



## wwPDB EM Validation Summary Report ⓘ

Aug 26, 2021 – 06:20 am BST

PDB ID : 7OKY  
EMDB ID : EMD-12973  
Title : Structure of active transcription elongation complex Pol II-DSIF-ELL2-EAF1  
(composite structure)  
Authors : Chen, Y.; Vos, S.M.; Dienemann, C.; Ninov, M.; Urlaub, H.; Cramer, P.  
Deposited on : 2021-05-18  
Resolution : 4.14 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](mailto: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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

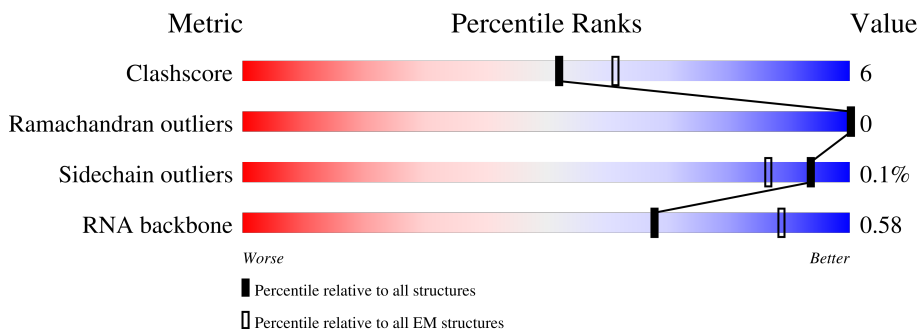
EMDB validation analysis : 0.0.0.dev97  
MolProbity : 4.02b-467  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.23.1

# 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 4.14 Å.

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



Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826
RNA backbone	4643	859

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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	A	1970	60% (green), 11% (yellow), 29% (grey)
2	B	1251	76% (green), 14% (yellow), 10% (grey)
3	C	275	77% (green), 16% (yellow), 6% (grey)
4	D	184	57% (green), 8% (yellow), 36% (grey); 7% (red) above bar
5	E	210	90% (green), 10% (yellow)
6	F	127	55% (green), 6% (yellow), 39% (grey); 7% (red) above bar
7	G	172	81% (green), 18% (yellow), 1% (grey); 7% (red) above bar

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
8	H	150	81% 17%
9	I	125	75% 18% 7%
10	J	67	81% 18%
11	K	117	92% 6%
12	L	58	69% 7% 24%
13	M	640	18% 78%
14	N	48	40% 17% 44%
15	O	268	34% 5% 60%
16	P	47	21% 6% 72%
17	T	48	48% 33% 19%
18	Z	1087	8% 28% 6% 66%
19	Y	117	21% 79% 20%

## 2 Entry composition [i](#)

There are 21 unique types of molecules in this entry. The entry contains 38341 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 DNA-directed RNA polymerase II subunit RPB1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	1407	11149	7018	1998	2062	71	0	0

- Molecule 2 is a protein called DNA-directed RNA polymerase subunit beta.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	B	1131	9047	5721	1592	1670	64	0	0

- Molecule 3 is a protein called DNA-directed RNA polymerase II subunit RPB3.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	C	258	2072	1300	356	410	6	0	0

- Molecule 4 is a protein called RNA polymerase II subunit D.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	D	118	967	608	167	188	4	0	0

- Molecule 5 is a protein called DNA-directed RNA polymerase II subunit E.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
5	E	209	1721	1089	300	324	8	0	0

- Molecule 6 is a protein called DNA-directed RNA polymerase II subunit F.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
6	F	78	626	401	106	114	5	0	0

- Molecule 7 is a protein called DNA-directed RNA polymerase II subunit RPB7.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
7	G	171	1347	872	218	249	8	0	0

- Molecule 8 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC3.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
8	H	148	1186	750	194	237	5	0	0

- Molecule 9 is a protein called DNA-directed RNA polymerase II subunit RPB9.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
9	I	116	932	577	165	179	11	0	0

- Molecule 10 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC5.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
10	J	66	524	339	88	91	6	0	0

- Molecule 11 is a protein called RNA\_pol\_L\_2 domain-containing protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
11	K	115	920	593	152	173	2	0	0

- Molecule 12 is a protein called RNA polymerase II subunit K.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
12	L	44	367	228	69	64	6	0	0

- Molecule 13 is a protein called RNA polymerase II elongation factor ELL2.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
13	M	139	1106	686	201	213	6	0	0

- Molecule 14 is a DNA chain called Non-template DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
14	N	27	Total	C	N	O	P	0	0
			549	264	96	162	27		

- Molecule 15 is a protein called ELL-associated factor 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
15	O	106	Total	C	N	O	S	0	0
			832	523	149	156	4		

- Molecule 16 is a RNA chain called RNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
16	P	13	Total	C	N	O	P	0	0
			279	125	54	87	13		

- Molecule 17 is a DNA chain called Template DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
17	T	39	Total	C	N	O	P	0	0
			808	385	149	235	39		

- Molecule 18 is a protein called Transcription elongation factor SPT5.

Mol	Chain	Residues	Atoms					AltConf	Trace
18	Z	374	Total	C	N	O	S	2	0
			2979	1889	527	548	15		

- Molecule 19 is a protein called Transcription elongation factor SPT4.

Mol	Chain	Residues	Atoms					AltConf	Trace
19	Y	116	Total	C	N	O	S	3	0
			920	578	159	174	9		

- Molecule 20 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		AltConf
20	A	1	Total	Mg	0
			1	1	

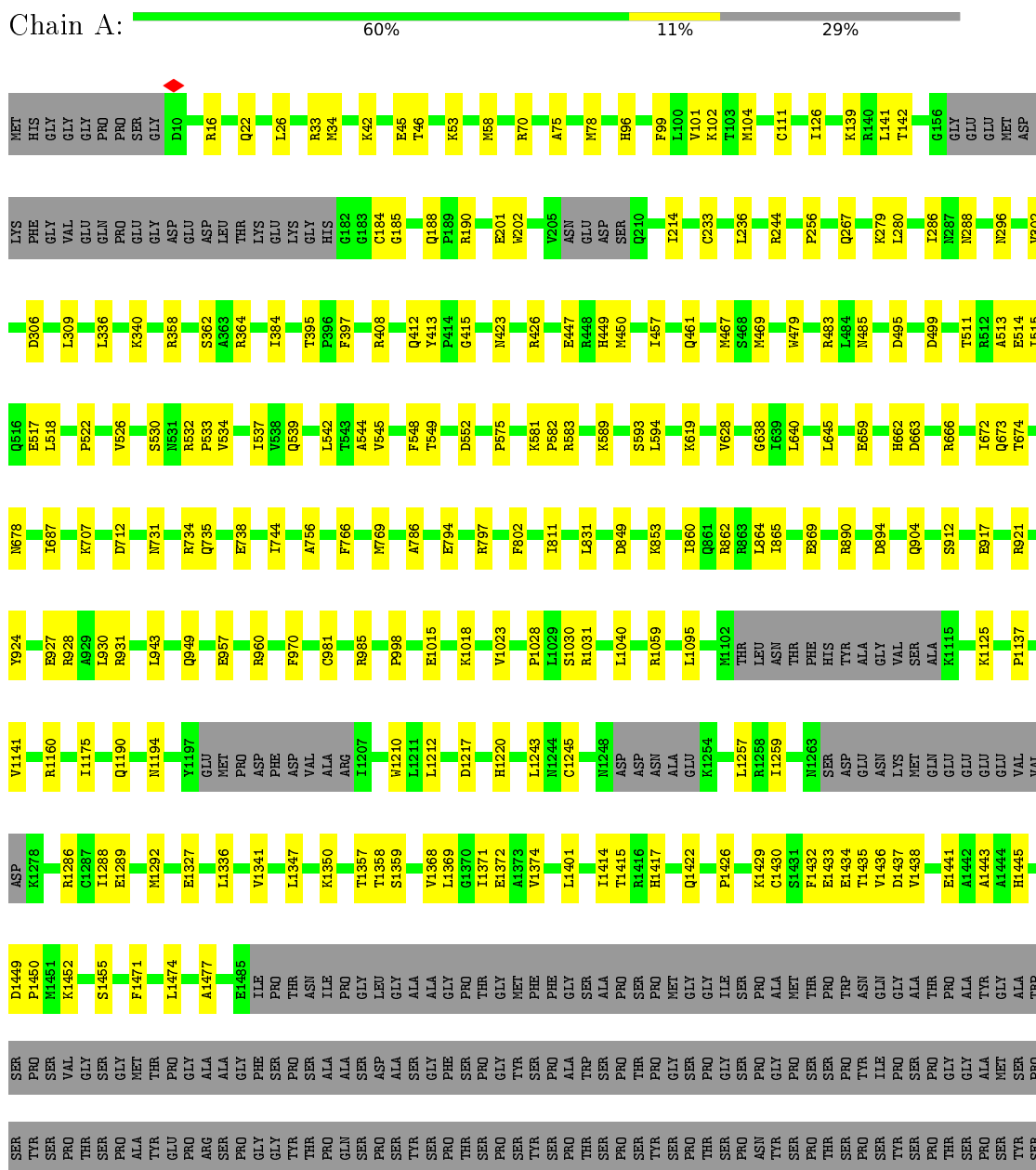
- Molecule 21 is ZINC ION (three-letter code: ZN) (formula: Zn).

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>AltConf</b>
21	A	2	Total 2	Zn 2	0
21	B	1	Total 1	Zn 1	0
21	C	1	Total 1	Zn 1	0
21	I	2	Total 2	Zn 2	0
21	J	1	Total 1	Zn 1	0
21	L	1	Total 1	Zn 1	0
21	Y	1	Total 1	Zn 1	0

### 3 Residue-property plots

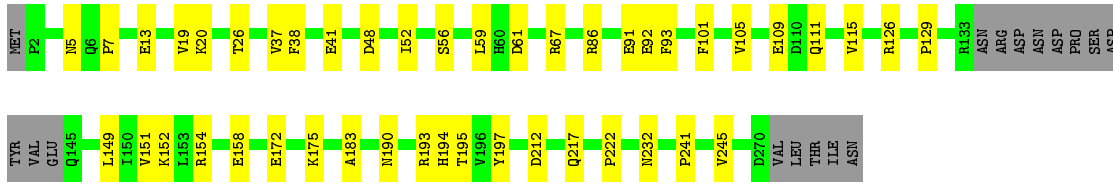
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and atom inclusion 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: DNA-directed RNA polymerase II subunit RPB1

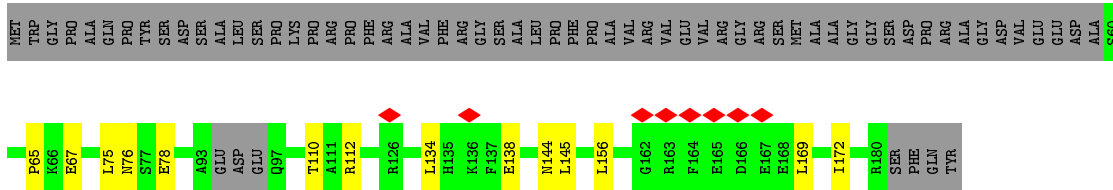




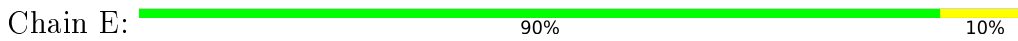




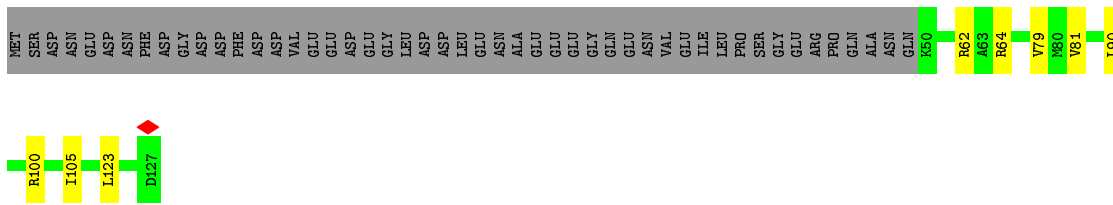
• Molecule 4: RNA polymerase II subunit D



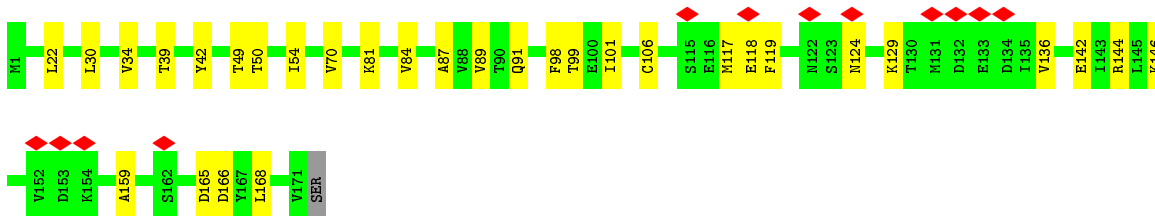
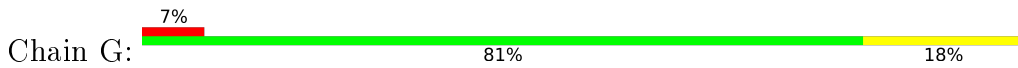
• Molecule 5: DNA-directed RNA polymerase II subunit E



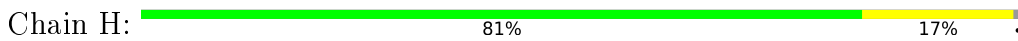
• Molecule 6: DNA-directed RNA polymerase II subunit F



• Molecule 7: DNA-directed RNA polymerase II subunit RPB7



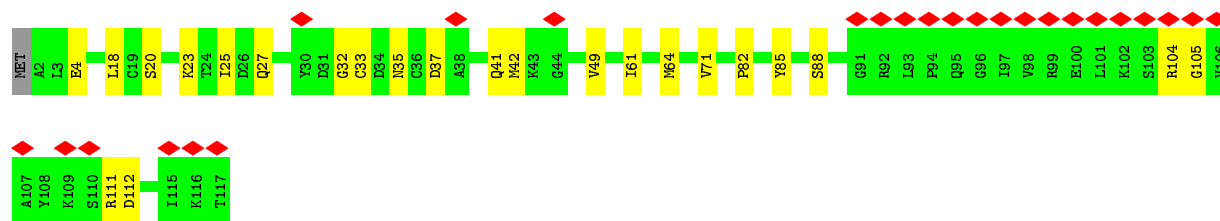
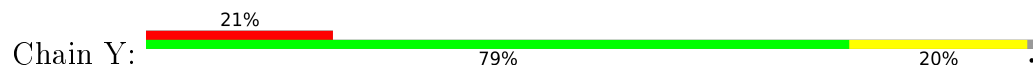
• Molecule 8: DNA-directed RNA polymerases I, II, and III subunit RPABC3











## 4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	31910	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 ( $e^-/\text{\AA}^2$ )	43.21	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K3 BIOQUANTUM (6k x 4k)	Depositor
Maximum map value	21.548	Depositor
Minimum map value	-7.865	Depositor
Average map value	0.001	Depositor
Map value standard deviation	0.990	Depositor
Recommended contour level	2.85	Depositor
Map size ( $\text{\AA}$ )	315.0, 315.0, 315.0	wwPDB
Map dimensions	300, 300, 300	wwPDB
Map angles ( $^\circ$ )	90.0, 90.0, 90.0	wwPDB
Pixel spacing ( $\text{\AA}$ )	1.05, 1.05, 1.05	Depositor

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

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

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	0.28	0/11350	0.49	0/15314
2	B	0.29	0/9227	0.49	1/12454 (0.0%)
3	C	0.29	0/2115	0.50	1/2873 (0.0%)
4	D	0.31	0/979	0.59	0/1312
5	E	0.27	0/1752	0.50	0/2366
6	F	0.29	0/636	0.51	0/859
7	G	0.37	0/1378	0.61	0/1870
8	H	0.29	0/1207	0.51	0/1628
9	I	0.31	0/954	0.57	0/1293
10	J	0.32	0/533	0.47	0/719
11	K	0.29	0/939	0.45	0/1271
12	L	0.31	0/372	0.57	0/493
13	M	0.29	0/1121	0.60	2/1507 (0.1%)
14	N	0.55	0/613	0.94	0/940
15	O	0.32	0/849	0.62	1/1145 (0.1%)
16	P	0.23	0/312	0.78	0/484
17	T	0.55	0/907	0.92	0/1400
18	Z	0.27	0/3035	0.51	0/4094
19	Y	0.30	0/950	0.61	0/1282
All	All	0.30	0/39229	0.54	5/53304 (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
7	G	0	2

There are no bond length outliers.



All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
15	O	36	ASP	CB-CG-OD1	6.34	124.01	118.30
2	B	640	ASP	CB-CG-OD1	5.97	123.67	118.30
3	C	126	ARG	NE-CZ-NH1	5.13	122.86	120.30
13	M	120	ASP	CB-CG-OD1	5.10	122.89	118.30
13	M	17	LEU	CA-CB-CG	5.08	126.99	115.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
7	G	117	MET	Peptide
7	G	124	ASN	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	11149	0	11300	147	0
2	B	9047	0	9080	122	0
3	C	2072	0	2019	33	0
4	D	967	0	973	11	0
5	E	1721	0	1737	13	0
6	F	626	0	657	6	0
7	G	1347	0	1347	18	0
8	H	1186	0	1147	16	0
9	I	932	0	856	17	0
10	J	524	0	540	11	0
11	K	920	0	942	5	0
12	L	367	0	367	3	0
13	M	1106	0	1082	18	0
14	N	549	0	308	4	0
15	O	832	0	832	13	0
16	P	279	0	142	3	0
17	T	808	0	442	13	0
18	Z	2979	0	3031	43	0
19	Y	920	0	915	14	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
20	A	1	0	0	0	0
21	A	2	0	0	0	0
21	B	1	0	0	0	0
21	C	1	0	0	0	0
21	I	2	0	0	0	0
21	J	1	0	0	0	0
21	L	1	0	0	0	0
21	Y	1	0	0	0	0
All	All	38341	0	37717	431	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 6.

The worst 5 of 431 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:C:91:GLU:HG2	3:C:92:GLU:OE1	1.46	1.15
15:O:31:HIS:ND1	15:O:104:SER:OG	2.10	0.83
3:C:91:GLU:CG	3:C:92:GLU:OE1	2.29	0.80
3:C:92:GLU:HG2	3:C:93:PHE:N	1.97	0.77
18:Z:440:ILE:HG12	18:Z:450:ILE:HD11	1.70	0.74

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	1393/1970 (71%)	1360 (98%)	33 (2%)	0	100	100
2	B	1123/1251 (90%)	1093 (97%)	30 (3%)	0	100	100
3	C	254/275 (92%)	249 (98%)	5 (2%)	0	100	100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
4	D	114/184 (62%)	109 (96%)	5 (4%)	0	100	100
5	E	207/210 (99%)	204 (99%)	3 (1%)	0	100	100
6	F	76/127 (60%)	76 (100%)	0	0	100	100
7	G	169/172 (98%)	154 (91%)	15 (9%)	0	100	100
8	H	146/150 (97%)	144 (99%)	2 (1%)	0	100	100
9	I	114/125 (91%)	109 (96%)	5 (4%)	0	100	100
10	J	64/67 (96%)	64 (100%)	0	0	100	100
11	K	113/117 (97%)	111 (98%)	2 (2%)	0	100	100
12	L	42/58 (72%)	41 (98%)	1 (2%)	0	100	100
13	M	131/640 (20%)	128 (98%)	3 (2%)	0	100	100
15	O	104/268 (39%)	99 (95%)	5 (5%)	0	100	100
18	Z	370/1087 (34%)	362 (98%)	8 (2%)	0	100	100
19	Y	117/117 (100%)	110 (94%)	7 (6%)	0	100	100
All	All	4537/6818 (66%)	4413 (97%)	124 (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	
1	A	1239/1749 (71%)	1239 (100%)	0	100	100
2	B	991/1084 (91%)	988 (100%)	3 (0%)	92	95
3	C	235/252 (93%)	235 (100%)	0	100	100
4	D	109/160 (68%)	109 (100%)	0	100	100
5	E	191/192 (100%)	191 (100%)	0	100	100
6	F	68/111 (61%)	68 (100%)	0	100	100
7	G	151/153 (99%)	150 (99%)	1 (1%)	84	90
8	H	129/131 (98%)	129 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
9	I	102/112 (91%)	102 (100%)	0	100	100
10	J	55/56 (98%)	55 (100%)	0	100	100
11	K	104/106 (98%)	104 (100%)	0	100	100
12	L	40/55 (73%)	40 (100%)	0	100	100
13	M	124/575 (22%)	124 (100%)	0	100	100
15	O	93/240 (39%)	93 (100%)	0	100	100
18	Z	331/940 (35%)	331 (100%)	0	100	100
19	Y	104/103 (101%)	104 (100%)	0	100	100
All	All	4066/6019 (68%)	4062 (100%)	4 (0%)	93	97

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	188	ASN
2	B	465	TYR
2	B	936	ARG
7	G	81	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
19	Y	41	GLN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
16	P	12/47 (25%)	1 (8%)	0

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
16	P	36	G

There are no RNA pucker outliers to report.

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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

## 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 10 ligands modelled in this entry, 10 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

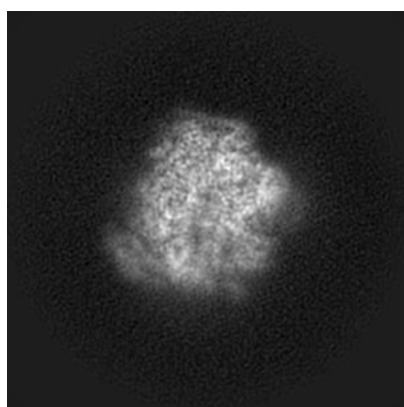
## 6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-12973. These allow visual inspection of the internal detail of the map and identification of artifacts.

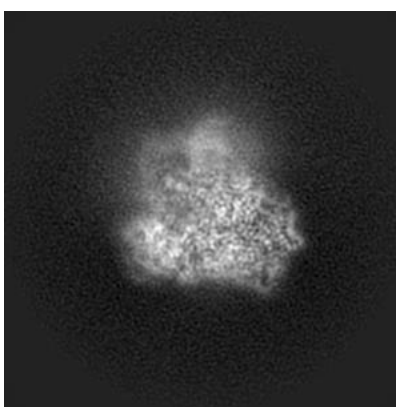
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

### 6.1 Orthogonal projections [i](#)

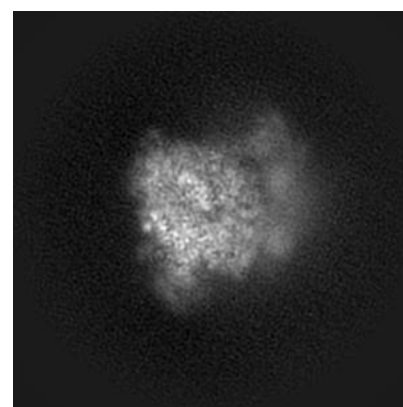
#### 6.1.1 Primary map



X



Y

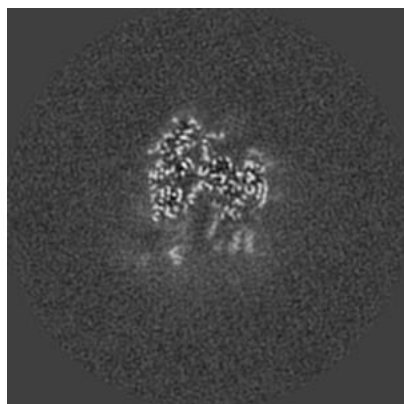


Z

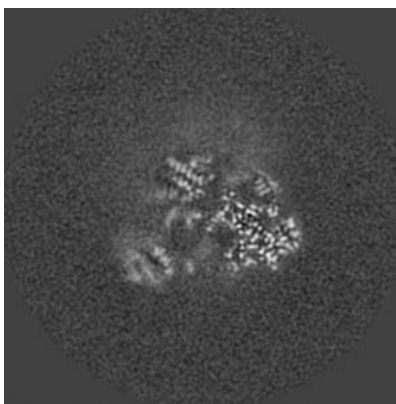
The images above show the map projected in three orthogonal directions.

### 6.2 Central slices [i](#)

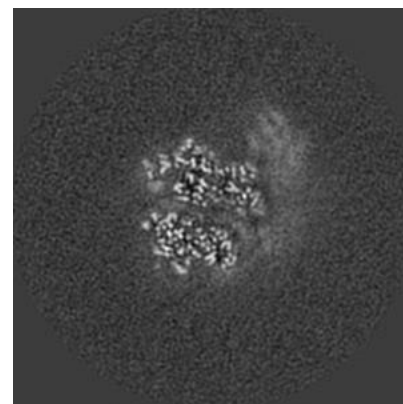
#### 6.2.1 Primary map



X Index: 150



Y Index: 150

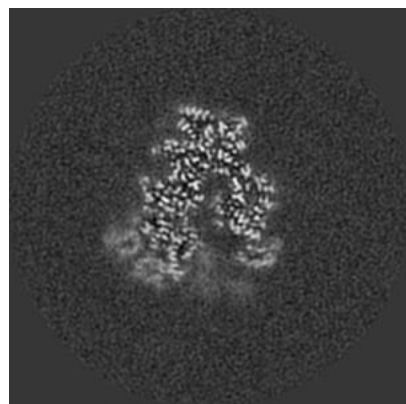


Z Index: 150

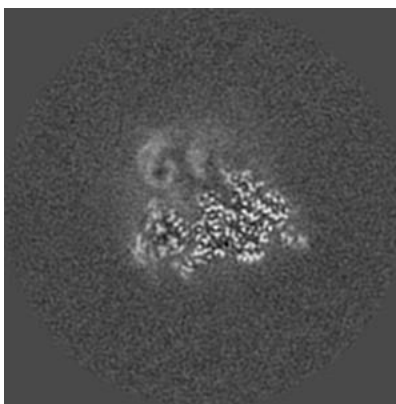
The images above show central slices of the map in three orthogonal directions.

## 6.3 Largest variance slices [i](#)

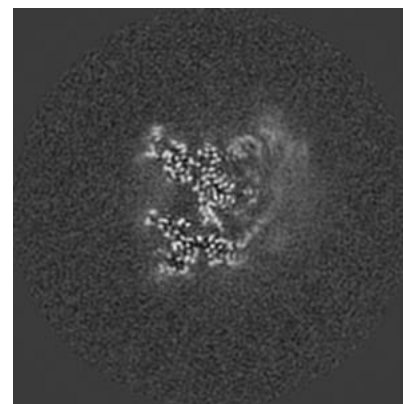
### 6.3.1 Primary map



X Index: 127



Y Index: 129



Z Index: 164

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

## 6.5 Mask visualisation

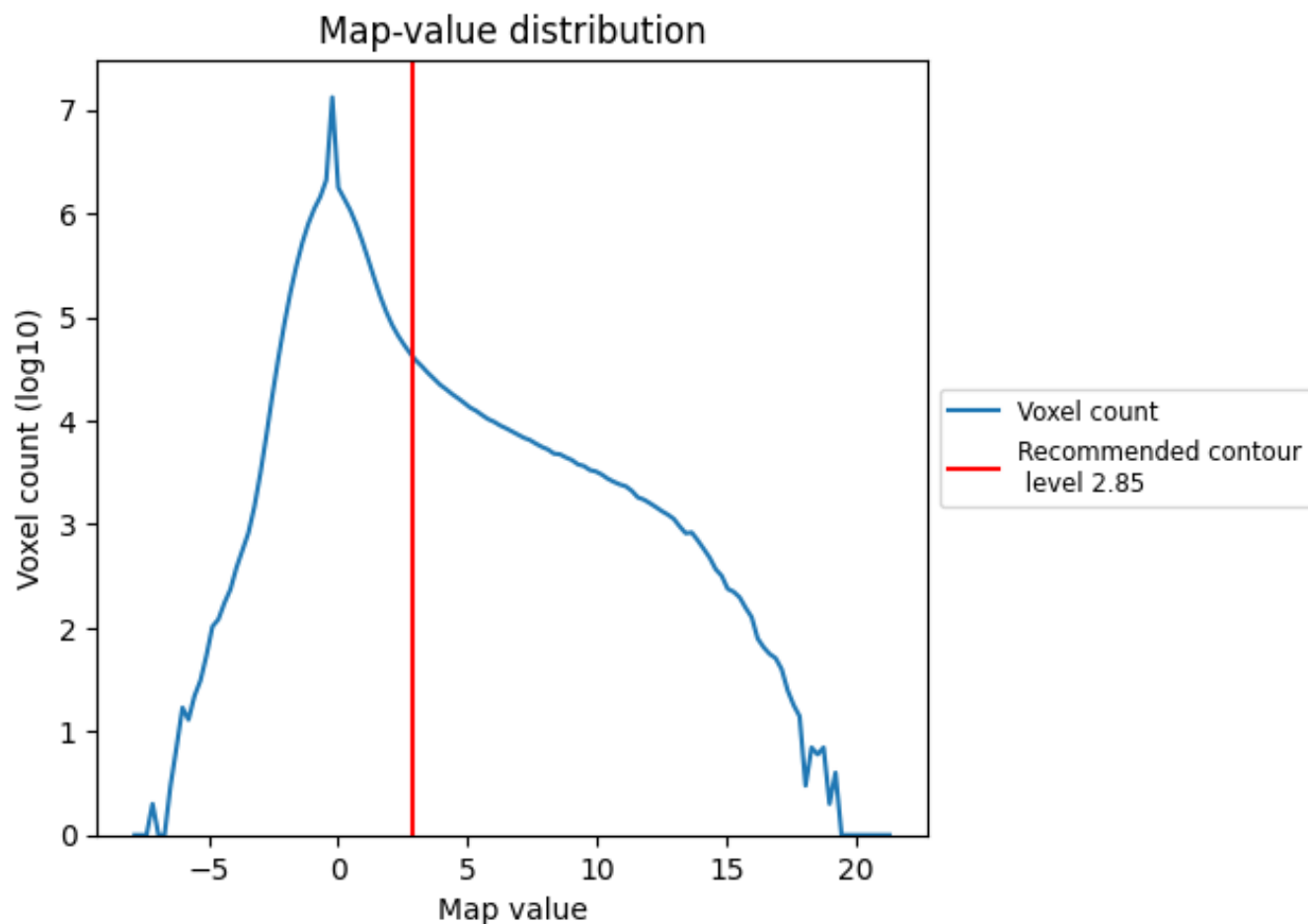
This section was not generated. No masks/segmentation were deposited.



## 7 Map analysis [i](#)

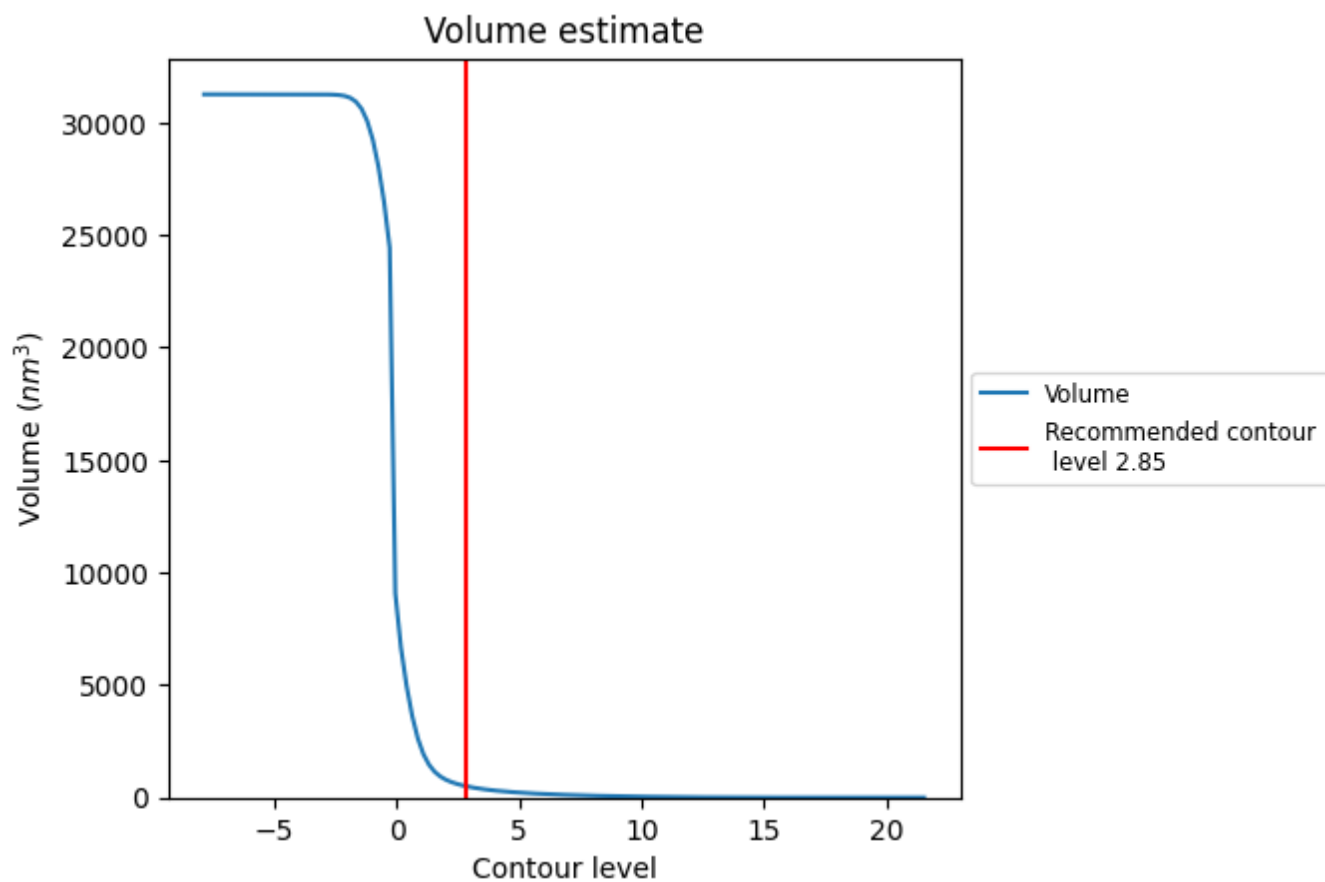
This section contains the results of statistical analysis of the map.

### 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.

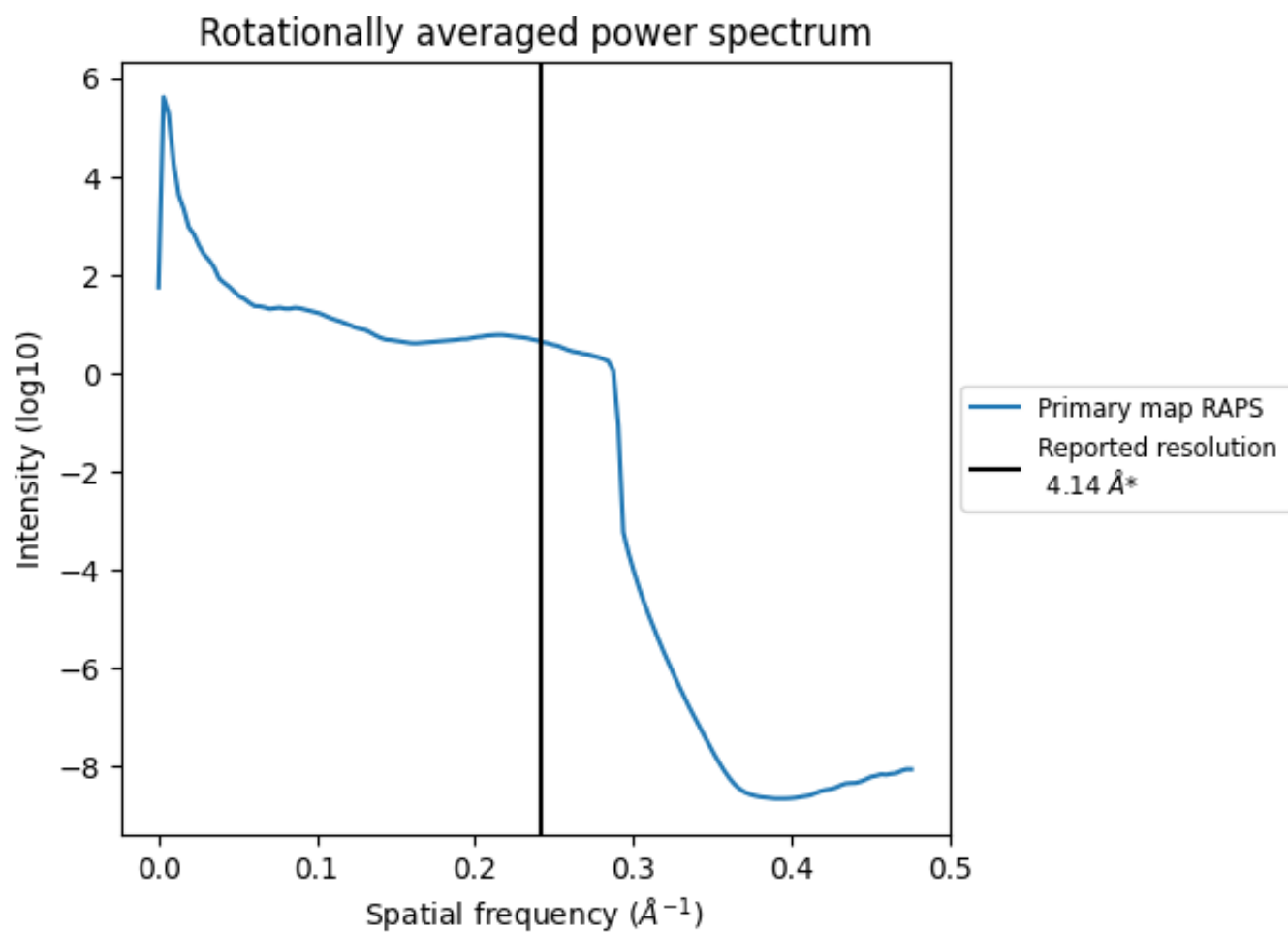
## 7.2 Volume estimate [i](#)



The volume at the recommended contour level is 503  $\text{nm}^3$ ; this corresponds to an approximate mass of 455 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  $0.242 \text{\AA}^{-1}$

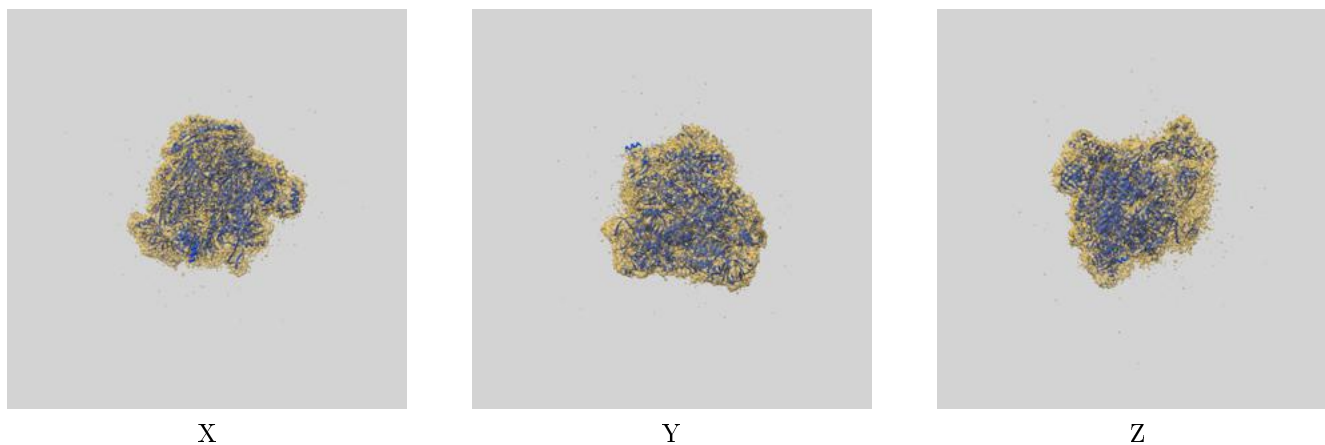
## 8 Fourier-Shell correlation

This section was not generated. No FSC curve or half-maps provided.

## 9 Map-model fit [i](#)

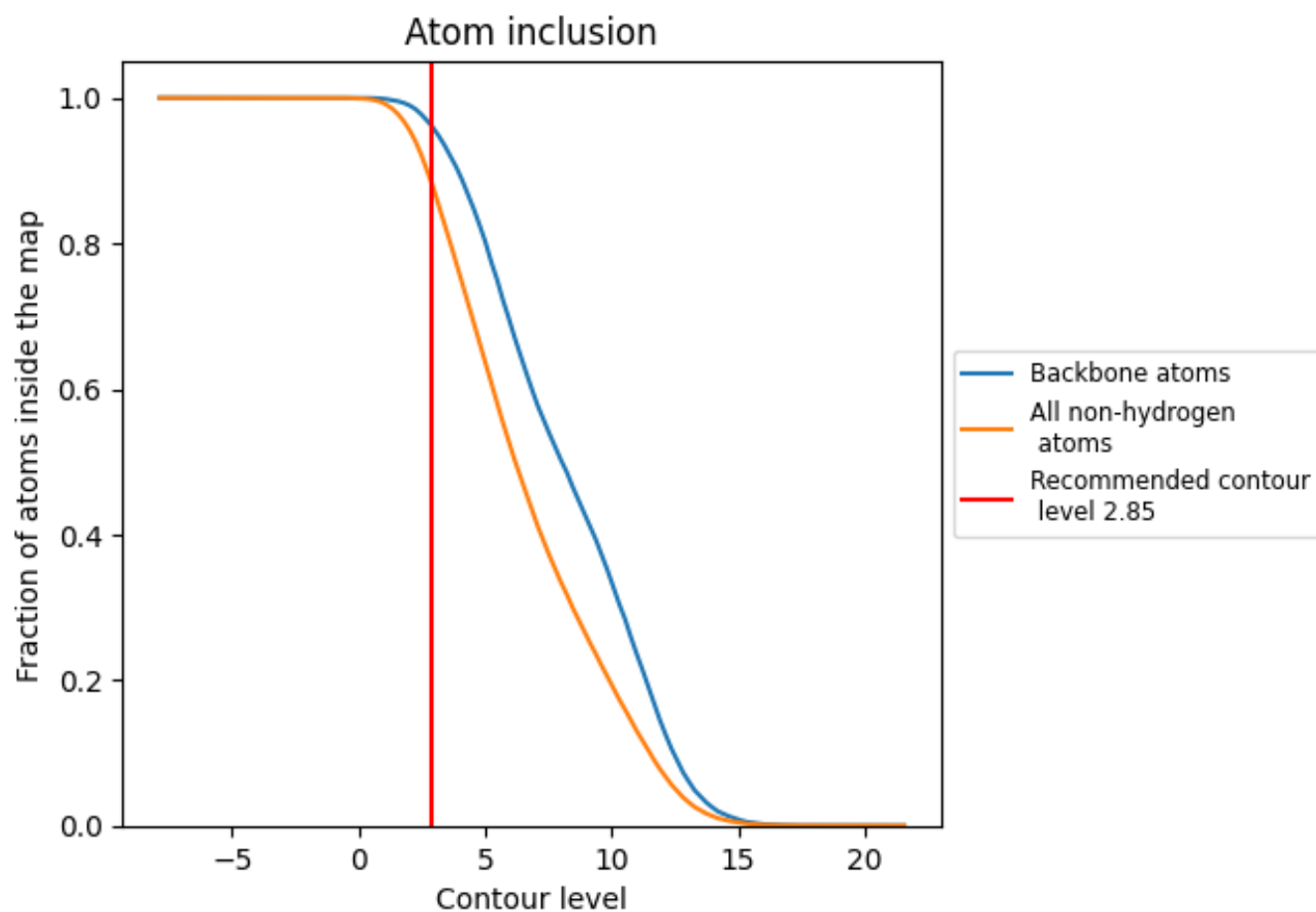
This section contains information regarding the fit between EMDB map EMD-12973 and PDB model 7OKY. Per-residue inclusion information can be found in section [3](#) on page [8](#).

### 9.1 Map-model overlay [i](#)



The images above show the 3D surface view of the map at the recommended contour level 2.85 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 Atom inclusion [i](#)



At the recommended contour level, 96% of all backbone atoms, 89% of all non-hydrogen atoms, are inside the map.