wwPDB EM Validation Summary Report

Mar 13, 2023 – 12:25 PM EDT

PDB ID	:	8DY9
EMDB ID	:	EMD-27778
Title	:	Streptomyces venezuelae RNAP unconstrained open promoter complex with WhiA and WhiB transcription factors
Authors	:	Lilic, M.; Campbell, E.A.
Deposited on	:	2022-08-03
Resolution	:	3.12 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:

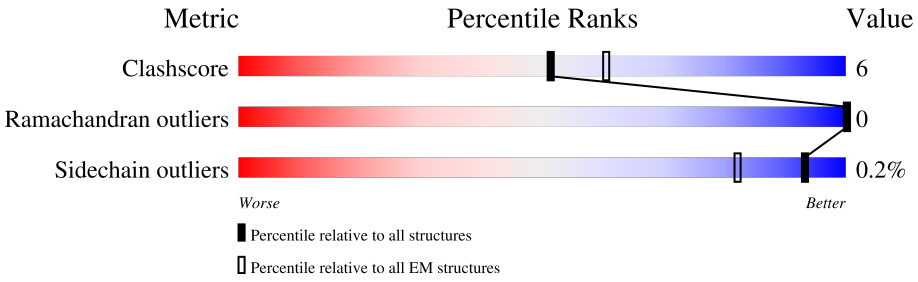
EMDB validation analysis	:	0.0.1.dev43
Mogul	:	1.8.5 (274361), CSD as541be (2020)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.9
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.32.1

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 3.12 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	340	60% 6% 34%
1	B	340	83% 11% 6%
2	C	1178	83% 12% 5%
3	D	1298	86% 10%
4	E	90	81% 14%
5	F	516	51% 8% 41%
6	G	124	13% 71% 13% 16%
7	H	87	71% 15% 14%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
8	I	327	71% 21% 7%
9	O	49	43% 35% 22%
9	Q	49	6% 51% 18% 8% 22%
10	P	49	47% 31% 22%
10	R	49	6% 61% 39%

2 Entry composition (i)

There are 13 unique types of molecules in this entry. The entry contains 63435 atoms, of which 31140 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 subunit alpha.

Mol	Chain	Residues	Atoms						AltConf	Trace
1	A	225	Total	C	H	N	O	S	0	0
			3519	1096	1785	301	333	4		
1	B	319	Total	C	H	N	O	S	0	0
			4908	1528	2483	416	473	8		

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

Mol	Chain	Residues	Atoms						AltConf	Trace
2	C	1118	Total	C	H	N	O	S	0	0
			17368	5465	8655	1516	1701	31		

- Molecule 3 is a protein called DNA-directed RNA polymerase subunit beta'.

Mol	Chain	Residues	Atoms						AltConf	Trace
3	D	1253	Total	C	H	N	O	S	0	0
			19789	6143	9980	1779	1847	40		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	2	ASP	LEU	conflict	UNP A0A5P2AAC9

- Molecule 4 is a protein called DNA-directed RNA polymerase subunit omega.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	E	77	Total	C	H	N	O	0	0
			1202	382	605	98	117		

- Molecule 5 is a protein called RNA polymerase sigma factor SigA.

Mol	Chain	Residues	Atoms						AltConf	Trace
5	F	303	Total	C	H	N	O	S	0	0
			4873	1515	2466	428	457	7		

- Molecule 6 is a protein called RNA polymerase-binding protein RbpA.

Mol	Chain	Residues	Atoms						AltConf	Trace
6	G	104	Total	C	H	N	O	S	0	0
			1657	518	825	150	158	6		

- Molecule 7 is a protein called Transcriptional regulator WhiB.

Mol	Chain	Residues	Atoms						AltConf	Trace
7	H	75	Total	C	H	N	O	S	0	0
			1158	366	562	110	116	4		

- Molecule 8 is a protein called Probable cell division protein WhiA.

Mol	Chain	Residues	Atoms						AltConf	Trace
8	I	303	Total	C	H	N	O	S	0	0
			4496	1367	2276	435	408	10		

- Molecule 9 is a DNA chain called DNA (38-MER).

Mol	Chain	Residues	Atoms						AltConf	Trace
9	O	38	Total	C	H	N	O	P	0	0
			1197	367	424	146	223	37		
9	Q	38	Total	C	H	N	O	P	0	0
			1096	370	315	143	230	38		

- Molecule 10 is a DNA chain called DNA (38-MER).

Mol	Chain	Residues	Atoms						AltConf	Trace
10	P	38	Total	C	H	N	O	P	0	0
			1209	370	427	143	231	38		
10	R	30	Total	C	H	N	O	P	0	0
			951	291	337	114	179	30		

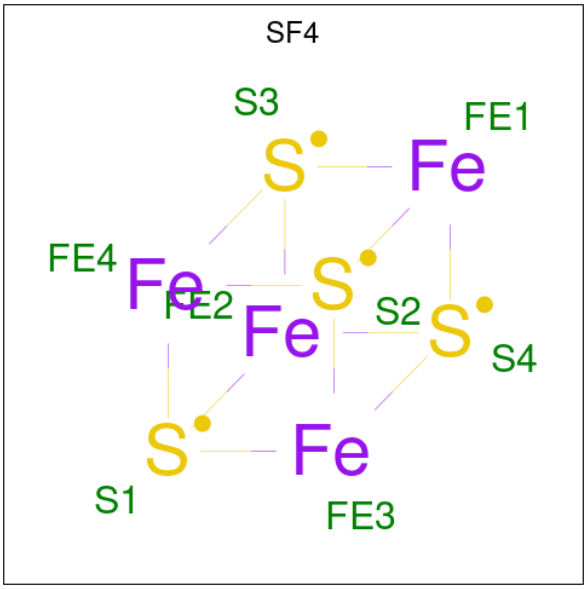
- Molecule 11 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		AltConf
11	D	2	Total	Zn	0
			2	2	
11	I	1	Total	Zn	0
			1	1	

- Molecule 12 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	AltConf
12	D	1	Total Mg 1 1	0

- Molecule 13 is IRON/SULFUR CLUSTER (three-letter code: SF4) (formula: Fe_4S_4).



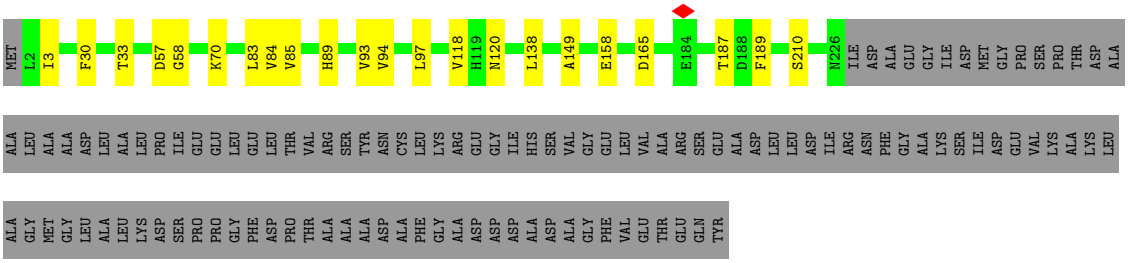
Mol	Chain	Residues	Atoms	AltConf
13	H	1	Total Fe S 8 4 4	0

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.

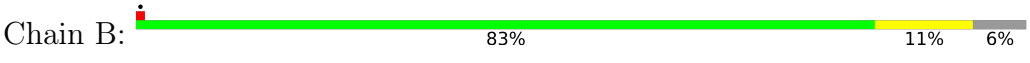
• Molecule 1: DNA-directed RNA polymerase subunit alpha

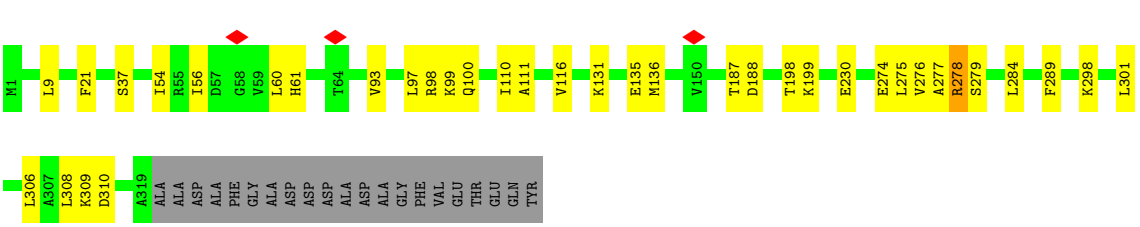
Chain A:



• Molecule 1: DNA-directed RNA polymerase subunit alpha

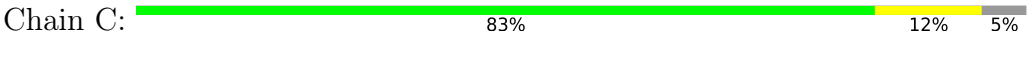
Chain B:

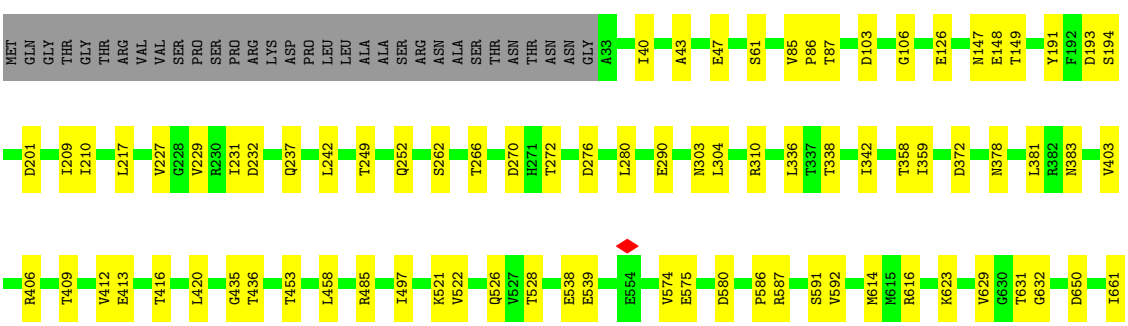




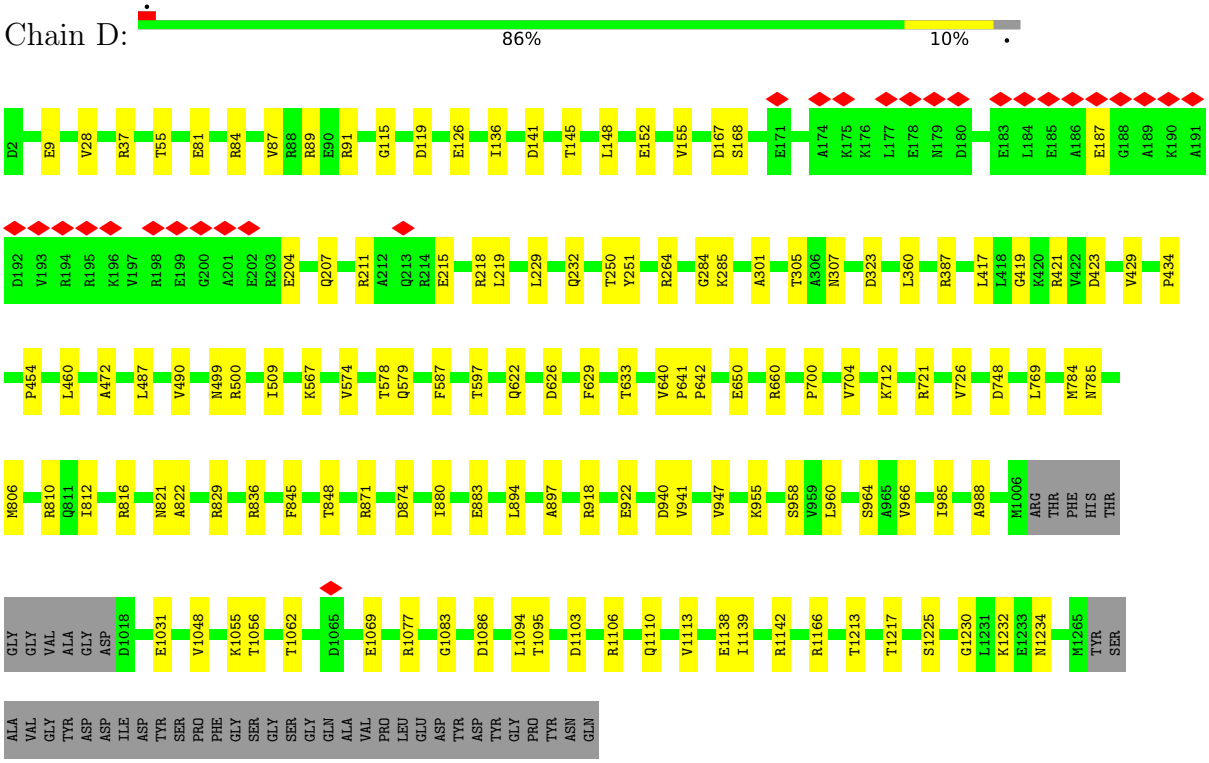
• Molecule 2: DNA-directed RNA polymerase subunit beta

Chain C:

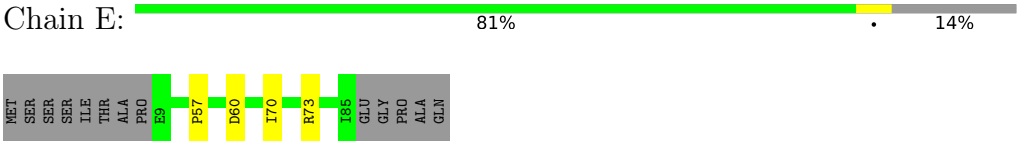




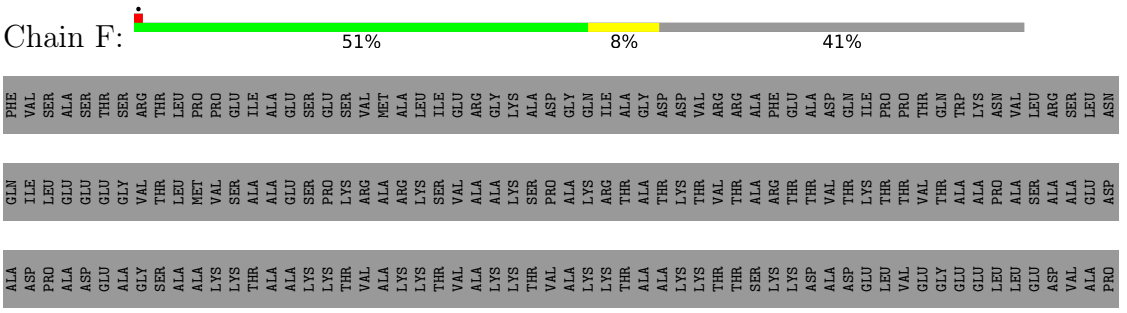
• Molecule 3: DNA-directed RNA polymerase subunit beta'

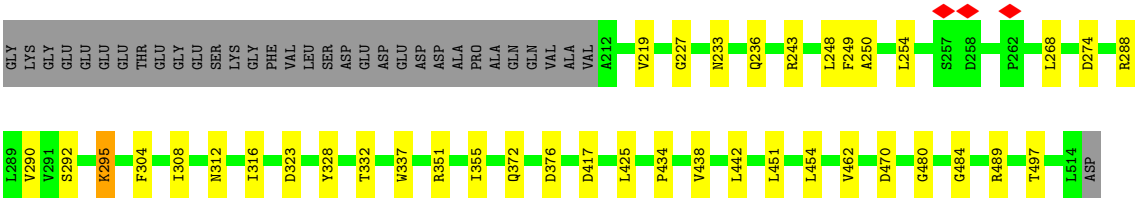


• Molecule 4: DNA-directed RNA polymerase subunit omega



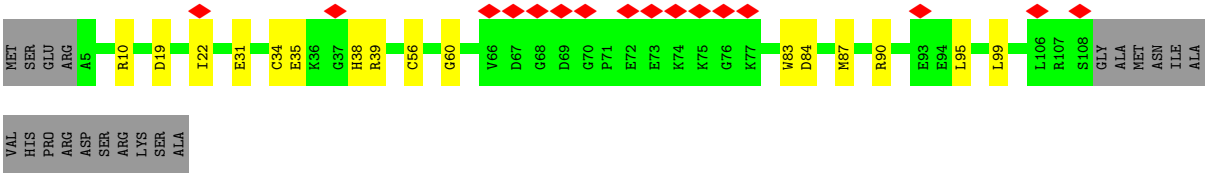
• Molecule 5: RNA polymerase sigma factor SigA





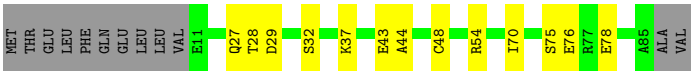
• Molecule 6: RNA polymerase-binding protein RbpA

Chain G:



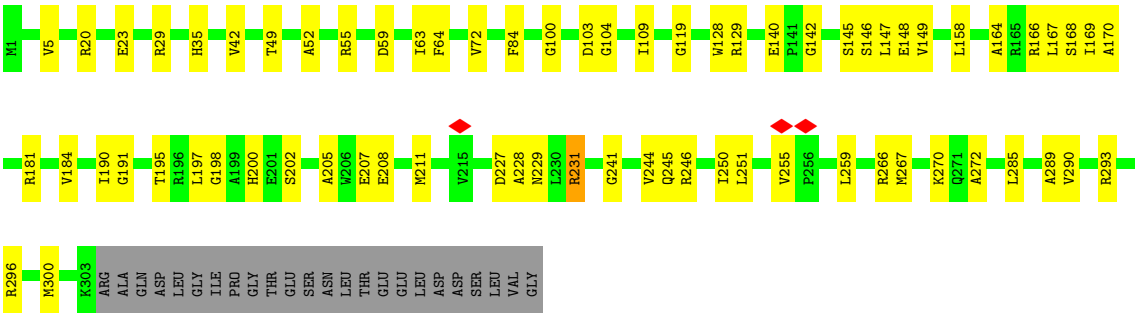
• Molecule 7: Transcriptional regulator WhiB

Chain H:



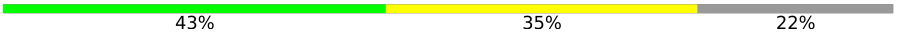
• Molecule 8: Probable cell division protein WhiA

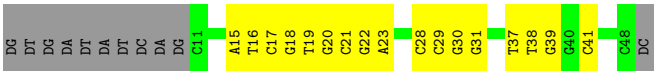
Chain I:



• Molecule 9: DNA (38-MER)

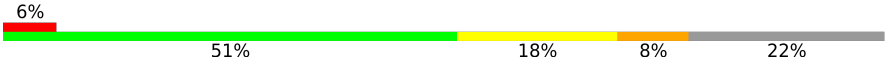
Chain O:

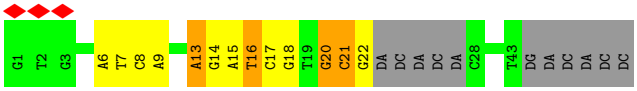




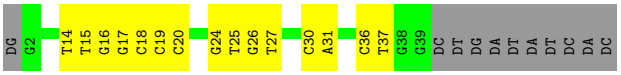
• Molecule 9: DNA (38-MER)

Chain Q:





• Molecule 10: DNA (38-MER)



• Molecule 10: DNA (38-MER)

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	118181	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 ($e^-/Å^2$)	66	Depositor
Minimum defocus (nm)	500	Depositor
Maximum defocus (nm)	2500	Depositor
Magnification	Not provided	
Image detector	GATAN K3 (6k x 4k)	Depositor
Maximum map value	44.974	Depositor
Minimum map value	-17.397	Depositor
Average map value	-0.004	Depositor
Map value standard deviation	1.038	Depositor
Recommended contour level	3.15	Depositor
Map size (Å)	324.9, 324.9, 324.9	wwPDB
Map dimensions	300, 300, 300	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.083, 1.083, 1.083	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, MG, ZN

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	0.24	0/1760	0.52	0/2389
1	B	0.25	0/2461	0.52	0/3339
2	C	0.25	0/8869	0.51	0/12011
3	D	0.24	0/9958	0.52	0/13438
4	E	0.25	0/607	0.45	0/826
5	F	0.24	0/2442	0.49	0/3292
6	G	0.24	0/847	0.56	0/1142
7	H	0.25	0/605	0.53	0/809
8	I	0.25	0/2245	0.56	0/3033
9	O	0.58	0/867	0.91	0/1335
9	Q	0.61	0/874	1.13	4/1345 (0.3%)
10	P	0.56	0/876	0.95	0/1351
10	R	0.51	0/687	0.88	0/1055
All	All	0.29	0/33098	0.59	4/45365 (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})
9	Q	21	DC	P-O3'-C3'	-12.77	104.38	119.70
9	Q	16	DT	P-O3'-C3'	-12.57	104.62	119.70
9	Q	13	DA	P-O3'-C3'	-8.13	109.95	119.70
9	Q	20	DG	P-O3'-C3'	-7.94	110.17	119.70

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (i)

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

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1734	1785	1785	18	0
1	B	2425	2483	2483	23	0
2	C	8713	8655	8654	97	0
3	D	9809	9980	9979	87	0
4	E	597	605	604	2	0
5	F	2407	2466	2466	39	0
6	G	832	825	824	11	0
7	H	596	562	565	15	0
8	I	2220	2276	2256	54	0
9	O	773	424	426	20	0
9	Q	781	315	429	44	0
10	P	782	427	428	13	0
10	R	614	337	338	0	0
11	D	2	0	0	0	0
11	I	1	0	0	0	0
12	D	1	0	0	0	0
13	H	8	0	0	0	0
All	All	32295	31140	31237	368	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 368 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)
9:Q:16:DT:H2”	9:Q:17:DC:H5’	1.22	1.10
9:Q:16:DT:H2’	9:Q:17:DC:C5	1.90	1.05
9:Q:14:DG:H3’	9:Q:15:DA:H5”	1.40	1.03
9:Q:15:DA:H2”	9:Q:16:DT:H5’	1.43	1.00
9:Q:16:DT:H2”	9:Q:17:DC:C5’	1.97	0.95

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	223/340 (66%)	216 (97%)	7 (3%)	0	100	100
1	B	317/340 (93%)	296 (93%)	21 (7%)	0	100	100
2	C	1116/1178 (95%)	1055 (94%)	61 (6%)	0	100	100
3	D	1249/1298 (96%)	1209 (97%)	40 (3%)	0	100	100
4	E	75/90 (83%)	72 (96%)	3 (4%)	0	100	100
5	F	301/516 (58%)	296 (98%)	5 (2%)	0	100	100
6	G	102/124 (82%)	98 (96%)	4 (4%)	0	100	100
7	H	73/87 (84%)	69 (94%)	4 (6%)	0	100	100
8	I	301/327 (92%)	276 (92%)	25 (8%)	0	100	100
All	All	3757/4300 (87%)	3587 (96%)	170 (4%)	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	193/279 (69%)	193 (100%)	0	100	100
1	B	266/279 (95%)	265 (100%)	1 (0%)	91	96
2	C	945/996 (95%)	945 (100%)	0	100	100
3	D	1047/1081 (97%)	1046 (100%)	1 (0%)	93	98
4	E	64/74 (86%)	64 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
5	F	259/425 (61%)	257 (99%)	2 (1%)	81 92
6	G	88/104 (85%)	88 (100%)	0	100 100
7	H	60/73 (82%)	60 (100%)	0	100 100
8	I	214/245 (87%)	213 (100%)	1 (0%)	88 94
All	All	3136/3556 (88%)	3131 (100%)	5 (0%)	93 97

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	278	ARG
3	D	816	ARG
5	F	288	ARG
5	F	295	LYS
8	I	231	ARG

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

Mol	Chain	Res	Type
8	I	245	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)

Of 5 ligands modelled in this entry, 4 are monoatomic - leaving 1 for Mogul analysis.

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	$\#\|Z\| > 2$
13	SF4	H	101	7	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
13	SF4	H	101	7	-	-	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)

There are no chain breaks in this entry.

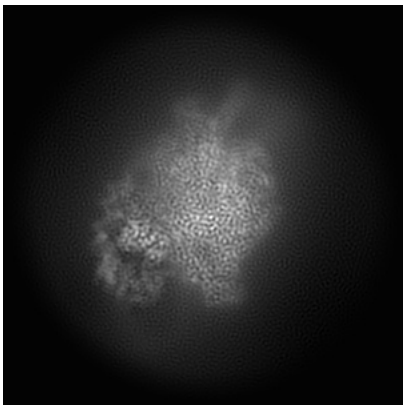
6 Map visualisation

This section contains visualisations of the EMDB entry EMD-27778. These allow visual inspection of the internal detail of the map and identification of artifacts.

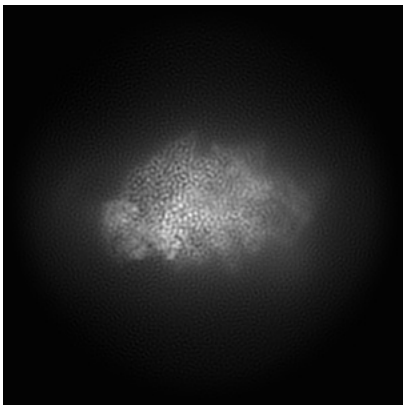
Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

6.1 Orthogonal projections

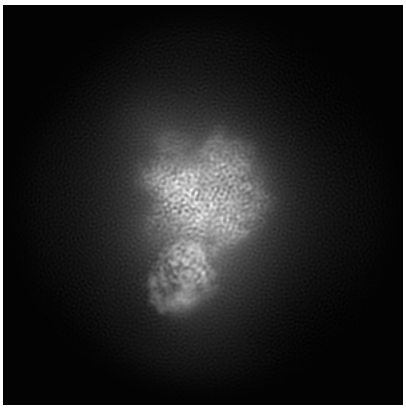
6.1.1 Primary map



X

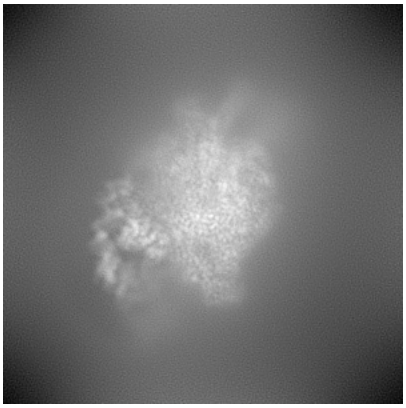


Y

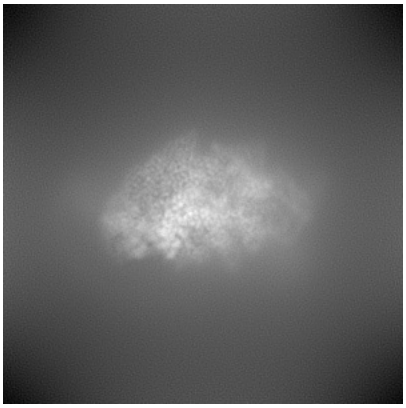


Z

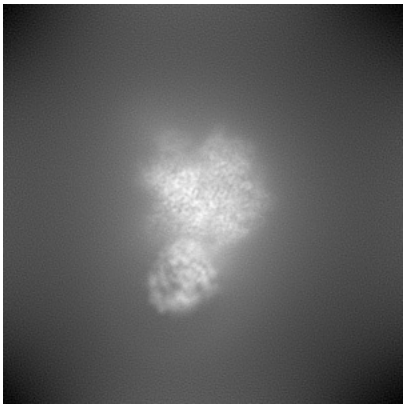
6.1.2 Raw map



X



Y

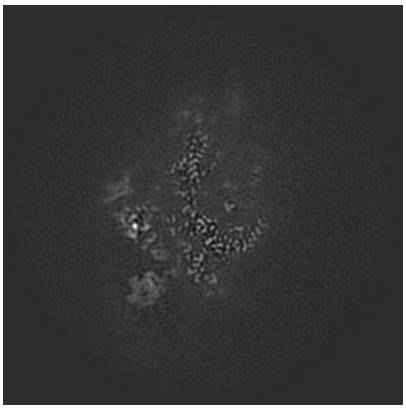


Z

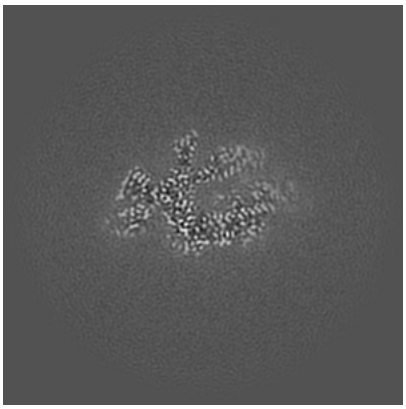
The images above show the map projected in three orthogonal directions.

6.2 Central slices (i)

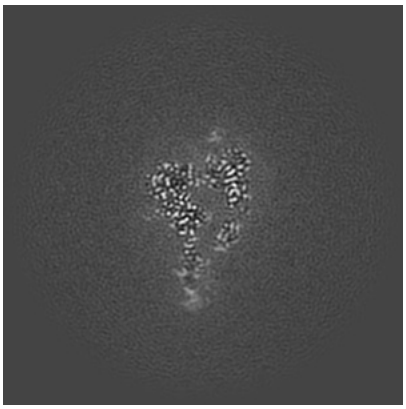
6.2.1 Primary map



X Index: 150

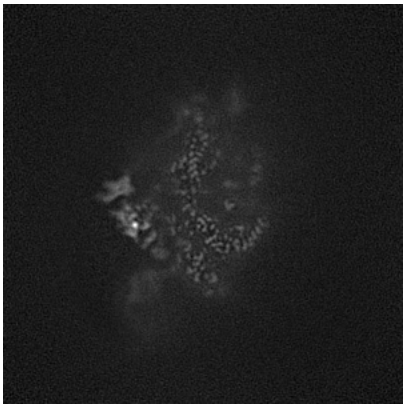


Y Index: 150

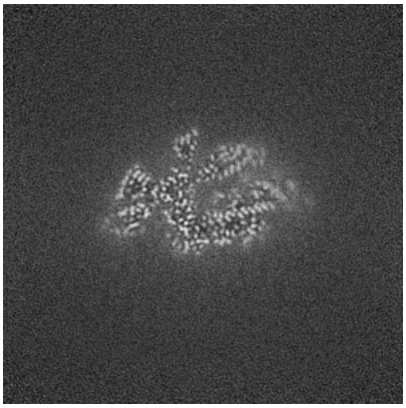


Z Index: 150

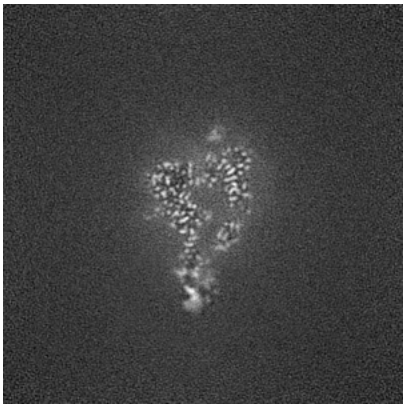
6.2.2 Raw map



X Index: 150



Y Index: 150

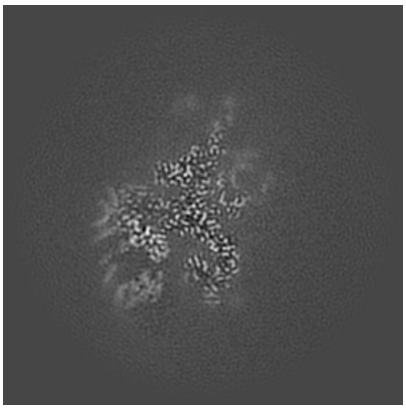


Z Index: 150

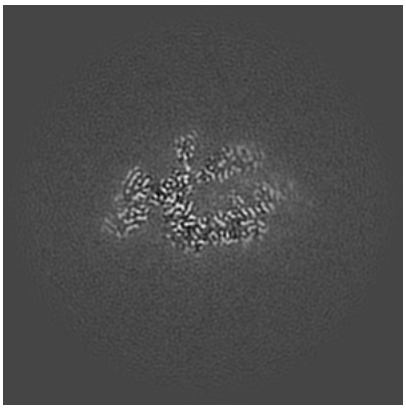
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

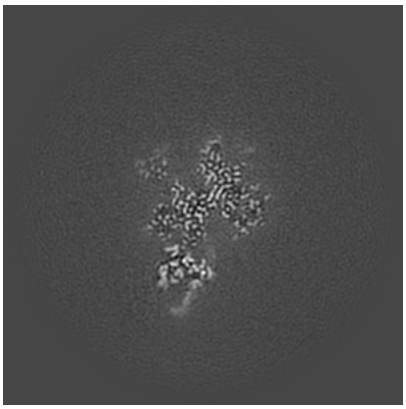
6.3.1 Primary map



X Index: 140

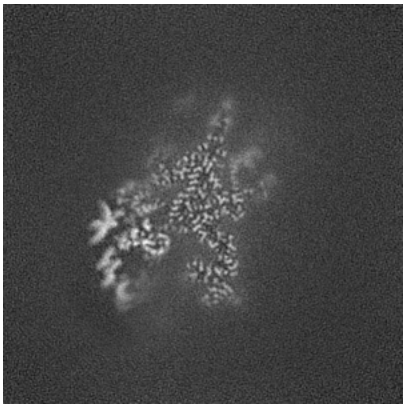


Y Index: 151

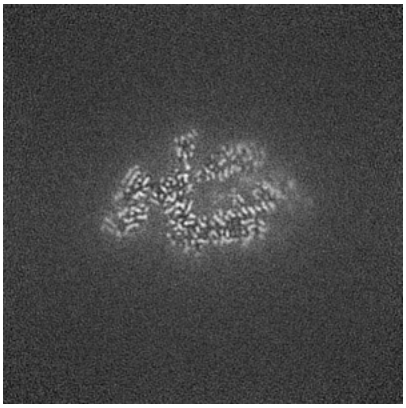


Z Index: 132

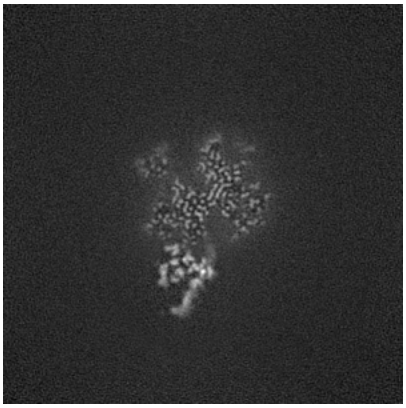
6.3.2 Raw map



X Index: 136



Y Index: 151



Z Index: 132

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 3.15. 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.

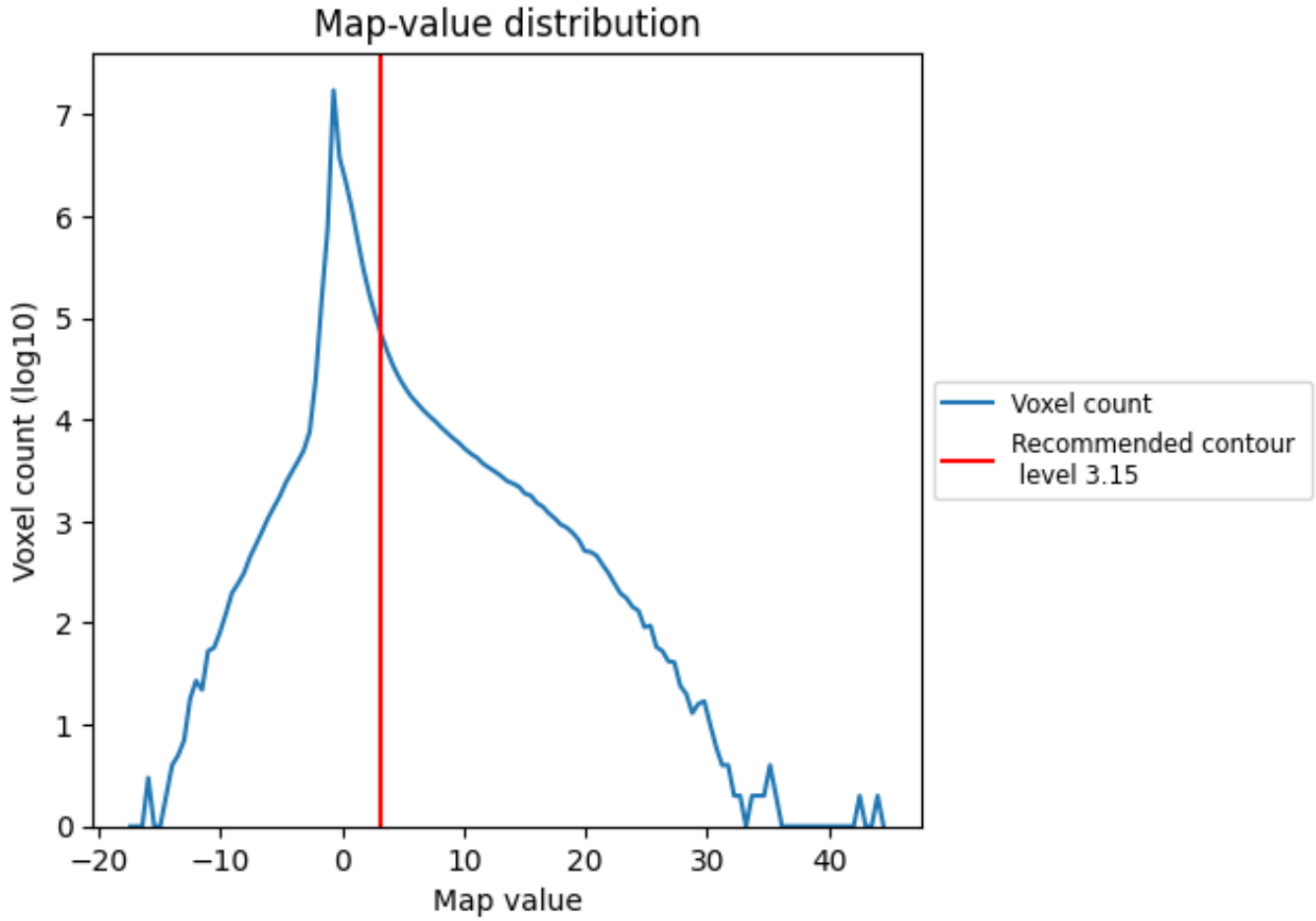
6.5 Mask visualisation (i)

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

7 Map analysis

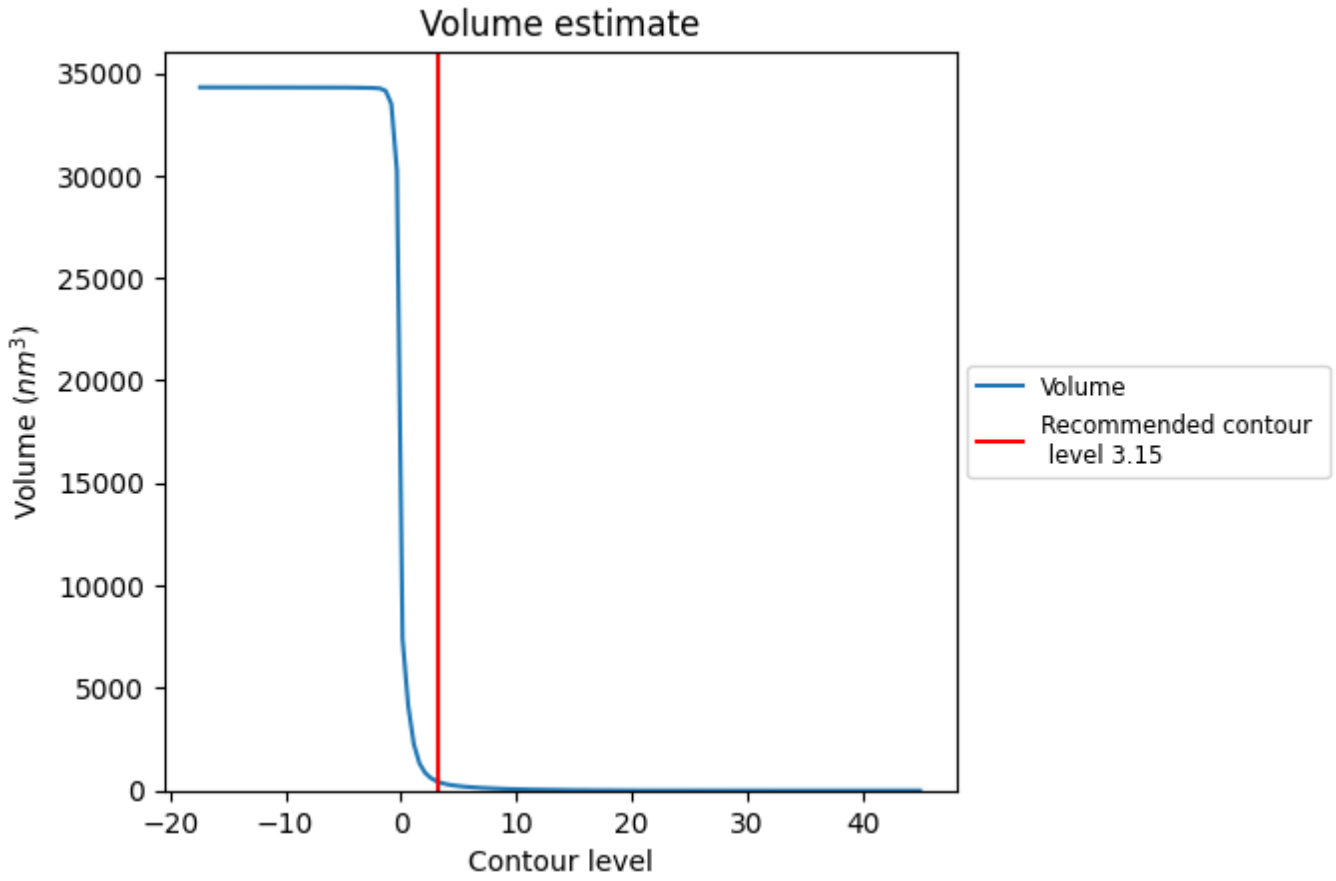
This section contains the results of statistical analysis of the map.

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.

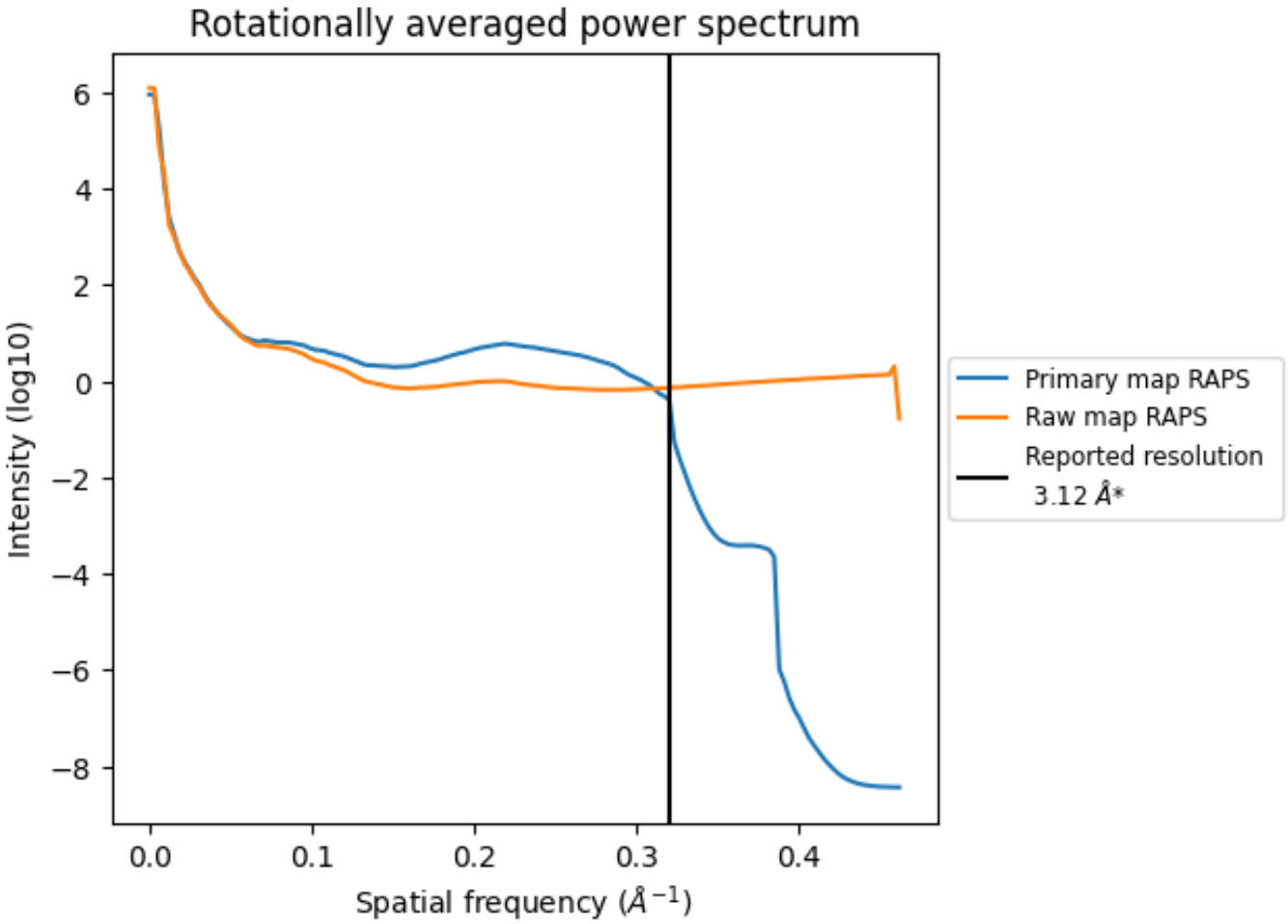
7.2 Volume estimate (i)



The volume at the recommended contour level is 447 nm^3; this corresponds to an approximate mass of 403 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)

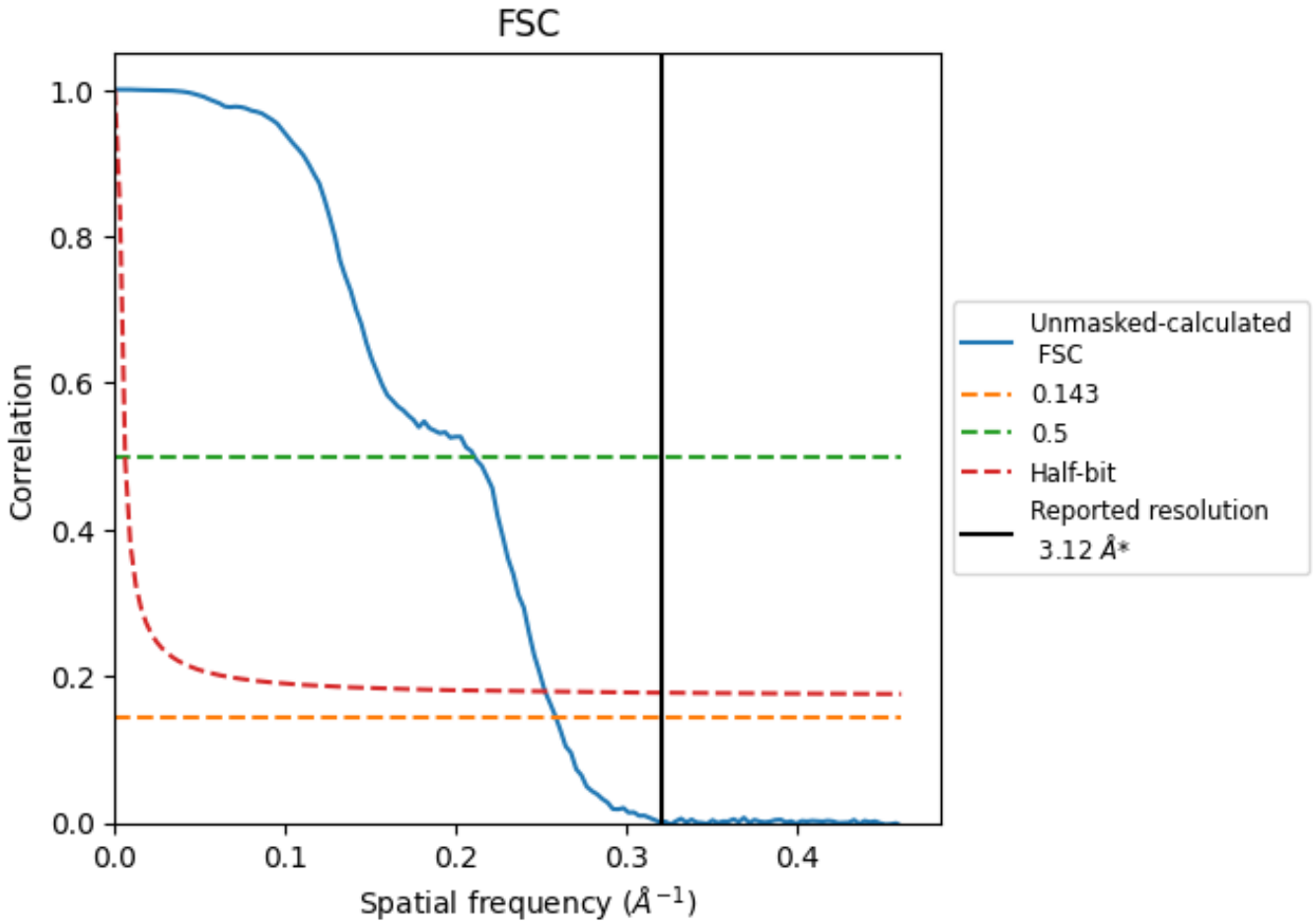


*Reported resolution corresponds to spatial frequency of 0.321 Å^{-1}

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 0.321 Å^{-1}

8.2 Resolution estimates (i)

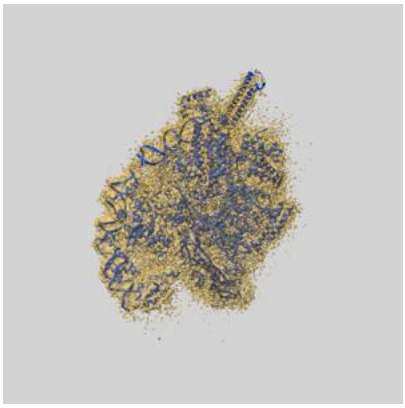
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.12	-	-
Author-provided FSC curve	-	-	-
Unmasked-calculated*	3.86	4.74	3.96

*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 3.86 differs from the reported value 3.12 by more than 10 %

9 Map-model fit

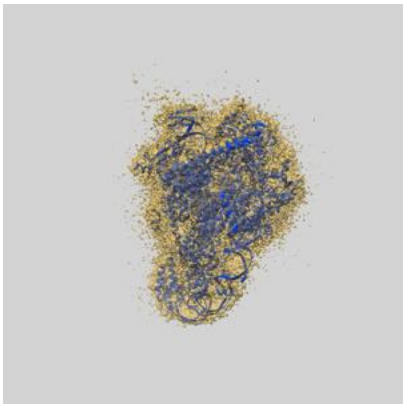
This section contains information regarding the fit between EMDB map EMD-27778 and PDB model 8DY9. Per-residue inclusion information can be found in section 3 on page 7.

9.1 Map-model overlay



X

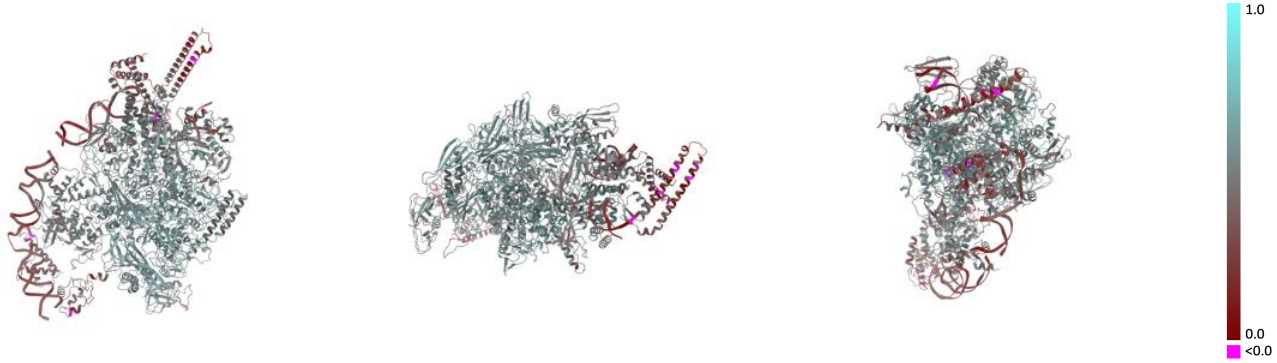
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 3.15 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.

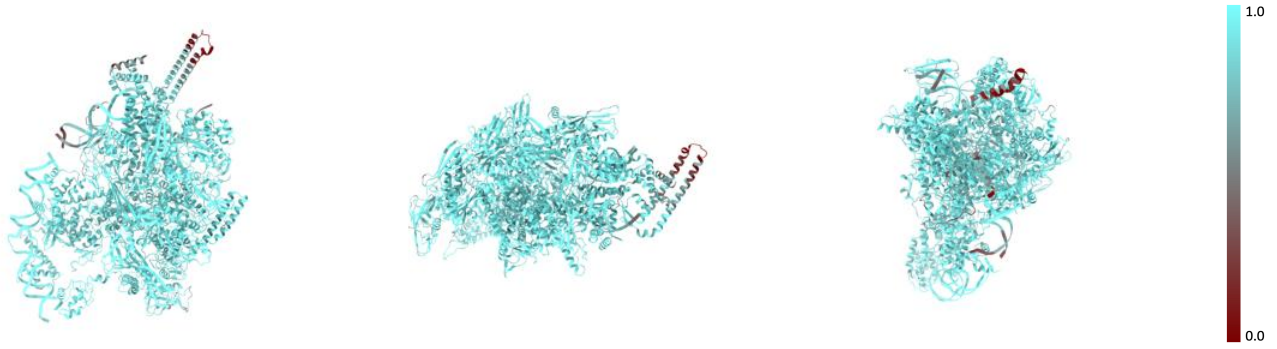
9.2 Q-score mapped to coordinate model



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.

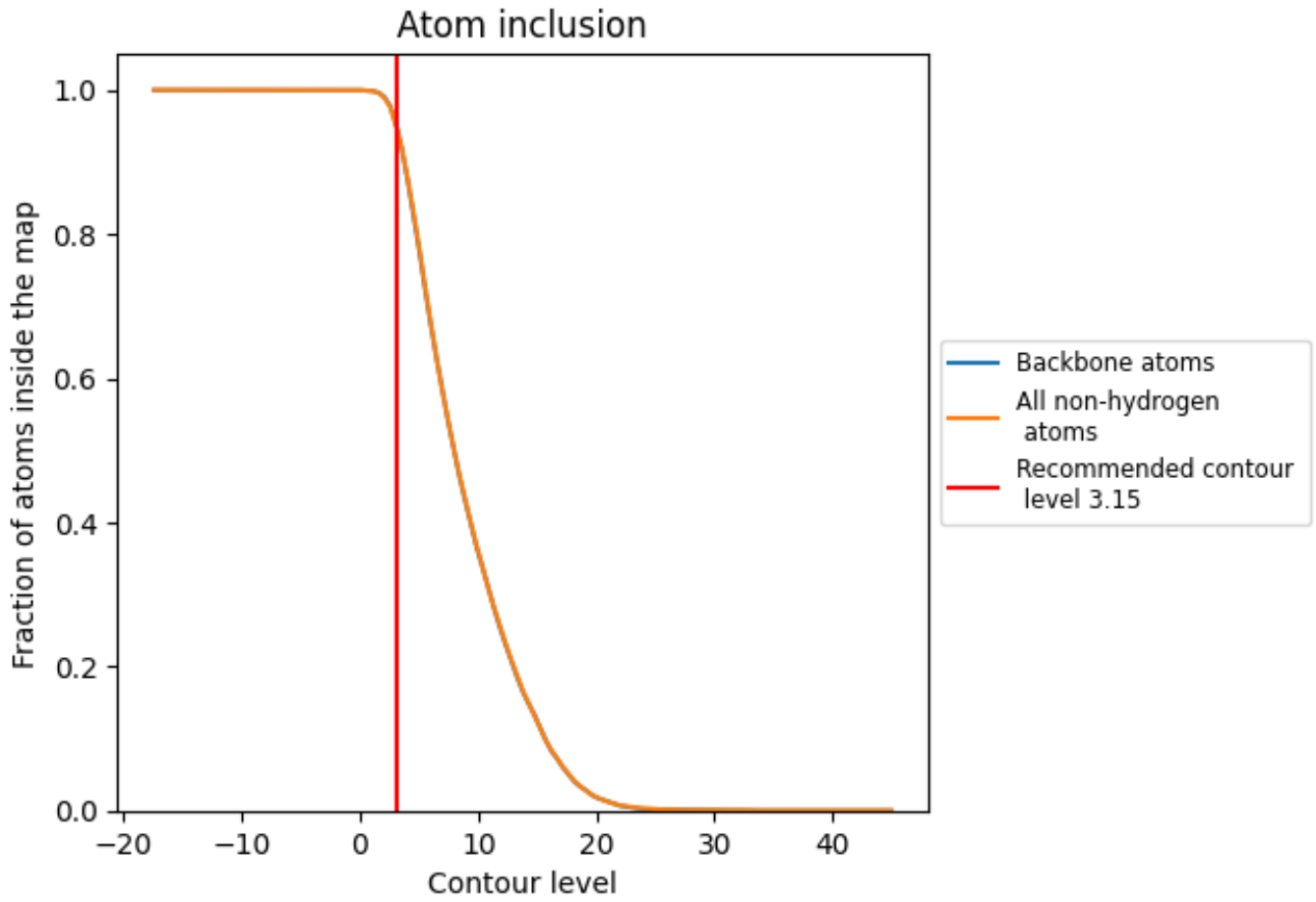
9.3 Atom inclusion mapped to coordinate model



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 (3.15).

9.4 Atom inclusion (i)



At the recommended contour level, 95% of all backbone atoms, 95% 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 (3.15) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.9471	0.4920
A	0.9635	0.5520
B	0.9475	0.4660
C	0.9729	0.5460
D	0.9435	0.5310
E	0.9625	0.5520
F	0.9425	0.4580
G	0.7060	0.3730
H	0.9863	0.4880
I	0.9772	0.4290
O	0.9677	0.2620
P	0.9629	0.2780
Q	0.8297	0.3110
R	0.8290	0.3150