wwPDB EM Validation Summary Report (i)

Nov 19, 2022 – 07:35 AM EST

PDB ID	:	1Z8Y
EMDB ID	:	EMD-1121
Title	:	Mapping the E2 Glycoprotein of Alphaviruses
Authors	:	Mukhopadhyay, S.; Zhang, W.; Gabler, S.; Chipman, P.R.; Strauss, E.G.; Strauss, J.H.; Baker, T.S.; Kuhn, R.J.; Rossmann, M.G.
Deposited on	:	2005-03-31
Resolution	:	9.00 Å(reported)
Based on initial models	:	1I9W, 1WYK

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:

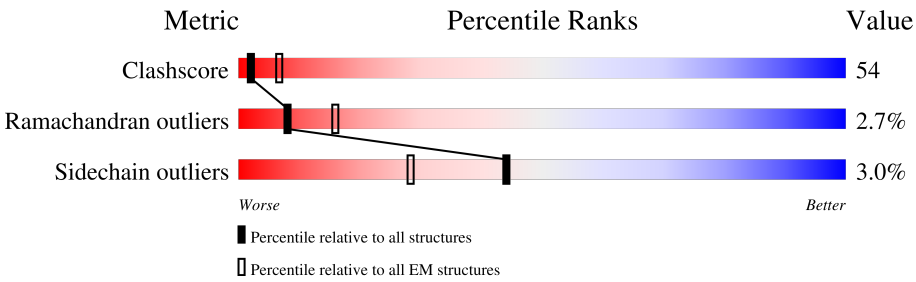
EMDB validation analysis	:	0.0.1.dev43
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.9
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.3

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 9.00 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	290	16% 34% 62% •
1	C	290	18% 34% 61% •
1	E	290	14% 34% 61% •
1	G	290	16% 35% 60% •
2	B	89	8% 24% 65% 10% •
2	D	89	8% 25% 64% 10% •
2	F	89	9% 24% 65% 10% •
2	H	89	12% 22% 66% 10% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	I	31	42% 87% 13%
3	K	31	52% 87% 13%
3	M	31	58% 87% 13%
3	O	31	65% 87% 13%
4	J	36	31% 69% 31%
4	L	36	50% 75% 25%
4	N	36	53% 75% 25%
4	P	36	53% 75% 25%
5	Q	151	16% 80% 17%
5	R	151	17% 83% 15%
5	S	151	17% 83% 15%
5	T	151	21% 84% 14%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 18071 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 Spike glycoprotein E1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	290	2204	1405	364	419	16	0	0
1	C	290	2205	1405	365	419	16	0	0
1	E	290	2205	1405	365	419	16	0	0
1	G	290	2205	1405	365	419	16	0	0

- Molecule 2 is a protein called Spike glycoprotein E1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	B	89	657	404	111	135	7	0	0
2	D	89	657	404	111	135	7	0	0
2	F	89	657	404	111	135	7	0	0
2	H	89	657	404	111	135	7	0	0

- Molecule 3 is a protein called Spike glycoprotein E1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	I	31	237	157	38	38	4	0	0
3	K	31	237	157	38	38	4	0	0
3	M	31	237	157	38	38	4	0	0
3	O	31	237	157	38	38	4	0	0

- Molecule 4 is a protein called Spike glycoprotein E2.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	J	36	257	164	44	44	5	0	0
4	L	36	257	164	44	44	5	0	0
4	N	36	257	164	44	44	5	0	0
4	P	36	257	164	44	44	5	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
J	29	LEU	LYS	SEE REMARK 999	UNP P11259
L	29	LEU	LYS	SEE REMARK 999	UNP P11259
N	29	LEU	LYS	SEE REMARK 999	UNP P11259
P	29	LEU	LYS	SEE REMARK 999	UNP P11259

- Molecule 5 is a protein called Capsid protein C.

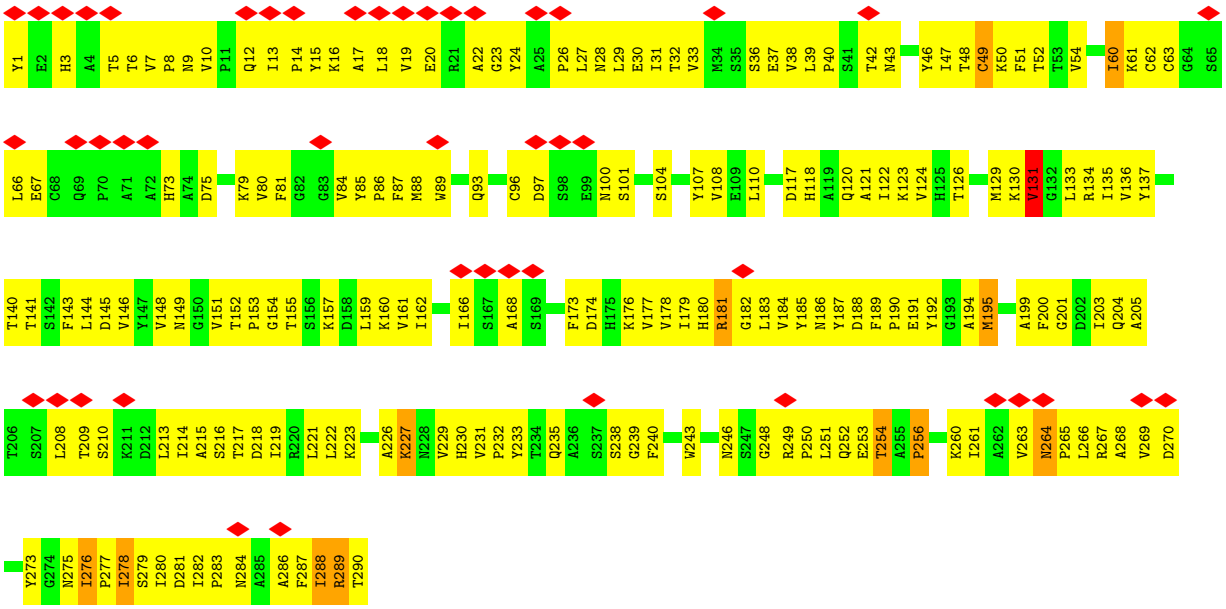
Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
5	Q	151	1162	731	207	219	5	0	0
5	R	151	1162	731	207	219	5	0	0
5	S	151	1162	731	207	219	5	0	0
5	T	151	1162	731	207	219	5	0	0

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.

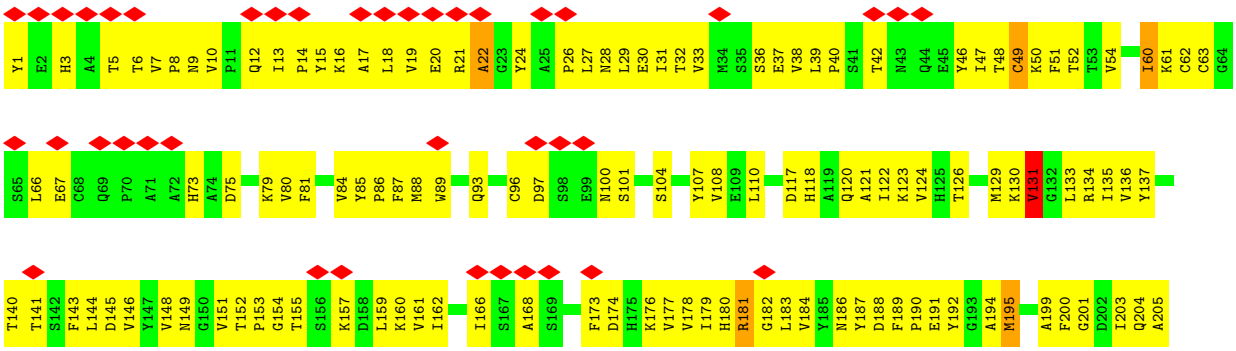
- Molecule 1: Spike glycoprotein E1

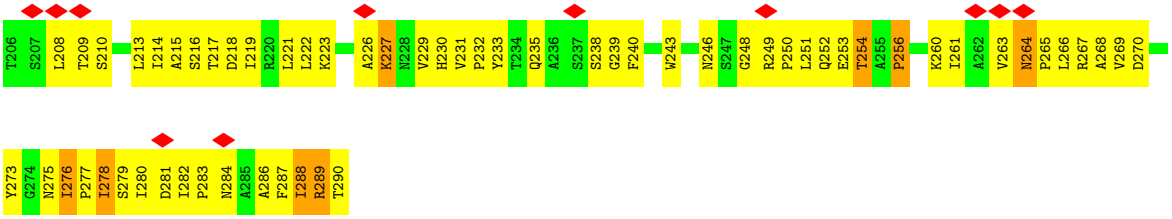
Chain A:



- Molecule 1: Spike glycoprotein E1

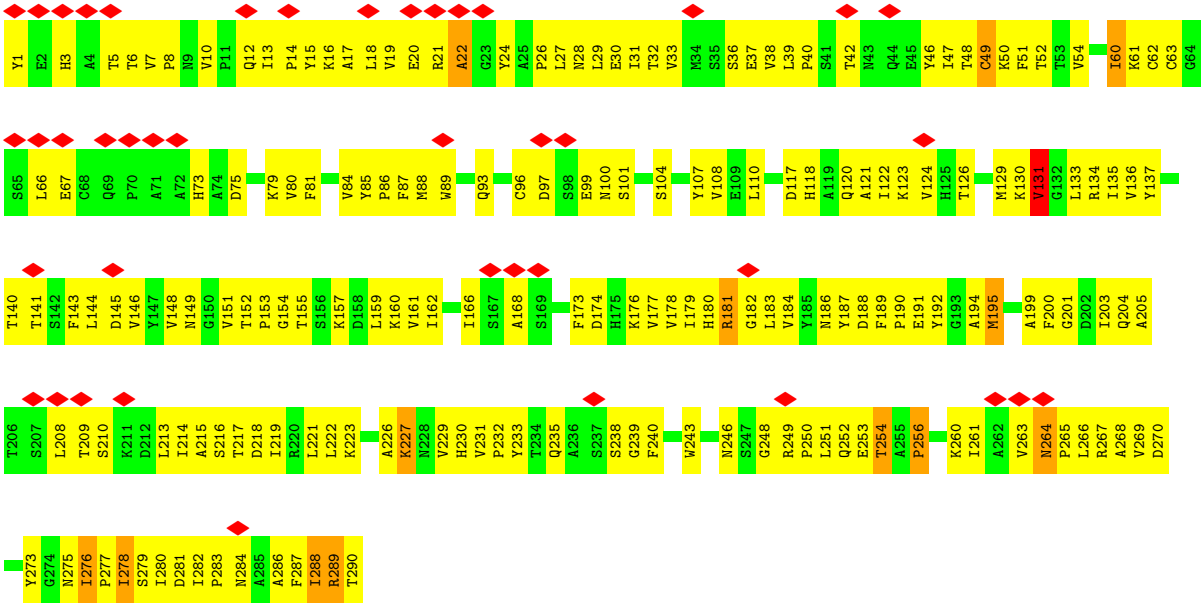
Chain C:





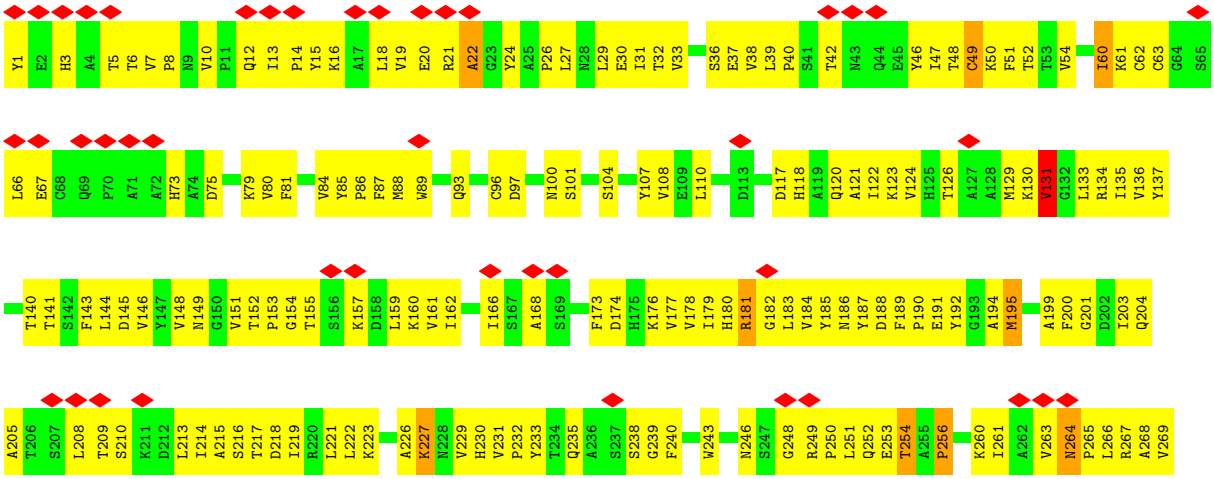
• Molecule 1: Spike glycoprotein E1

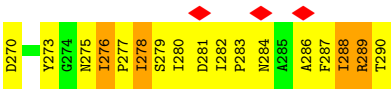
Chain E:



• Molecule 1: Spike glycoprotein E1

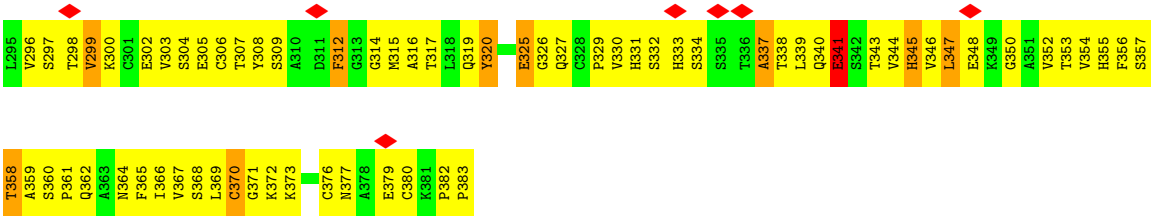
Chain G:





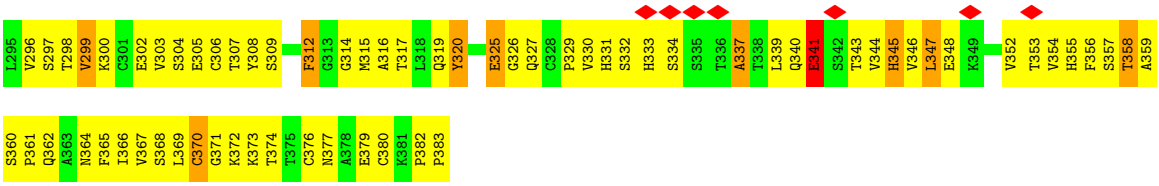
• Molecule 2: Spike glycoprotein E1

Chain B:



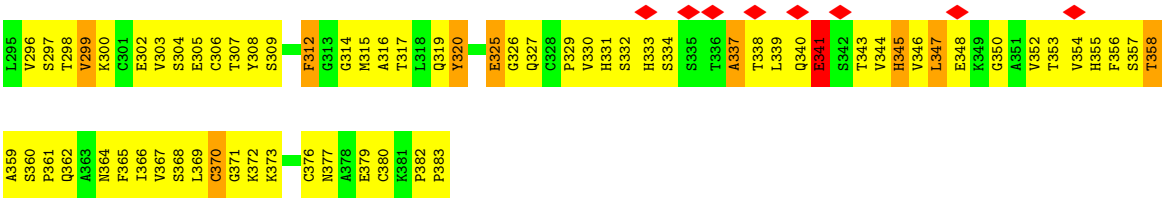
• Molecule 2: Spike glycoprotein E1

Chain D:



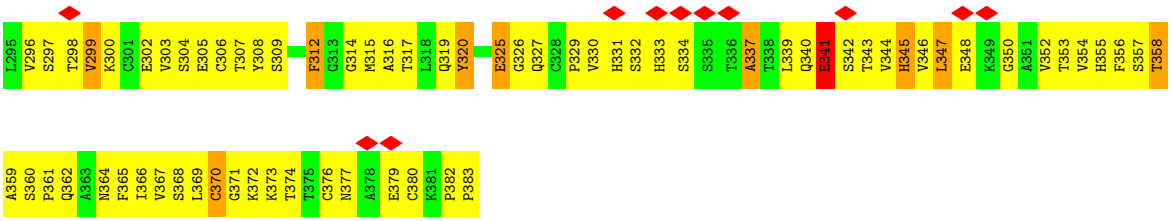
• Molecule 2: Spike glycoprotein E1

Chain F:



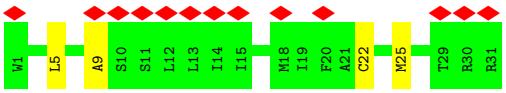
• Molecule 2: Spike glycoprotein E1

Chain H:



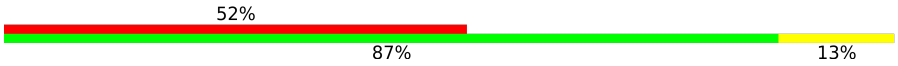
• Molecule 3: Spike glycoprotein E1

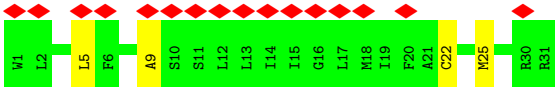
Chain I:



• Molecule 3: Spike glycoprotein E1

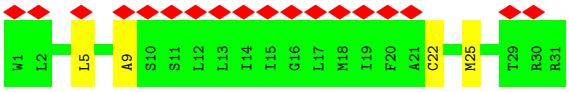
Chain K:





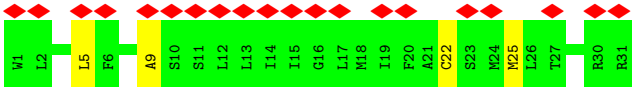
• Molecule 3: Spike glycoprotein E1

Chain M:



• Molecule 3: Spike glycoprotein E1

Chain O:

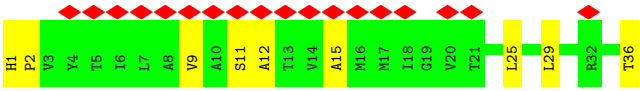


• Molecule 4: Spike glycoprotein E2

Chain J:

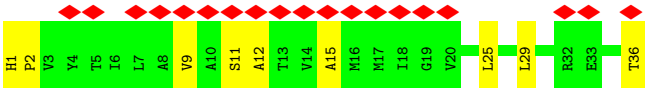
• Molecule 4: Spike glycoprotein E2

Chain L:



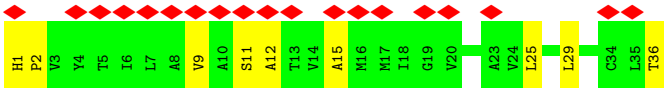
• Molecule 4: Spike glycoprotein E2

Chain N:



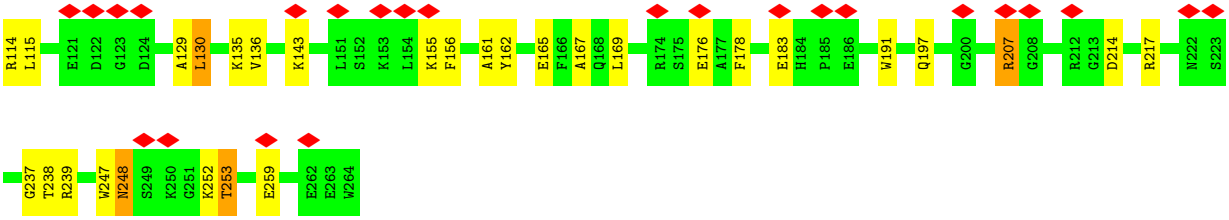
• Molecule 4: Spike glycoprotein E2

Chain P: 53% 75% 25%



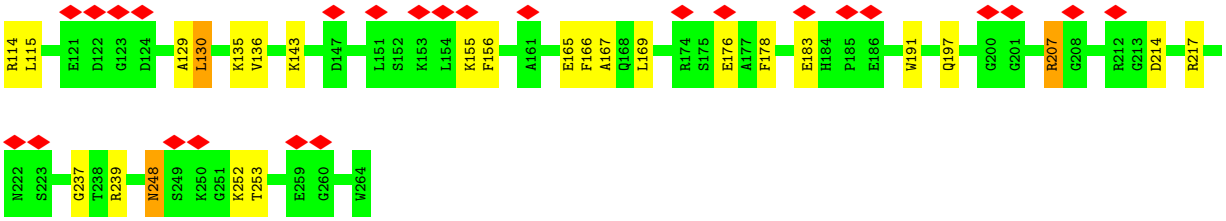
• Molecule 5: Capsid protein C

Chain Q: 16% 80% 17%



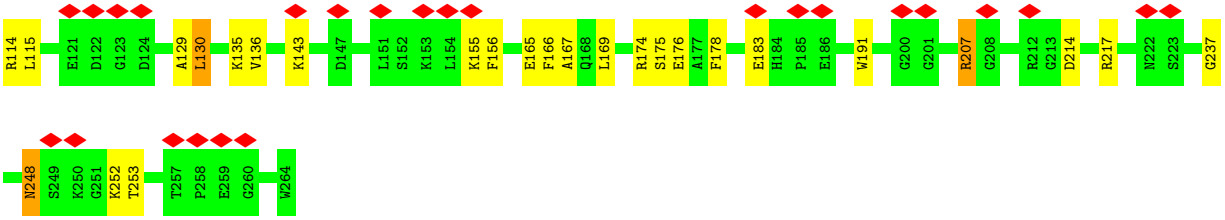
• Molecule 5: Capsid protein C

Chain R: 17% 83% 15%



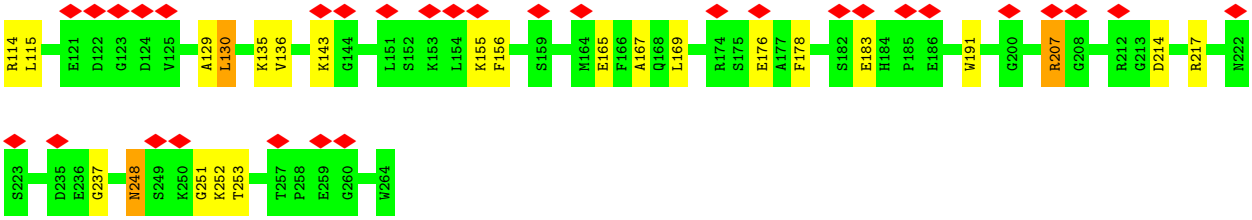
• Molecule 5: Capsid protein C

Chain S: 17% 83% 15%



• Molecule 5: Capsid protein C

Chain T: 21% 84% 14%



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, I	Depositor
Number of particles used	7085	Depositor
Resolution determination method	Not provided	
CTF correction method	Fourier transform of each image was modified	Depositor
Microscope	FEI/PHILIPS CM200FEG	Depositor
Voltage (kV)	200	Depositor
Electron dose ($e^-/Å^2$)	18	Depositor
Minimum defocus (nm)	1100	Depositor
Maximum defocus (nm)	2580	Depositor
Magnification	38000	Depositor
Image detector	KODAK SO-163 FILM	Depositor
Maximum map value	402.470	Depositor
Minimum map value	-297.650	Depositor
Average map value	7.032	Depositor
Map value standard deviation	65.091	Depositor
Recommended contour level	138.0	Depositor
Map size (Å)	787.18, 787.18, 787.18	wwPDB
Map dimensions	441, 441, 441	wwPDB
Map angles (°)	90, 90, 90	wwPDB
Pixel spacing (Å)	1.78499, 1.78499, 1.78499	Depositor

5 Model quality (i)

5.1 Standard geometry (i)

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	0.91	0/2261	1.12	3/3083 (0.1%)
1	C	0.91	0/2262	1.12	3/3085 (0.1%)
1	E	0.91	0/2262	1.12	3/3085 (0.1%)
1	G	0.91	0/2262	1.12	3/3085 (0.1%)
2	B	0.83	0/669	1.16	2/908 (0.2%)
2	D	0.83	0/669	1.16	2/908 (0.2%)
2	F	0.83	0/669	1.16	2/908 (0.2%)
2	H	0.83	0/669	1.16	2/908 (0.2%)
3	I	0.37	0/241	0.46	0/321
3	K	0.37	0/241	0.46	0/321
3	M	0.38	0/241	0.46	0/321
3	O	0.37	0/241	0.46	0/321
4	J	0.28	0/259	0.43	0/354
4	L	0.28	0/259	0.43	0/354
4	N	0.28	0/259	0.43	0/354
4	P	0.28	0/259	0.43	0/354
5	Q	0.52	0/1190	0.81	1/1608 (0.1%)
5	R	0.52	0/1190	0.81	1/1608 (0.1%)
5	S	0.52	0/1190	0.81	1/1608 (0.1%)
5	T	0.52	0/1190	0.81	1/1608 (0.1%)
All	All	0.77	0/18483	1.00	24/25102 (0.1%)

There are no bond length outliers.

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	G	254	THR	C-N-CA	5.47	135.38	121.70
1	A	254	THR	C-N-CA	5.47	135.38	121.70
1	E	254	THR	C-N-CA	5.47	135.37	121.70
1	C	254	THR	C-N-CA	5.46	135.35	121.70
1	C	131	VAL	C-N-CA	5.45	133.74	122.30

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	2204	0	2157	368	0
1	C	2205	0	2159	369	0
1	E	2205	0	2161	359	0
1	G	2205	0	2161	328	0
2	B	657	0	630	152	0
2	D	657	0	630	149	0
2	F	657	0	630	144	0
2	H	657	0	630	116	0
3	I	237	0	254	2	0
3	K	237	0	254	2	0
3	M	237	0	254	2	0
3	O	237	0	254	2	0
4	J	257	0	281	18	0
4	L	257	0	281	8	0
4	N	257	0	281	6	0
4	P	257	0	281	7	0
5	Q	1162	0	1131	28	0
5	R	1162	0	1131	14	0
5	S	1162	0	1131	13	0
5	T	1162	0	1131	13	0
All	All	18071	0	17822	1929	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 54.

The worst 5 of 1929 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)
1:A:18:LEU:HD23	2:B:331:HIS:CD2	1.49	1.46
1:C:18:LEU:HB2	2:D:331:HIS:NE2	1.35	1.41
1:C:28:ASN:CG	2:D:331:HIS:ND1	1.76	1.37
1:A:16:LYS:HZ2	2:B:338:THR:CG2	1.38	1.34
1:A:18:LEU:HB2	2:B:331:HIS:NE2	1.42	1.33

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	288/290 (99%)	250 (87%)	30 (10%)	8 (3%)	5	30
1	C	288/290 (99%)	250 (87%)	30 (10%)	8 (3%)	5	30
1	E	288/290 (99%)	250 (87%)	30 (10%)	8 (3%)	5	30
1	G	288/290 (99%)	250 (87%)	30 (10%)	8 (3%)	5	30
2	B	87/89 (98%)	63 (72%)	16 (18%)	8 (9%)	1	11
2	D	87/89 (98%)	63 (72%)	16 (18%)	8 (9%)	1	11
2	F	87/89 (98%)	63 (72%)	16 (18%)	8 (9%)	1	11
2	H	87/89 (98%)	63 (72%)	16 (18%)	8 (9%)	1	11
3	I	29/31 (94%)	28 (97%)	1 (3%)	0	100	100
3	K	29/31 (94%)	28 (97%)	1 (3%)	0	100	100
3	M	29/31 (94%)	28 (97%)	1 (3%)	0	100	100
3	O	29/31 (94%)	28 (97%)	1 (3%)	0	100	100
4	J	34/36 (94%)	33 (97%)	1 (3%)	0	100	100
4	L	34/36 (94%)	33 (97%)	1 (3%)	0	100	100
4	N	34/36 (94%)	33 (97%)	1 (3%)	0	100	100
4	P	34/36 (94%)	33 (97%)	1 (3%)	0	100	100
5	Q	149/151 (99%)	144 (97%)	5 (3%)	0	100	100
5	R	149/151 (99%)	144 (97%)	5 (3%)	0	100	100
5	S	149/151 (99%)	144 (97%)	5 (3%)	0	100	100
5	T	149/151 (99%)	144 (97%)	5 (3%)	0	100	100
All	All	2348/2388 (98%)	2072 (88%)	212 (9%)	64 (3%)	8	31

5 of 64 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	131	VAL
1	A	195	MET
1	A	227	LYS
1	A	288	ILE
2	B	299	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	242/243 (100%)	237 (98%)	5 (2%)	53	72
1	C	243/243 (100%)	238 (98%)	5 (2%)	53	72
1	E	243/243 (100%)	238 (98%)	5 (2%)	53	72
1	G	243/243 (100%)	238 (98%)	5 (2%)	53	72
2	B	77/77 (100%)	74 (96%)	3 (4%)	32	56
2	D	77/77 (100%)	74 (96%)	3 (4%)	32	56
2	F	77/77 (100%)	74 (96%)	3 (4%)	32	56
2	H	77/77 (100%)	74 (96%)	3 (4%)	32	56
3	I	25/25 (100%)	25 (100%)	0	100	100
3	K	25/25 (100%)	25 (100%)	0	100	100
3	M	25/25 (100%)	25 (100%)	0	100	100
3	O	25/25 (100%)	25 (100%)	0	100	100
4	J	28/28 (100%)	28 (100%)	0	100	100
4	L	28/28 (100%)	28 (100%)	0	100	100
4	N	28/28 (100%)	28 (100%)	0	100	100
4	P	28/28 (100%)	28 (100%)	0	100	100
5	Q	122/123 (99%)	115 (94%)	7 (6%)	20	45
5	R	122/123 (99%)	115 (94%)	7 (6%)	20	45
5	S	122/123 (99%)	115 (94%)	7 (6%)	20	45
5	T	122/123 (99%)	115 (94%)	7 (6%)	20	45

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1979/1984 (100%)	1919 (97%)	60 (3%)	44 63

5 of 60 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	289	ARG
5	T	183	GLU
5	Q	191	TRP
5	T	155	LYS
5	T	253	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 62 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	100	ASN
5	S	128	HIS
1	G	9	ASN
5	R	248	ASN
5	T	190	ASN

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.

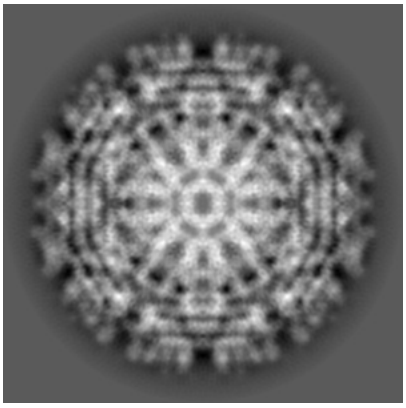
6 Map visualisation (i)

This section contains visualisations of the EMDB entry EMD-1121. These allow visual inspection of the internal detail of the map and identification of artifacts.

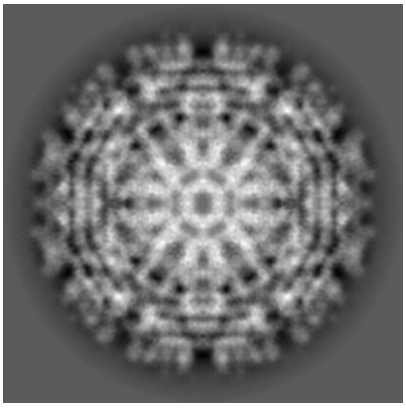
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections (i)

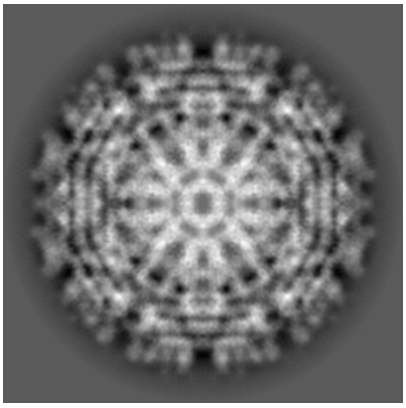
6.1.1 Primary map



X



Y

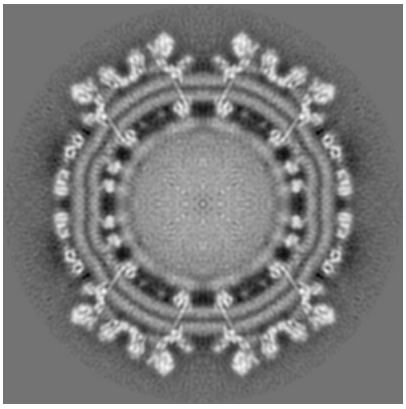


Z

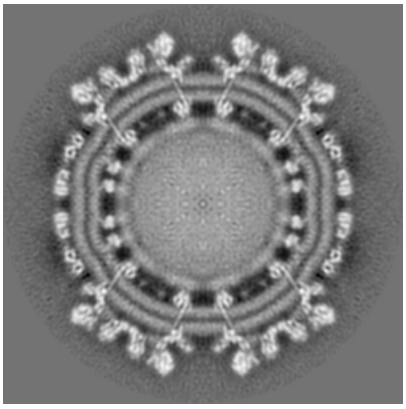
The images above show the map projected in three orthogonal directions.

6.2 Central slices (i)

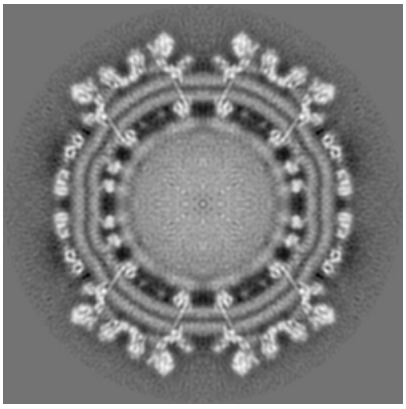
6.2.1 Primary map



X Index: 220



Y Index: 220

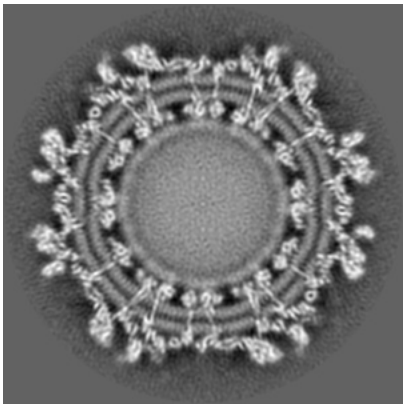


Z Index: 220

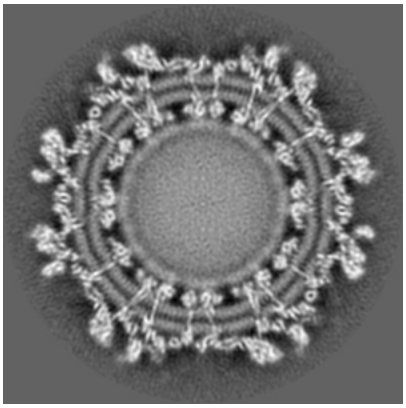
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

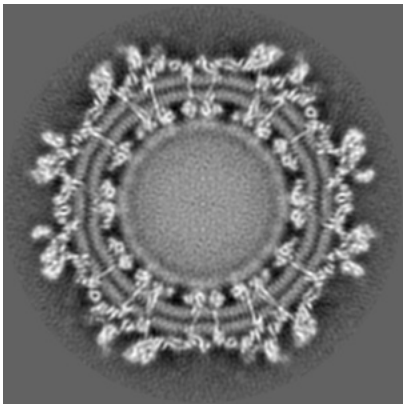
6.3.1 Primary map



X Index: 202



Y Index: 202



Z Index: 238

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

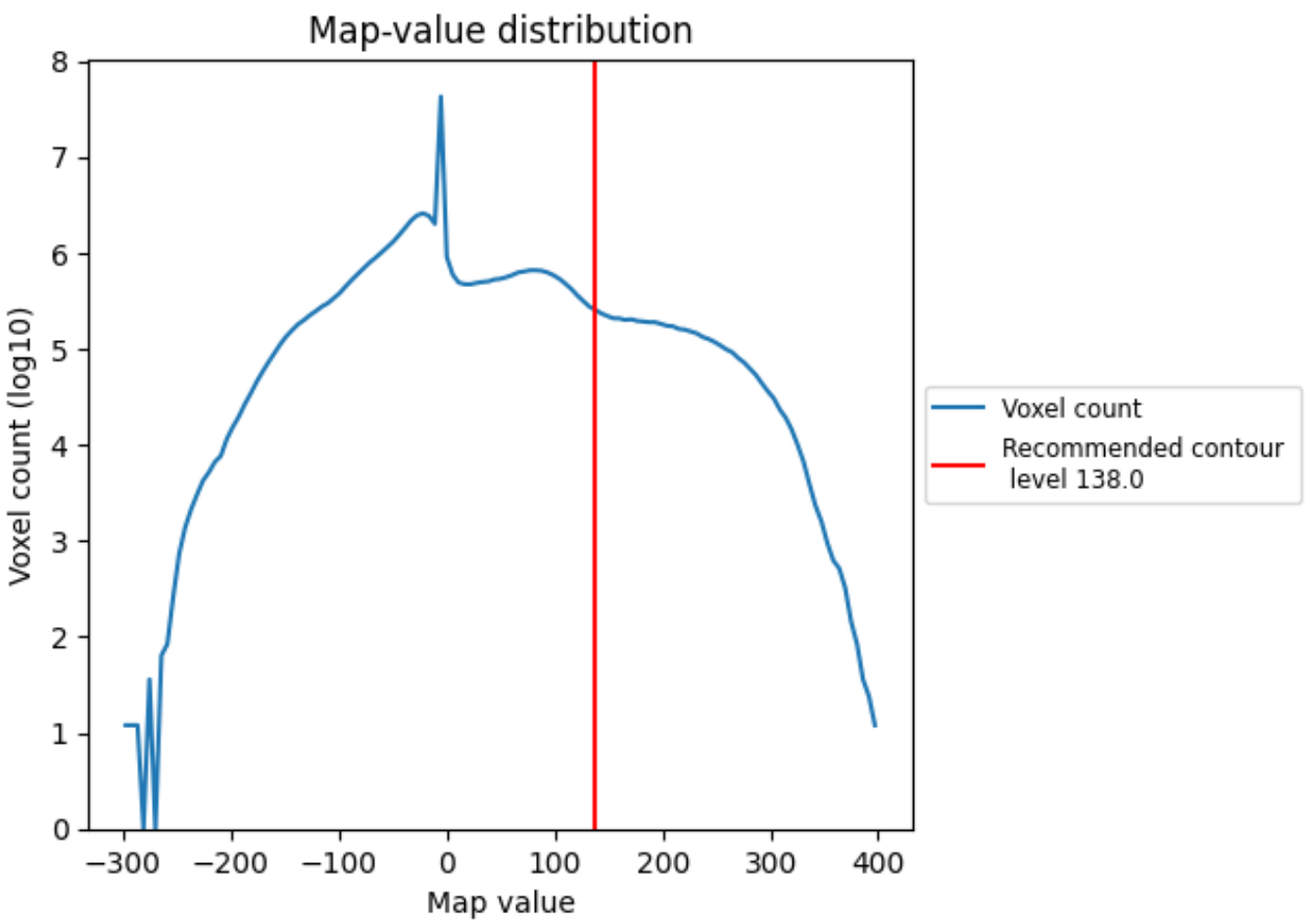
6.5 Mask visualisation (i)

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

7 Map analysis (i)

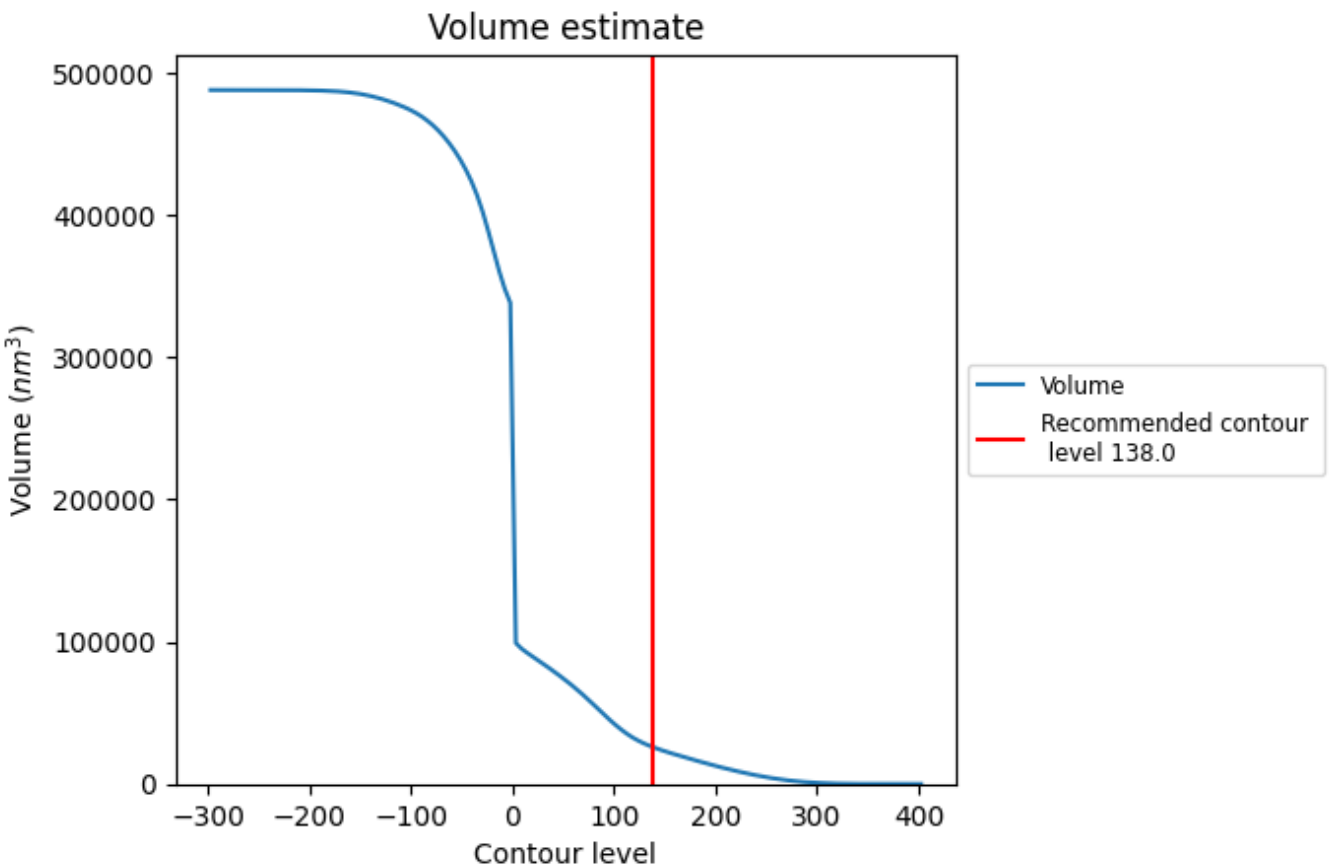
This section contains the results of statistical analysis of the map.

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.

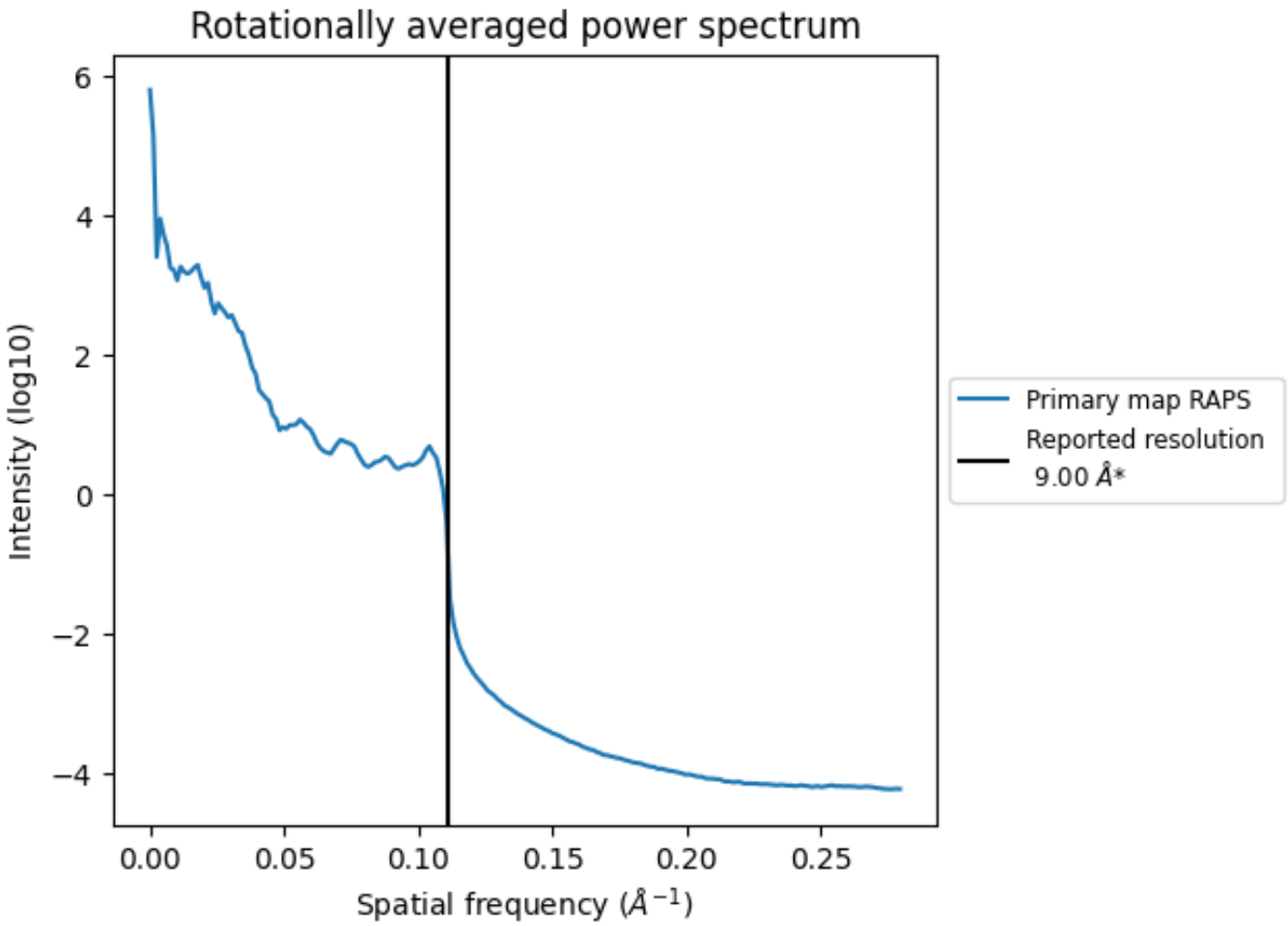
7.2 Volume estimate (i)



The volume at the recommended contour level is 26000 nm^3; this corresponds to an approximate mass of 23486 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 0.111 Å^{-1}

8 Fourier-Shell correlation (i)

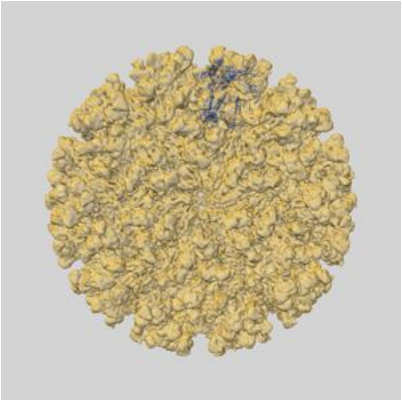
This section was not generated. No FSC curve or half-maps provided.

9 Map-model fit (i)

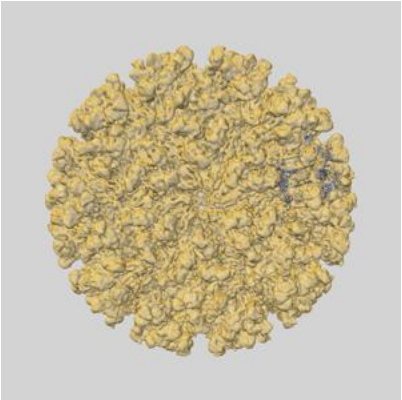
This section contains information regarding the fit between EMDB map EMD-1121 and PDB model 1Z8Y. Per-residue inclusion information can be found in section 3 on page 6.

9.1 Map-model overlays

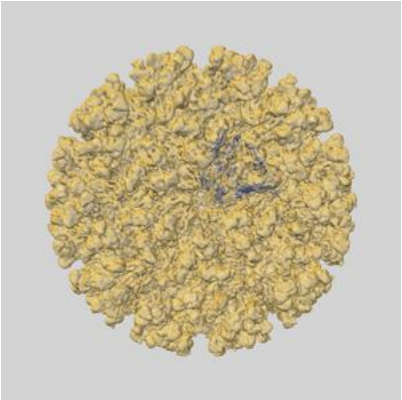
9.1.1 Map-model overlay (i)



X

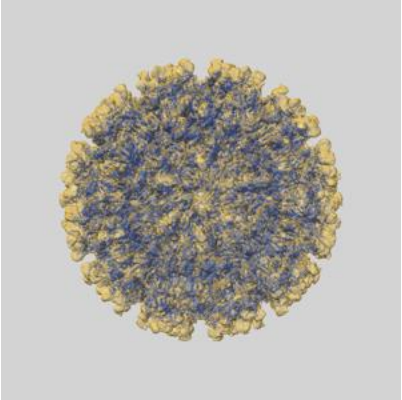


Y

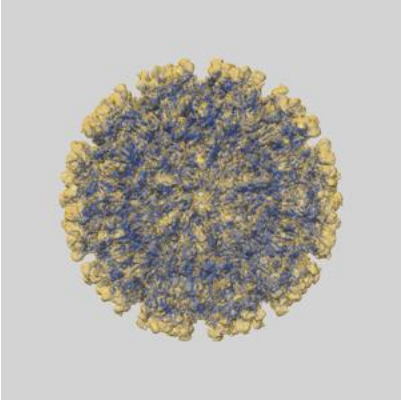


Z

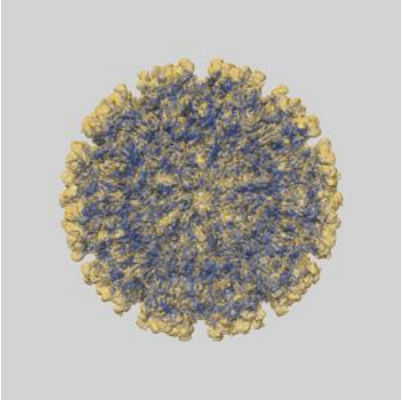
9.1.2 Map-model assembly overlay (i)



X



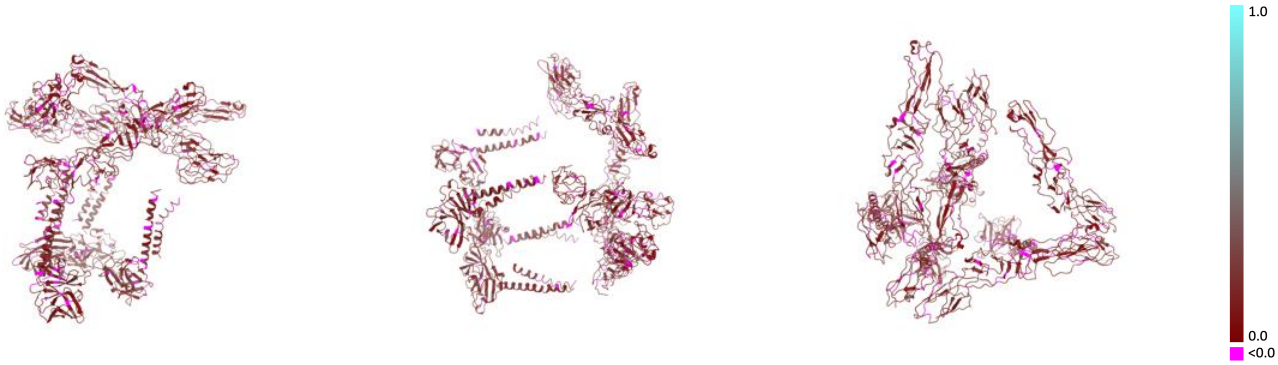
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 138.0 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.

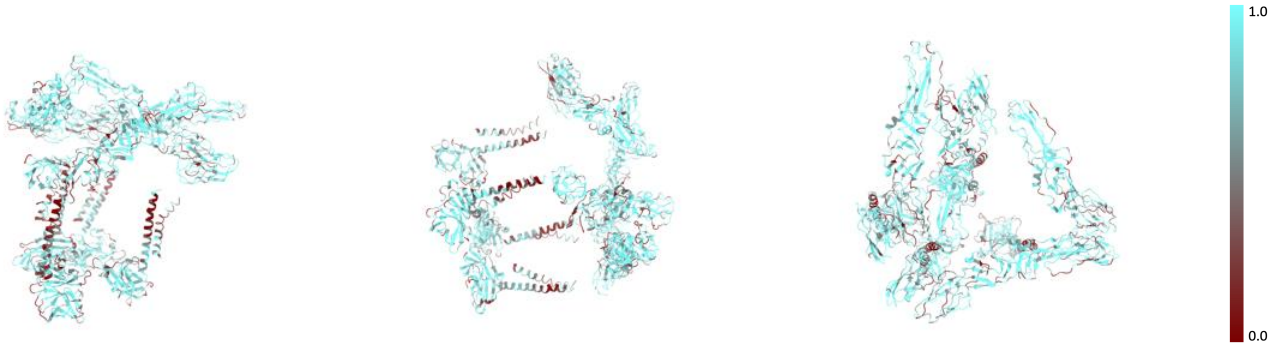
9.2 Q-score mapped to coordinate model (i)



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.

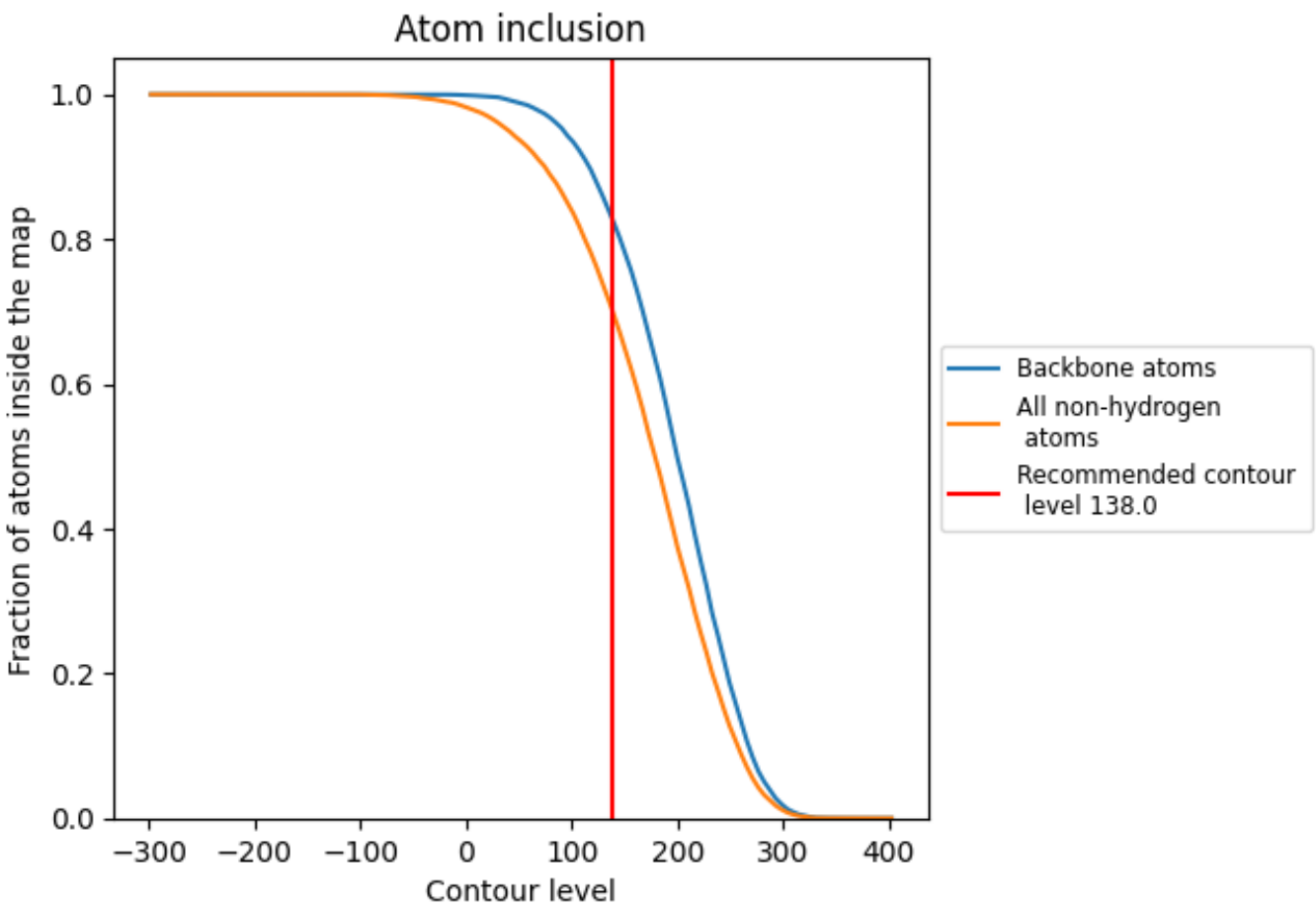
9.3 Atom inclusion mapped to coordinate model (i)



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 (138.0).

9.4 Atom inclusion (i)



At the recommended contour level, 83% of all backbone atoms, 70% of all non-hydrogen atoms, are inside the map.

9.5 Map-model fit summary (i)

The table lists the average atom inclusion at the recommended contour level (138.0) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.7039	0.1320
A	0.7249	0.1270
B	0.7994	0.1380
C	0.7136	0.1310
D	0.8239	0.1500
E	0.7439	0.1370
F	0.8101	0.1320
G	0.7329	0.1320
H	0.7443	0.1360
I	0.4784	0.1270
J	0.5278	0.1500
K	0.4181	0.1110
L	0.4524	0.1160
M	0.3534	0.1070
N	0.3929	0.1050
O	0.3060	0.0640
P	0.4127	0.1330
Q	0.7612	0.1390
R	0.7296	0.1400
S	0.7419	0.1430
T	0.6699	0.1280