



wwPDB EM Validation Summary Report ⓘ

Nov 2, 2022 – 08:33 PM EDT

PDB ID : 5U07
EMDB ID : EMD-8477
Title : CRISPR RNA-guided surveillance complex
Authors : Xiao, Y.; Luo, M.; Hayes, R.P.; Kim, J.; Ng, S.; Ding, F.; Liao, M.; Ke, A.
Deposited on : 2016-11-23
Resolution : 3.80 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
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) : 2.31.2

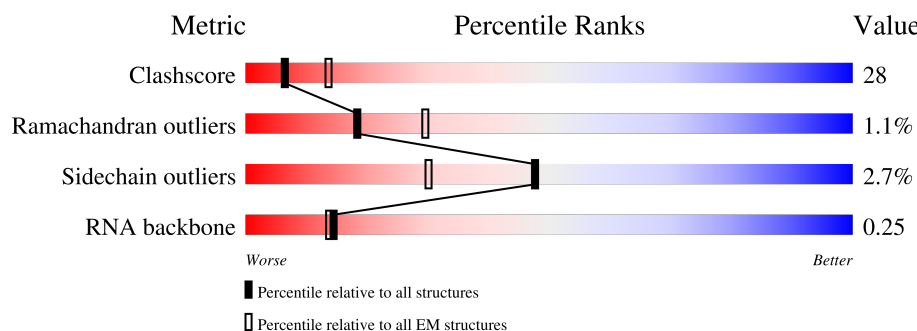
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 3.80 Å.

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



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

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the EM map (all-atom inclusion $< 40\%$). The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	232	<div> <div>82%</div> <div>34% 41% 6% 18%</div> </div>
2	B	244	<div> <div>66%</div> <div>27% 35% 34%</div> </div>
2	J	244	<div> <div>69%</div> <div>30% 33% 7% 30%</div> </div>
3	C	549	<div> <div>88%</div> <div>43% 43% 10%</div> </div>
4	D	373	<div> <div>36%</div> <div>40% 29% 30%</div> </div>
4	E	373	<div> <div>59%</div> <div>55% 42% ..</div> </div>
4	F	373	<div> <div>46%</div> <div>56% 39% ..</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	G	373	
4	H	373	
4	I	373	
5	K	61	
6	M	21	
7	N	254	
8	O	13	

2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 27681 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 CRISPR-associated protein, Cse3 family.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	190	Total	C	N	O	S	0	0
			1471	909	303	258	1		

- Molecule 2 is a protein called Cse2.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	161	Total	C	N	O	S	0	0
			1285	811	256	216	2		
2	J	170	Total	C	N	O	S	0	0
			1319	831	261	225	2		

- Molecule 3 is a protein called CRISPR-associated protein, Cse1 family.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	493	Total	C	N	O	S	0	0
			3880	2466	709	696	9		

- Molecule 4 is a protein called CRISPR-associated protein, Cse4 family.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	D	260	Total	C	N	O	S	0	0
			2017	1266	356	390	5		
4	E	367	Total	C	N	O	S	0	0
			2828	1775	507	541	5		
4	F	367	Total	C	N	O	S	0	0
			2834	1777	509	543	5		
4	G	366	Total	C	N	O	S	0	0
			2823	1769	505	544	5		
4	H	366	Total	C	N	O	S	0	0
			2829	1772	508	544	5		
4	I	329	Total	C	N	O	S	0	0
			2544	1595	459	487	3		

- Molecule 5 is a RNA chain called crRNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	K	59	Total	C	N	O	P	0	0
			1267	565	235	409	58		

- Molecule 6 is a DNA chain called Target Strand.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	M	21	Total	C	N	O	P	0	0
			425	203	70	131	21		

- Molecule 7 is a protein called CRISPR-associated protein, Cas5e family.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	N	241	Total	C	N	O	S	0	0
			1891	1201	343	344	3		

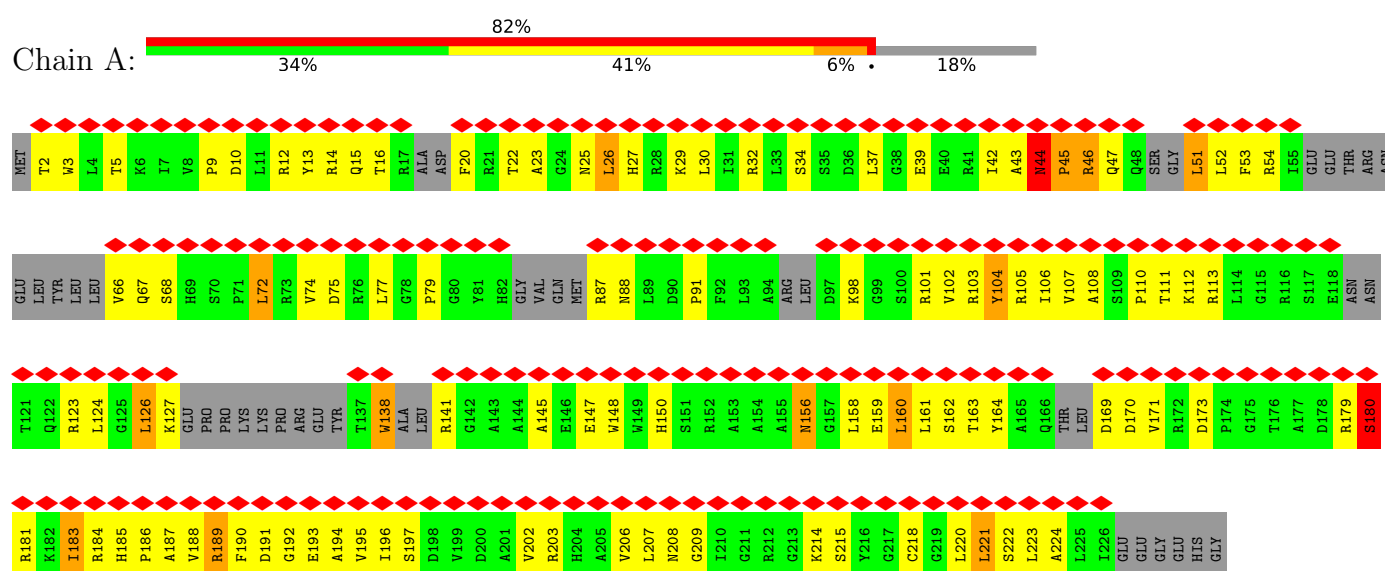
- Molecule 8 is a DNA chain called Nontarget Strand.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	O	13	Total	C	N	O	P	0	0
			268	125	55	75	13		

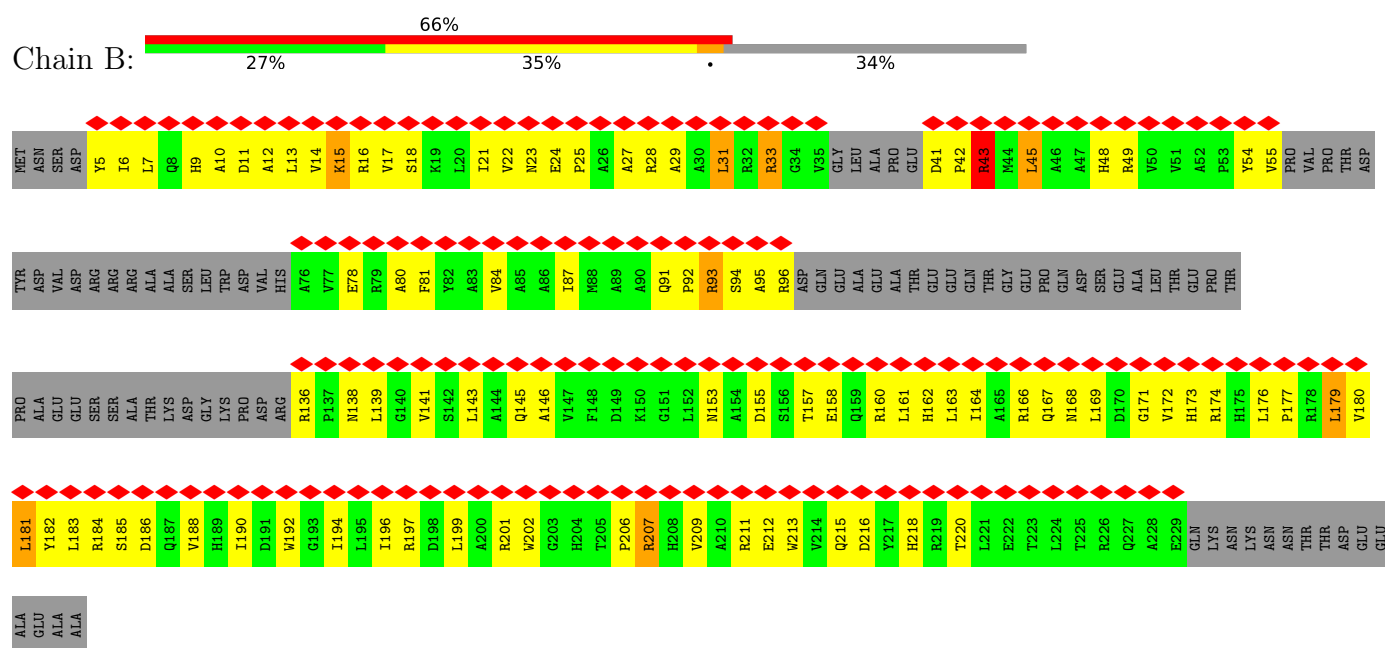
3 Residue-property plots

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and atom inclusion in map density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red diamond above a residue indicates a poor fit to the EM map for this residue (all-atom inclusion < 40%). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

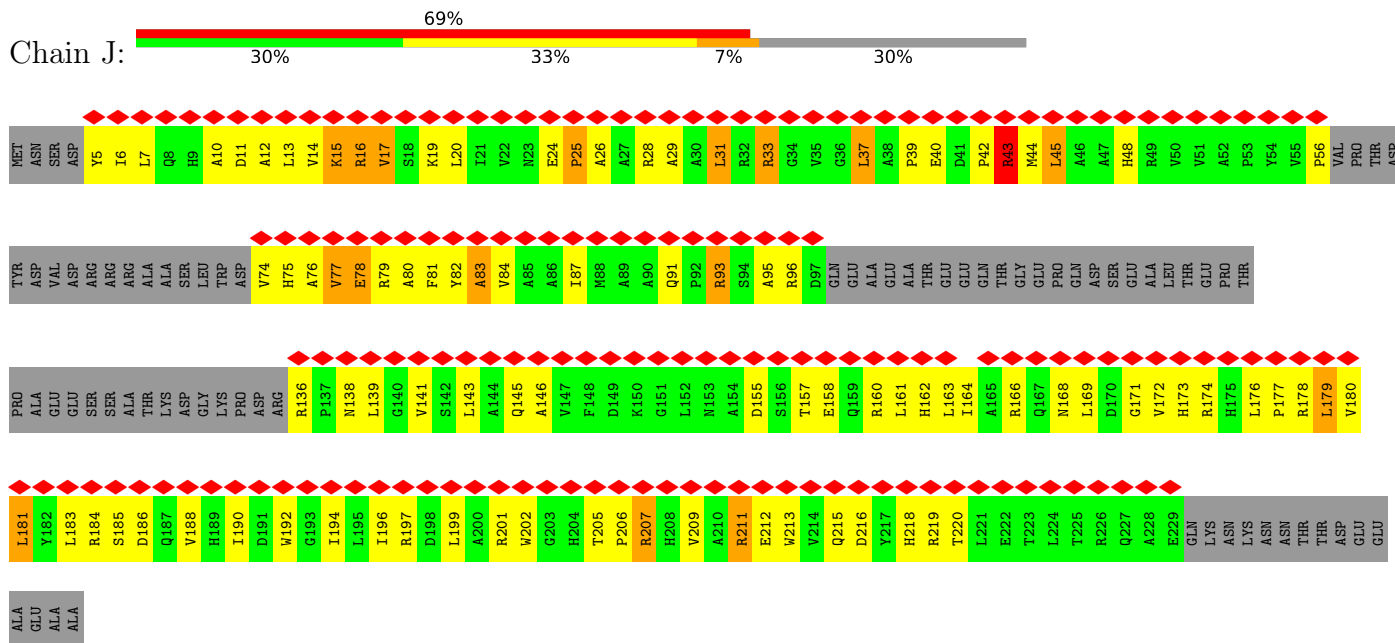
• Molecule 1: CRISPR-associated protein, Cse3 family



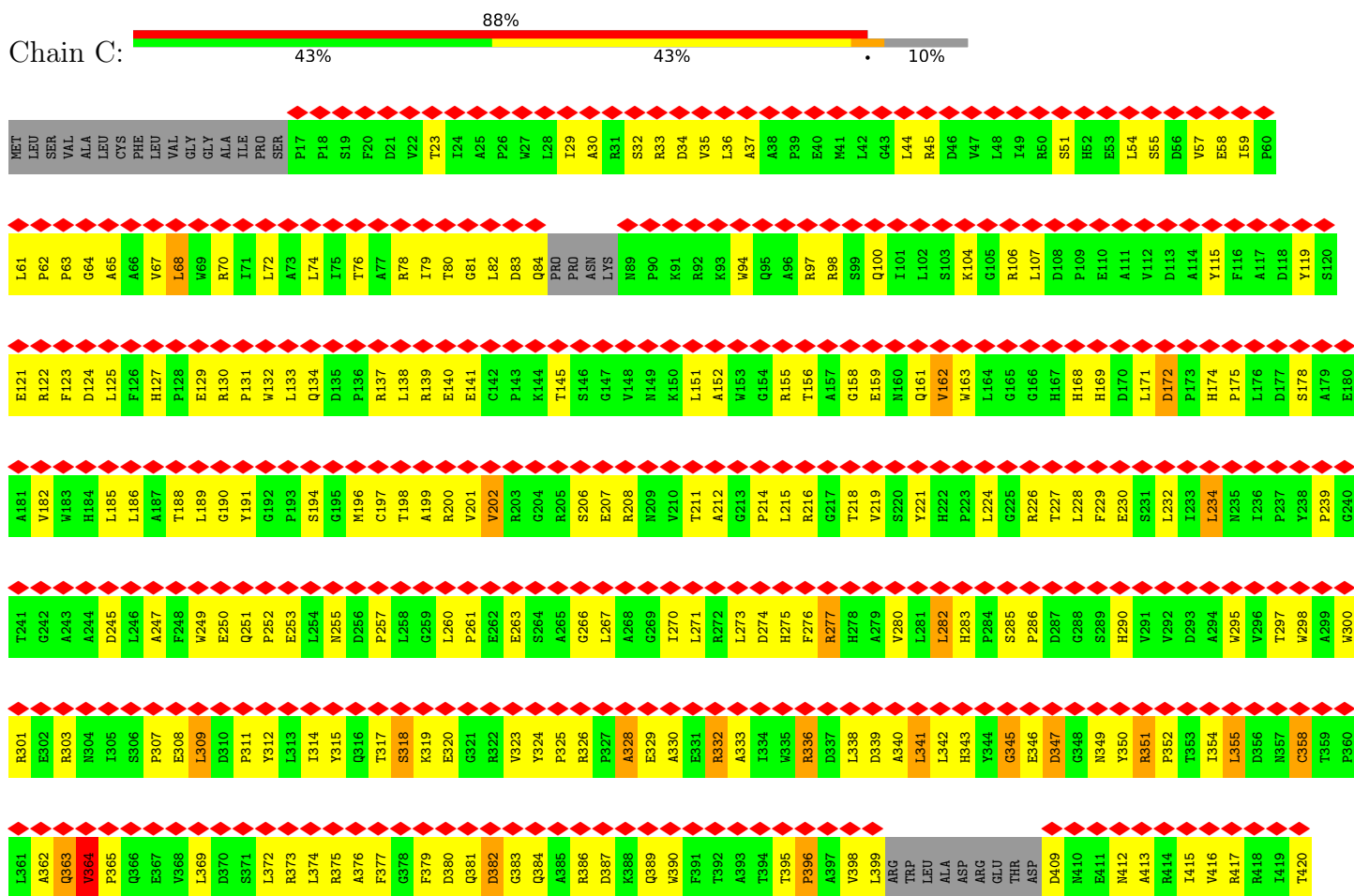
• Molecule 2: Cse2

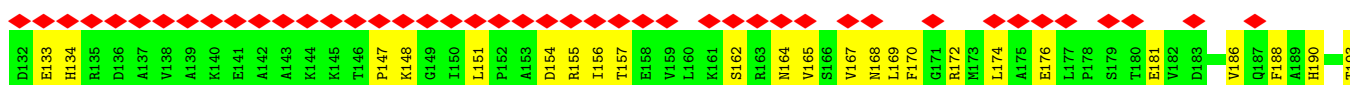


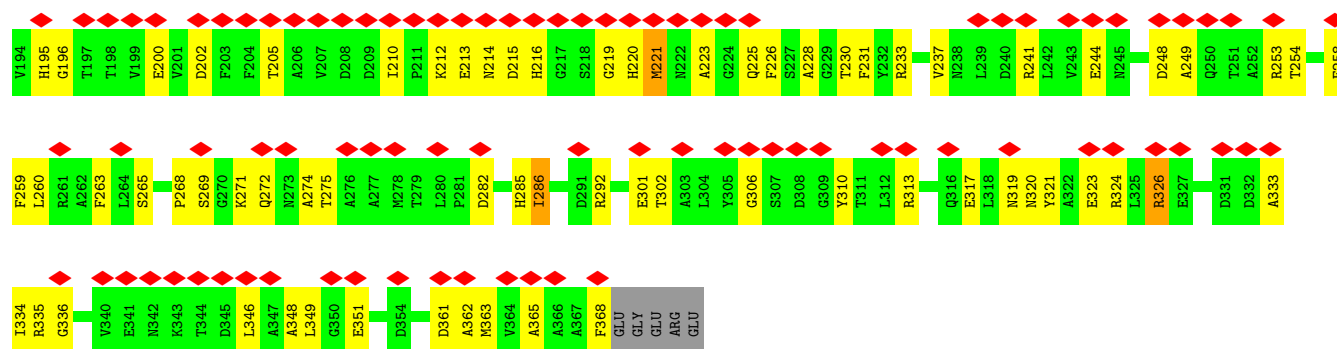
Chain J:



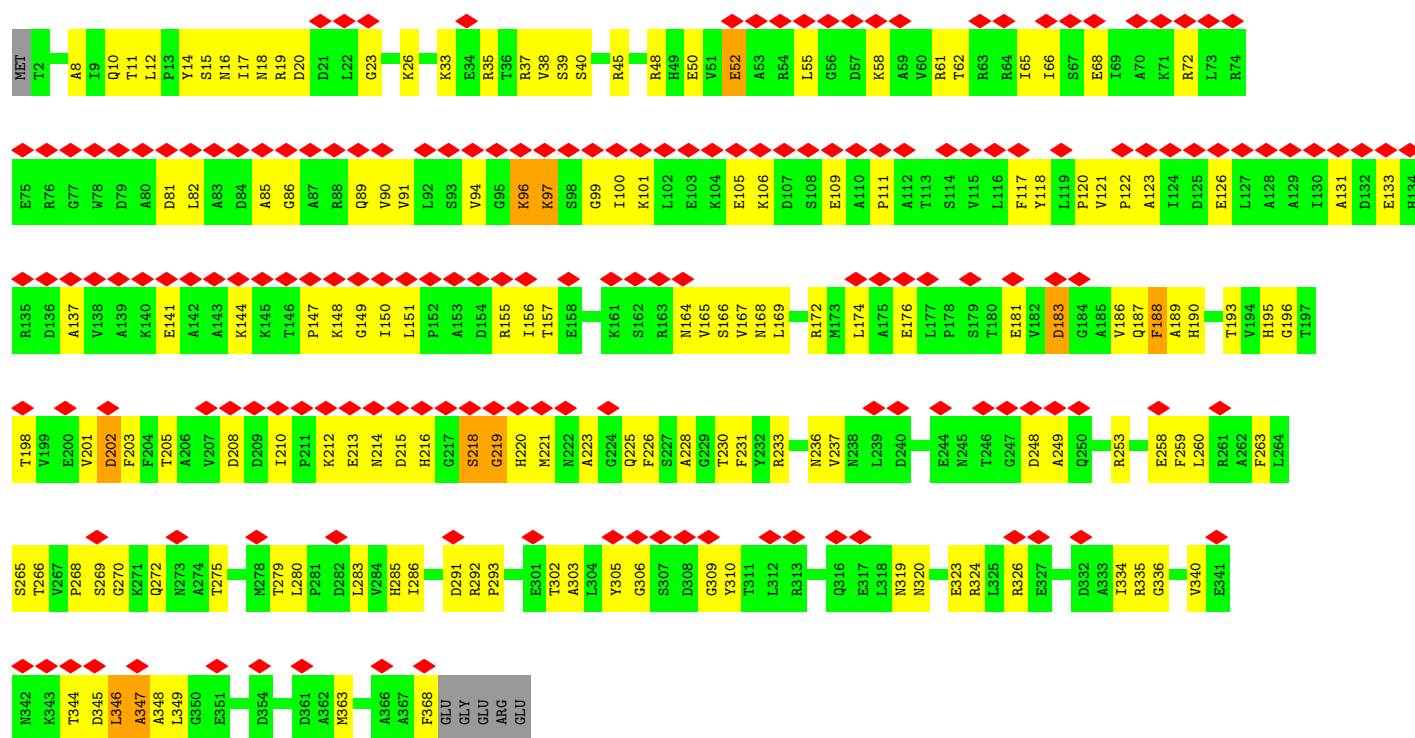
Chain C:



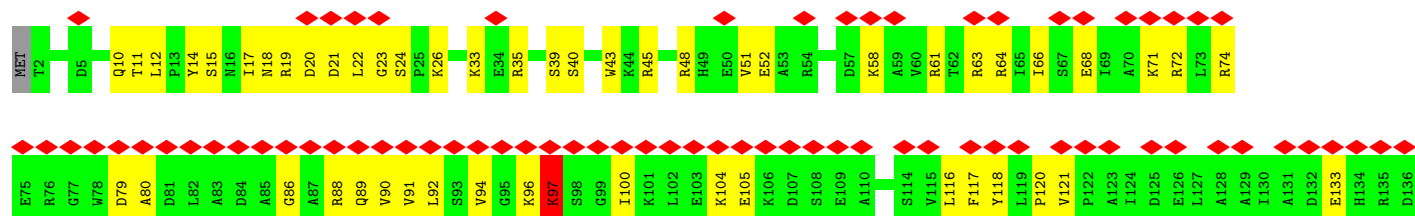


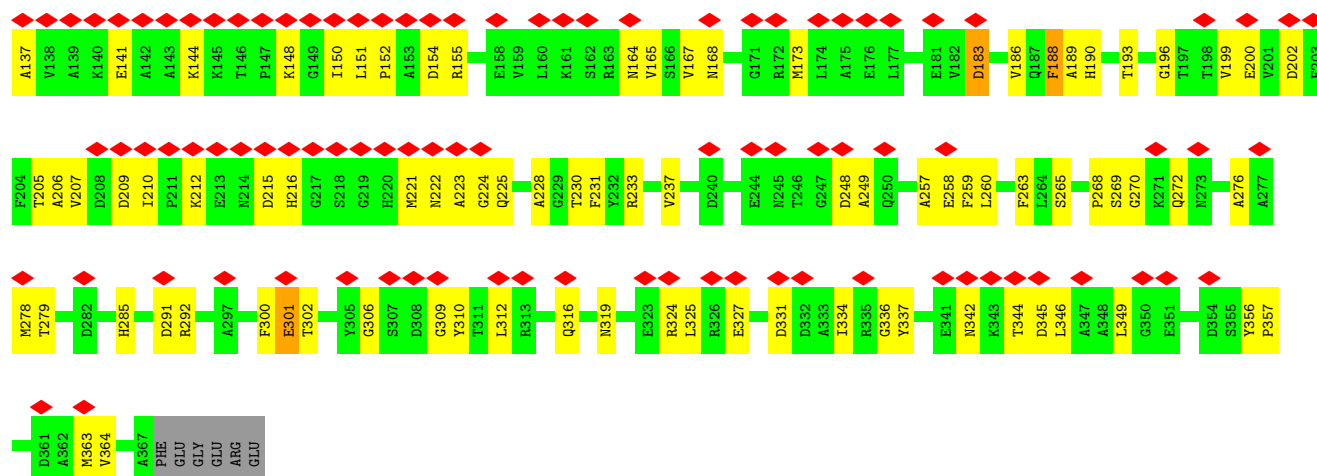


• Molecule 4: CRISPR-associated protein, Cse4 family



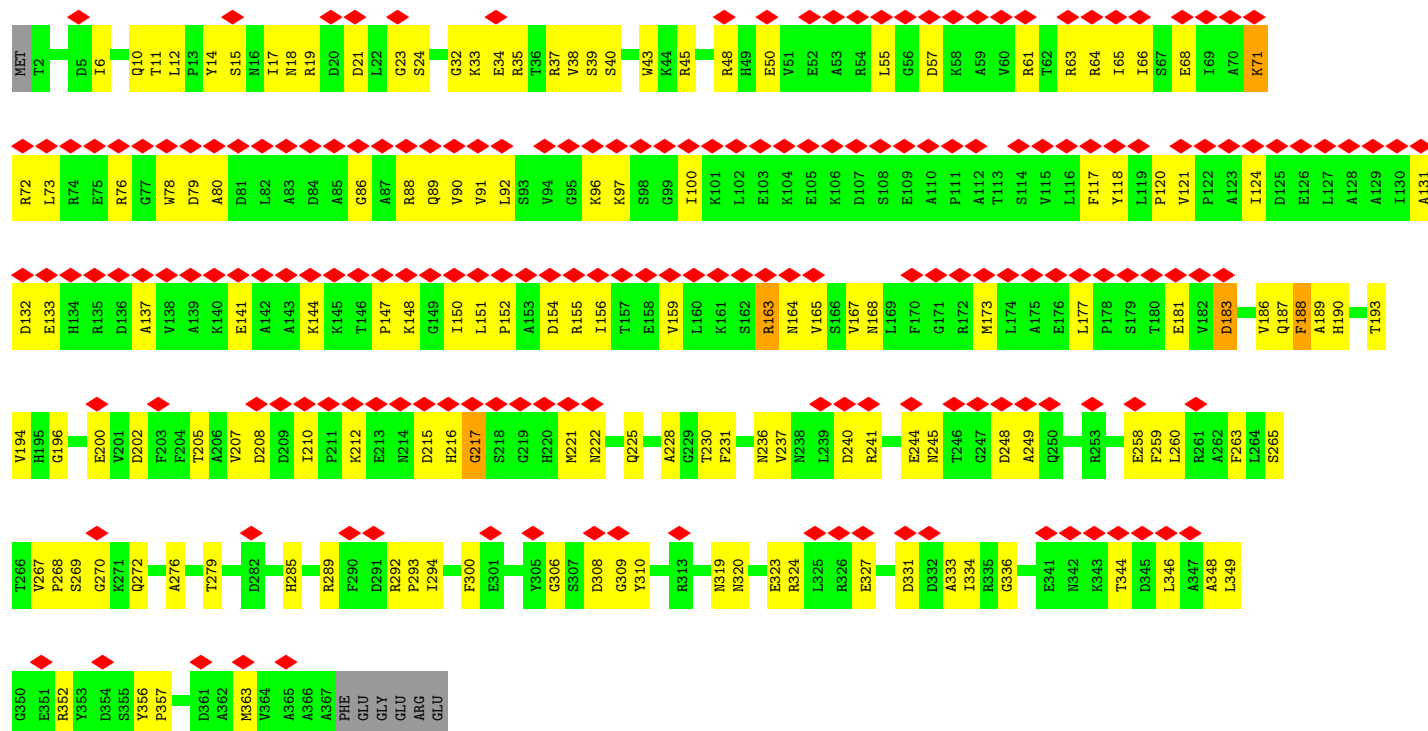
• Molecule 4: CRISPR-associated protein, Cse4 family





- Molecule 4: CRISPR-associated protein, Cse4 family

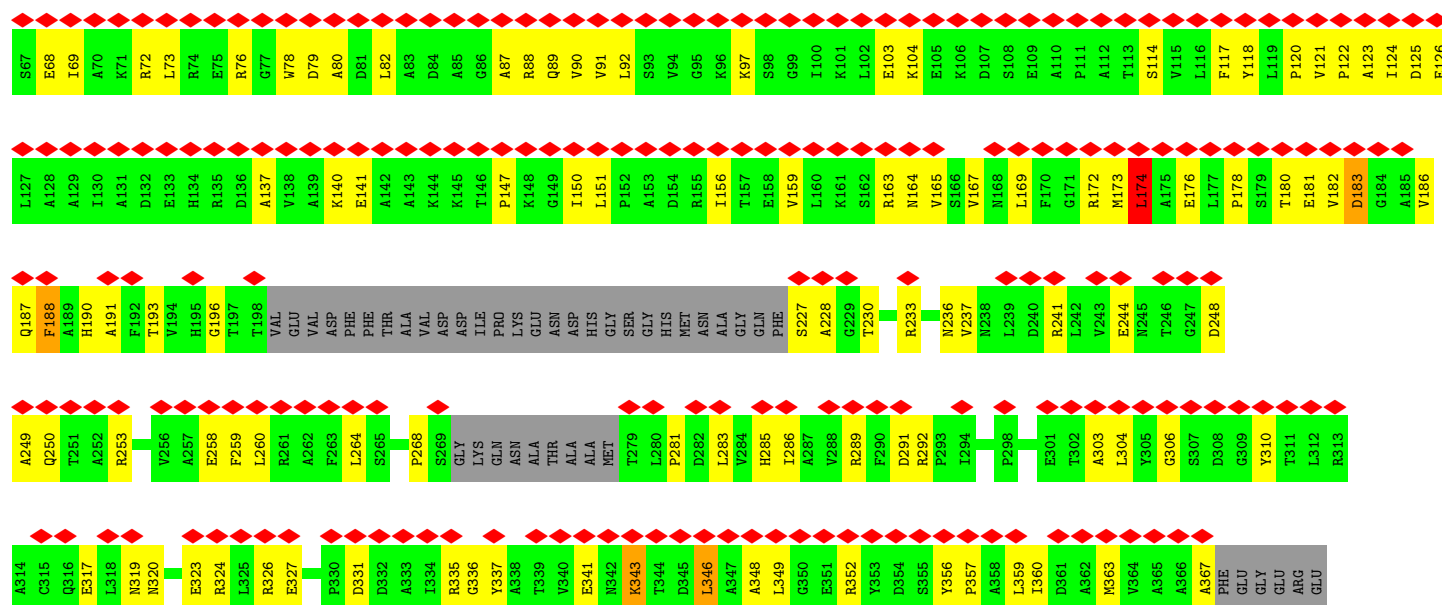
Chain H: 50% 58% 38% ..



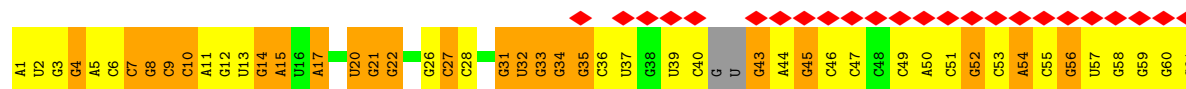
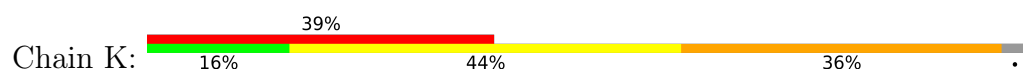
- Molecule 4: CRISPR-associated protein, Cse4 family

Chain I: 71% 51% 36% 12%

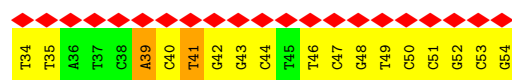




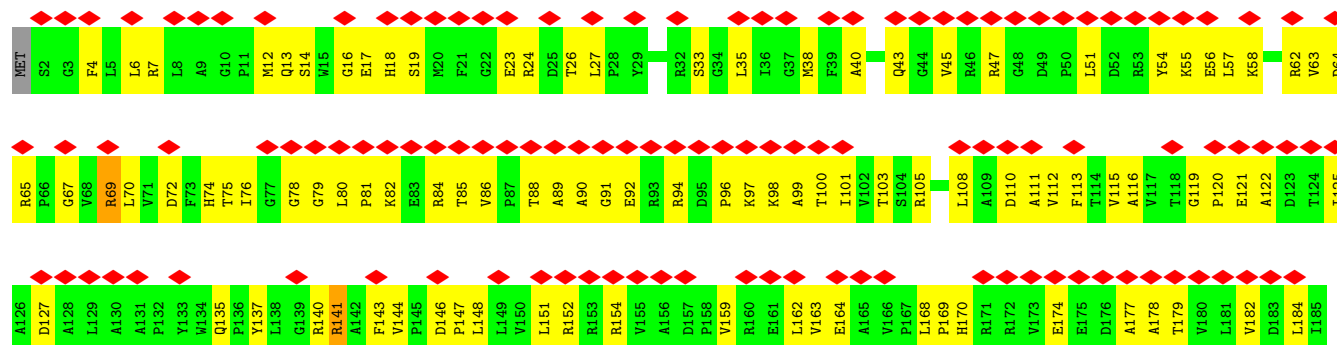
- Molecule 5: crRNA

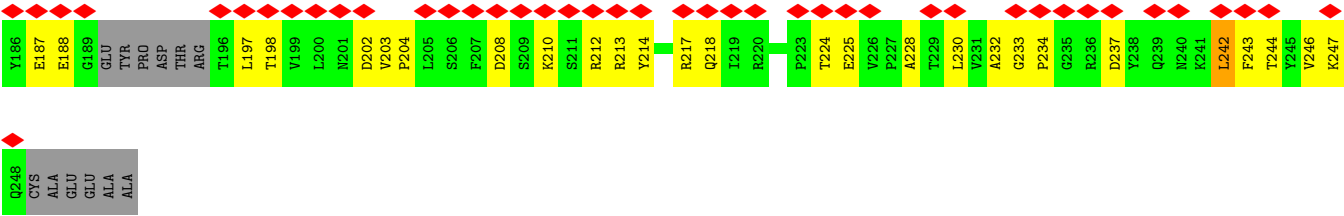


- Molecule 6: Target Strand

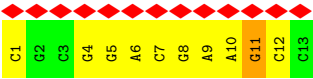


- Molecule 7: CRISPR-associated protein, Cas5e family





• Molecule 8: Nontarget Strand



4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	70222	Depositor
Resolution determination method	FSC 0.5 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI POLARA 300	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	8	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	2300	Depositor
Magnification	31000	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	0.214	Depositor
Minimum map value	-0.108	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.007	Depositor
Recommended contour level	0.07	Depositor
Map size (Å)	314.88, 314.88, 314.88	wwPDB
Map dimensions	256, 256, 256	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.23, 1.23, 1.23	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	0.36	0/1489	0.90	8/1995 (0.4%)
2	B	0.42	0/1310	0.80	3/1777 (0.2%)
2	J	0.41	0/1348	0.81	5/1836 (0.3%)
3	C	0.45	0/3991	0.75	2/5445 (0.0%)
4	D	0.60	0/2060	0.72	1/2802 (0.0%)
4	E	0.59	0/2881	0.70	1/3910 (0.0%)
4	F	0.61	0/2888	0.72	3/3919 (0.1%)
4	G	0.59	0/2876	0.68	1/3904 (0.0%)
4	H	0.57	0/2882	0.71	2/3911 (0.1%)
4	I	0.50	0/2587	0.72	3/3508 (0.1%)
5	K	0.83	0/1417	1.06	4/2208 (0.2%)
6	M	0.92	0/473	1.12	3/727 (0.4%)
7	N	0.54	0/1934	0.75	1/2631 (0.0%)
8	O	0.86	1/301 (0.3%)	1.02	3/462 (0.6%)
All	All	0.56	1/28437 (0.0%)	0.77	40/39035 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
8	O	11	DG	C3'-O3'	5.72	1.51	1.44

The worst 5 of 40 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	H	188	PHE	CB-CG-CD2	-9.57	114.10	120.80
4	I	188	PHE	CB-CG-CD2	-9.06	114.45	120.80
4	H	188	PHE	CB-CG-CD1	8.91	127.04	120.80
4	I	174	LEU	CA-CB-CG	8.21	134.19	115.30
8	O	11	DG	P-O3'-C3'	7.37	128.54	119.70

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	1471	0	1452	140	0
2	B	1285	0	1299	123	0
2	J	1319	0	1291	131	0
3	C	3880	0	3775	287	0
4	D	2017	0	1937	127	0
4	E	2828	0	2798	139	0
4	F	2834	0	2798	160	0
4	G	2823	0	2789	117	0
4	H	2829	0	2800	126	0
4	I	2544	0	2542	125	0
5	K	1267	0	642	94	0
6	M	425	0	239	28	0
7	N	1891	0	1914	129	0
8	O	268	0	144	15	0
All	All	27681	0	26420	1521	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 28.

The worst 5 of 1521 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:B:153:ASN:HD21	4:F:220:HIS:CD2	1.41	1.37
2:B:153:ASN:ND2	4:F:220:HIS:HD2	1.32	1.27
1:A:107:VAL:HA	1:A:188:VAL:O	1.41	1.18
3:C:211:THR:HG21	3:C:277:ARG:HG2	1.25	1.18
3:C:341:LEU:CD1	3:C:374:LEU:HD21	1.76	1.14

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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	170/232 (73%)	153 (90%)	9 (5%)	8 (5%)	2	24
2	B	153/244 (63%)	146 (95%)	6 (4%)	1 (1%)	22	60
2	J	164/244 (67%)	145 (88%)	12 (7%)	7 (4%)	2	26
3	C	485/549 (88%)	433 (89%)	40 (8%)	12 (2%)	5	36
4	D	256/373 (69%)	235 (92%)	19 (7%)	2 (1%)	19	57
4	E	365/373 (98%)	354 (97%)	11 (3%)	0	100	100
4	F	365/373 (98%)	353 (97%)	9 (2%)	3 (1%)	19	57
4	G	364/373 (98%)	353 (97%)	10 (3%)	1 (0%)	41	74
4	H	364/373 (98%)	349 (96%)	13 (4%)	2 (0%)	29	66
4	I	323/373 (87%)	313 (97%)	9 (3%)	1 (0%)	41	74
7	N	237/254 (93%)	218 (92%)	19 (8%)	0	100	100
All	All	3246/3761 (86%)	3052 (94%)	157 (5%)	37 (1%)	18	51

5 of 37 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	44	ASN
1	A	156	ASN
1	A	160	LEU
1	A	183	ILE
2	B	6	ILE

5.3.2 Protein sidechains

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	145/193 (75%)	139 (96%)	6 (4%)	30	59
2	B	128/200 (64%)	118 (92%)	10 (8%)	12	42
2	J	126/200 (63%)	117 (93%)	9 (7%)	14	45
3	C	395/452 (87%)	379 (96%)	16 (4%)	30	59
4	D	209/299 (70%)	207 (99%)	2 (1%)	76	86
4	E	291/299 (97%)	284 (98%)	7 (2%)	49	71
4	F	292/299 (98%)	286 (98%)	6 (2%)	53	74
4	G	292/299 (98%)	289 (99%)	3 (1%)	76	86
4	H	293/299 (98%)	288 (98%)	5 (2%)	60	78
4	I	264/299 (88%)	260 (98%)	4 (2%)	65	81
7	N	201/211 (95%)	197 (98%)	4 (2%)	55	75
All	All	2636/3050 (86%)	2564 (97%)	72 (3%)	48	69

5 of 72 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	I	183	ASP
7	N	242	LEU
2	J	15	LYS
2	J	183	LEU
3	C	336	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 54 such sidechains are listed below:

Mol	Chain	Res	Type
4	E	272	GLN
4	F	316	GLN
2	J	173	HIS
4	E	285	HIS
4	F	134	HIS

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
5	K	57/61 (93%)	25 (43%)	0

5 of 25 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
5	K	4	G
5	K	7	C
5	K	8	G
5	K	9	C
5	K	10	C

There are no RNA pucker outliers to report.

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

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

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

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.

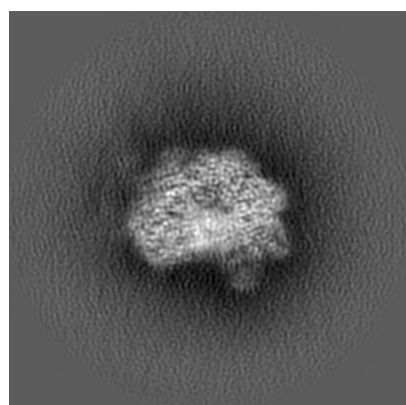
6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-8477. These allow visual inspection of the internal detail of the map and identification of artifacts.

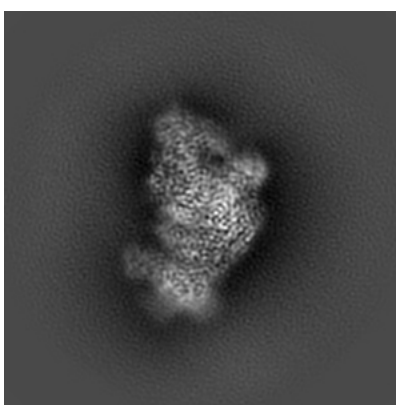
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections [i](#)

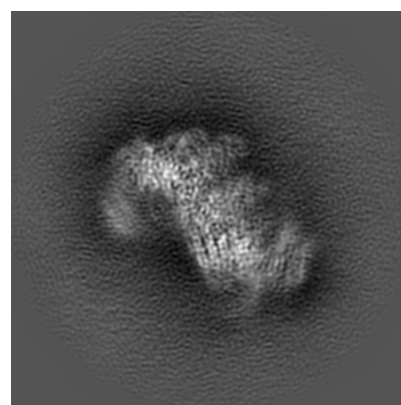
6.1.1 Primary map



X



Y

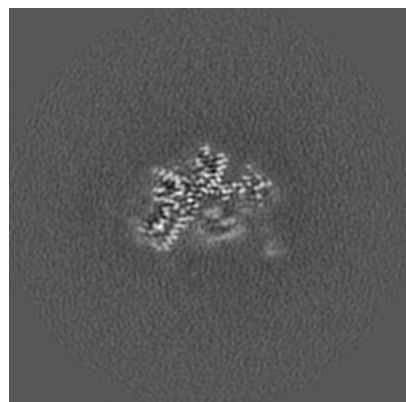


Z

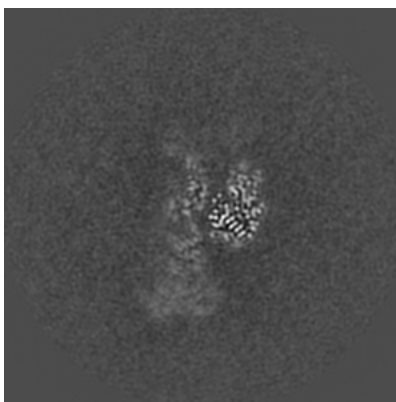
The images above show the map projected in three orthogonal directions.

6.2 Central slices [i](#)

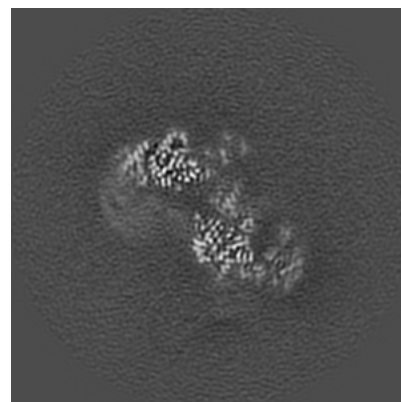
6.2.1 Primary map



X Index: 128



Y Index: 128

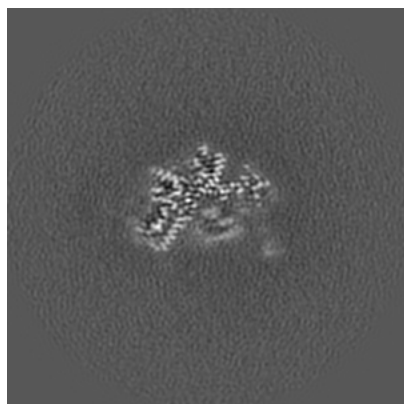


Z Index: 128

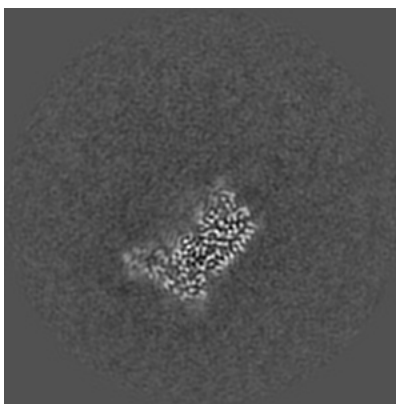
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices [i](#)

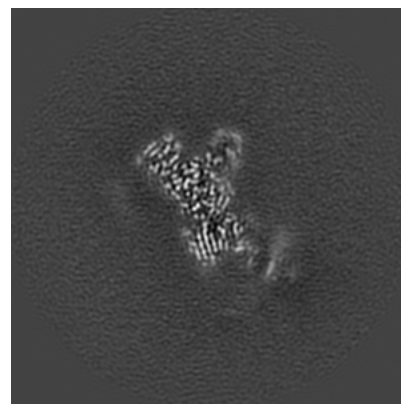
6.3.1 Primary map



X Index: 128



Y Index: 154



Z Index: 138

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

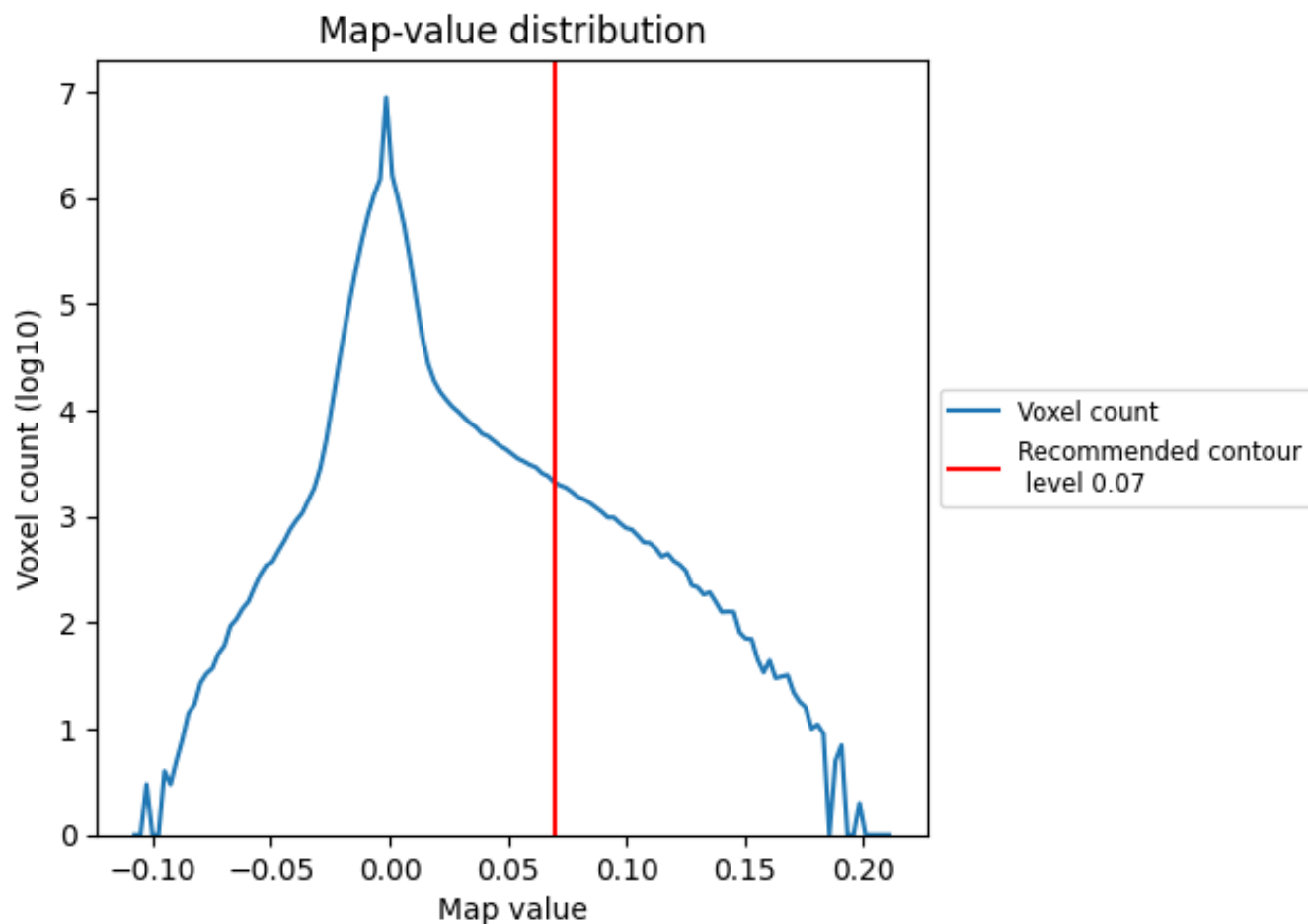
6.5 Mask visualisation

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

7 Map analysis [i](#)

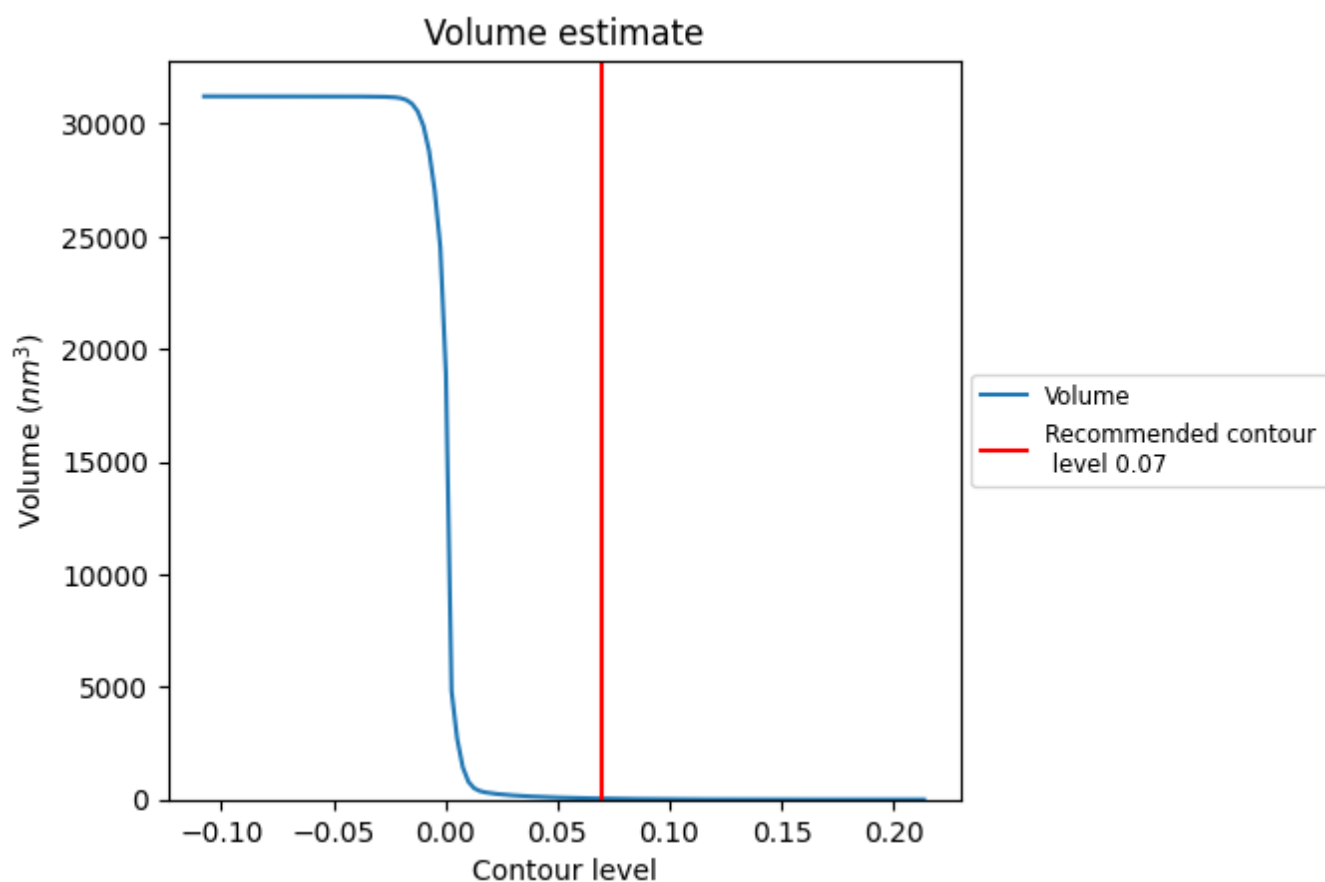
This section contains the results of statistical analysis of the map.

7.1 Map-value distribution [i](#)



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

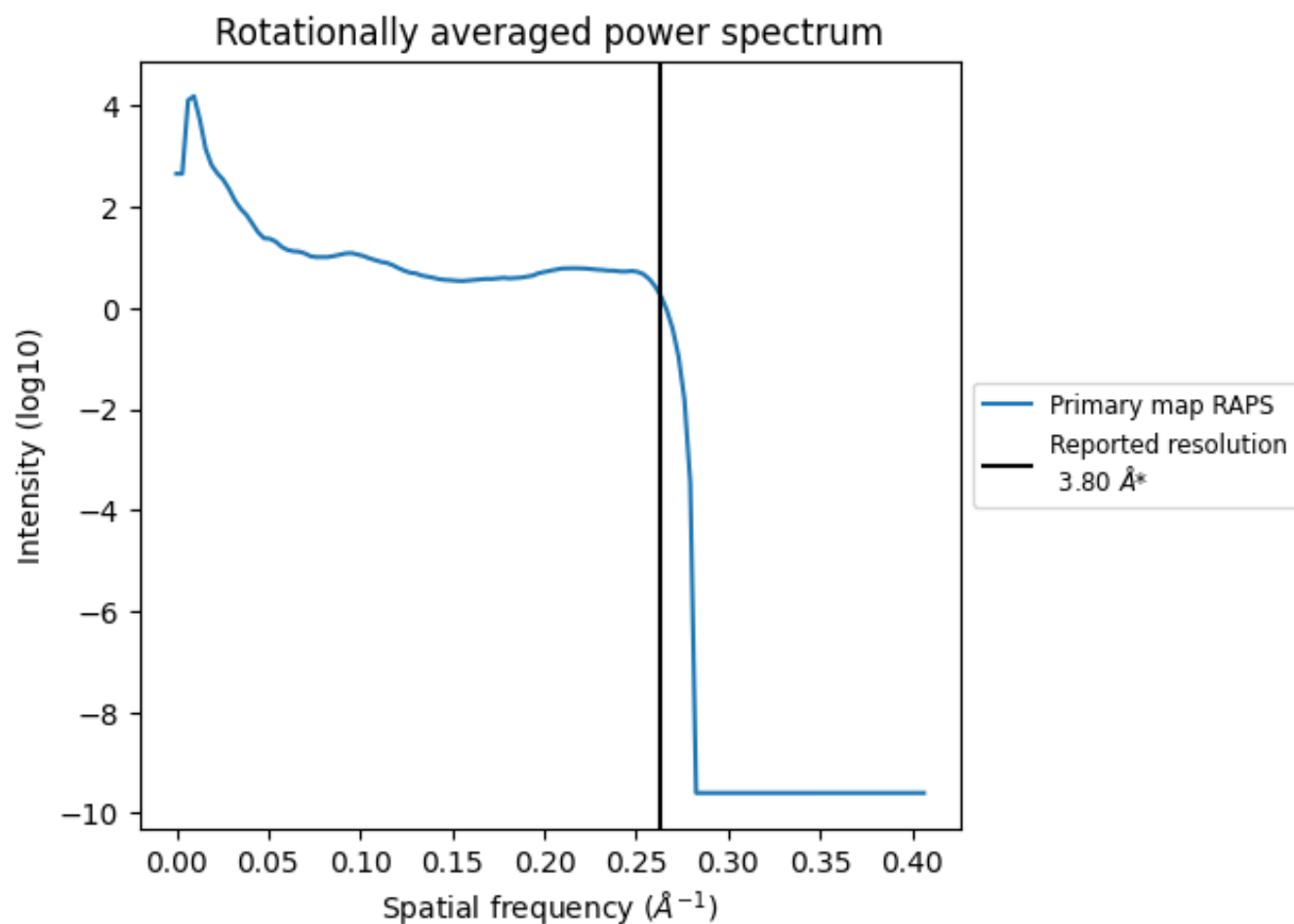
7.2 Volume estimate [i](#)



The volume at the recommended contour level is 45 nm³; this corresponds to an approximate mass of 41 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 ⓘ



*Reported resolution corresponds to spatial frequency of 0.263 \AA^{-1}

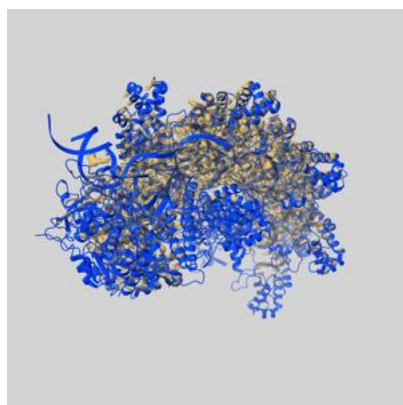
8 Fourier-Shell correlation

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

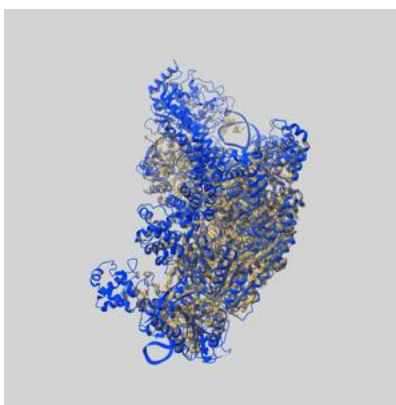
9 Map-model fit [i](#)

This section contains information regarding the fit between EMDB map EMD-8477 and PDB model 5U07. Per-residue inclusion information can be found in section [3](#) on page [6](#).

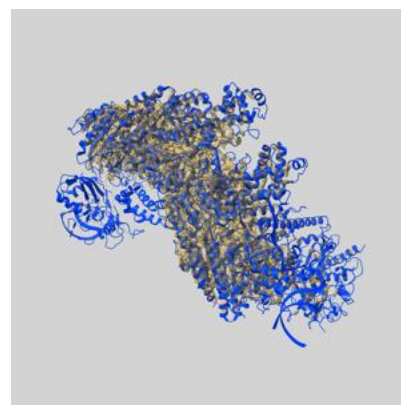
9.1 Map-model overlay [i](#)



X



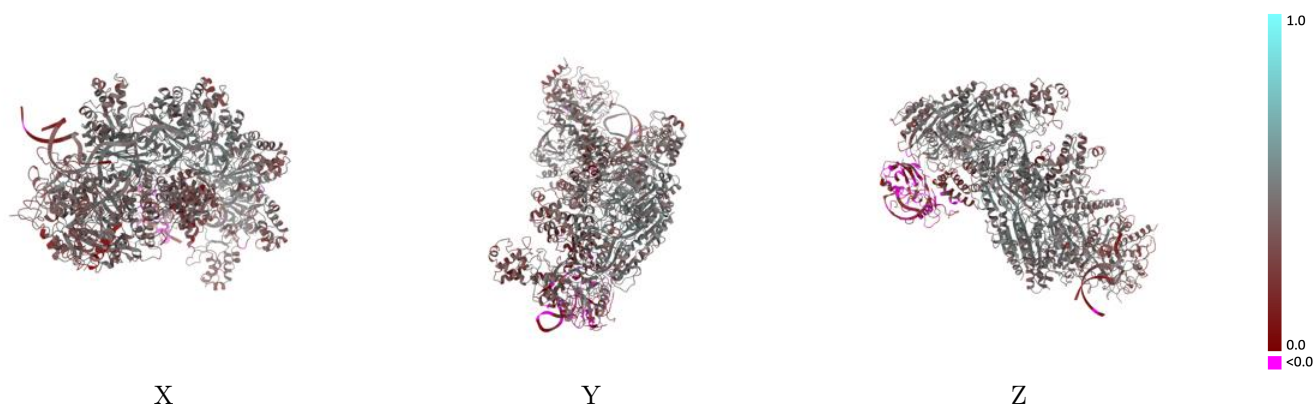
Y



Z

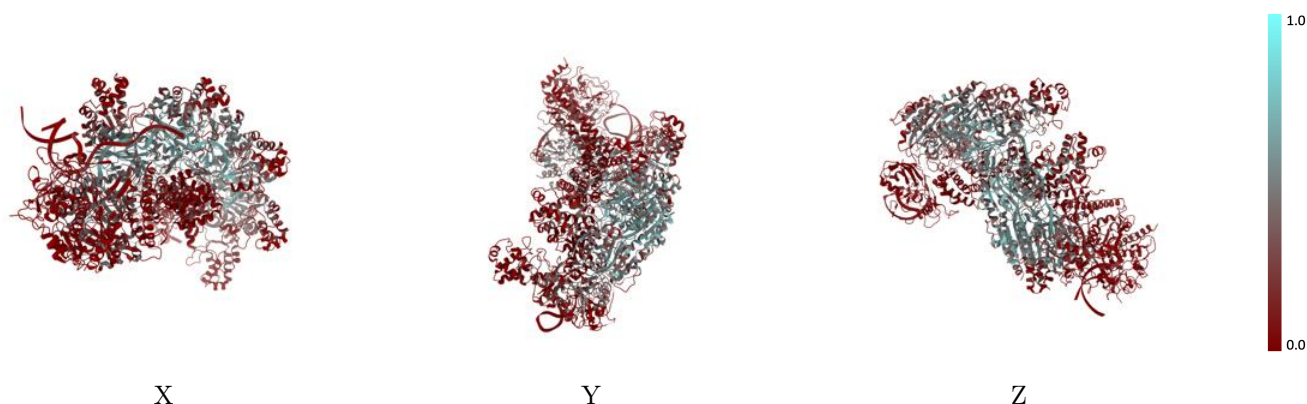
The images above show the 3D surface view of the map at the recommended contour level 0.07 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

9.2 Q-score mapped to coordinate model [i](#)



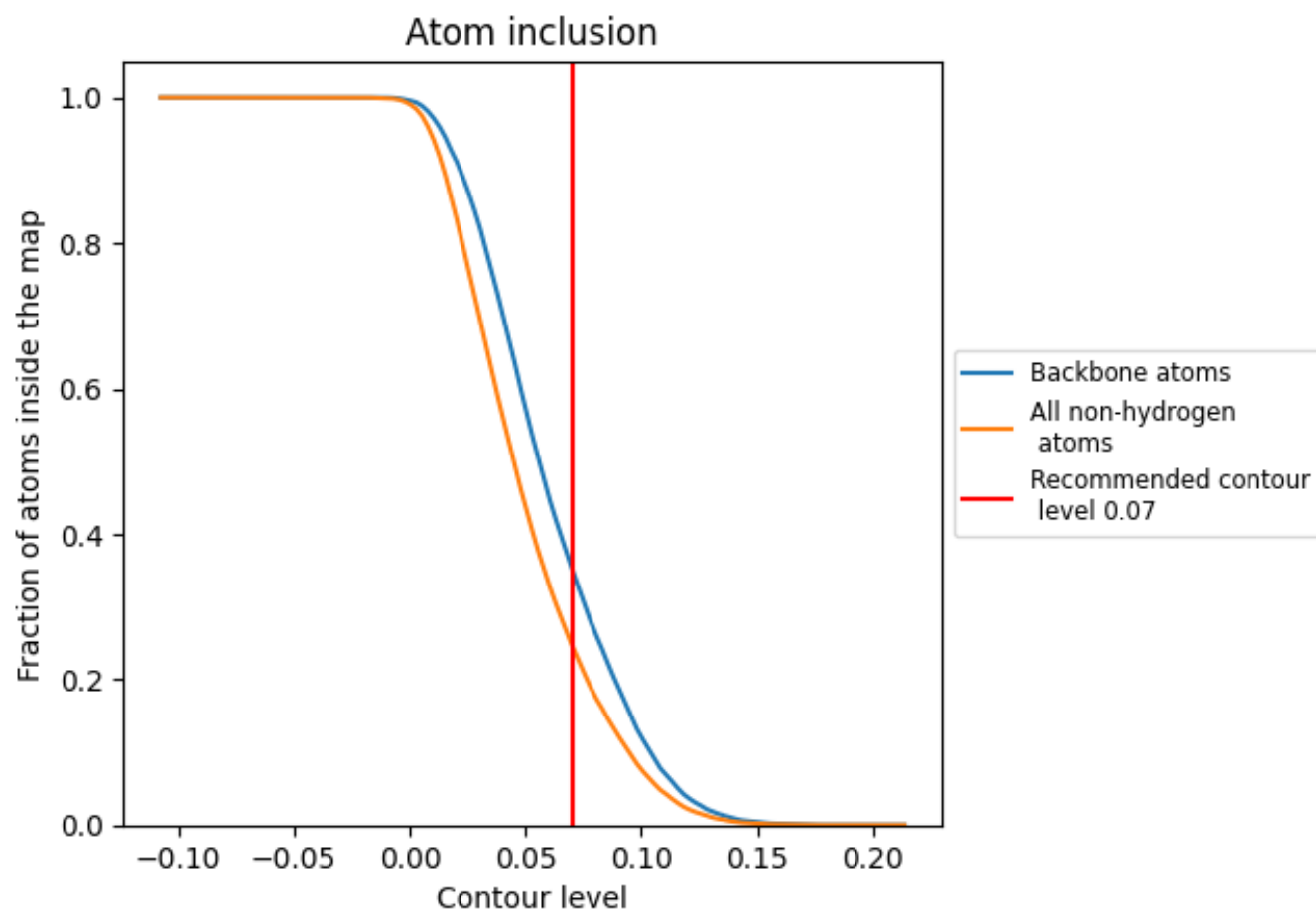
The images above show the model with each residue coloured according to its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions 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](#)



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 (0.07).































9.4 Atom inclusion [i](#)



At the recommended contour level, 35% of all backbone atoms, 25% 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 (0.07) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	 0.2462	 0.3850
A	 0.0000	 0.0560
B	 0.0049	 0.2560
C	 0.0417	 0.3710
D	 0.3662	 0.4470
E	 0.3328	 0.4240
F	 0.3954	 0.4460
G	 0.4127	 0.4590
H	 0.3541	 0.4470
I	 0.2032	 0.3750
J	 0.0424	 0.3400
K	 0.4112	 0.3450
M	 0.0635	 0.3590
N	 0.2812	 0.4370
O	 0.0299	 0.2130

