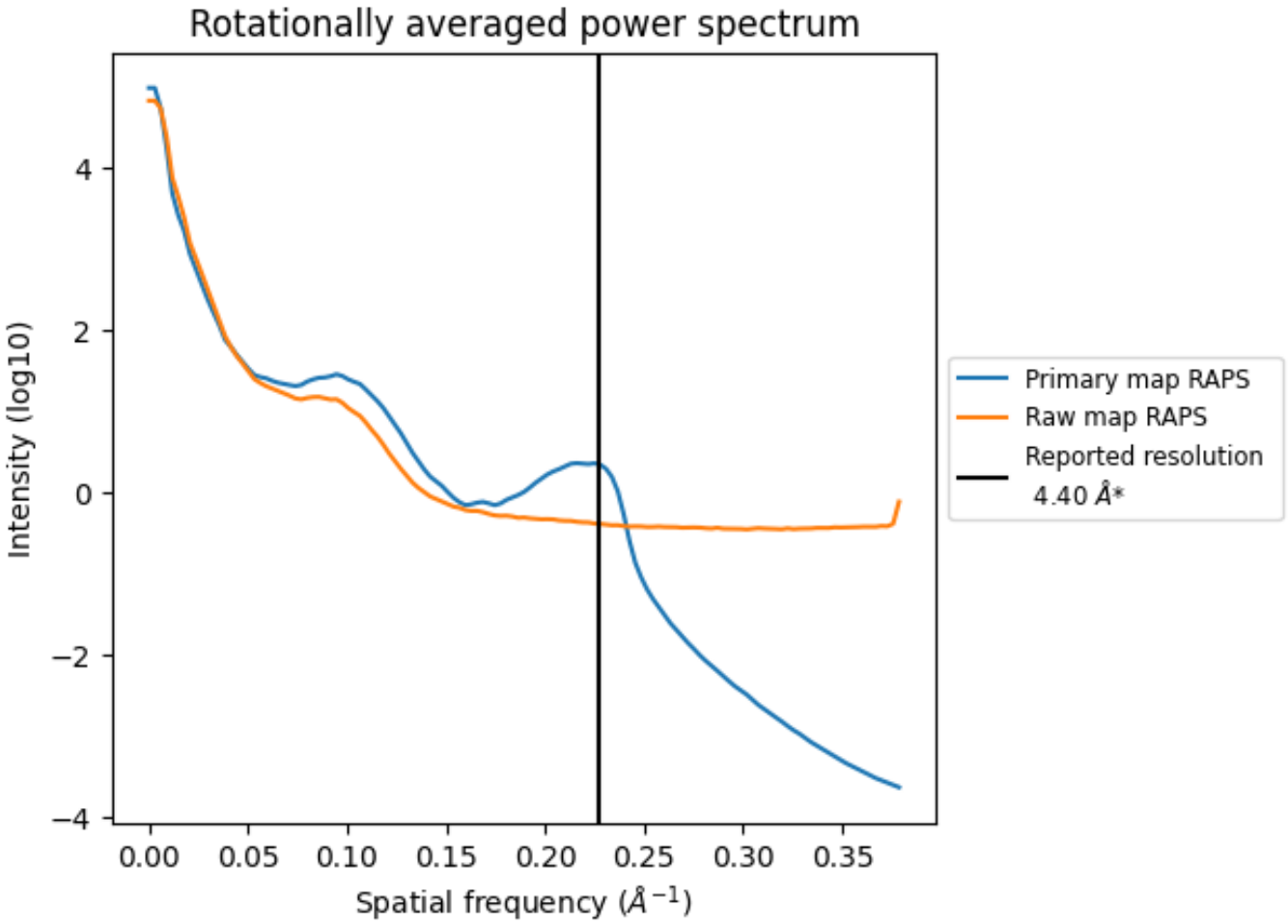
7.2 Volume estimate (i)



The volume at the recommended contour level is 163 nm^3; this corresponds to an approximate mass of 147 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)

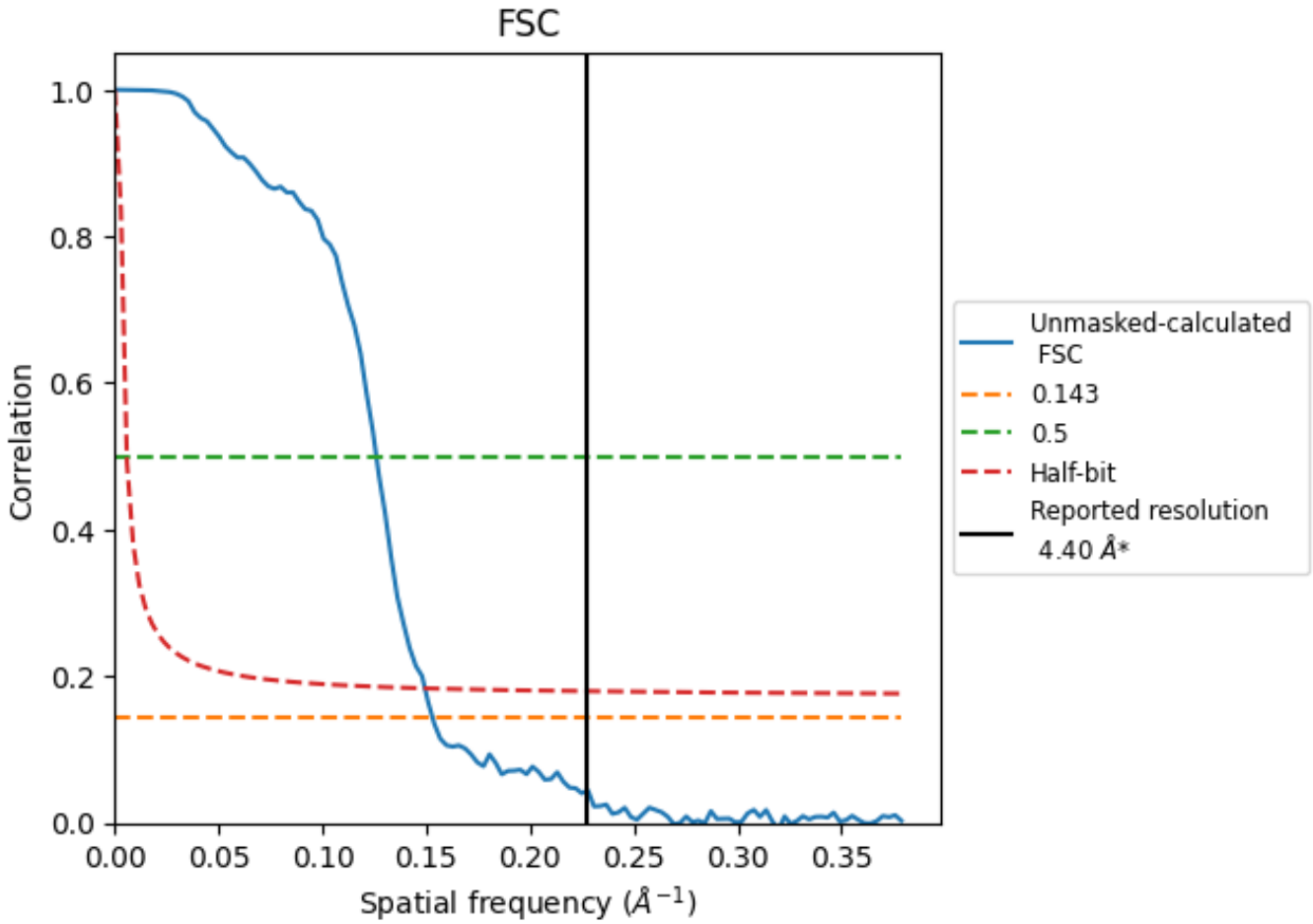


*Reported resolution corresponds to spatial frequency of 0.227 Å^{-1}

8 Fourier-Shell correlation

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



*Reported resolution corresponds to spatial frequency of 0.227 Å^{-1}

8.2 Resolution estimates (i)

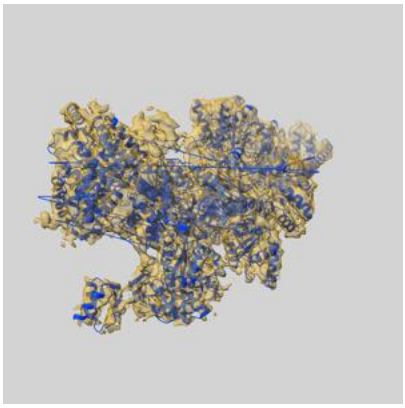
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	4.40	-	-
Author-provided FSC curve	-	-	-
Unmasked-calculated*	6.53	7.94	6.69

*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 6.53 differs from the reported value 4.4 by more than 10 %

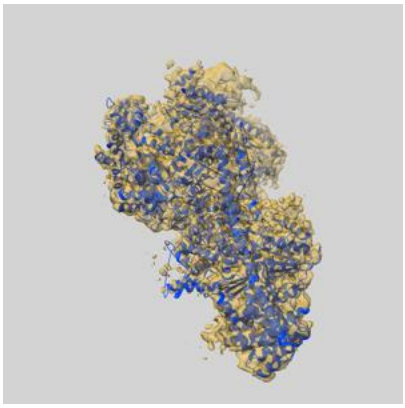
9 Map-model fit

This section contains information regarding the fit between EMDB map EMD-3802 and PDB model 5OF4. Per-residue inclusion information can be found in section 3 on page 7.

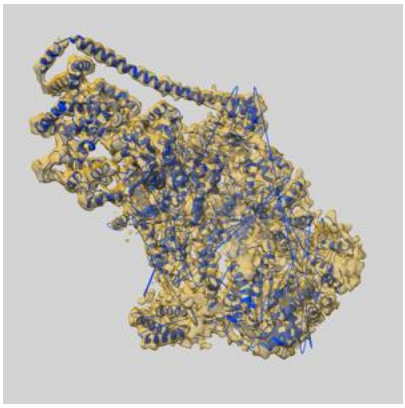
9.1 Map-model overlay



X



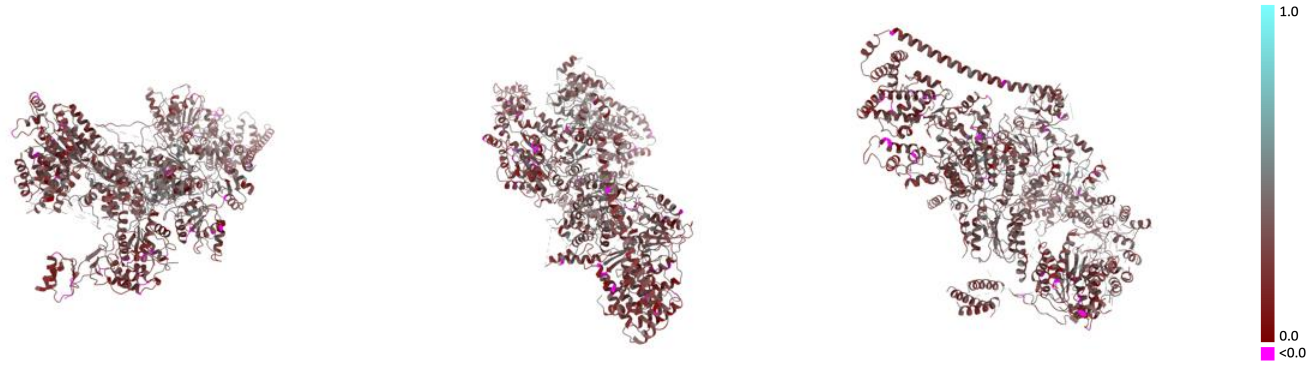
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.03 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.

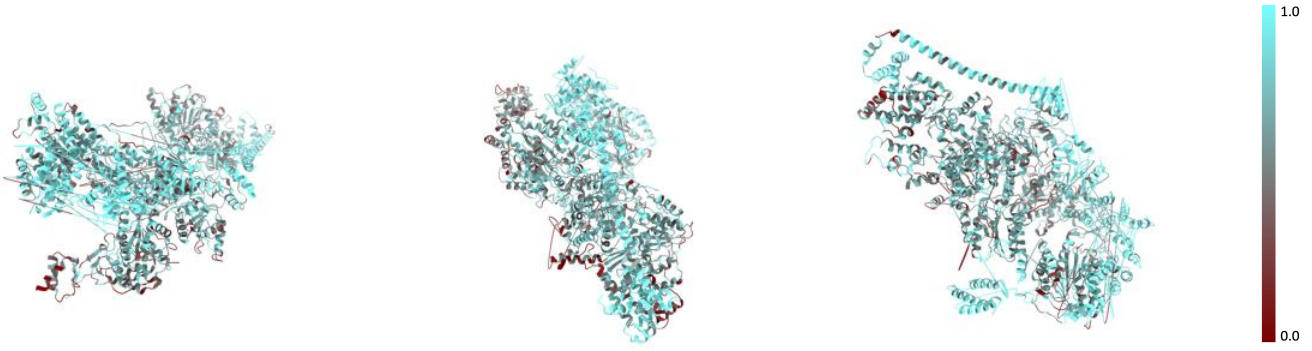
9.2 Q-score mapped to coordinate model (i)



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.

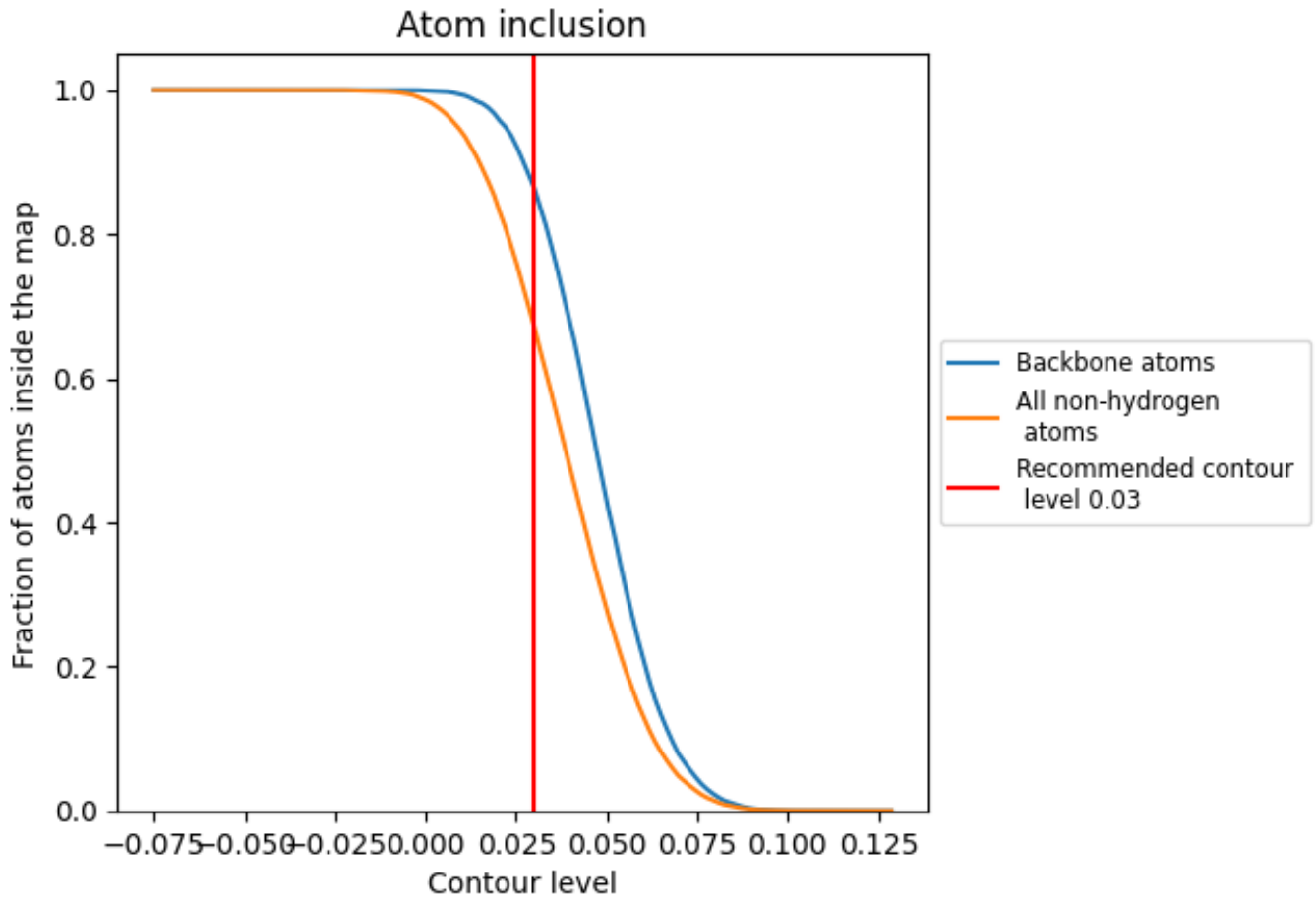
9.3 Atom inclusion mapped to coordinate model (i)



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 (0.03).

9.4 Atom inclusion (i)



At the recommended contour level, 86% of all backbone atoms, 67% 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 (0.03) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.6714	0.3080
A	0.6120	0.2980
B	0.6469	0.3020
D	0.5353	0.2680
E	0.6690	0.3630
F	0.6802	0.2860
G	0.5183	0.2640
H	0.8452	0.2810
X	0.9128	0.3870
Y	0.9009	0.3510
Z	0.7081	0.3200