



## wwPDB EM Validation Summary Report ⓘ

Nov 6, 2022 – 05:52 AM EST

PDB ID : 6B47  
EMDB ID : EMD-7051  
Title : Cryo-EM structure of Type I-F CRISPR crRNA-guided Csy surveillance complex with bound anti-CRISPR protein AcrF2  
Authors : Guo, T.W.; Bartesaghi, A.; Yang, H.; Falconieri, V.; Rao, P.; Merk, A.; Fox, T.; Earl, L.; Patel, D.J.; Subramaniam, S.  
Deposited on : 2017-09-25  
Resolution : 3.20 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](mailto: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

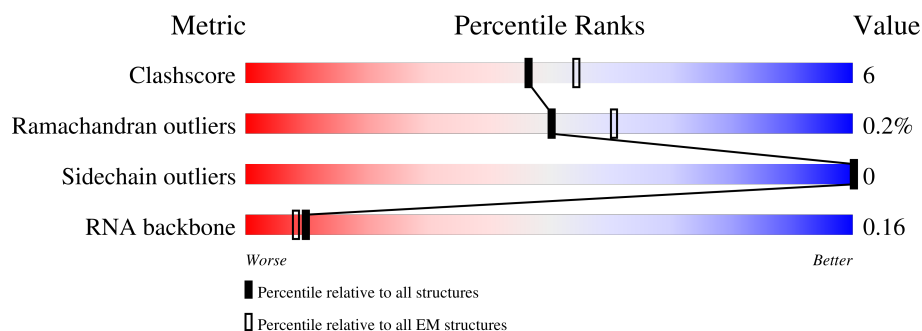
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 3.20 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
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  $\geq 3$ , 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	436	<div> <div>68%</div> <div>92%</div> <div>5%</div> <div>.</div> </div>
2	B	329	<div> <div>43%</div> <div>72%</div> <div>21%</div> <div>7%</div> </div>
3	C	344	<div> <div>30%</div> <div>69%</div> <div>16%</div> <div>15%</div> </div>
3	D	344	<div> <div>19%</div> <div>81%</div> <div>15%</div> <div>.</div> </div>
3	E	344	<div> <div>14%</div> <div>83%</div> <div>14%</div> <div>.</div> </div>
3	F	344	<div> <div>12%</div> <div>83%</div> <div>15%</div> <div>.</div> </div>
3	G	344	<div> <div>12%</div> <div>79%</div> <div>17%</div> <div>.</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	H	344	<div><div></div><div>17%</div><div>85%</div><div>12%</div><div></div></div>
4	K	92	<div><div></div><div>92%</div><div>68%</div><div>26%</div><div>5%</div></div>
5	L	189	<div><div></div><div>100%</div><div>94%</div><div>6%</div><div></div></div>
6	M	60	<div><div></div><div>35%</div><div>25%</div><div>52%</div><div>22%</div><div></div></div>

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 22213 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 Csy1.

Mol	Chain	Residues	Atoms				AltConf	Trace
1	A	424	Total	C	N	O	0	0
			2079	1145	473	461		

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP Q02ML9
A	0	SER	-	expression tag	UNP Q02ML9

- Molecule 2 is a protein called CRISPR-associated protein Csy2.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	305	Total	C	N	O	S	0	0
			2374	1504	440	425	5		

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-1	MET	-	initiating methionine	UNP Q02MM0
B	0	ALA	-	expression tag	UNP Q02MM0

- Molecule 3 is a protein called CRISPR-associated protein Csy3.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	293	Total	C	N	O	S	0	0
			2272	1430	409	431	2		
3	D	333	Total	C	N	O	S	0	0
			2554	1603	466	483	2		
3	E	334	Total	C	N	O	S	0	0
			2561	1611	466	482	2		
3	F	334	Total	C	N	O	S	0	0
			2561	1611	466	482	2		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					AltConf	Trace
3	G	333	Total	C	N	O	S	0	0
			2557	1608	466	481	2		
3	H	333	Total	C	N	O	S	0	0
			2557	1608	466	481	2		

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-1	MET	-	initiating methionine	UNP Q02MM1
C	0	ALA	-	expression tag	UNP Q02MM1
D	-1	MET	-	initiating methionine	UNP Q02MM1
D	0	ALA	-	expression tag	UNP Q02MM1
E	-1	MET	-	initiating methionine	UNP Q02MM1
E	0	ALA	-	expression tag	UNP Q02MM1
F	-1	MET	-	initiating methionine	UNP Q02MM1
F	0	ALA	-	expression tag	UNP Q02MM1
G	-1	MET	-	initiating methionine	UNP Q02MM1
G	0	ALA	-	expression tag	UNP Q02MM1
H	-1	MET	-	initiating methionine	UNP Q02MM1
H	0	ALA	-	expression tag	UNP Q02MM1

- Molecule 4 is a protein called Anti-CRISPR protein AcrF2.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	K	87	Total	C	N	O	S	0	0
			668	410	107	148	3		

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
K	-1	GLY	-	expression tag	UNP Q6TM72
K	0	SER	-	expression tag	UNP Q6TM72

- Molecule 5 is a protein called CRISPR-associated endonuclease Cas6/Csy4.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	L	189	Total	C	N	O		0	0
			758	380	189	189			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
L	-1	MET	-	initiating methionine	UNP Q02MM2
L	0	ALA	-	expression tag	UNP Q02MM2

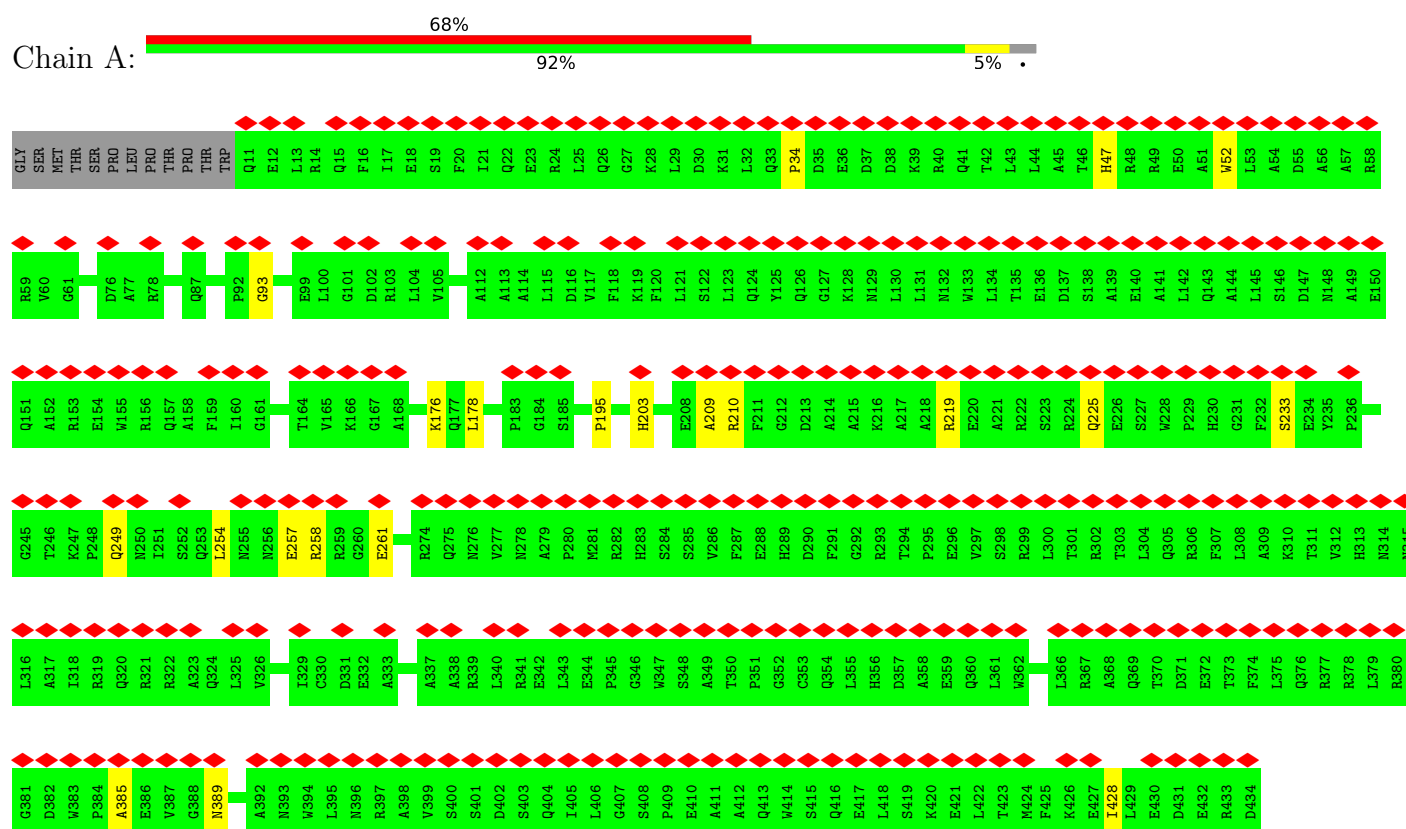
- Molecule 6 is a RNA chain called *Pseudomonas aeruginosa* strain SMC4485 CRISPR repeat sequence.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	M	60	Total	C	N	O	P	0	0
			1272	569	223	421	59		

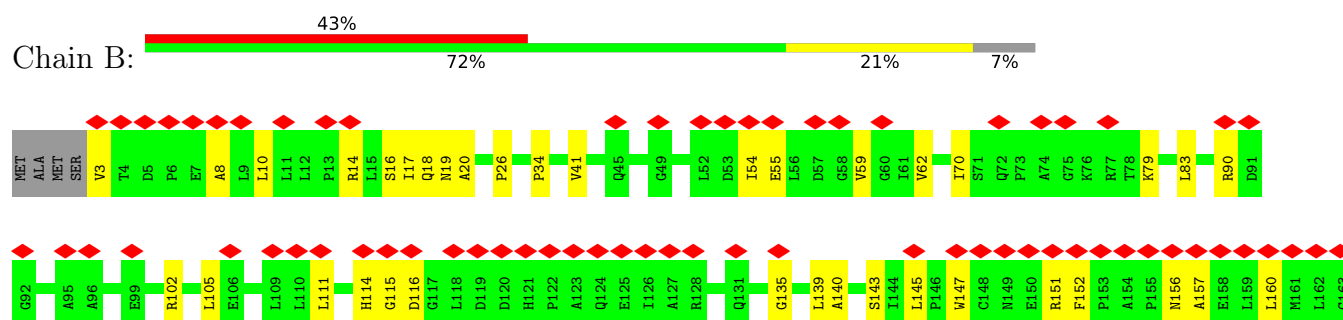
### 3 Residue-property plots

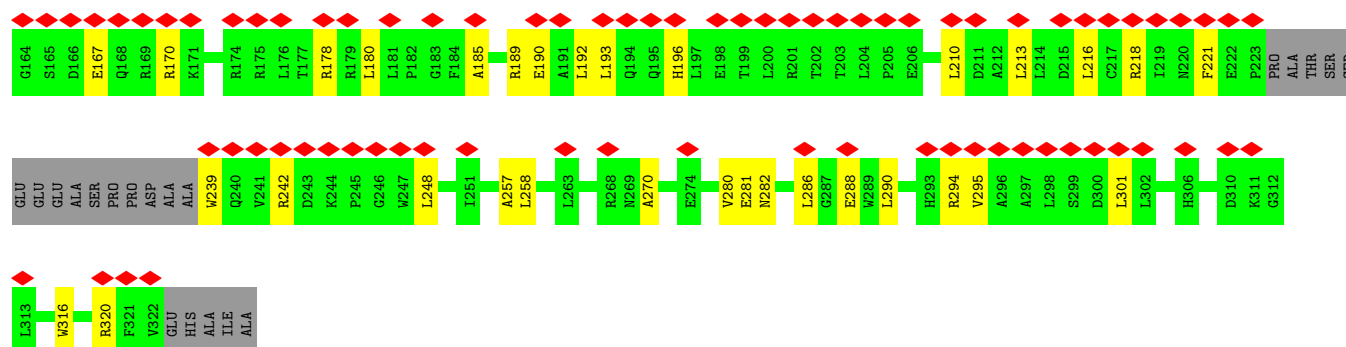
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and atom inclusion 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 Csy1

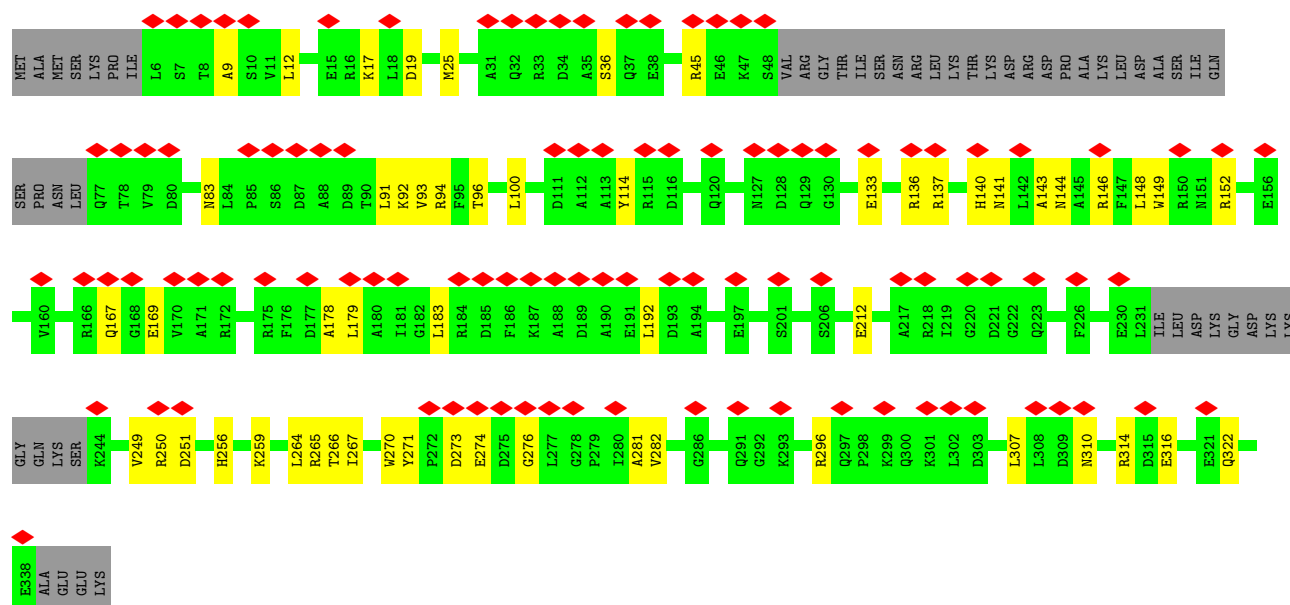


#### • Molecule 2: CRISPR-associated protein Csy2

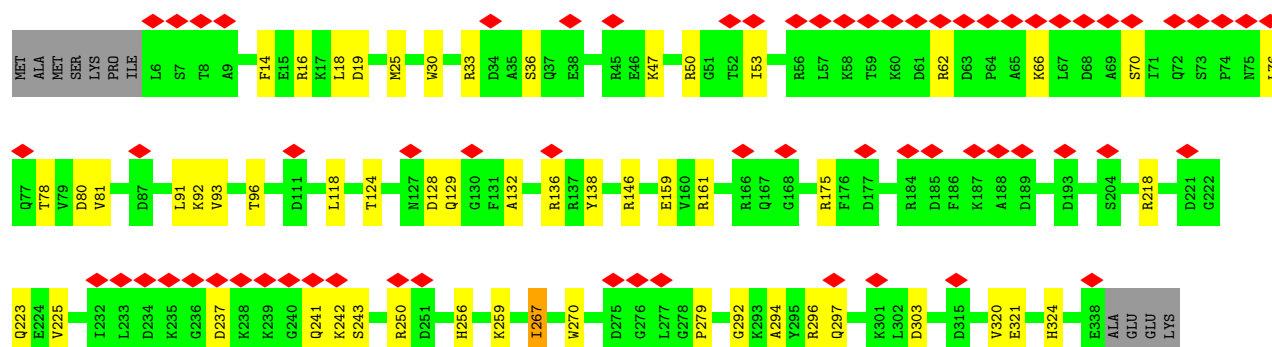
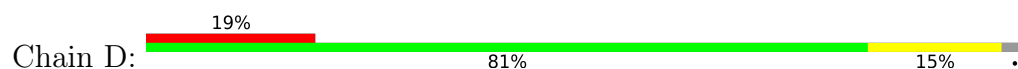




• Molecule 3: CRISPR-associated protein Csy3

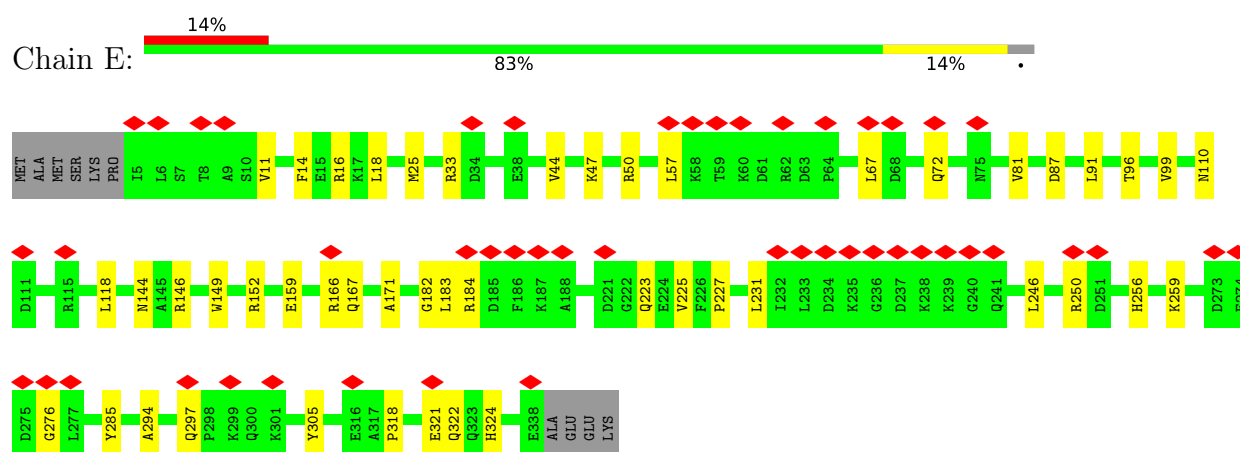


• Molecule 3: CRISPR-associated protein Csy3

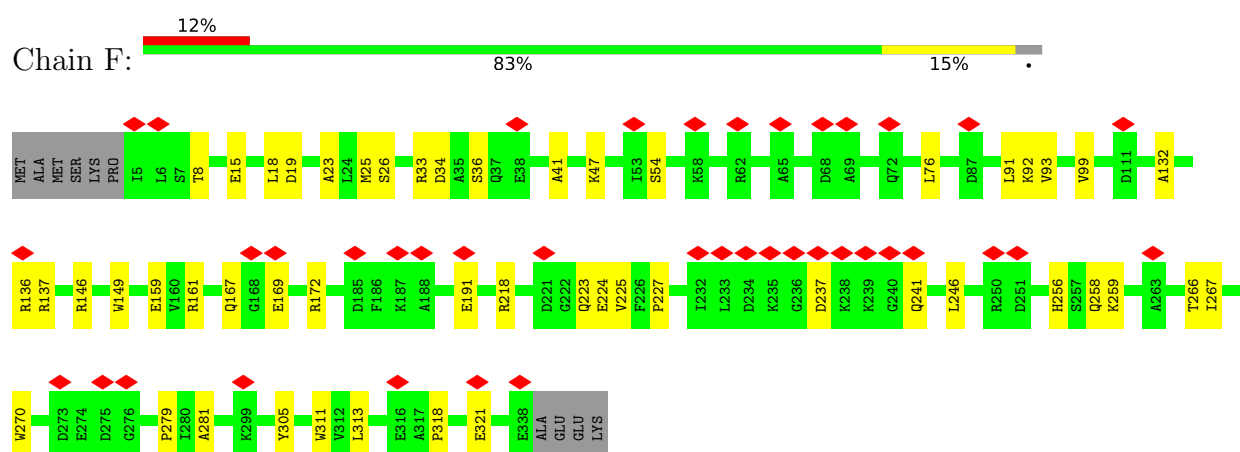


• Molecule 3: CRISPR-associated protein Csy3

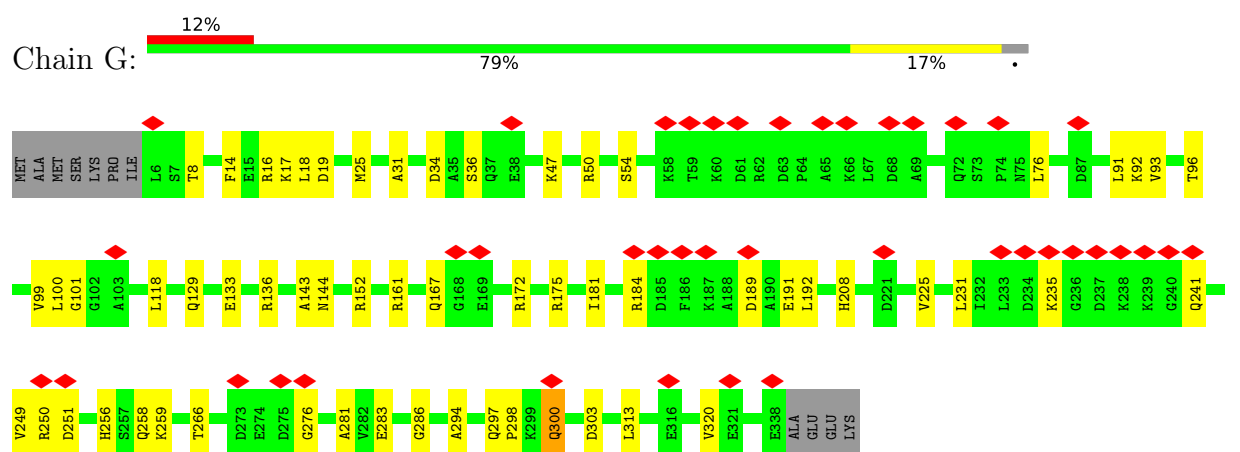




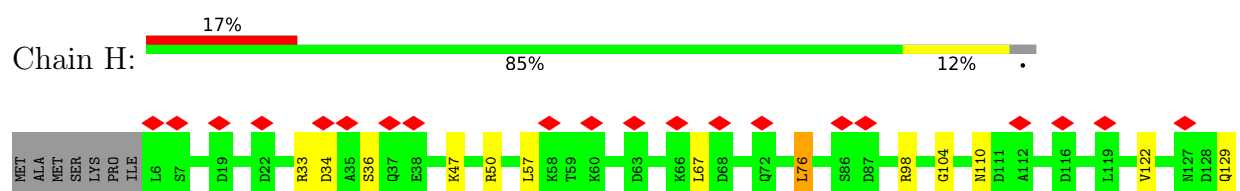
• Molecule 3: CRISPR-associated protein Csy3

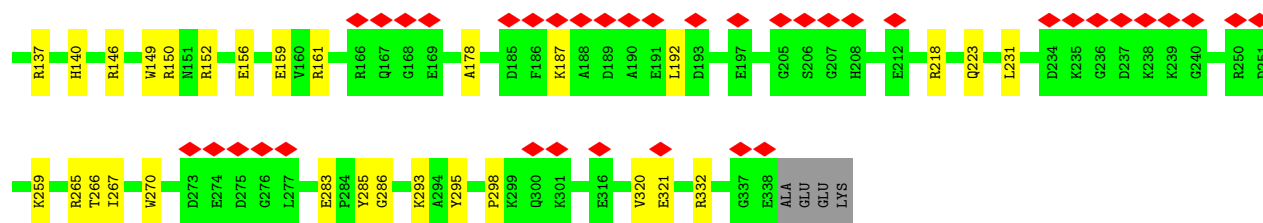


• Molecule 3: CRISPR-associated protein Csy3

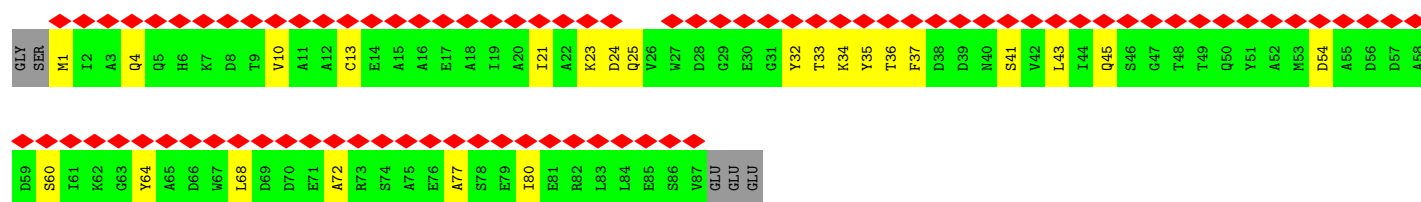


• Molecule 3: CRISPR-associated protein Csy3

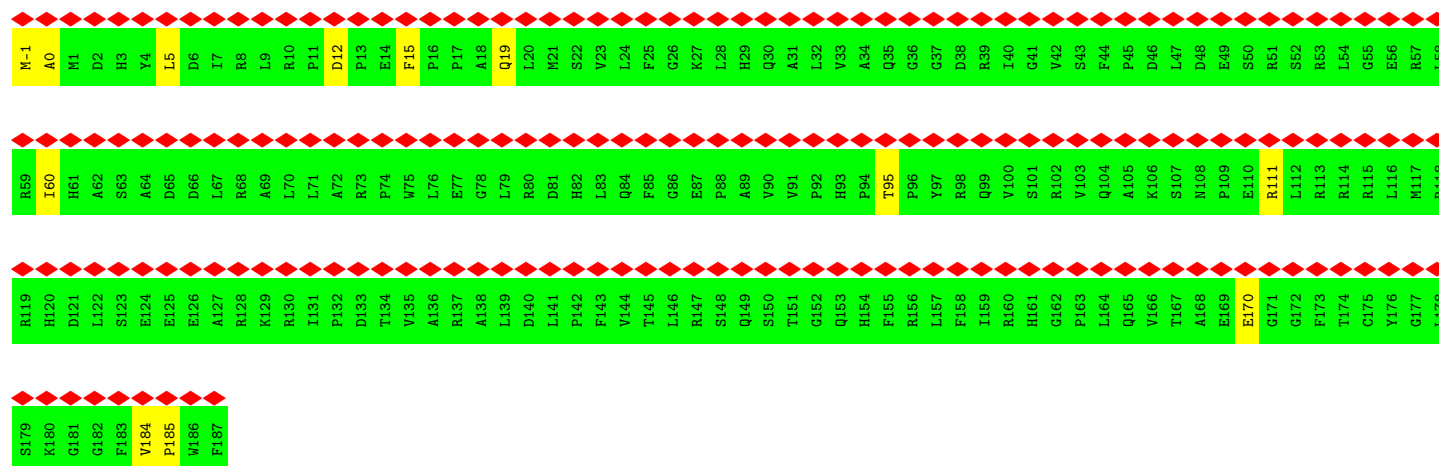




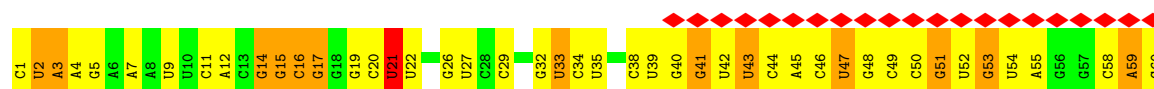
• Molecule 4: Anti-CRISPR protein AcrF2



• Molecule 5: CRISPR-associated endonuclease Cas6/Csy4



• Molecule 6: Pseudomonas aeruginosa strain SMC4485 CRISPR repeat sequence



## 4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	219718	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 ( $e^-/\text{\AA}^2$ )	84	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	0.335	Depositor
Minimum map value	-0.175	Depositor
Average map value	0.001	Depositor
Map value standard deviation	0.010	Depositor
Recommended contour level	0.07	Depositor
Map size (Å)	255.2312, 255.2312, 255.2312	wwPDB
Map dimensions	238, 238, 238	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.0724, 1.0724, 1.0724	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	0/2104	0.53	0/2715
2	B	0.38	1/2431 (0.0%)	0.64	0/3310
3	C	0.38	0/2315	0.60	0/3143
3	D	0.45	0/2601	0.61	2/3532 (0.1%)
3	E	0.50	0/2608	0.62	0/3540
3	F	0.50	0/2608	0.62	0/3540
3	G	0.50	0/2604	0.63	0/3533
3	H	0.47	0/2604	0.62	1/3533 (0.0%)
4	K	0.29	0/677	0.57	0/918
5	L	0.24	0/757	0.51	0/946
6	M	0.87	0/1420	1.15	8/2212 (0.4%)
All	All	0.47	1/22729 (0.0%)	0.66	11/30922 (0.0%)

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
3	D	0	1
3	E	0	1
3	G	0	2
3	H	0	1
All	All	0	5

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	152	PHE	C-N	5.80	1.45	1.34

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	M	1	C	O4'-C1'-N1	10.52	116.62	108.20
6	M	21	U	C2-N1-C1'	7.06	126.18	117.70
3	H	76	LEU	CB-CG-CD2	-6.85	99.35	111.00
6	M	1	C	N1-C2-O2	5.86	122.41	118.90
6	M	43	U	C2-N1-C1'	5.72	124.56	117.70

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	D	267	ILE	Peptide
3	E	285	TYR	Peptide
3	G	300	GLN	Peptide
3	G	31	ALA	Peptide
3	H	285	TYR	Peptide

## 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	2079	0	1094	16	0
2	B	2374	0	2345	47	0
3	C	2272	0	2232	34	0
3	D	2554	0	2522	32	0
3	E	2561	0	2542	31	0
3	F	2561	0	2542	35	0
3	G	2557	0	2542	36	0
3	H	2557	0	2542	32	0
4	K	668	0	615	21	0
5	L	758	0	209	16	0
6	M	1272	0	644	20	0
All	All	22213	0	19829	273	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 6.

The worst 5 of 273 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)
4:K:1:MET:CE	4:K:4:GLN:NE2	1.82	1.42
5:L:0:ALA:CB	5:L:170:GLU:H	1.31	1.39
4:K:1:MET:HE3	4:K:4:GLN:NE2	0.99	1.30
4:K:1:MET:HE3	4:K:4:GLN:CD	1.55	1.26
4:K:1:MET:CE	4:K:4:GLN:HE22	1.44	1.24

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	422/436 (97%)	357 (85%)	64 (15%)	1 (0%)	47	79
2	B	301/329 (92%)	273 (91%)	28 (9%)	0	100	100
3	C	287/344 (83%)	255 (89%)	31 (11%)	1 (0%)	41	74
3	D	331/344 (96%)	299 (90%)	31 (9%)	1 (0%)	41	74
3	E	332/344 (96%)	308 (93%)	23 (7%)	1 (0%)	41	74
3	F	332/344 (96%)	299 (90%)	33 (10%)	0	100	100
3	G	331/344 (96%)	306 (92%)	24 (7%)	1 (0%)	41	74
3	H	331/344 (96%)	299 (90%)	32 (10%)	0	100	100
4	K	85/92 (92%)	72 (85%)	13 (15%)	0	100	100
5	L	187/189 (99%)	153 (82%)	34 (18%)	0	100	100
All	All	2939/3110 (94%)	2621 (89%)	313 (11%)	5 (0%)	50	79

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	D	250	ARG
3	G	250	ARG
3	C	250	ARG

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	E	250	ARG
1	A	34	PRO

### 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	78/366 (21%)	78 (100%)	0	100	100
2	B	245/271 (90%)	245 (100%)	0	100	100
3	C	230/274 (84%)	230 (100%)	0	100	100
3	D	258/274 (94%)	258 (100%)	0	100	100
3	E	259/274 (94%)	259 (100%)	0	100	100
3	F	259/274 (94%)	259 (100%)	0	100	100
3	G	259/274 (94%)	259 (100%)	0	100	100
3	H	259/274 (94%)	259 (100%)	0	100	100
4	K	68/72 (94%)	68 (100%)	0	100	100
All	All	1915/2353 (81%)	1915 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 25 such sidechains are listed below:

Mol	Chain	Res	Type
3	E	329	ASN
3	G	208	HIS
4	K	45	GLN
3	G	164	HIS
3	H	72	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
6	M	59/60 (98%)	36 (61%)	1 (1%)

5 of 36 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
6	M	2	U
6	M	3	A
6	M	5	G
6	M	7	A
6	M	9	U

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
6	M	16	C

## 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.



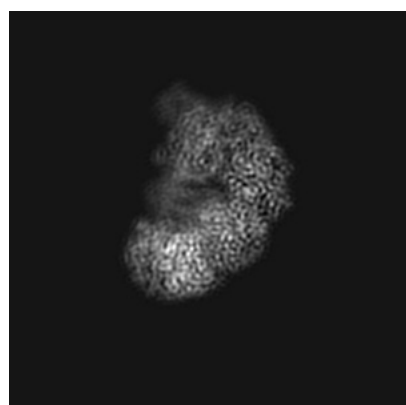
## 6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-7051. These allow visual inspection of the internal detail of the map and identification of artifacts.

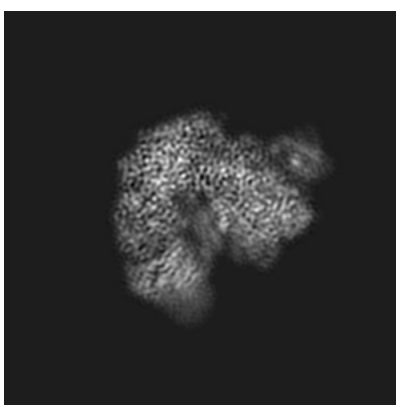
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

### 6.1 Orthogonal projections [i](#)

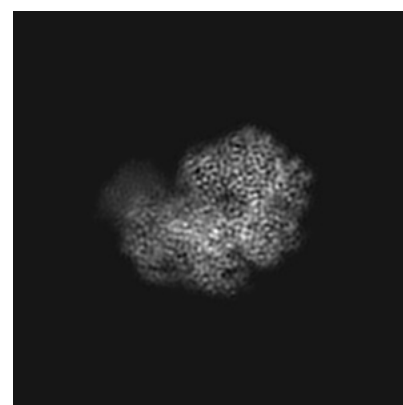
#### 6.1.1 Primary map



X



Y

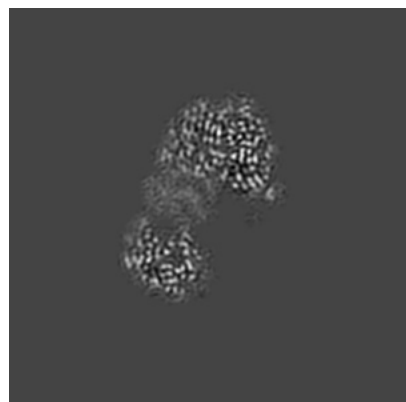


Z

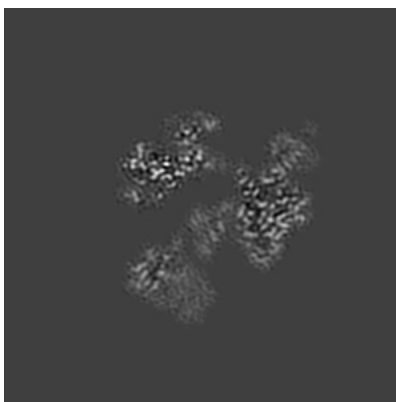
The images above show the map projected in three orthogonal directions.

### 6.2 Central slices [i](#)

#### 6.2.1 Primary map



X Index: 119



Y Index: 119

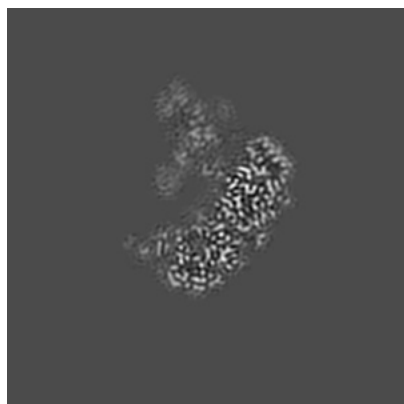


Z Index: 119

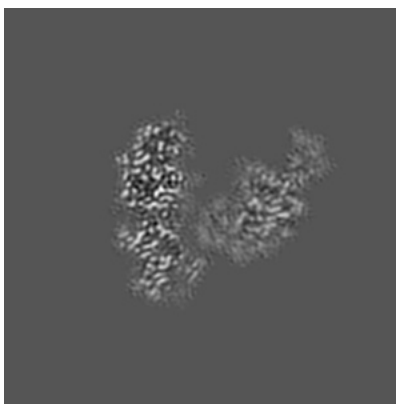
The images above show central slices of the map in three orthogonal directions.

## 6.3 Largest variance slices [i](#)

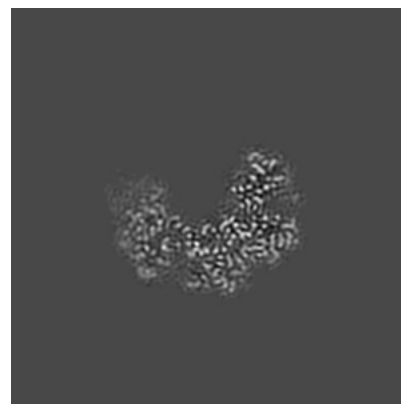
### 6.3.1 Primary map



X Index: 140



Y Index: 106

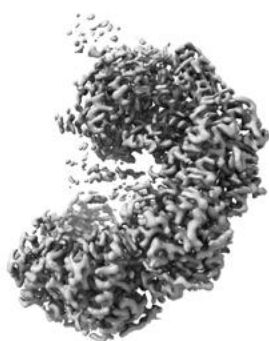


Z Index: 95

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

## 6.4 Orthogonal surface views [i](#)

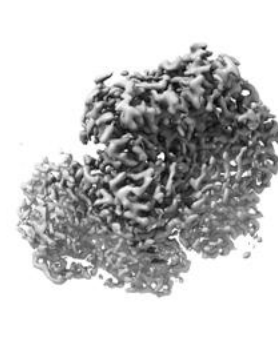
### 6.4.1 Primary map



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.

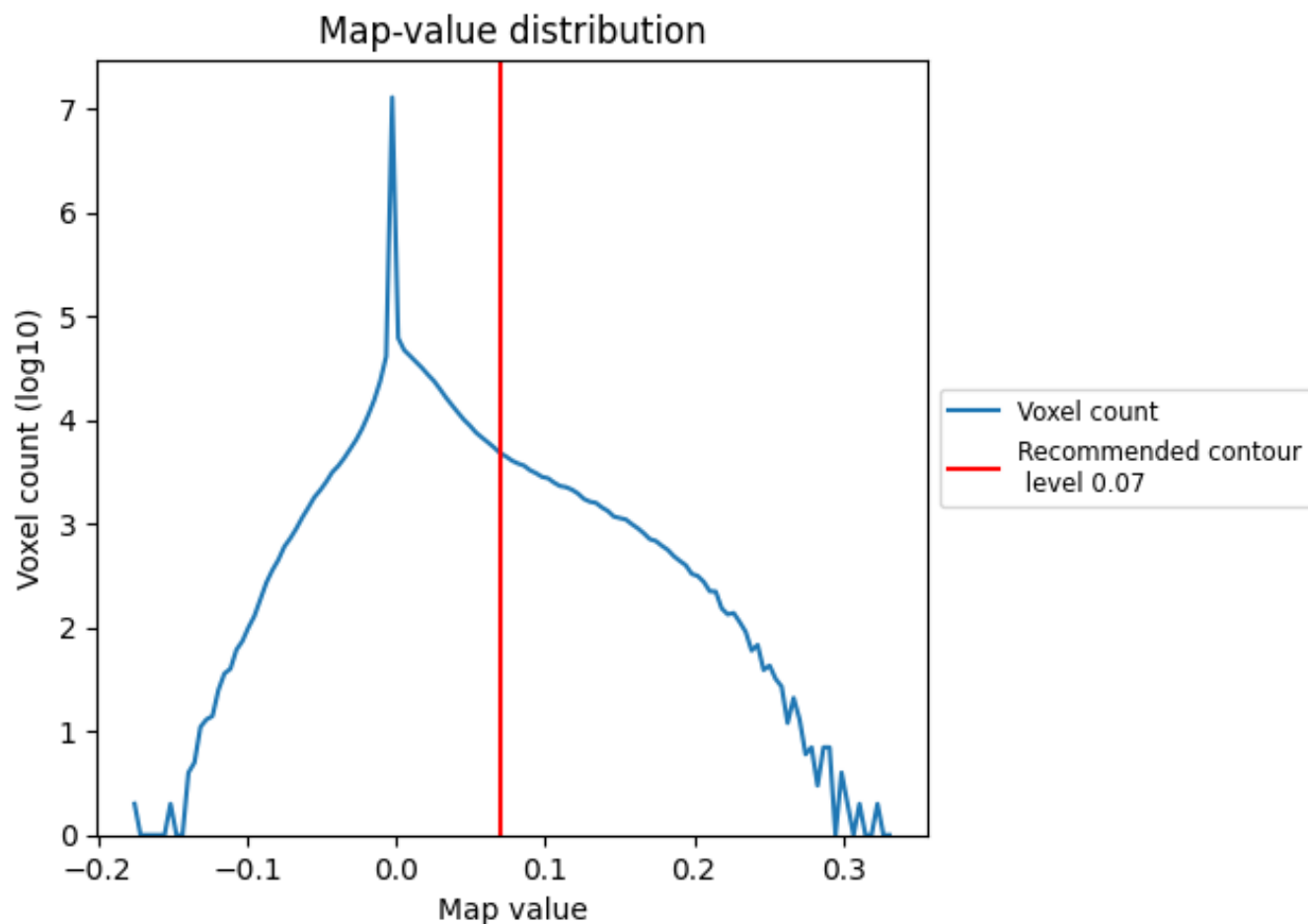
## 6.5 Mask visualisation

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

## 7 Map analysis [i](#)

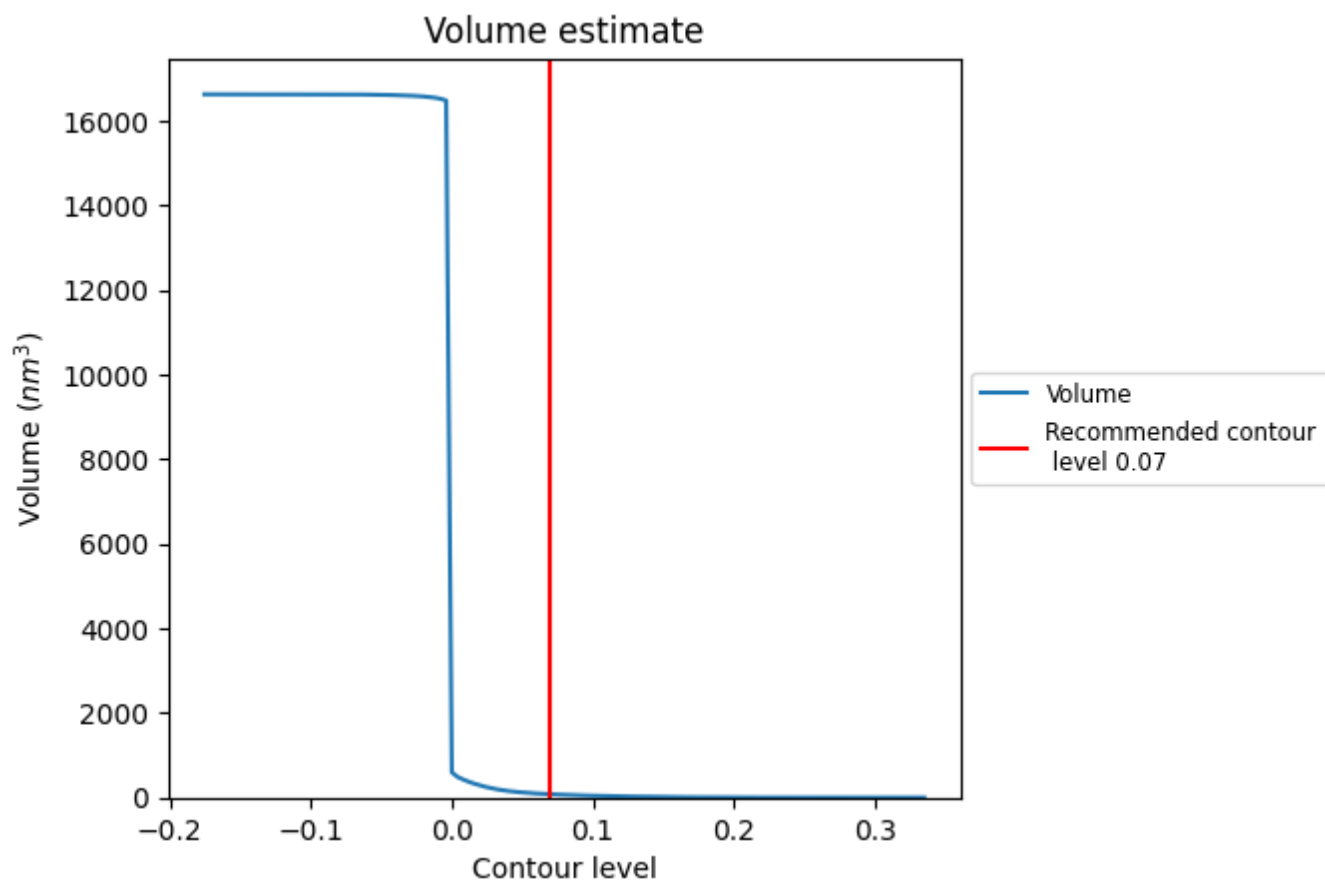
This section contains the results of statistical analysis of the map.

### 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.

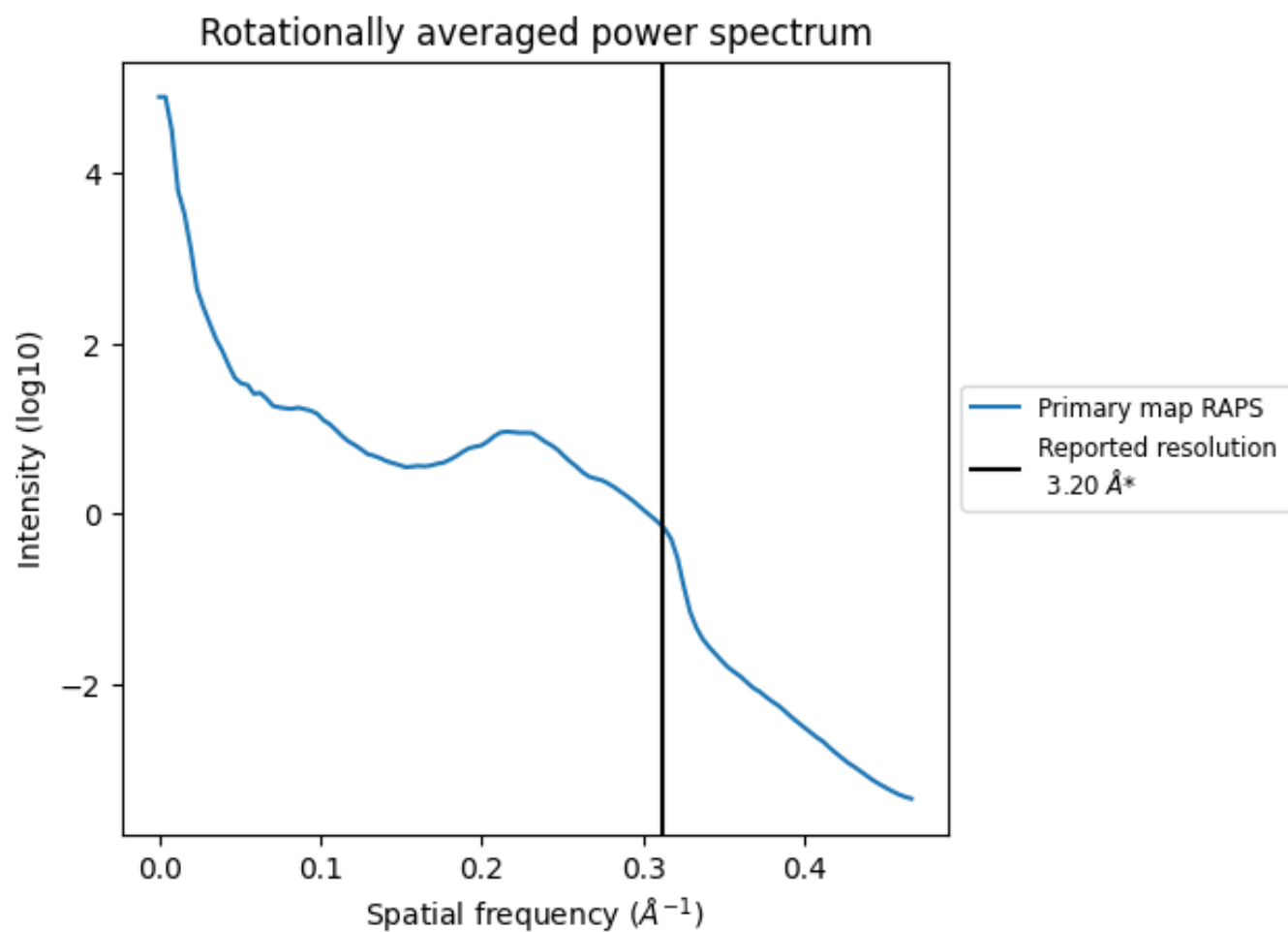
## 7.2 Volume estimate [i](#)



The volume at the recommended contour level is 79  $\text{nm}^3$ ; this corresponds to an approximate mass of 71 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<sup>-1</sup>

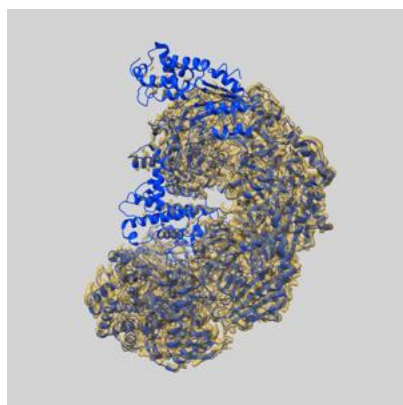
## 8 Fourier-Shell correlation

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

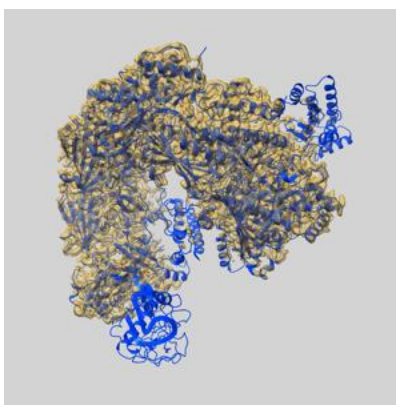
## 9 Map-model fit [i](#)

This section contains information regarding the fit between EMDB map EMD-7051 and PDB model 6B47. Per-residue inclusion information can be found in section [3](#) on page [7](#).

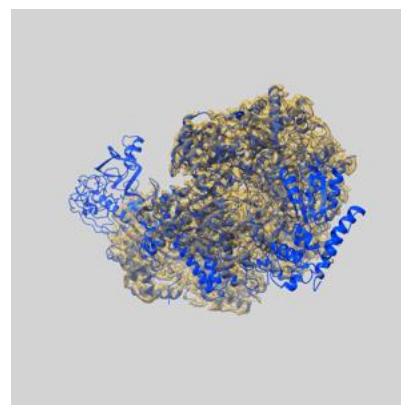
### 9.1 Map-model overlay [i](#)



X



Y

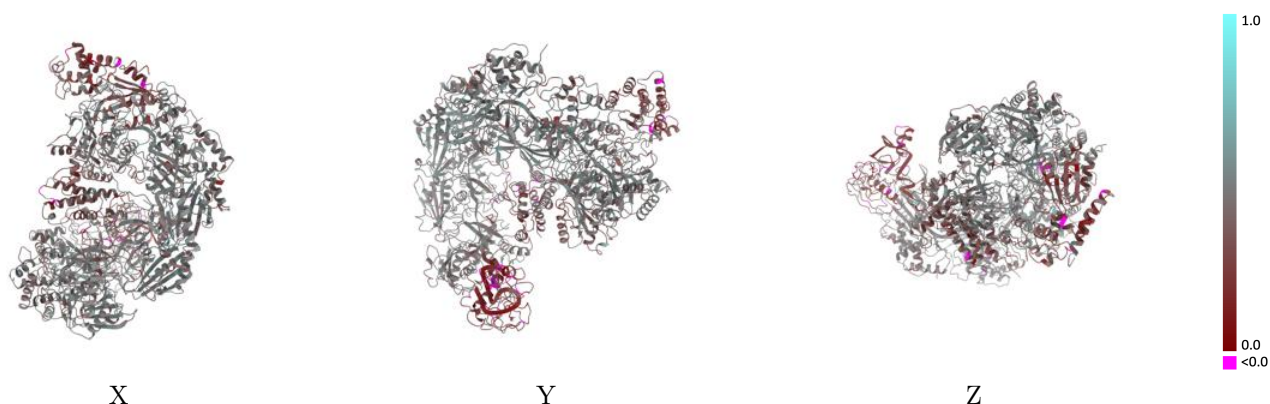


Z

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.

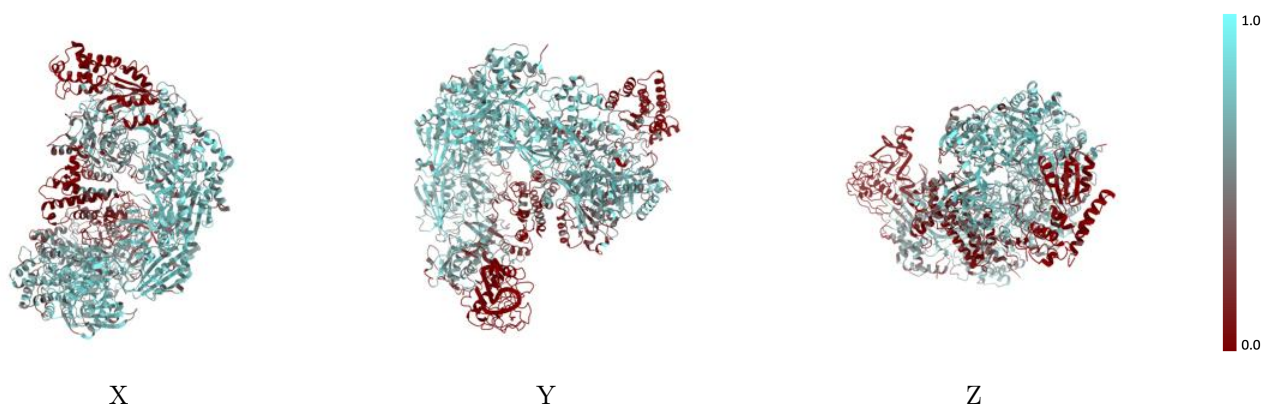


## 9.2 Q-score mapped to coordinate model [i](#)



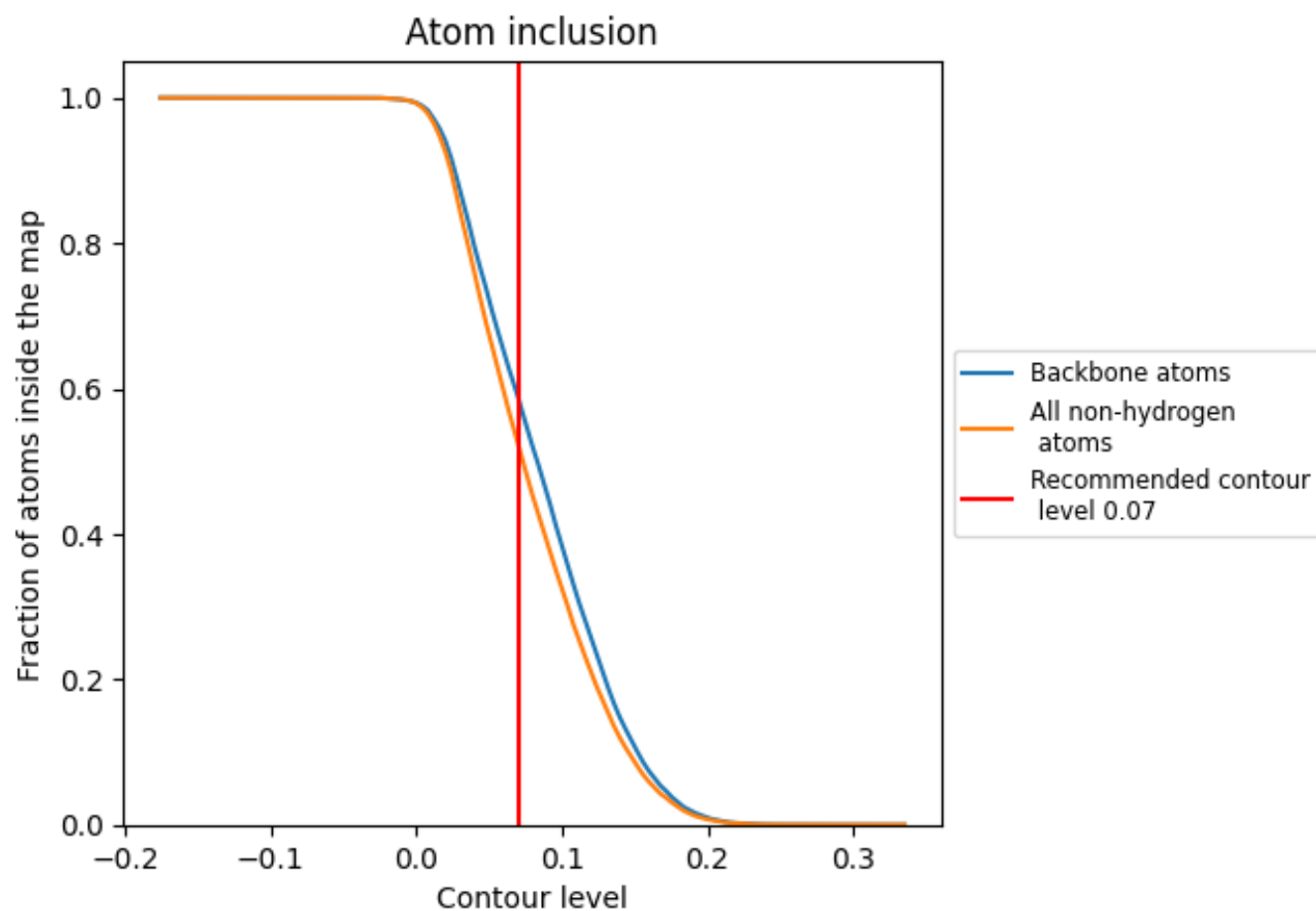
The images above show the model with each residue coloured according its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions 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, 59% of all backbone atoms, 52% 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	<div></div> 0.5235	<div></div> 0.4250
A	<div></div> 0.2646	<div></div> 0.3770
B	<div></div> 0.4288	<div></div> 0.3980
C	<div></div> 0.4993	<div></div> 0.4330
D	<div></div> 0.6123	<div></div> 0.4500
E	<div></div> 0.6545	<div></div> 0.4590
F	<div></div> 0.6693	<div></div> 0.4670
G	<div></div> 0.6664	<div></div> 0.4610
H	<div></div> 0.6279	<div></div> 0.4610
K	<div></div> 0.0410	<div></div> 0.2750
L	<div></div> 0.0026	<div></div> 0.2310
M	<div></div> 0.5142	<div></div> 0.3810

1.0

0.0

<0.0