



## wwPDB/EMDataBank EM Map/Model Validation Summary Report ⓘ

Sep 7, 2019 – 02:09 PM EDT

PDB ID : 6O7H  
EMDB ID: : EMD-0641  
Title : Cryo-EM structure of Csm-crRNA-target RNA ternary complex in complex with cA4 in type III-A CRISPR-Cas system  
Authors : Jia, N.; Patel, D.J.  
Deposited on : 2019-03-07  
Resolution : 2.90 Å(reported)

This is a wwPDB/EMDataBank EM Map/Model Validation Summary Report for a publicly released PDB/EMDB 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.

---

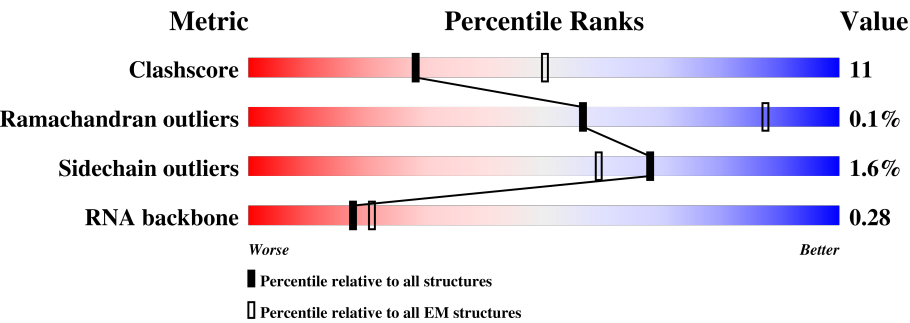
MolProbity : 4.02b-467  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.4

# 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 2.90 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	136327	1886
Ramachandran outliers	132723	1663
Sidechain outliers	132532	1531
RNA backbone	3747	458

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	791	<div><div>76%</div><div>21%</div><div>..</div></div>
2	B	188	<div><div>70%</div><div>10%</div><div>•</div><div>20%</div></div>
3	C	292	<div><div>75%</div><div>20%</div><div>•</div><div>5%</div></div>
3	D	292	<div><div>75%</div><div>23%</div><div>•</div></div>
4	E	289	<div><div>82%</div><div>16%</div><div>..</div></div>
5	G	38	<div><div>37%</div><div>24%</div><div>11%</div><div>29%</div></div>
6	H	40	<div><div>5%</div><div>35%</div><div>8%</div><div>53%</div></div>
7	F	378	<div><div>82%</div><div>13%</div><div>•</div></div>

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
8	K	4	 100%

## 2 Entry composition

There are 10 unique types of molecules in this entry. The entry contains 17576 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 system single-strand-specific deoxyribonuclease Cas10/Csm1 (subtype III-A).

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	771	Total	C	N	O	S	0	0
			6152	3943	1073	1117	19		

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-13	MET	-	initiating methionine	UNP B6YWB8
A	-12	GLY	-	expression tag	UNP B6YWB8
A	-11	SER	-	expression tag	UNP B6YWB8
A	-10	SER	-	expression tag	UNP B6YWB8
A	-9	HIS	-	expression tag	UNP B6YWB8
A	-8	HIS	-	expression tag	UNP B6YWB8
A	-7	HIS	-	expression tag	UNP B6YWB8
A	-6	HIS	-	expression tag	UNP B6YWB8
A	-5	HIS	-	expression tag	UNP B6YWB8
A	-4	HIS	-	expression tag	UNP B6YWB8
A	-3	SER	-	expression tag	UNP B6YWB8
A	-2	GLN	-	expression tag	UNP B6YWB8
A	-1	ASP	-	expression tag	UNP B6YWB8
A	0	PRO	-	expression tag	UNP B6YWB8

- Molecule 2 is a protein called Csm2.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	150	Total	C	N	O	S	0	0
			1047	668	183	192	4		

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-1	GLY	-	expression tag	UNP B6YWB9
B	0	SER	-	expression tag	UNP B6YWB9

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	VAL	-	expression tag	UNP B6YWB9

- Molecule 3 is a protein called Csm3.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	278	Total	C	N	O	S	0	0
			2214	1406	389	413	6		
3	D	285	Total	C	N	O	S	0	0
			2265	1437	396	425	7		

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-1	GLY	-	expression tag	UNP B6YWC0
C	0	SER	-	expression tag	UNP B6YWC0
C	36	ALA	ASP	conflict	UNP B6YWC0
D	-1	GLY	-	expression tag	UNP B6YWC0
D	0	SER	-	expression tag	UNP B6YWC0
D	36	ALA	ASP	conflict	UNP B6YWC0

- Molecule 4 is a protein called Csm4.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	E	286	Total	C	N	O	S	0	0
			2261	1467	380	410	4		

- Molecule 5 is a RNA chain called RNA (38-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
5	G	27	Total	C	N	O	P	0	0
			589	262	115	186	26		

- Molecule 6 is a RNA chain called RNA (40-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
6	H	19	Total	C	N	O	P	0	0
			398	178	69	132	19		

- Molecule 7 is a protein called Csm5.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	F	361	Total	C	N	O	S	0	0
			2536	1603	453	471	9		

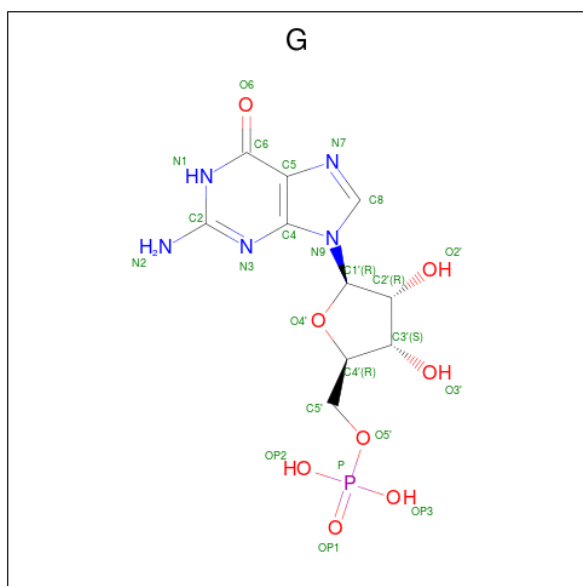
- Molecule 8 is a RNA chain called Cyclic DNA cA4.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	K	4	Total	C	N	O	P	0	0
			88	40	20	24	4		

- Molecule 9 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		AltConf
9	A	1	Total	Zn	0
			1	1	
9	D	1	Total	Zn	0
			1	1	
9	C	1	Total	Zn	0
			1	1	

- Molecule 10 is GUANOSINE-5'-MONOPHOSPHATE (three-letter code: G) (formula: C<sub>10</sub>H<sub>14</sub>N<sub>5</sub>O<sub>8</sub>P).

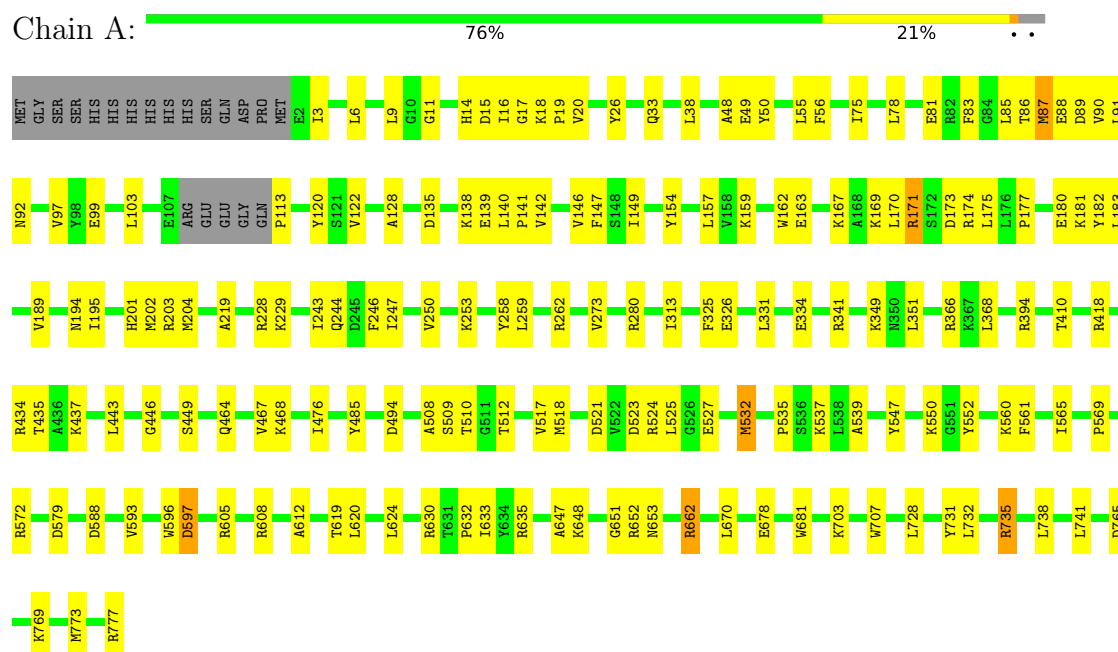


Mol	Chain	Residues	Atoms					AltConf
10	H	1	Total	C	N	O	P	0
			23	10	5	7	1	

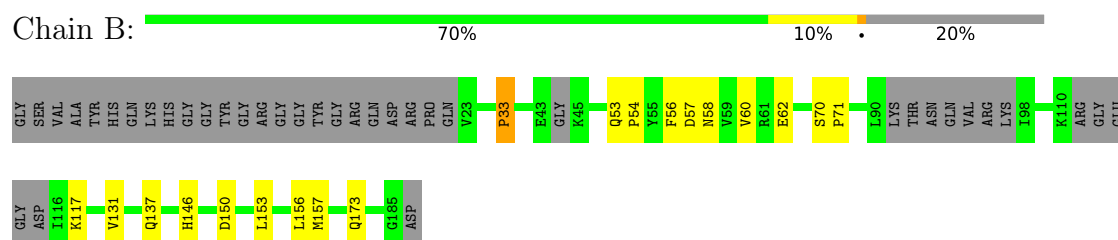
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA 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. 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. 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.

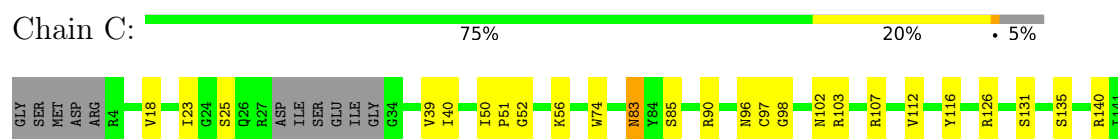
- Molecule 1: CRISPR system single-strand-specific deoxyribonuclease Cas10/Csm1 (subtype III-A)



- Molecule 2: Csm2



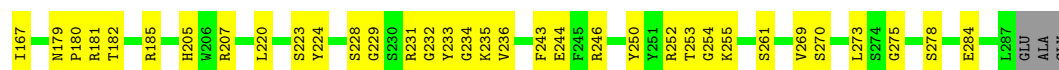
- Molecule 3: Csm3





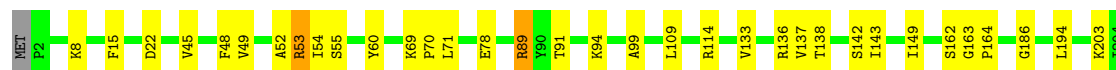
### • Molecule 3: Csm3

Chain D: 75% 23%



### • Molecule 4: Csm4

Chain E: 82% 16%



### • Molecule 5: RNA (38-MER)

Chain G: 37% 24% 11% 29%



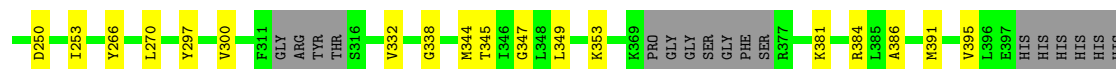
### • Molecule 6: RNA (40-MER)

Chain H: 5% 35% 8% 53%



### • Molecule 7: Csm5

Chain F: 82% 13%



### • Molecule 8: Cyclic DNA cA4

Chain K: 100%





## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	151715	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 ( $e^-/\text{\AA}^2$ )	1.35	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	$\# Z  > 2$	RMSZ	$\# Z  > 2$
1	A	0.37	0/6293	0.57	1/8493 (0.0%)
2	B	0.26	0/1059	0.55	3/1429 (0.2%)
3	C	0.35	0/2263	0.58	0/3051
3	D	0.35	0/2314	0.59	0/3119
4	E	0.32	0/2322	0.55	0/3146
5	G	0.42	0/661	0.84	0/1033
6	H	0.28	0/442	0.94	0/685
7	F	0.26	0/2122	0.55	0/2875
8	K	4.30	28/99 (28.3%)	7.27	56/152 (36.8%)
All	All	0.47	28/17575 (0.2%)	0.83	60/23983 (0.3%)

The worst 5 of 28 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
8	K	3	A	N7-C5	-10.78	1.32	1.39
8	K	4	A	N7-C5	-10.75	1.32	1.39
8	K	1	A	N7-C5	-10.13	1.33	1.39
8	K	2	A	N7-C5	-9.46	1.33	1.39
8	K	3	A	C5-C6	-9.15	1.32	1.41

The worst 5 of 60 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
8	K	3	A	O5'-P-OP2	-22.67	83.50	110.70
8	K	3	A	N1-C2-N3	-22.59	118.01	129.30
8	K	2	A	N1-C2-N3	-22.30	118.15	129.30
8	K	4	A	N1-C2-N3	-22.30	118.15	129.30
8	K	1	A	N1-C2-N3	-22.27	118.16	129.30

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	6152	0	6141	158	0
2	B	1047	0	939	9	0
3	C	2214	0	2184	67	0
3	D	2265	0	2236	76	0
4	E	2261	0	2265	53	0
5	G	589	0	295	12	0
6	H	398	0	207	12	0
7	F	2536	0	2139	33	0
8	K	88	0	44	25	0
9	A	1	0	0	0	0
9	C	1	0	0	0	0
9	D	1	0	0	0	0
10	H	23	0	12	2	0
All	All	17576	0	16462	388	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 11.

The worst 5 of 388 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:E:218:ALA:HB3	4:E:247:ILE:CG1	1.64	1.26
1:A:183:LEU:HD12	1:A:202:MET:HE1	1.24	1.16
4:E:52:ALA:HB2	4:E:163:GLY:CA	1.77	1.15
1:A:334:GLU:OE2	1:A:366:ARG:HD3	1.45	1.14
4:E:52:ALA:HB2	4:E:163:GLY:HA3	1.18	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	767/791 (97%)	729 (95%)	38 (5%)	0	100	100
2	B	142/188 (76%)	128 (90%)	12 (8%)	2 (1%)	12	39
3	C	274/292 (94%)	256 (93%)	18 (7%)	0	100	100
3	D	281/292 (96%)	265 (94%)	16 (6%)	0	100	100
4	E	284/289 (98%)	271 (95%)	13 (5%)	0	100	100
7	F	259/378 (68%)	249 (96%)	10 (4%)	0	100	100
All	All	2007/2230 (90%)	1898 (95%)	107 (5%)	2 (0%)	56	84

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	33	PRO
2	B	70	SER

### 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	643/665 (97%)	630 (98%)	13 (2%)	58	85
2	B	85/154 (55%)	85 (100%)	0	100	100
3	C	237/247 (96%)	233 (98%)	4 (2%)	63	88
3	D	243/247 (98%)	237 (98%)	6 (2%)	50	81
4	E	237/240 (99%)	235 (99%)	2 (1%)	83	95

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
7	F	212/245 (86%)	211 (100%)	1 (0%)	90	97
All	All	1657/1798 (92%)	1631 (98%)	26 (2%)	68	88

5 of 26 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	735	ARG
3	C	181	ARG
4	E	89	ARG
1	A	777	ARG
3	C	83	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	179	ASN
3	D	106	ASN
3	D	212	ASN
3	C	83	ASN
3	D	179	ASN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
5	G	26/38 (68%)	8 (30%)	2 (7%)
6	H	18/40 (45%)	13 (72%)	0
8	K	4/4 (100%)	3 (75%)	2 (50%)
All	All	48/82 (58%)	24 (50%)	4 (8%)

5 of 24 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
5	G	-7	U
5	G	-2	A
5	G	2	G
5	G	5	C
5	G	6	A

All (4) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
5	G	5	C
5	G	11	C
8	K	1	A
8	K	2	A

## 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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 3 are monoatomic - leaving 1 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
7	F	8

The worst 5 of 8 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	F	169:UNK	C	175:UNK	N	19.01
1	F	91:UNK	C	95:ARG	N	8.94
1	F	143:UNK	C	150:UNK	N	8.23
1	F	191:UNK	C	197:ILE	N	6.64
1	F	156:UNK	C	159:UNK	N	6.27