wwPDB EM Validation Summary Report

Jul 3, 2024 – 02:45 am BST

PDB ID	:	7BEG
EMDB ID	:	EMD-12157
Title	:	Structures of class I bacterial transcription complexes
Authors	:	Ye, F.Z.; Hao, M.; Zhang, X.D.
Deposited on	:	2020-12-23
Resolution	:	4.20 Å(reported)
Based on initial model	:	4YLP

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:

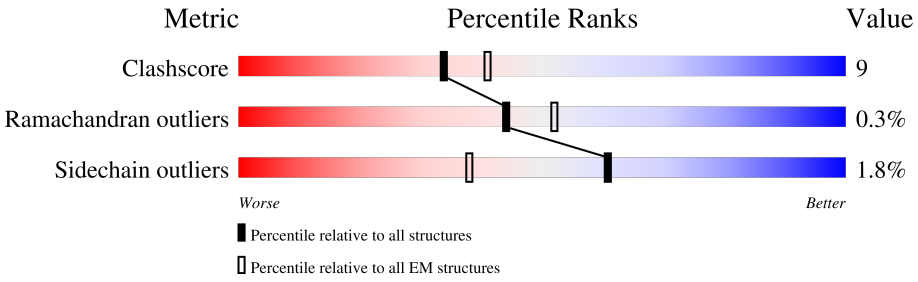
EMDB validation analysis	:	0.0.1.dev92
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.13
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 4.20 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.



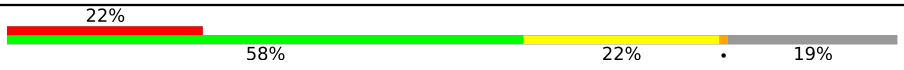
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

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	329	12% 77% 15% 7%
1	B	329	78% 13% 8%
2	C	1342	84% 15%
3	D	1407	80% 16%
4	E	91	89% 10%
5	F	630	69% 9% 21%
6	N	94	9% 29% 44% 27%
7	T	94	6% 26% 50% 23%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
8	G	130	

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 35246 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 subunit alpha.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	305	Total	C	N	O	S	0	0
			2372	1485	418	461	8		
1	B	302	Total	C	N	O	S	0	0
			2350	1471	413	458	8		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
2	C	1341	Total	C	N	O	S	0	0
			10570	6633	1839	2055	43		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
3	D	1362	Total	C	N	O	S	0	0
			10548	6622	1883	1996	47		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
4	E	90	Total	C	N	O	S	0	0
			708	430	136	141	1		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
5	F	496	Total	C	N	O	S	0	0
			3996	2499	708	766	23		

There are 17 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	-16	MET	-	initiating methionine	UNP Q0P6L9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	-15	ALA	-	expression tag	UNP Q0P6L9
F	-14	HIS	-	expression tag	UNP Q0P6L9
F	-13	HIS	-	expression tag	UNP Q0P6L9
F	-12	HIS	-	expression tag	UNP Q0P6L9
F	-11	HIS	-	expression tag	UNP Q0P6L9
F	-10	HIS	-	expression tag	UNP Q0P6L9
F	-9	HIS	-	expression tag	UNP Q0P6L9
F	-8	SER	-	expression tag	UNP Q0P6L9
F	-7	SER	-	expression tag	UNP Q0P6L9
F	-6	GLY	-	expression tag	UNP Q0P6L9
F	-5	LEU	-	expression tag	UNP Q0P6L9
F	-4	GLU	-	expression tag	UNP Q0P6L9
F	-3	VAL	-	expression tag	UNP Q0P6L9
F	-2	LEU	-	expression tag	UNP Q0P6L9
F	-1	PHE	-	expression tag	UNP Q0P6L9
F	0	GLN	-	expression tag	UNP Q0P6L9

- Molecule 6 is a DNA chain called Class I pacrA promoter non-template DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	N	94	Total	C	N	O	P	0	0
			1934	927	345	568	94		

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
N	?	-	DA	deletion	GB 1866584534
N	-5	DA	DC	conflict	GB 1866584534
N	-1	DT	-	insertion	GB 1866584534

- Molecule 7 is a DNA chain called Class I pacrA promoter template DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	T	94	Total	C	N	O	P	0	0
			1920	921	345	560	94		

- Molecule 8 is a protein called Transcriptional activator RamA.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	G	105	Total	C	N	O	S	0	0
			848	541	148	158	1		

There are 17 discrepancies between the modelled and reference sequences:

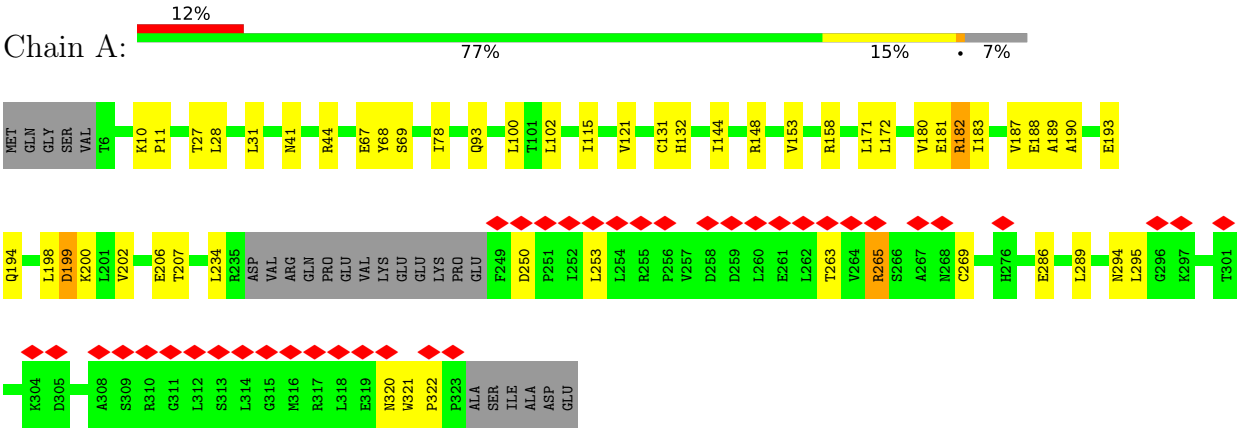
Chain	Residue	Modelled	Actual	Comment	Reference
G	-16	MET	-	initiating methionine	UNP Q48413
G	-15	ALA	-	expression tag	UNP Q48413
G	-14	HIS	-	expression tag	UNP Q48413
G	-13	HIS	-	expression tag	UNP Q48413
G	-12	HIS	-	expression tag	UNP Q48413
G	-11	HIS	-	expression tag	UNP Q48413
G	-10	HIS	-	expression tag	UNP Q48413
G	-9	HIS	-	expression tag	UNP Q48413
G	-8	SER	-	expression tag	UNP Q48413
G	-7	SER	-	expression tag	UNP Q48413
G	-6	GLY	-	expression tag	UNP Q48413
G	-5	LEU	-	expression tag	UNP Q48413
G	-4	GLU	-	expression tag	UNP Q48413
G	-3	VAL	-	expression tag	UNP Q48413
G	-2	LEU	-	expression tag	UNP Q48413
G	-1	PHE	-	expression tag	UNP Q48413
G	0	GLN	-	expression tag	UNP Q48413

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.

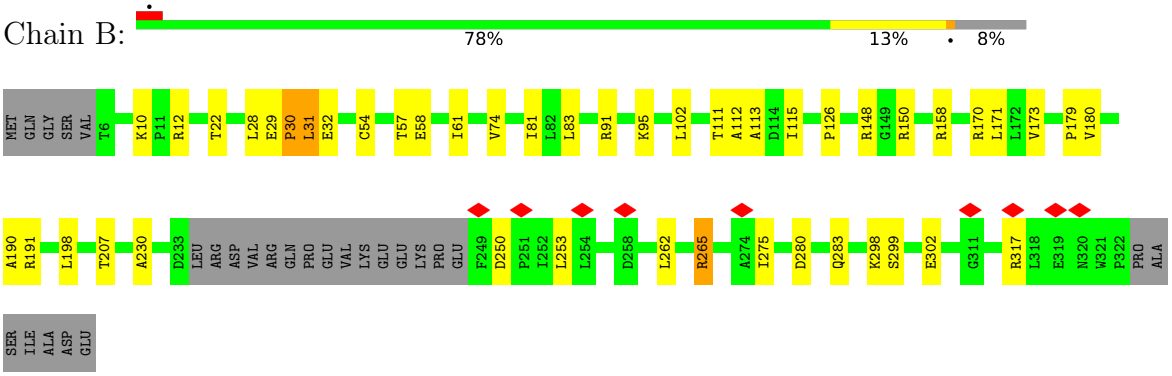
• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain A:



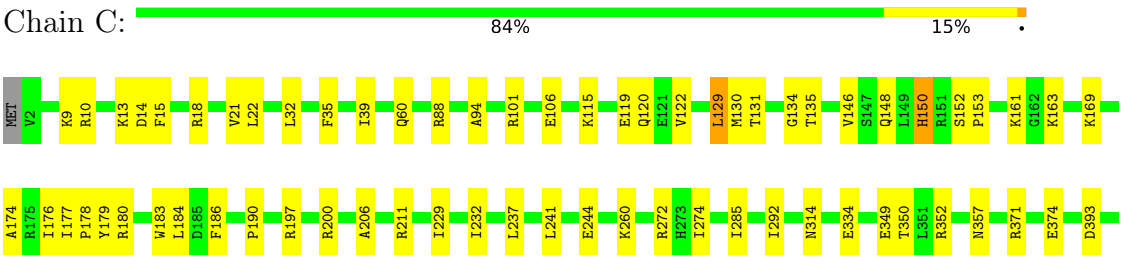
• Molecule 1: DNA-directed RNA polymerase subunit alpha

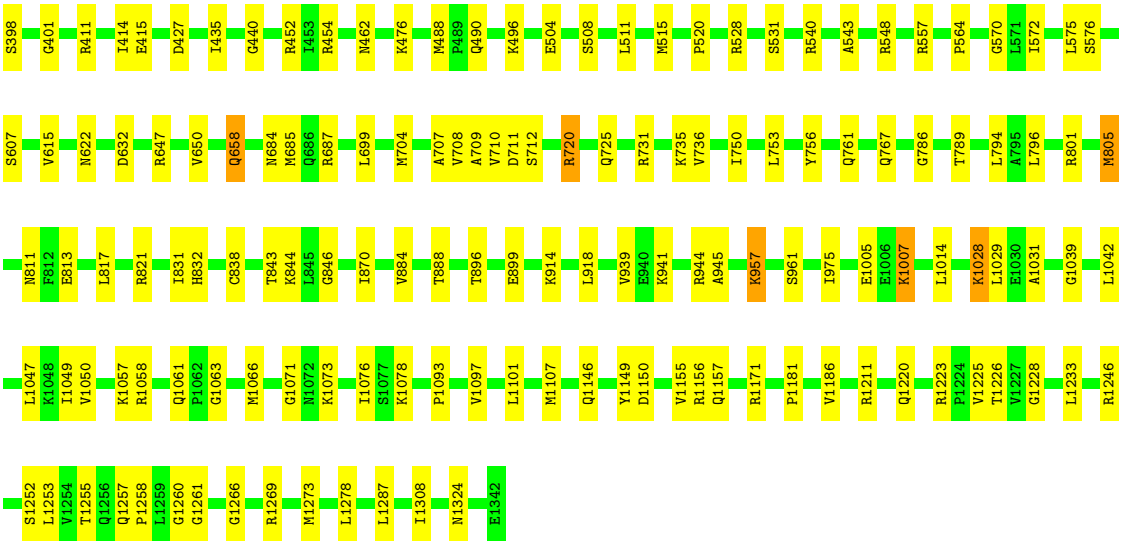
Chain B:



• Molecule 2: DNA-directed RNA polymerase subunit beta

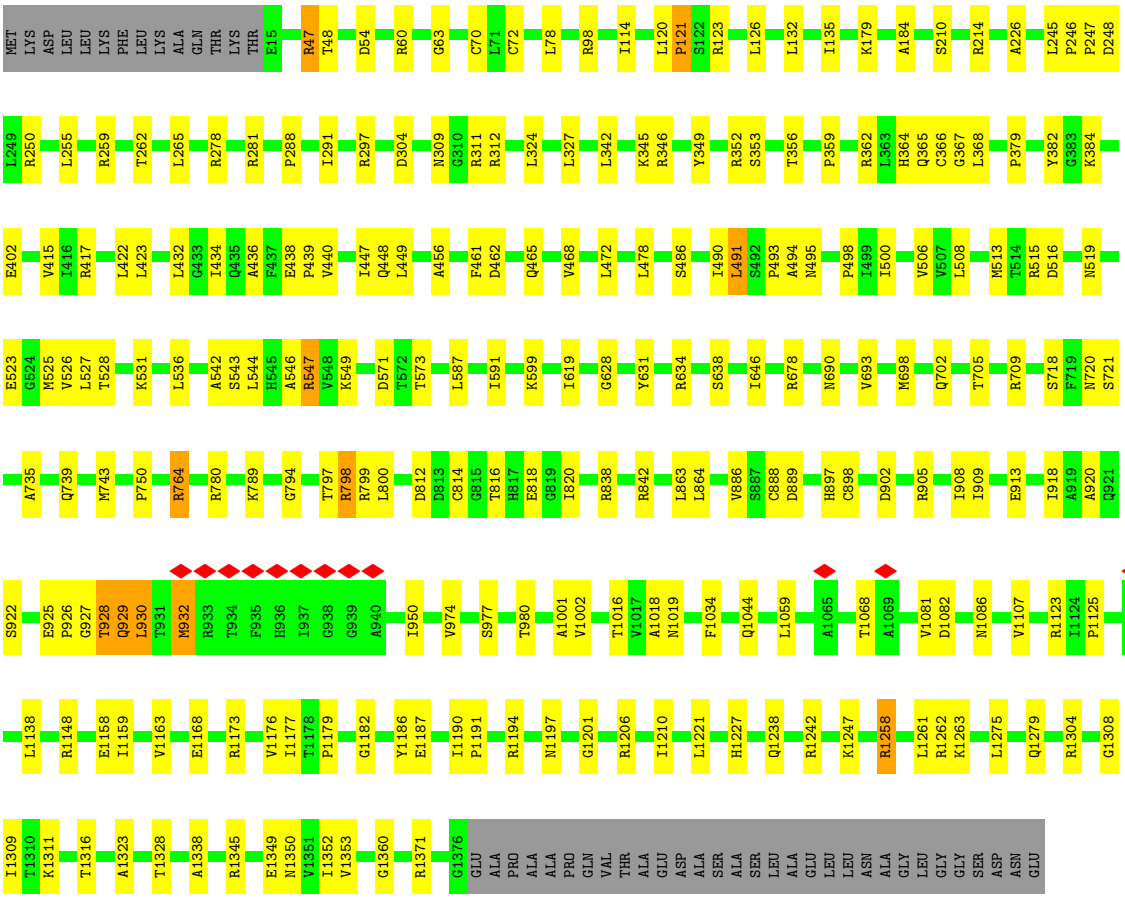
Chain C:





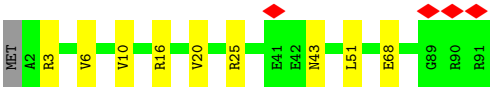
• Molecule 3: DNA-directed RNA polymerase subunit beta'

Chain D:

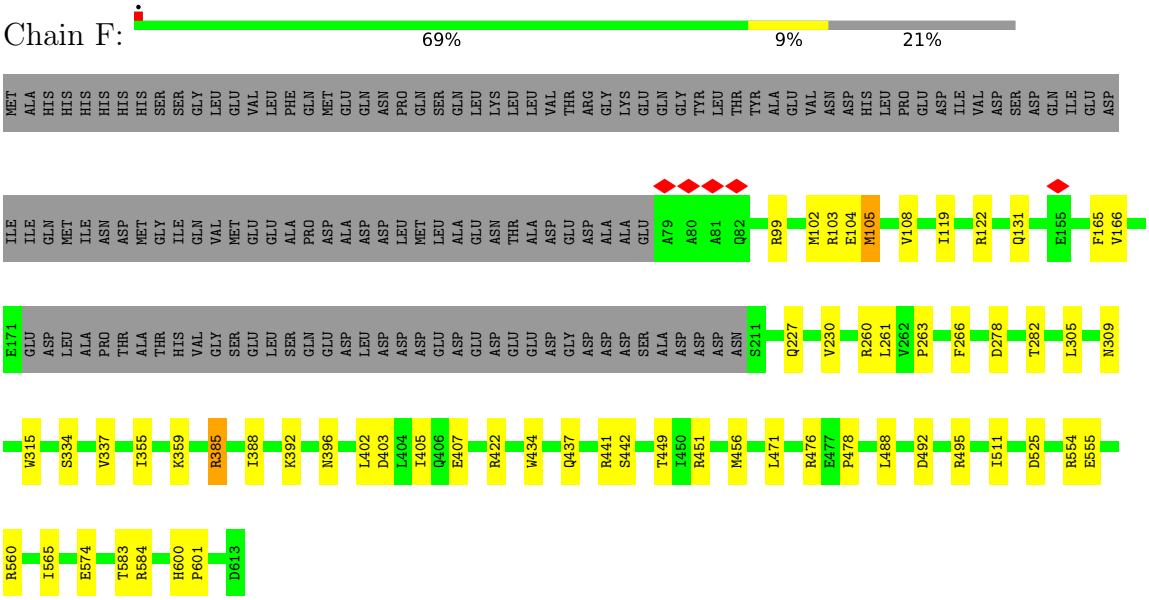


• Molecule 4: DNA-directed RNA polymerase subunit omega

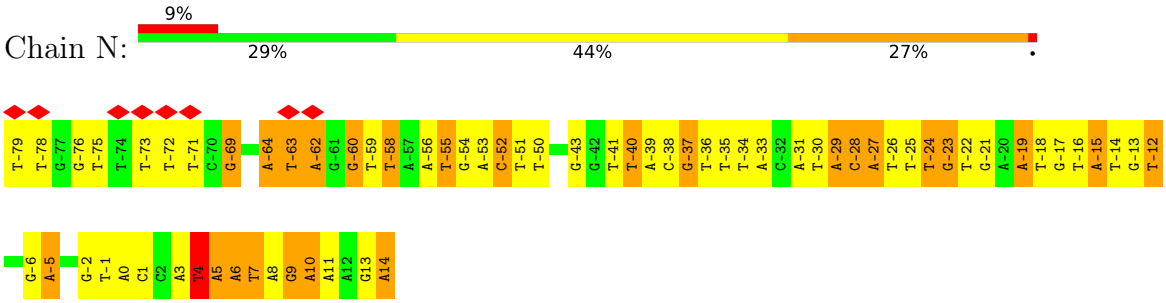
Chain E:



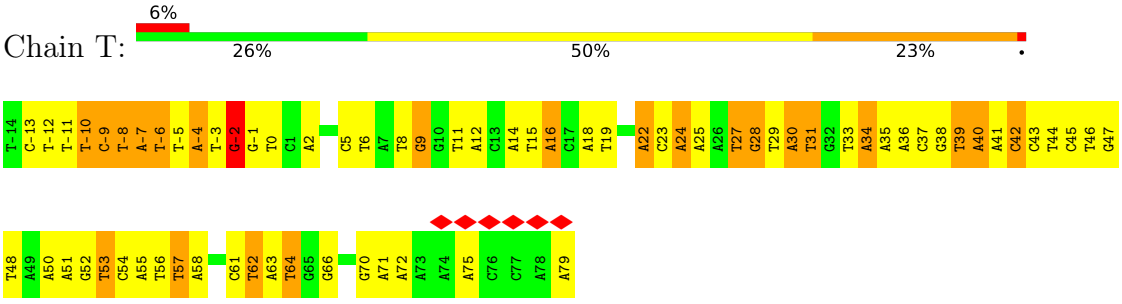
• Molecule 5: RNA polymerase sigma factor RpoD



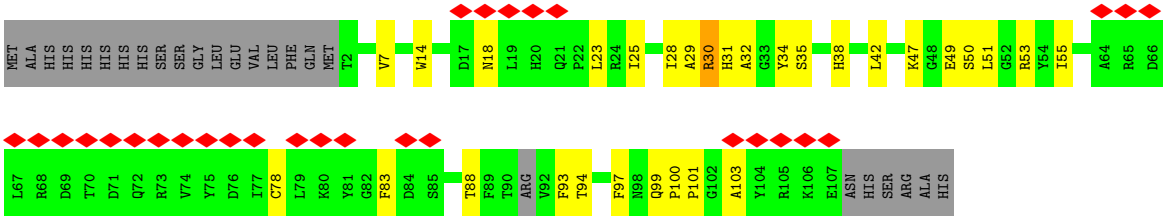
• Molecule 6: Class I pacrA promoter non-template DNA



• Molecule 7: Class I pacrA promoter template DNA



• Molecule 8: Transcriptional activator RamA


MET ALA HIS HIS HIS HIS HIS SER SER GLY LEU GLU VAL LEU PHE GLN MET T2 V7 W14 D17 N18 L19 H20 Q21 P22 L23 R24 I25 I28 A29 R30 H31 A32 G33 Y34 S35 H38 L42 K47 G48 E49 S50 L51 G52 R53 Y54 I55 A64 R65 D66
L67 R68 D69 T70 D71 Q72 R73 V74 Y75 D76 I77 C78 L79 K80 Y81 G82 F83 D84 S85 T88 F89 T90 ARG V92 F93 T94 F97 N98 Q99 P100 P101 G102 A103 Y104 R105 K106 E107 ASN HIS SER ARG ALA HIS

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	39164	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 ($e^-/Å^2$)	55	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	75000	Depositor
Image detector	FEI FALCON III (4k x 4k)	Depositor
Maximum map value	0.217	Depositor
Minimum map value	-0.094	Depositor
Average map value	-0.000	Depositor
Map value standard deviation	0.006	Depositor
Recommended contour level	0.01	Depositor
Map size (Å)	399.28003, 399.28003, 399.28003	wwPDB
Map dimensions	368, 368, 368	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.085, 1.085, 1.085	Depositor

5 Model quality

5.1 Standard geometry

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

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$\|Z\|$ >5	RMSZ	#$\|Z\|$ >5
1	A	0.29	2/2403 (0.1%)	0.49	1/3258 (0.0%)
1	B	0.25	0/2380	0.50	0/3225
2	C	0.26	0/10739	0.52	0/14492
3	D	0.26	0/10708	0.54	1/14461 (0.0%)
4	E	0.24	0/710	0.44	0/956
5	F	0.25	0/4050	0.49	1/5450 (0.0%)
6	N	1.42	26/2169 (1.2%)	1.31	32/3348 (1.0%)
7	T	1.13	20/2153 (0.9%)	1.26	27/3318 (0.8%)
8	G	0.25	0/865	0.45	0/1172
All	All	0.51	48/36177 (0.1%)	0.67	62/49680 (0.1%)

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
6	N	0	1
7	T	0	1
All	All	0	2

The worst 5 of 48 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	N	7	DT	C4'-O4'	20.78	1.65	1.45
6	N	-15	DA	O3'-P	-16.40	1.41	1.61
6	N	1	DC	O3'-P	-15.87	1.42	1.61
6	N	6	DA	C4'-O4'	14.36	1.59	1.45
6	N	-40	DT	O3'-P	-13.94	1.44	1.61

The worst 5 of 62 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
7	T	-4	DA	P-O3'-C3'	14.49	137.08	119.70
6	N	-15	DA	P-O3'-C3'	13.94	136.43	119.70
6	N	1	DC	P-O3'-C3'	13.09	135.41	119.70
6	N	-63	DT	O3'-P-O5'	11.45	125.76	104.00
6	N	-40	DT	P-O3'-C3'	11.45	133.43	119.70

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
6	N	4	DT	Sidechain
7	T	-2	DG	Sidechain

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	2372	0	2420	40	0
1	B	2350	0	2400	36	0
2	C	10570	0	10580	132	0
3	D	10548	0	10754	156	0
4	E	708	0	719	4	0
5	F	3996	0	4044	40	0
6	N	1934	0	1068	123	0
7	T	1920	0	1064	130	0
8	G	848	0	811	41	0
All	All	35246	0	33860	607	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 9.

The worst 5 of 607 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)
6:N:-75:DT:C7	8:G:94:THR:HG21	1.45	1.47
6:N:-60:DG:H2"	6:N:-59:DT:C7	1.47	1.40

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
6:N:7:DT:O4'	6:N:7:DT:C4'	1.65	1.32
7:T:18:DA:H2"	7:T:19:DT:C7	1.64	1.27
6:N:-53:DA:C2'	6:N:-52:DC:H5"	1.68	1.23

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	301/329 (92%)	289 (96%)	12 (4%)	0	100	100
1	B	298/329 (91%)	280 (94%)	17 (6%)	1 (0%)	41	76
2	C	1339/1342 (100%)	1228 (92%)	106 (8%)	5 (0%)	34	72
3	D	1360/1407 (97%)	1247 (92%)	107 (8%)	6 (0%)	34	72
4	E	88/91 (97%)	85 (97%)	3 (3%)	0	100	100
5	F	492/630 (78%)	463 (94%)	29 (6%)	0	100	100
8	G	101/130 (78%)	92 (91%)	8 (8%)	1 (1%)	15	54
All	All	3979/4258 (93%)	3684 (93%)	282 (7%)	13 (0%)	44	76

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	658	GLN
2	C	709	ALA
8	G	30	ARG
2	C	957	LYS
3	D	78	LEU

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	264/286 (92%)	260 (98%)	4 (2%)	65 80
1	B	262/286 (92%)	257 (98%)	5 (2%)	57 74
2	C	1155/1157 (100%)	1137 (98%)	18 (2%)	62 79
3	D	1129/1168 (97%)	1106 (98%)	23 (2%)	55 73
4	E	74/75 (99%)	72 (97%)	2 (3%)	44 66
5	F	435/555 (78%)	424 (98%)	11 (2%)	47 68
8	G	87/116 (75%)	87 (100%)	0	100 100
All	All	3406/3643 (94%)	3343 (98%)	63 (2%)	61 76

5 of 63 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	297	ARG
5	F	309	ASN
3	D	525	MET
5	F	105	MET
5	F	451	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 36 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	1019	ASN
8	G	38	HIS
3	D	1227	HIS
3	D	1350	ASN
2	C	1061	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)

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)

There are no chain breaks in this entry.

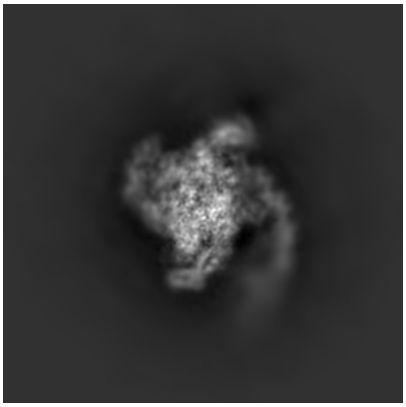
6 Map visualisation

This section contains visualisations of the EMDB entry EMD-12157. These allow visual inspection of the internal detail of the map and identification of artifacts.

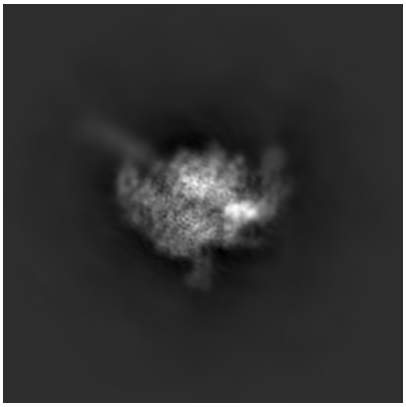
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections

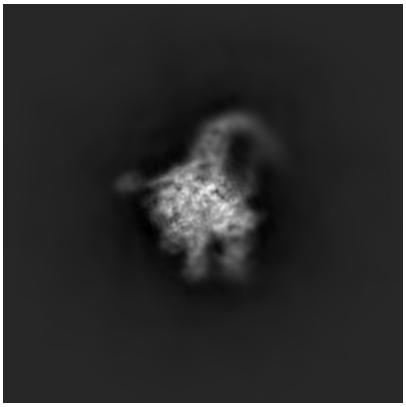
6.1.1 Primary map



X



Y

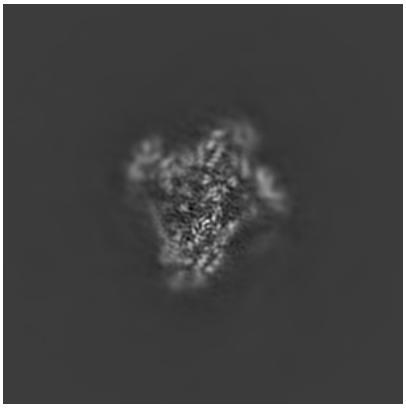


Z

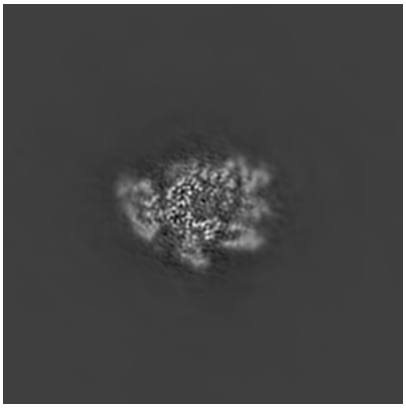
The images above show the map projected in three orthogonal directions.

6.2 Central slices

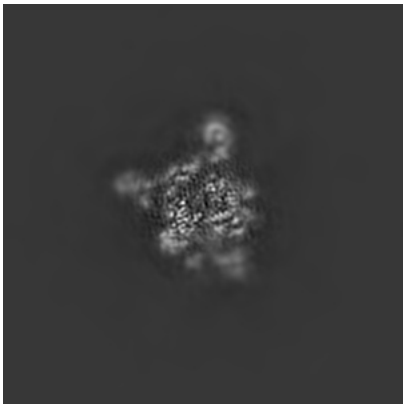
6.2.1 Primary map



X Index: 184



Y Index: 184

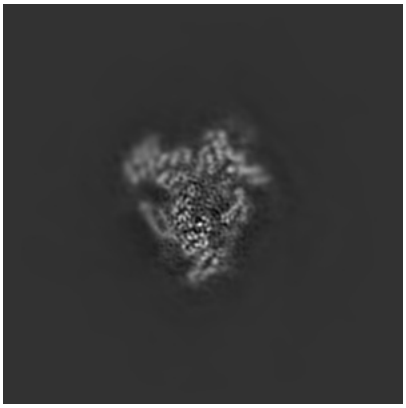


Z Index: 184

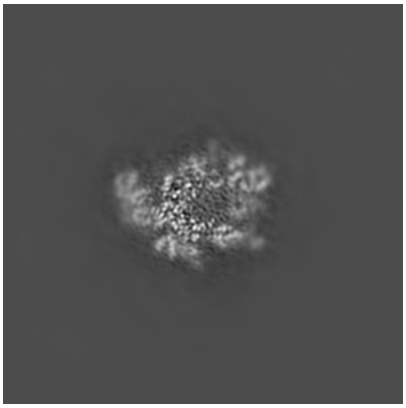
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

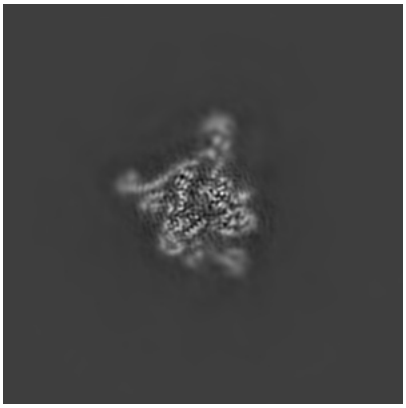
6.3.1 Primary map



X Index: 174



Y Index: 178

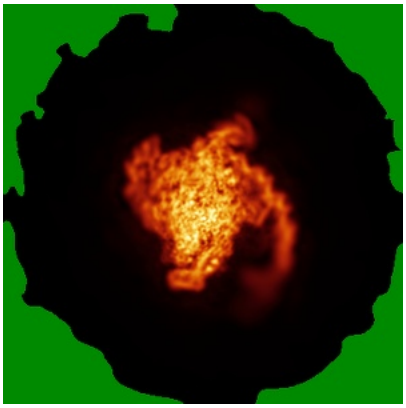


Z Index: 178

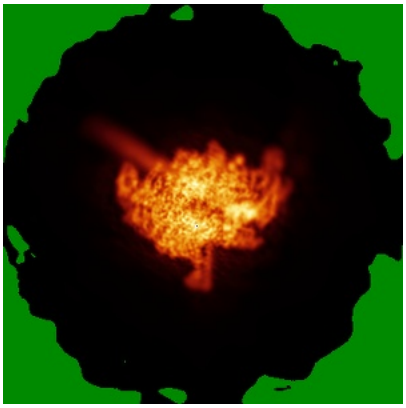
The images above show the largest variance slices of the map in three orthogonal directions.

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

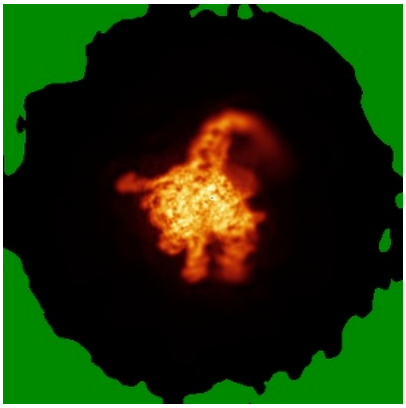
6.4.1 Primary map



X



Y

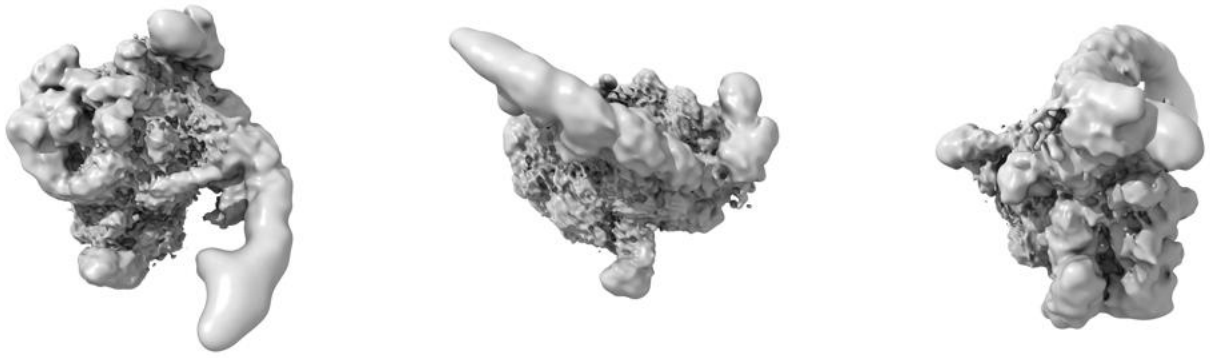


Z

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



X Y Z

The images above show the 3D surface view of the map at the recommended contour level 0.01. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

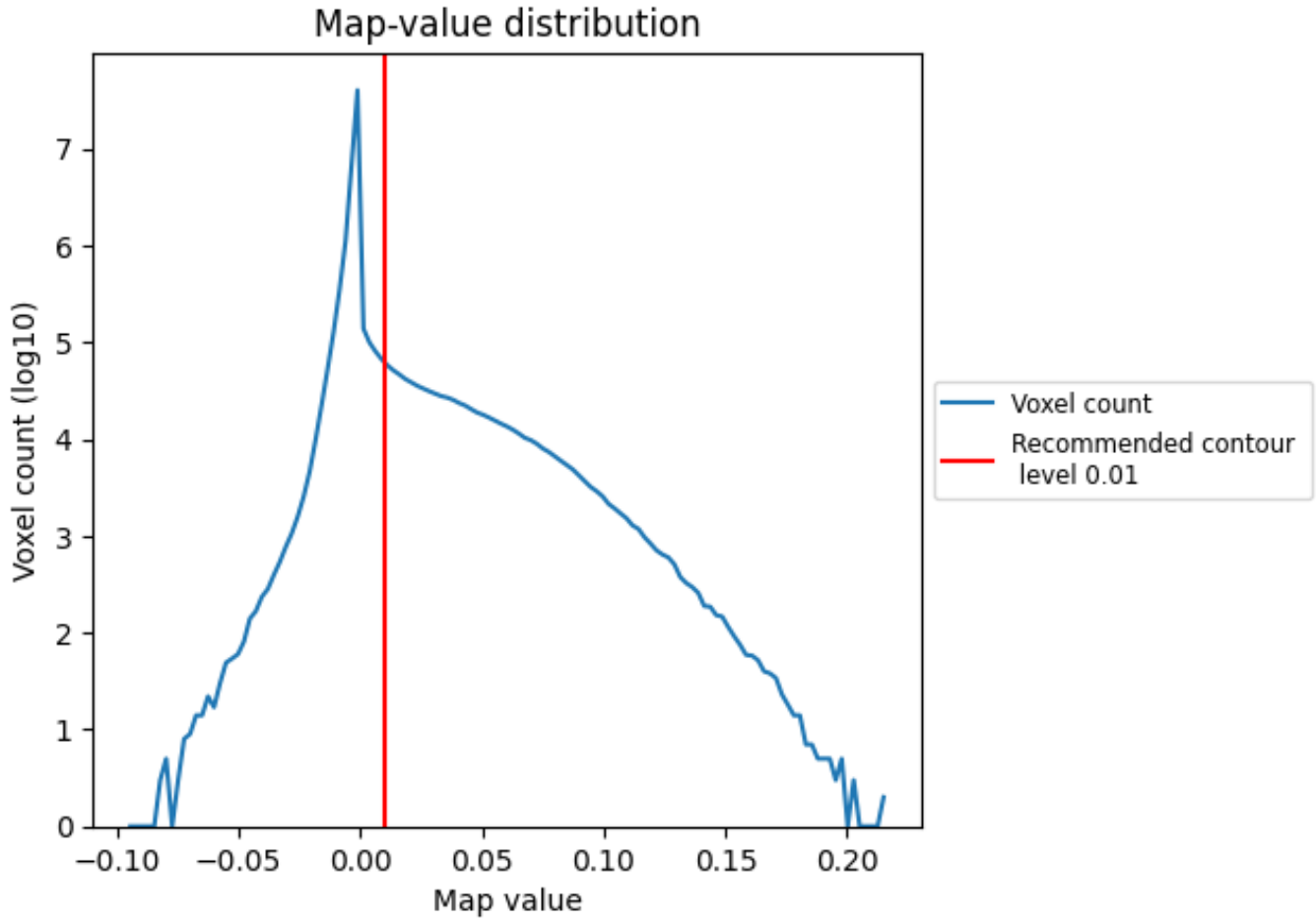
6.6 Mask visualisation (i)

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

7 Map analysis

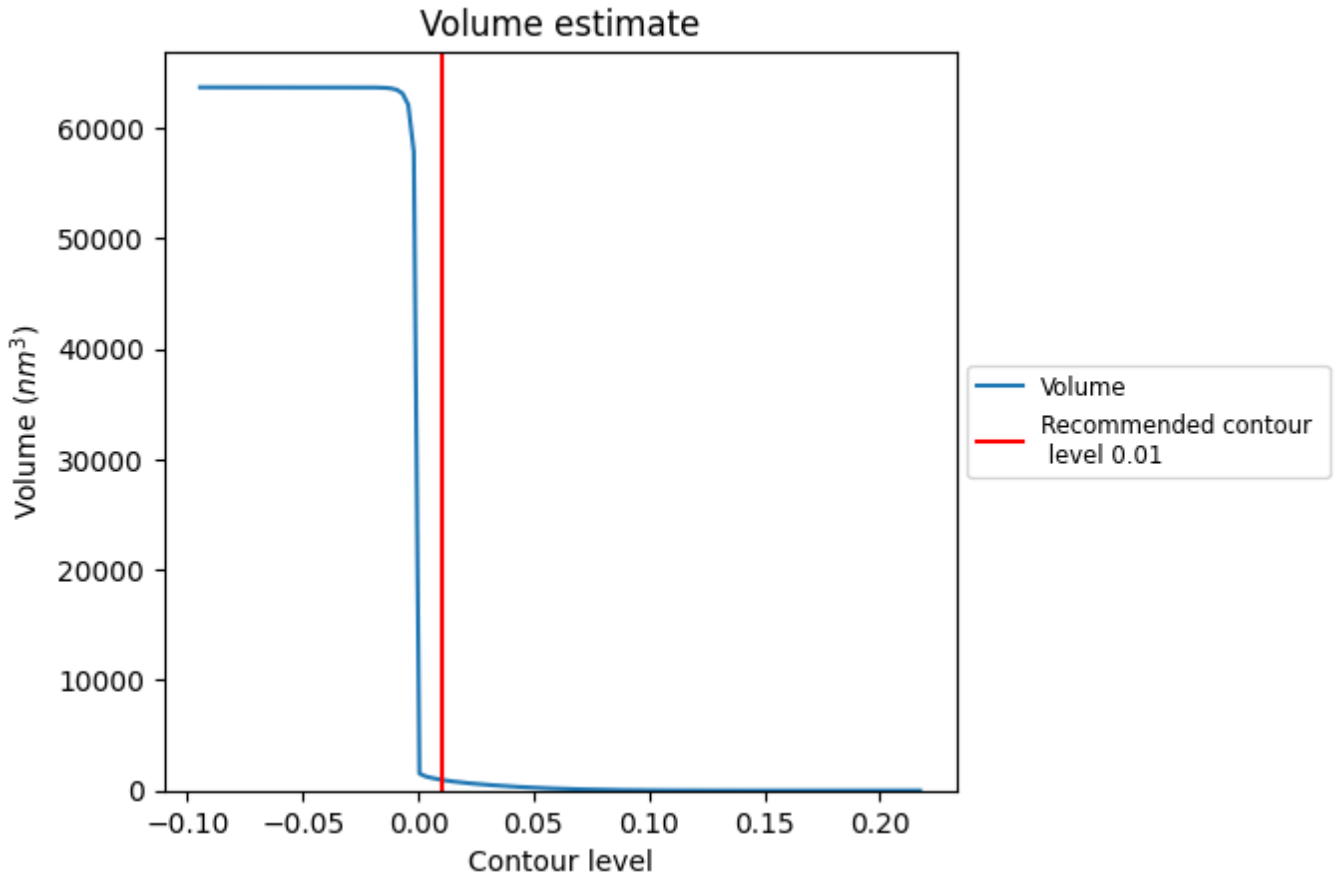
This section contains the results of statistical analysis of the map.

7.1 Map-value distribution



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

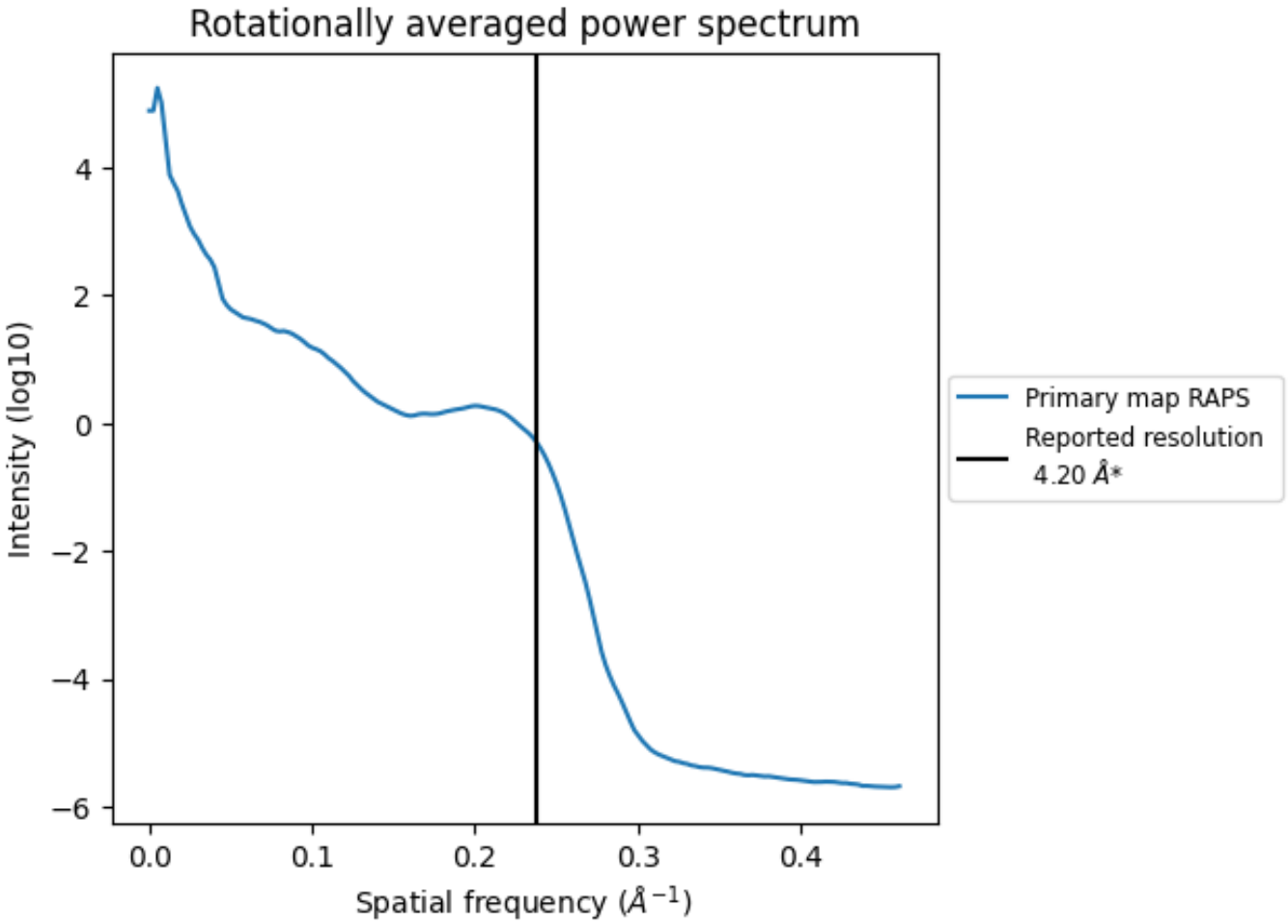
7.2 Volume estimate (i)



The volume at the recommended contour level is 976 nm^3; this corresponds to an approximate mass of 882 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 0.238 Å^{-1}

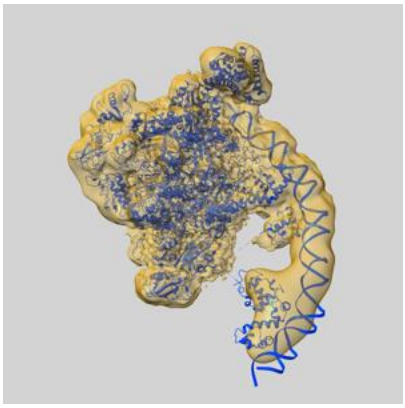
8 Fourier-Shell correlation (i)

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

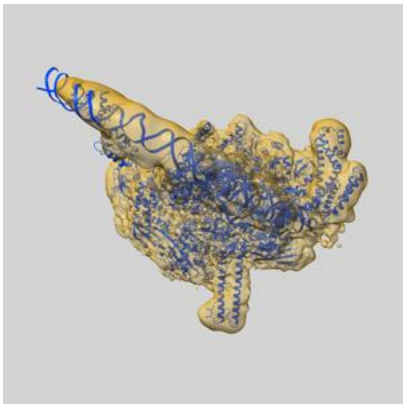
9 Map-model fit

This section contains information regarding the fit between EMDB map EMD-12157 and PDB model 7BEG. Per-residue inclusion information can be found in section 3 on page 7.

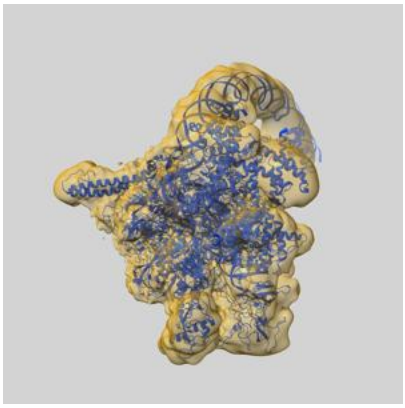
9.1 Map-model overlay



X



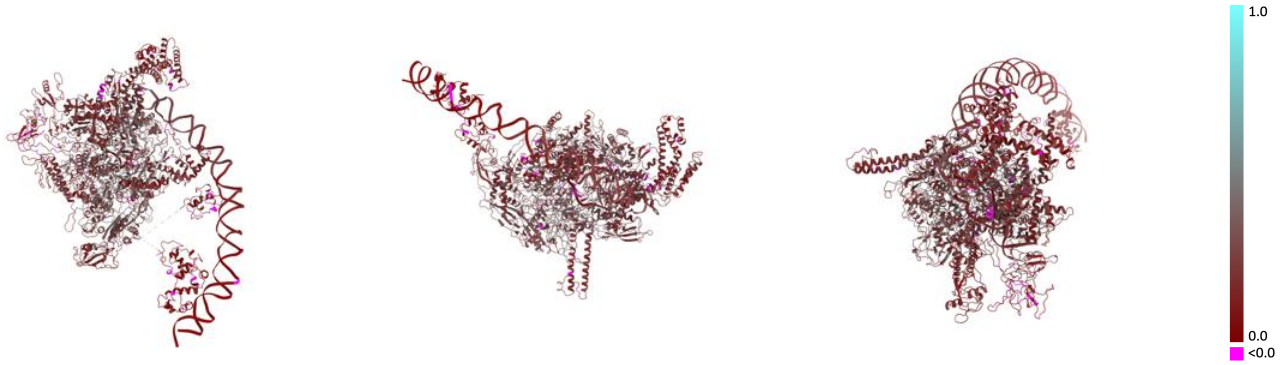
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.01 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

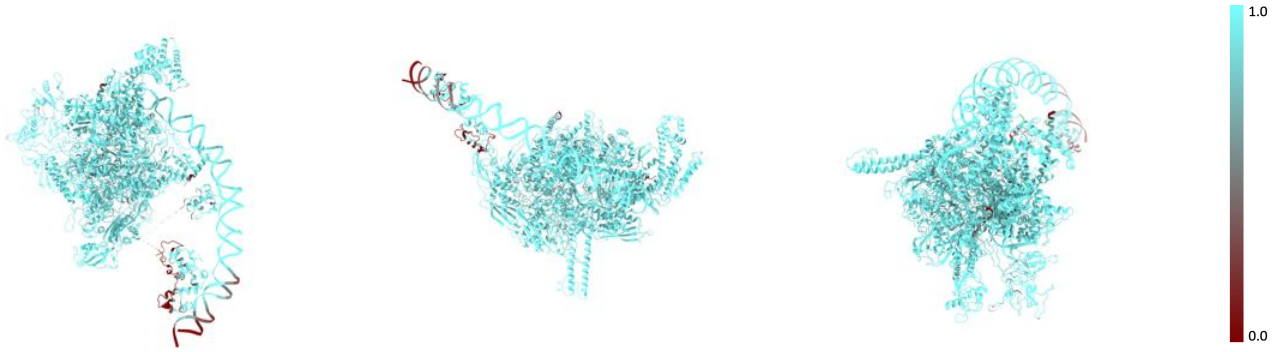
9.2 Q-score mapped to coordinate model



X Y Z

The images above show the model with each residue coloured according its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions at which atoms can be resolved. Low Q-score values may therefore be expected for many entries.

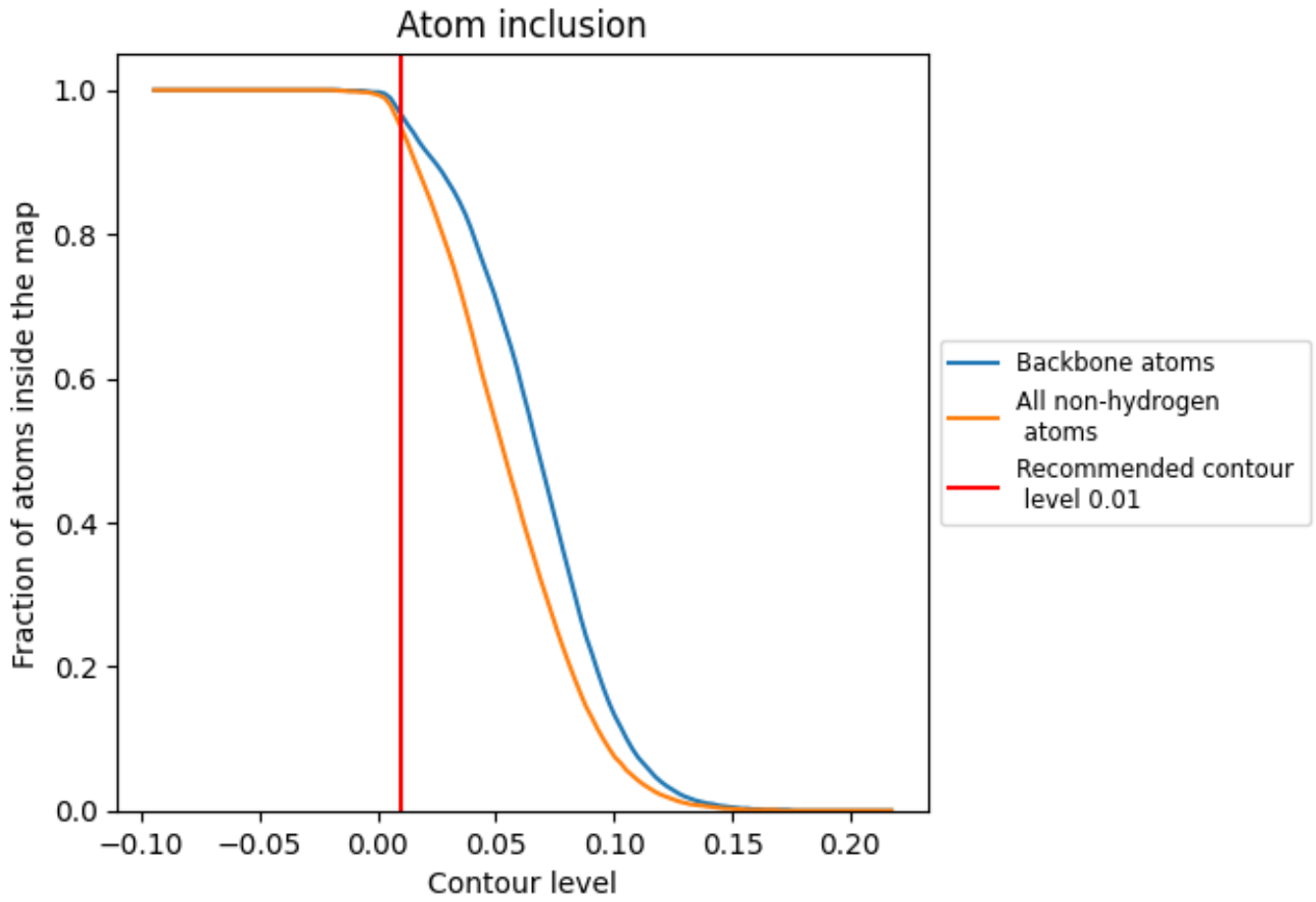
9.3 Atom inclusion mapped to coordinate model



X Y Z

The images above show the model with each residue coloured according to its atom inclusion. This shows to what extent they are inside the map at the recommended contour level (0.01).

9.4 Atom inclusion (i)



At the recommended contour level, 97% 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 (0.01) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.9480	0.2210
A	0.8590	0.2270
B	0.9540	0.1880
C	0.9840	0.2710
D	0.9710	0.2340
E	0.8550	0.2030
F	0.9720	0.1760
G	0.6840	0.0640
N	0.8950	0.1640
T	0.8830	0.1430