wwPDB EM Validation Summary Report (i)

May 20, 2024 – 10:26 am BST

PDB ID	:	8R1J
EMDB ID	:	EMD-18818
Title	:	Structure of avian H5N1 influenza A polymerase dimer in complex with human ANP32B.
Authors	:	Carrique, L.; Staller, E.; Keown, J.R.; Fan, H.; Fodor, E.; Grimes, J.M.
Deposited on	:	2023-11-02
Resolution	:	3.20 Å(reported)
Based on initial models	:	., 6RR7

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 (i) 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 (i)) were used in the production of this report:

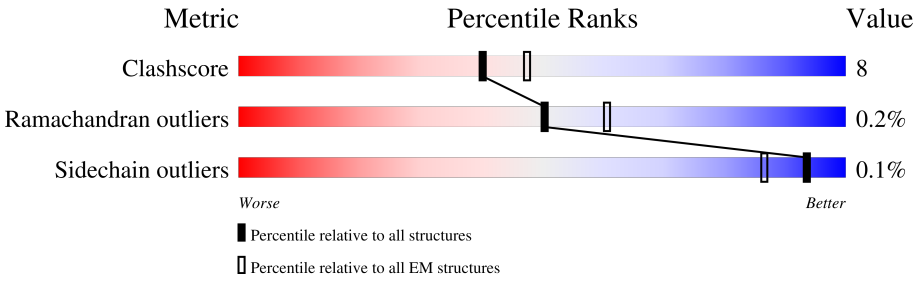
EMDB validation analysis	:	0.0.1.dev92
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.13
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.20 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	716	10% 83% 14%
1	D	716	15% 82% 16%
2	B	757	73% 18% 9%
2	E	757	63% 17% 20%
3	C	898	23% 68% 15% 17%
3	F	898	11% 60% 10% 30%
4	G	251	6% 49% 14% 37%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 67775 atoms, of which 33874 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 Polymerase acidic protein.

Mol	Chain	Residues	Atoms						AltConf	Trace
1	A	699	Total	C	H	N	O	S	0	0
			11256	3589	5596	958	1072	41		
1	D	707	Total	C	H	N	O	S	0	0
			11373	3621	5658	968	1085	41		

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	variant	UNP A5Z236
A	2	GLU	-	variant	UNP A5Z236
A	3	ASP	-	variant	UNP A5Z236
A	556	ARG	GLN	engineered mutation	UNP A5Z236
D	1	MET	-	variant	UNP A5Z236
D	2	GLU	-	variant	UNP A5Z236
D	3	ASP	-	variant	UNP A5Z236
D	556	ARG	GLN	engineered mutation	UNP A5Z236

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

Mol	Chain	Residues	Atoms						AltConf	Trace
2	B	689	Total	C	H	N	O	S	0	0
			10982	3480	5468	950	1042	42		
2	E	607	Total	C	H	N	O	S	0	0
			9585	3052	4752	828	917	36		

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	MET	-	variant	UNP A5Z231
B	2	ASP	-	variant	UNP A5Z231
B	3	VAL	-	variant	UNP A5Z231
B	577	GLU	LYS	engineered mutation	UNP A5Z231
E	1	MET	-	variant	UNP A5Z231

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	2	ASP	-	variant	UNP A5Z231
E	3	VAL	-	variant	UNP A5Z231
E	577	GLU	LYS	engineered mutation	UNP A5Z231

- Molecule 3 is a protein called Polymerase basic protein 2.

Mol	Chain	Residues	Atoms						AltConf	Trace
3	C	743	Total	C	H	N	O	S	0	0
			11937	3707	6033	1063	1095	39		
3	F	633	Total	C	H	N	O	S	0	0
			10071	3160	5064	894	919	34		

- Molecule 4 is a protein called Acidic leucine-rich nuclear phosphoprotein 32 family member B.

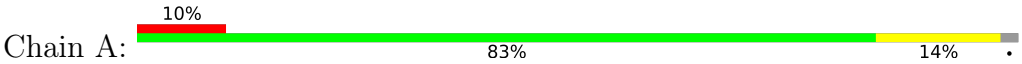
Mol	Chain	Residues	Atoms						AltConf	Trace
4	G	159	Total	C	H	N	O	S	0	0
			2571	801	1303	216	245	6		

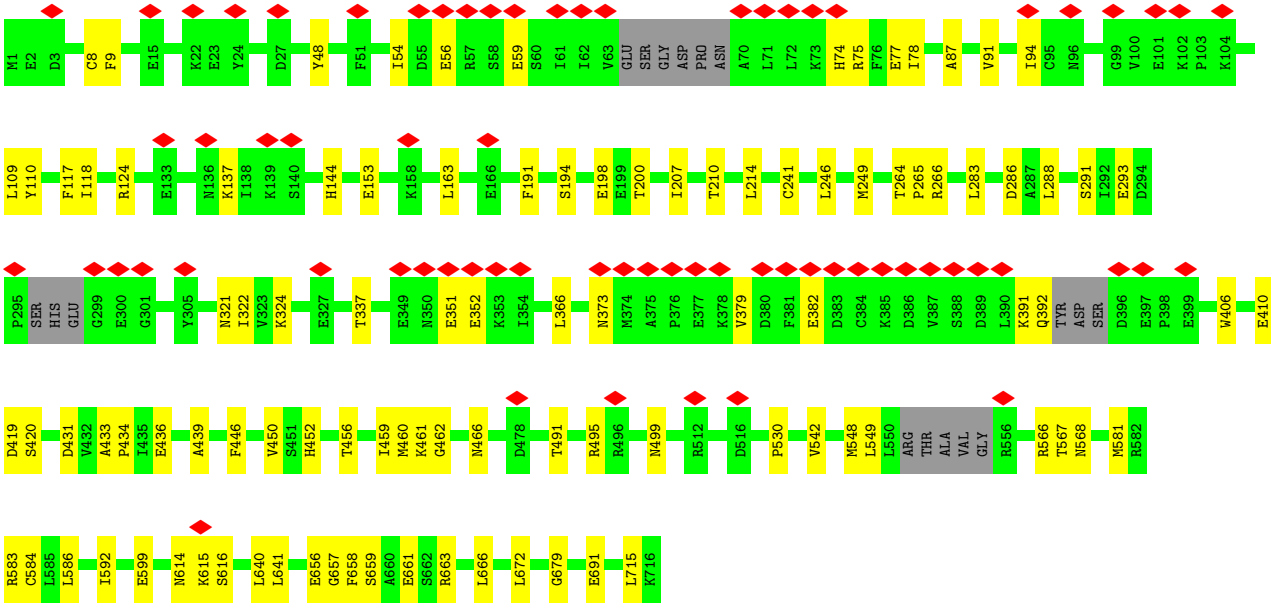
3 Residue-property plots (i)

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: Polymerase acidic protein

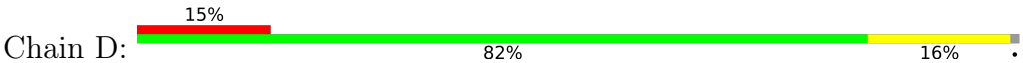
Chain A:

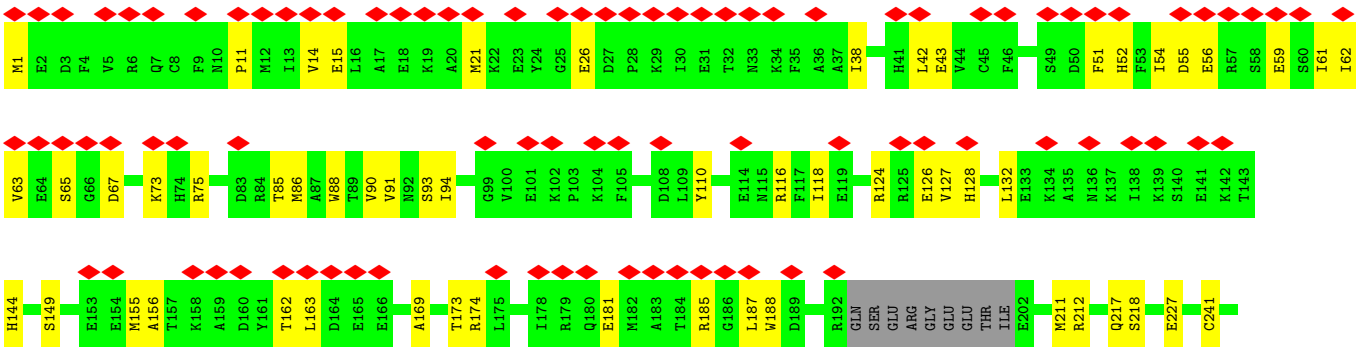


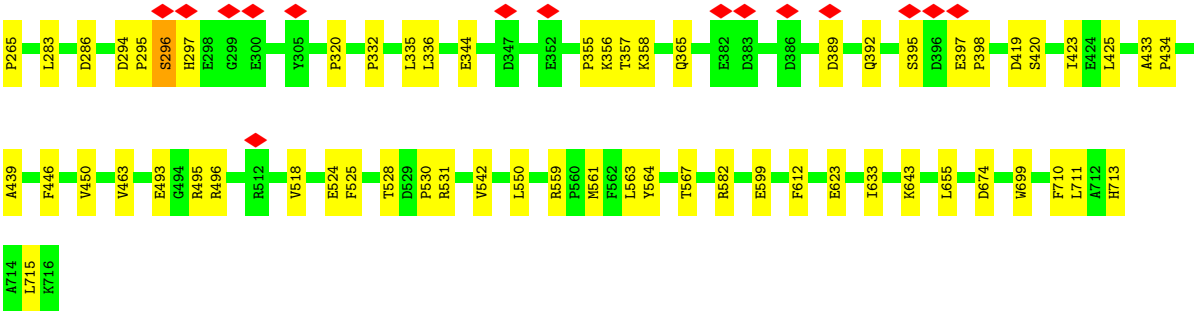


• Molecule 1: Polymerase acidic protein

Chain D:

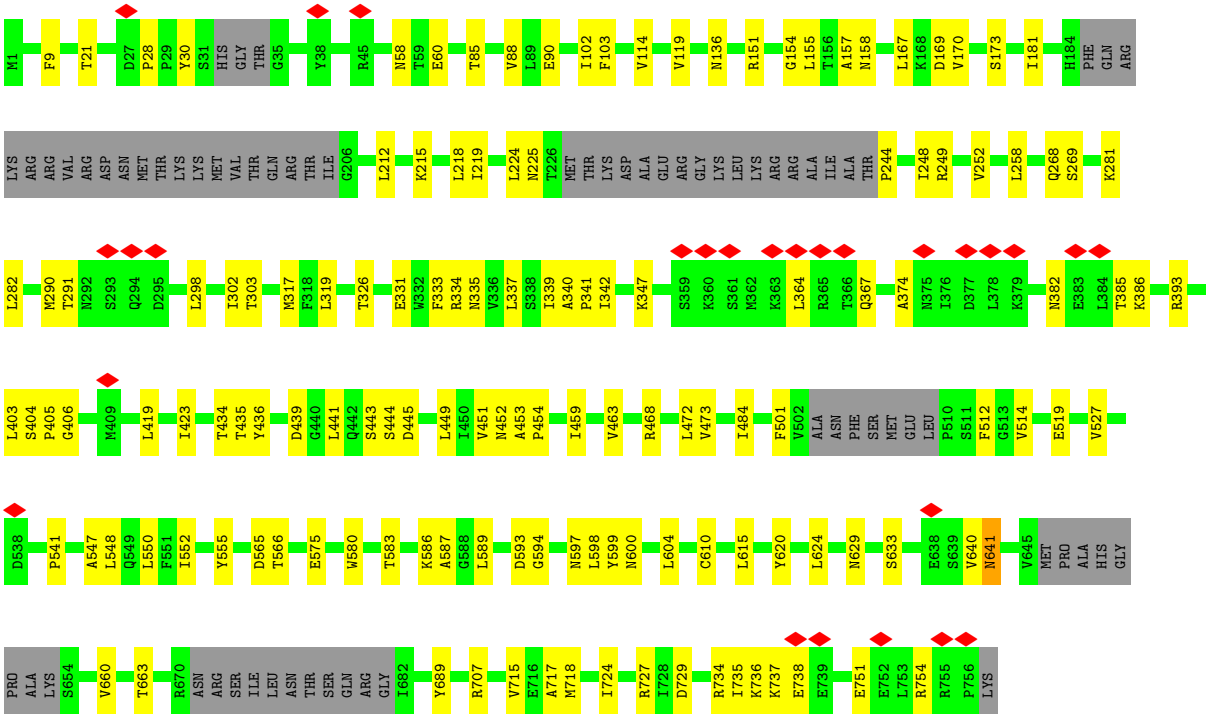






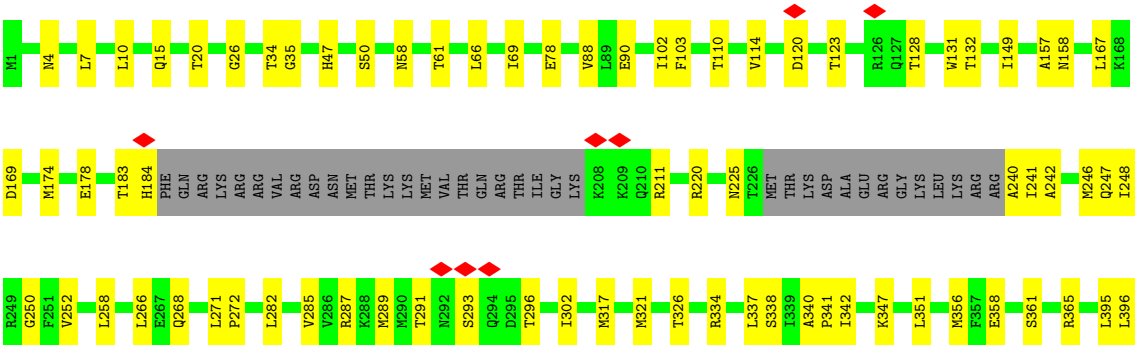
• Molecule 2: RNA-directed RNA polymerase catalytic subunit

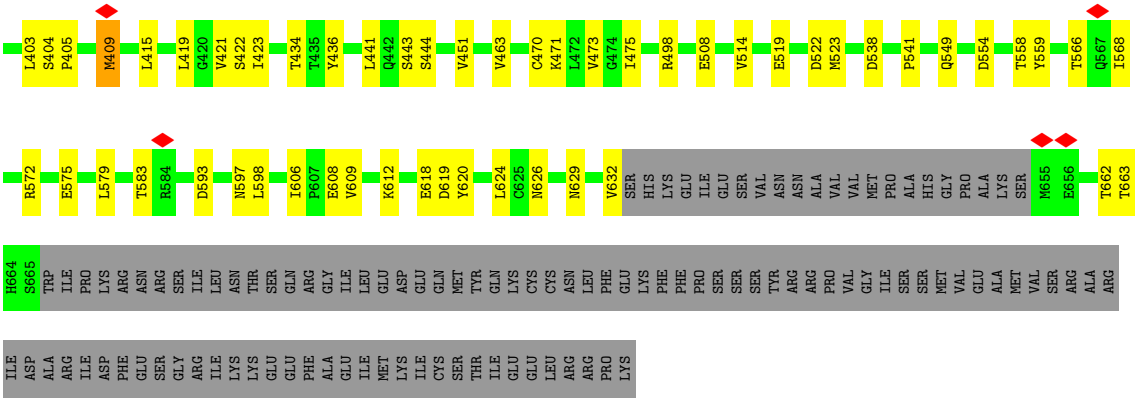
Chain B:



• Molecule 2: RNA-directed RNA polymerase catalytic subunit

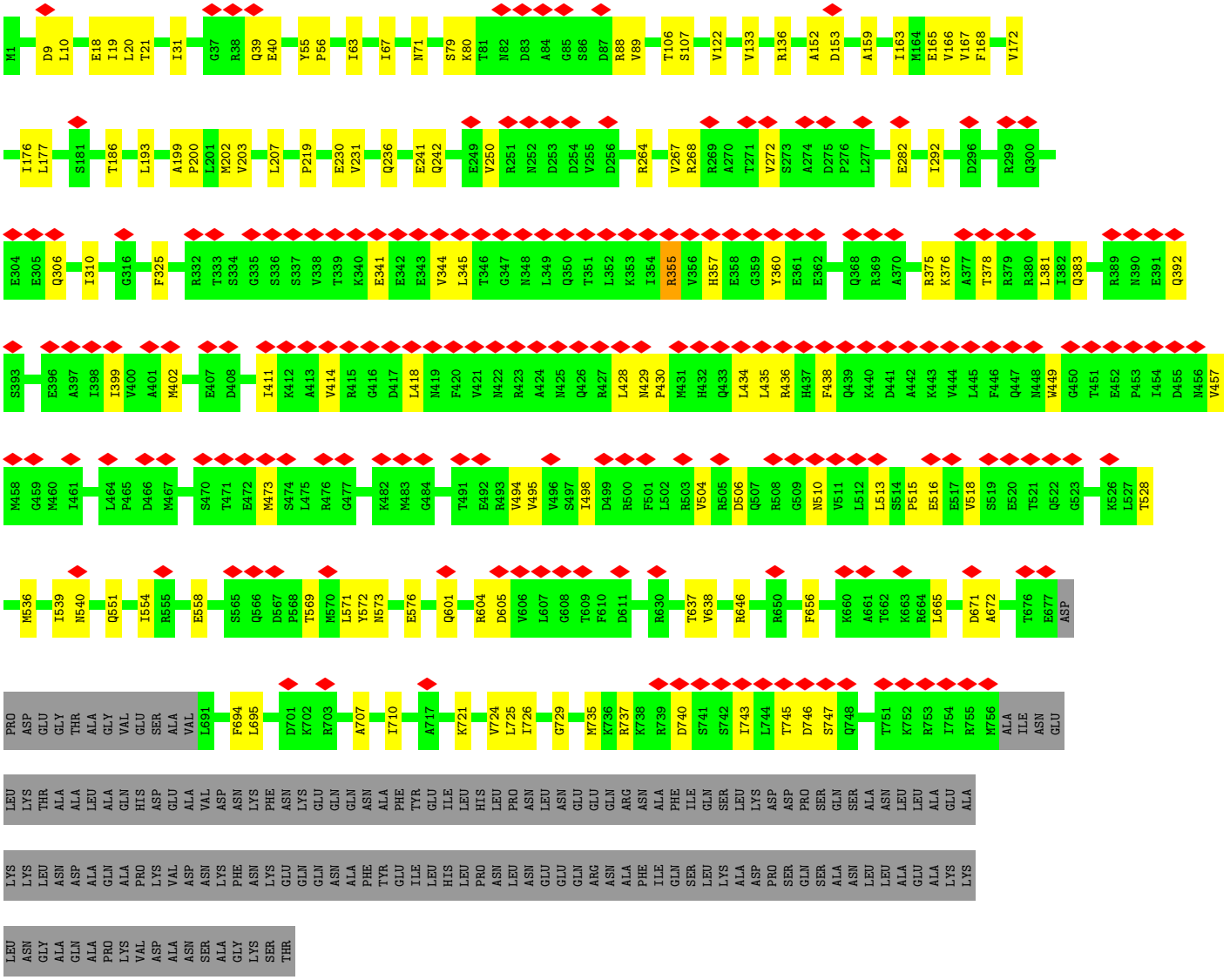
Chain E:





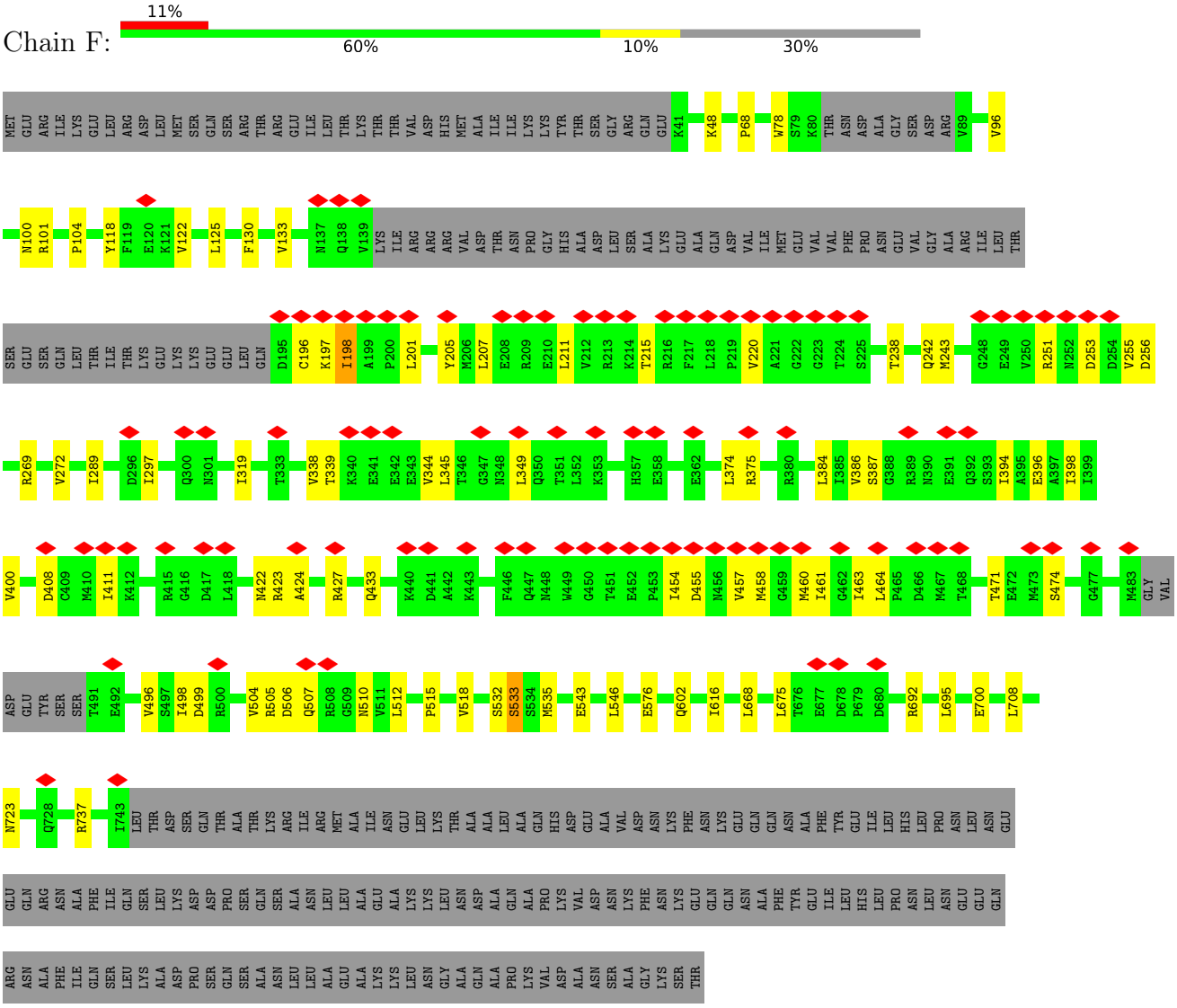
• Molecule 3: Polymerase basic protein 2

Chain C:



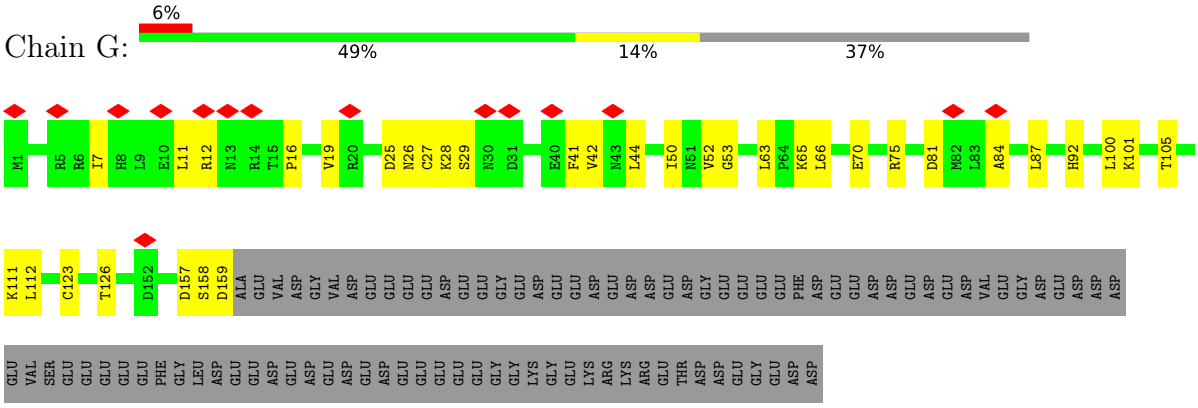
• Molecule 3: Polymerase basic protein 2

Chain F:



• Molecule 4: Acidic leucine-rich nuclear phosphoprotein 32 family member B

Chain G:



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	105000	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	TFS KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	50	Depositor
Minimum defocus (nm)	5000	Depositor
Maximum defocus (nm)	24000	Depositor
Magnification	Not provided	
Image detector	TFS FALCON 4i (4k x 4k)	Depositor
Maximum map value	43.208	Depositor
Minimum map value	-29.042	Depositor
Average map value	0.002	Depositor
Map value standard deviation	0.969	Depositor
Recommended contour level	5.84	Depositor
Map size (Å)	372.81, 372.81, 372.81	wwPDB
Map dimensions	300, 300, 300	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.2427, 1.2427, 1.2427	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	0.28	0/5775	0.49	0/7774
1	D	0.27	0/5836	0.49	0/7864
2	B	0.27	0/5620	0.50	0/7576
2	E	0.27	0/4930	0.48	0/6662
3	C	0.25	0/6000	0.53	0/8090
3	F	0.25	0/5095	0.51	0/6877
4	G	0.25	0/1284	0.50	0/1730
All	All	0.26	0/34540	0.50	0/46573

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

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	5660	5596	5617	83	0
1	D	5715	5658	5659	83	0
2	B	5514	5468	5487	120	0
2	E	4833	4752	4793	90	0
3	C	5904	6033	6043	102	0
3	F	5007	5064	5101	69	0
4	G	1268	1303	1310	23	0
All	All	33901	33874	34010	516	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 8.

The worst 5 of 516 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:E:522:ASP:OD2	2:E:559:TYR:OH	1.90	0.88
1:D:217:GLN:OE1	2:E:58:ASN:ND2	2.10	0.84
2:B:519:GLU:O	2:B:555:TYR:OH	1.95	0.83
2:E:26:GLY:N	2:E:508:GLU:OE1	2.12	0.83
1:A:420:SER:O	1:A:452:HIS:ND1	2.13	0.82

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	689/716 (96%)	635 (92%)	54 (8%)	0	100	100
1	D	703/716 (98%)	653 (93%)	48 (7%)	2 (0%)	41	74
2	B	675/757 (89%)	629 (93%)	44 (6%)	2 (0%)	41	74
2	E	599/757 (79%)	572 (96%)	27 (4%)	0	100	100
3	C	739/898 (82%)	673 (91%)	66 (9%)	0	100	100
3	F	625/898 (70%)	580 (93%)	42 (7%)	3 (0%)	29	67
4	G	157/251 (62%)	141 (90%)	16 (10%)	0	100	100
All	All	4187/4993 (84%)	3883 (93%)	297 (7%)	7 (0%)	50	79

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	641	ASN
1	D	296	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	F	198	ILE
3	F	533	SER
3	F	220	VAL

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	627/641 (98%)	627 (100%)	0	100	100
1	D	633/641 (99%)	632 (100%)	1 (0%)	93	98
2	B	612/670 (91%)	612 (100%)	0	100	100
2	E	534/670 (80%)	533 (100%)	1 (0%)	93	98
3	C	655/782 (84%)	653 (100%)	2 (0%)	92	96
3	F	552/782 (71%)	552 (100%)	0	100	100
4	G	147/229 (64%)	147 (100%)	0	100	100
All	All	3760/4415 (85%)	3756 (100%)	4 (0%)	93	98

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

Mol	Chain	Res	Type
3	C	355	ARG
3	C	376	LYS
1	D	531	ARG
2	E	409	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	144	HIS
1	A	499	ASN
3	C	601	GLN
1	D	146	HIS
2	E	47	HIS

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

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

5.5 Carbohydrates

There are no monosaccharides in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

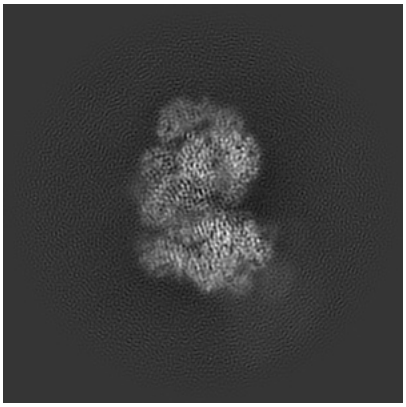
6 Map visualisation

This section contains visualisations of the EMDB entry EMD-18818. These allow visual inspection of the internal detail of the map and identification of artifacts.

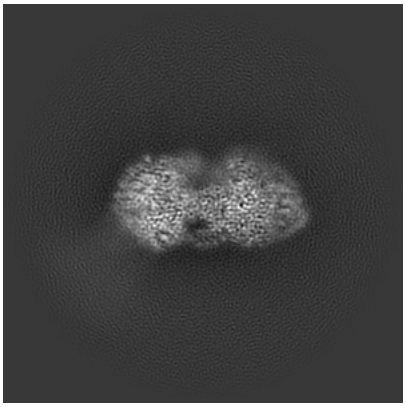
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections

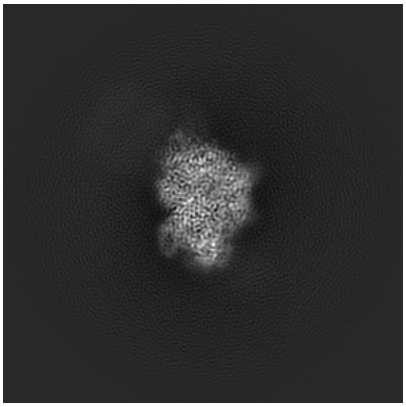
6.1.1 Primary map



X



Y

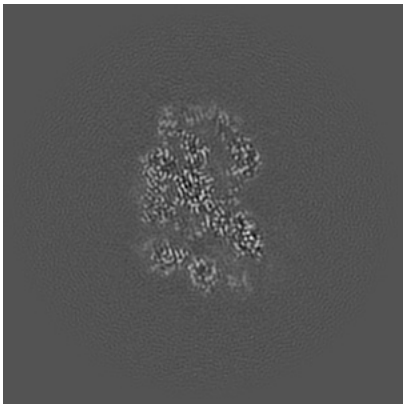


Z

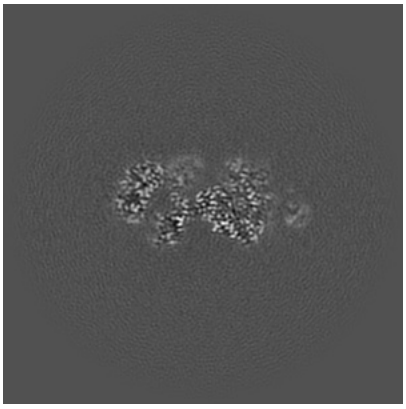
The images above show the map projected in three orthogonal directions.

6.2 Central slices

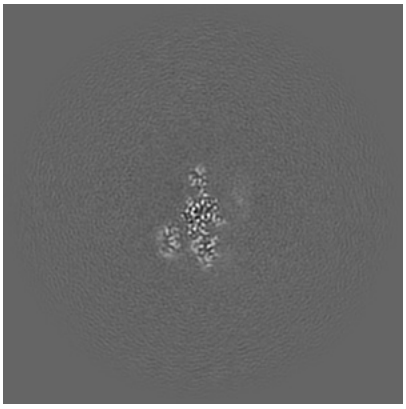
6.2.1 Primary map



X Index: 150



Y Index: 150

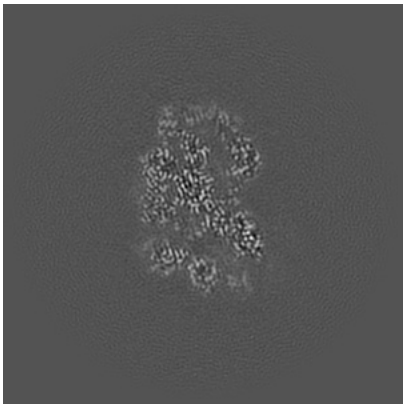


Z Index: 150

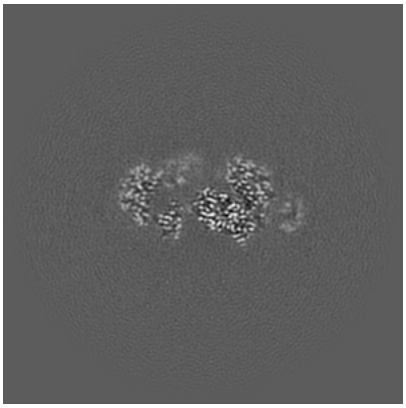
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

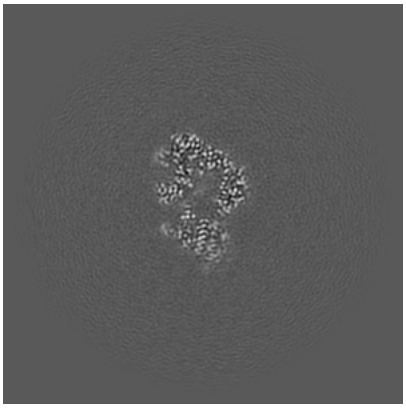
6.3.1 Primary map



X Index: 150



Y Index: 146

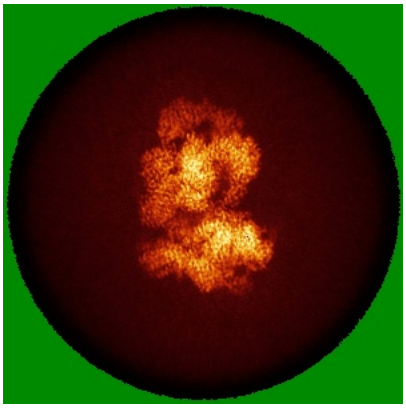


Z Index: 114

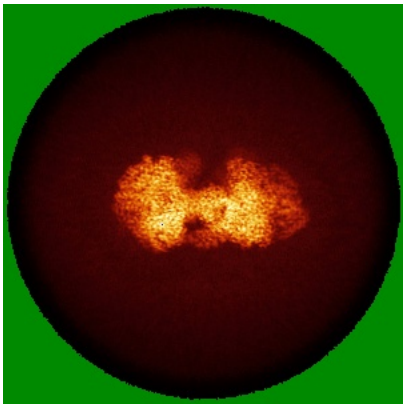
The images above show the largest variance slices of the map in three orthogonal directions.

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

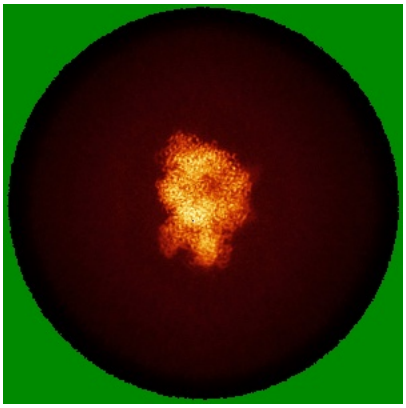
6.4.1 Primary map



X



Y

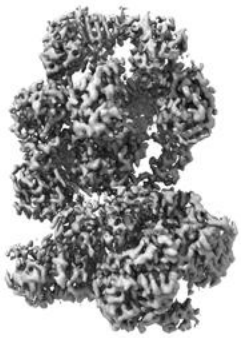


Z

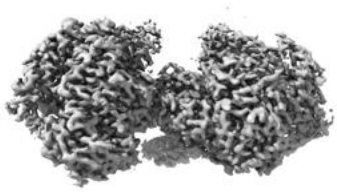
The images above show the map standard deviation projections with false color in three orthogonal directions. Minimum values are shown in green, max in blue, and dark to light orange shades represent small to large values respectively.

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

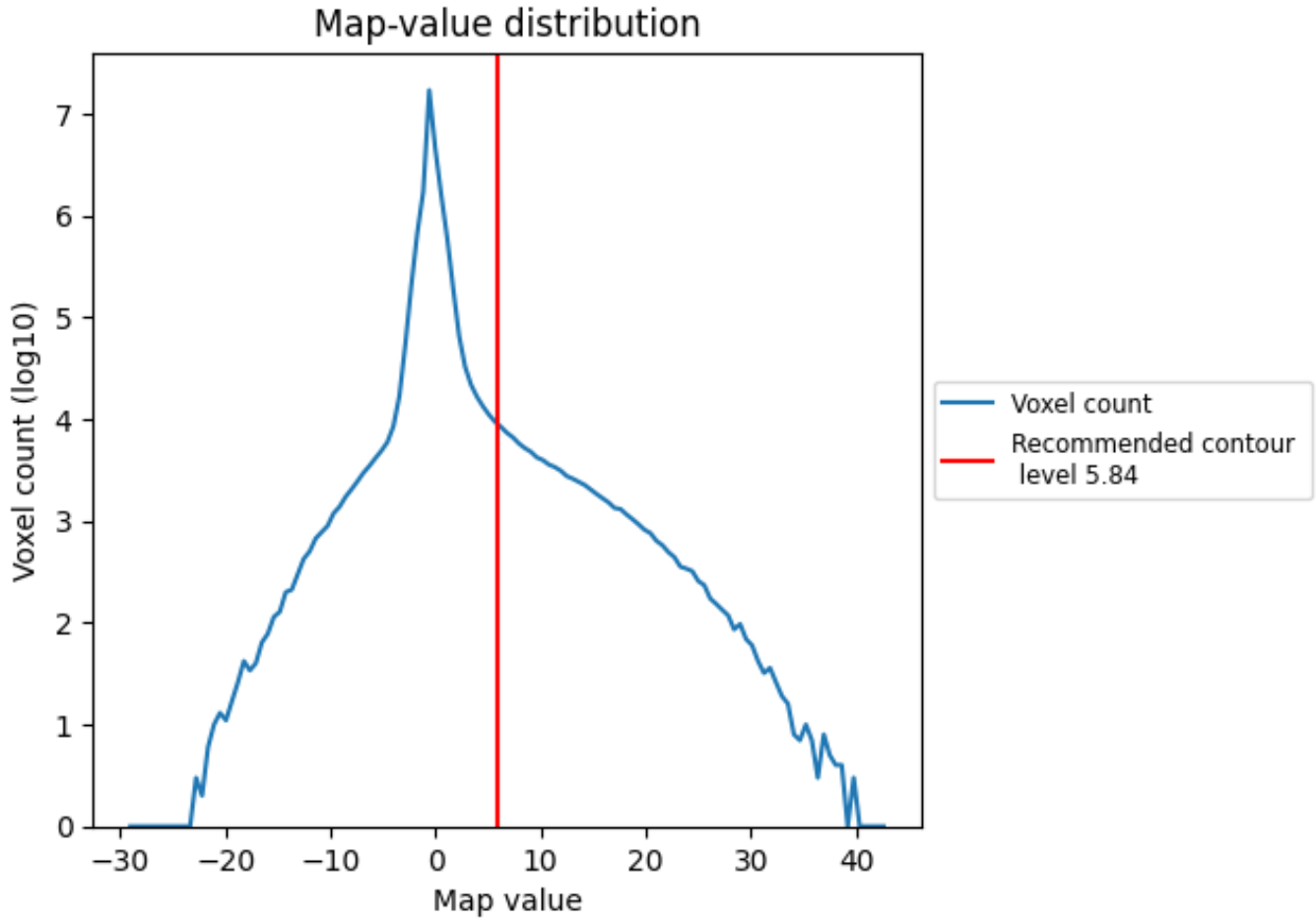
6.6 Mask visualisation (i)

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

7 Map analysis

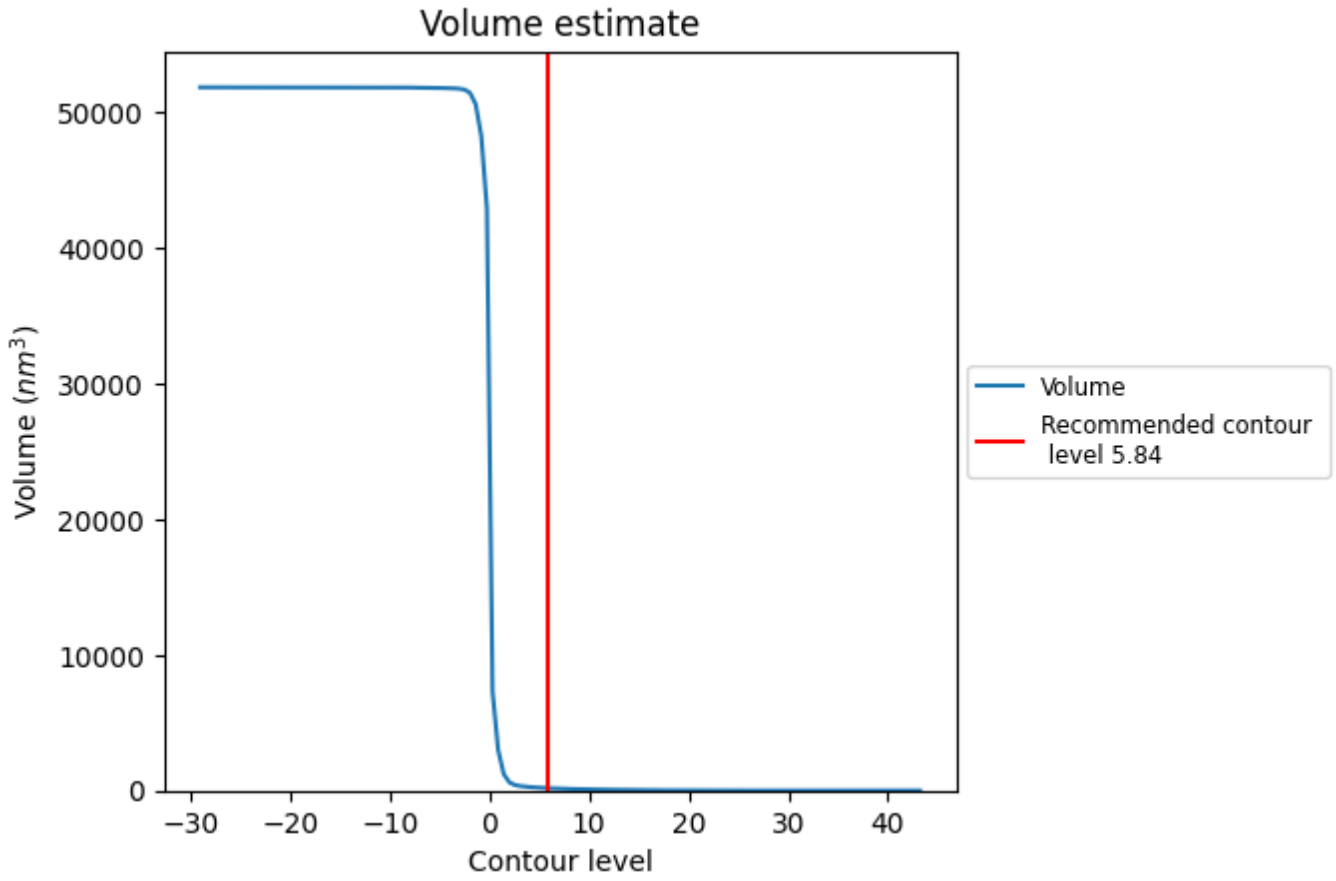
This section contains the results of statistical analysis of the map.

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.

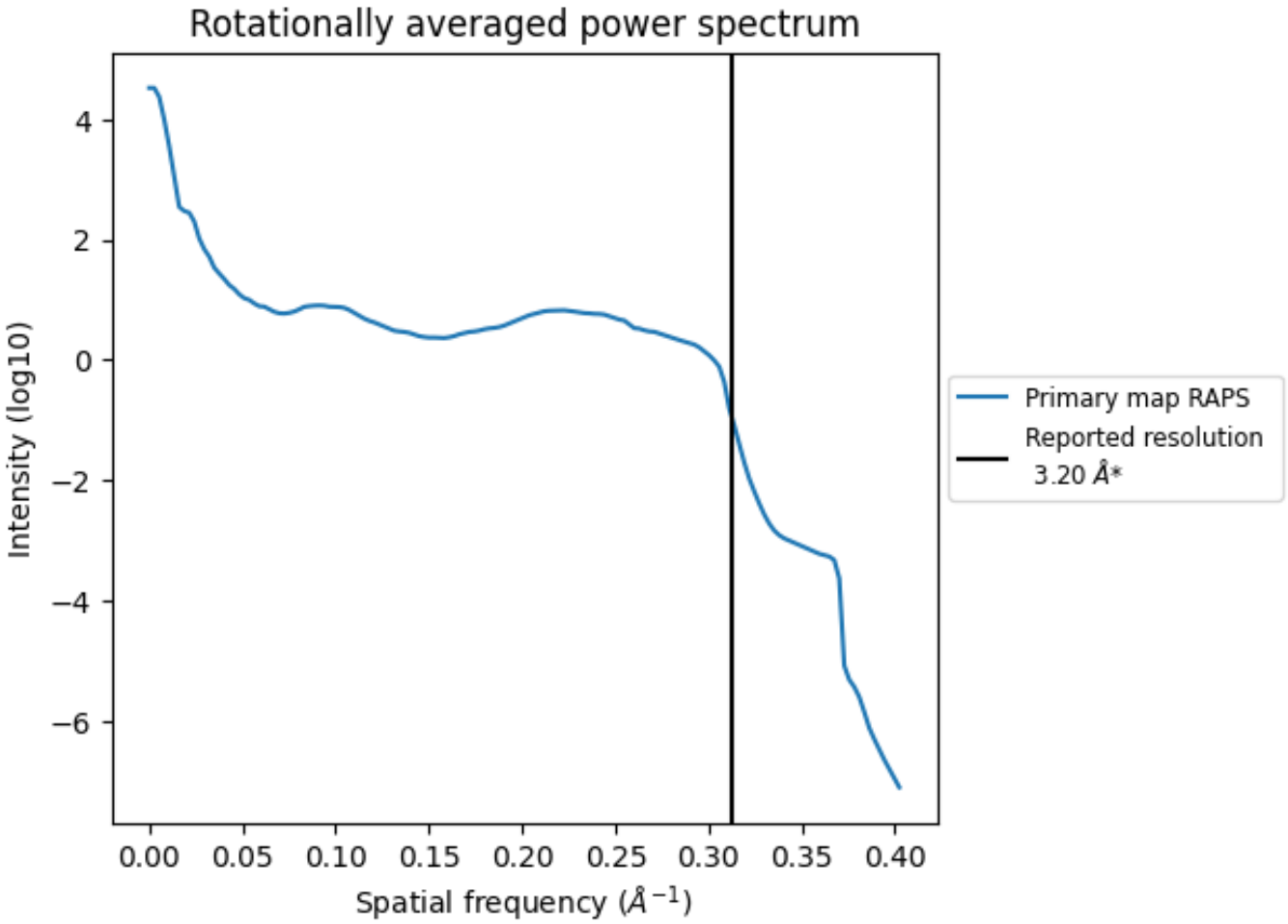
7.2 Volume estimate (i)



The volume at the recommended contour level is 177 nm^3; this corresponds to an approximate mass of 159 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 0.312 Å^{-1}

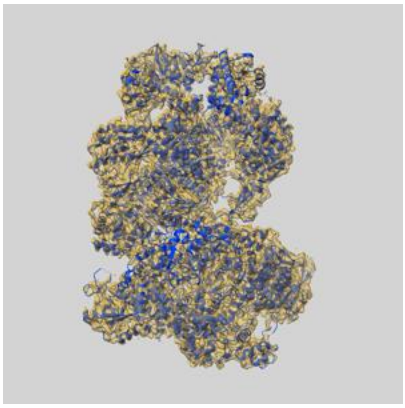
8 Fourier-Shell correlation (i)

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

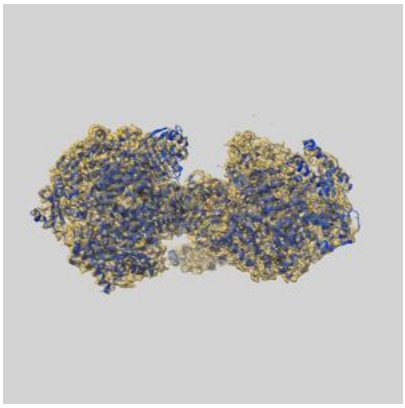
9 Map-model fit

This section contains information regarding the fit between EMDB map EMD-18818 and PDB model 8R1J. Per-residue inclusion information can be found in section 3 on page 5.

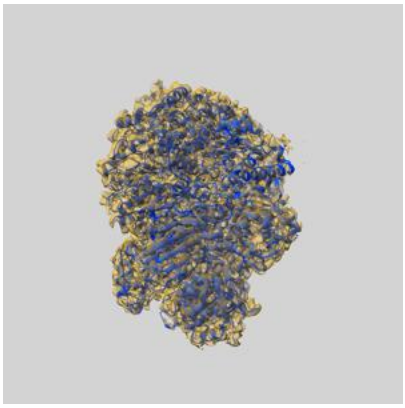
9.1 Map-model overlay



X



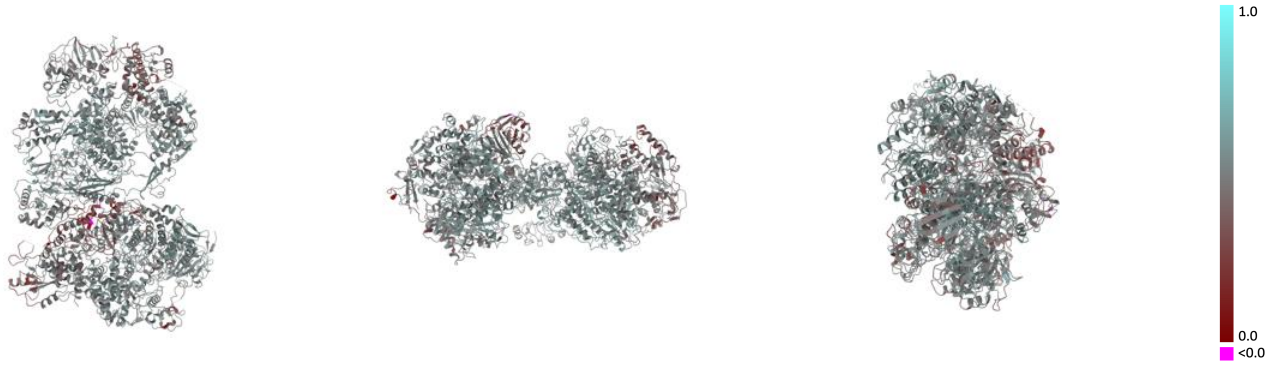
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 5.84 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.

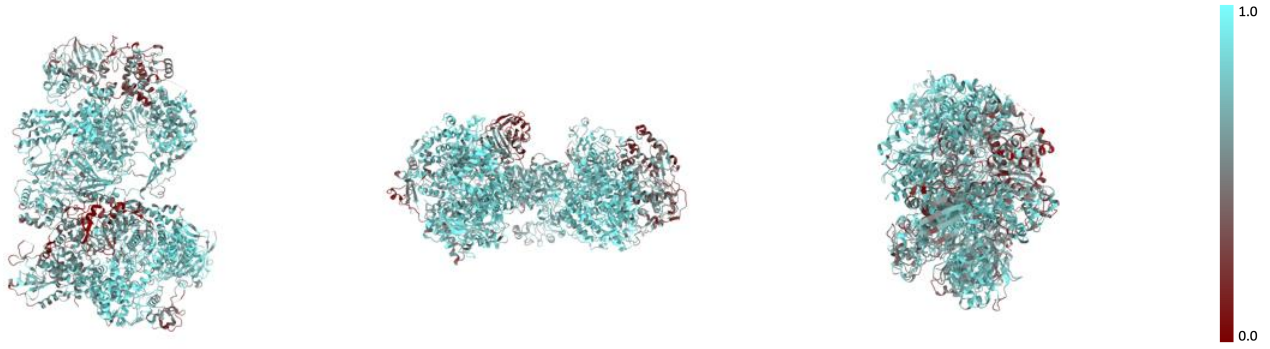
9.2 Q-score mapped to coordinate model (i)



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.

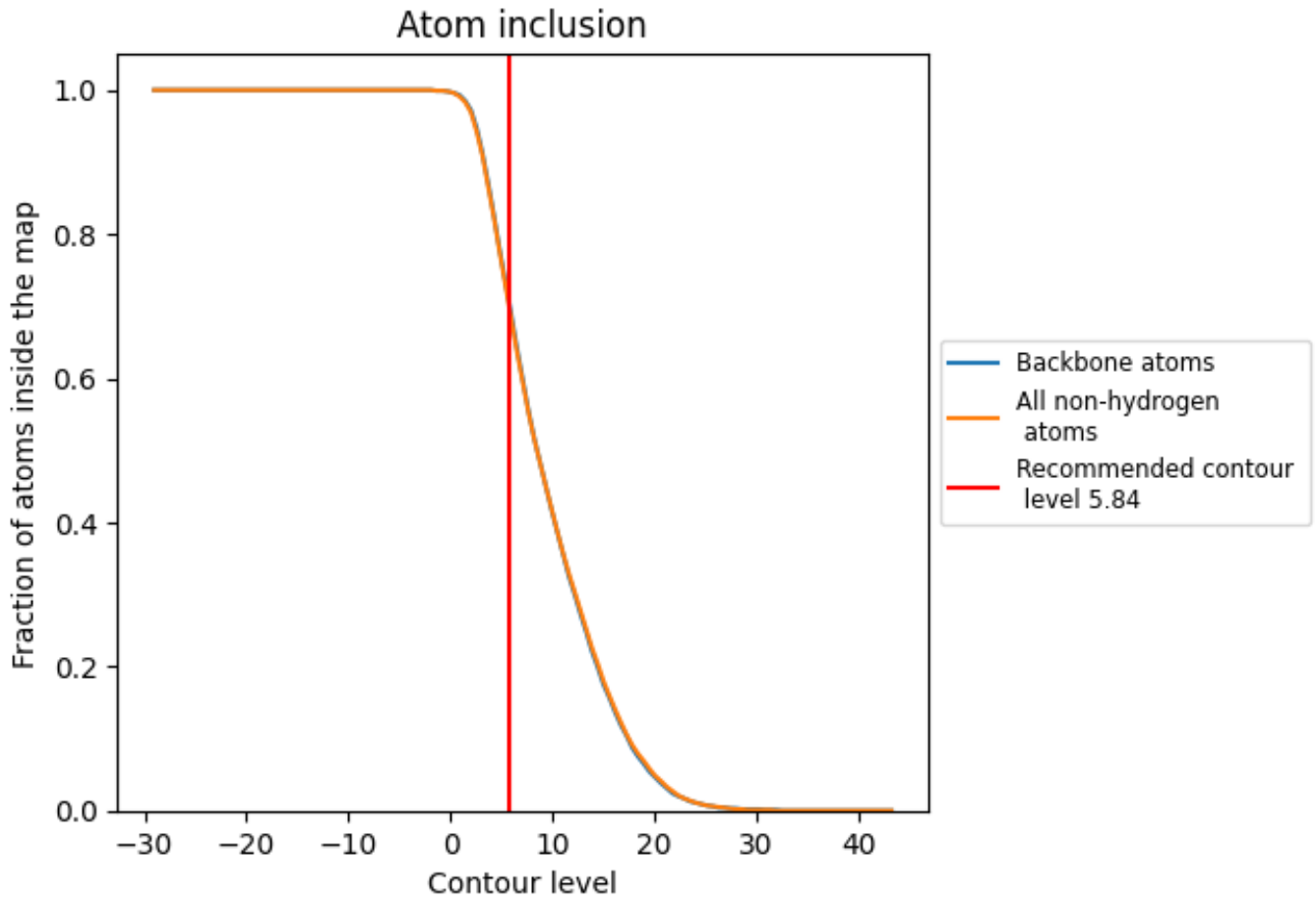
9.3 Atom inclusion mapped to coordinate model (i)



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 (5.84).

9.4 Atom inclusion (i)



At the recommended contour level, 70% of all backbone atoms, 69% 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 (5.84) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.6920	0.5010
A	0.7300	0.4950
B	0.7740	0.5180
C	0.5780	0.4650
D	0.6950	0.5130
E	0.8040	0.5350
F	0.6380	0.4890
G	0.6870	0.4980