



wwPDB/EMDataBank EM Map/Model Validation Summary Report ⓘ

Aug 7, 2017 – 05:49 PM EDT

PDB ID : 5U0A
EMDB ID: : EMD-8478
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 : unknown
Resolution : 3.30 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
A user guide is available at
<http://wwpdb.org/validation/2016/EMValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

MolProbity : 4.02b-467
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20029824

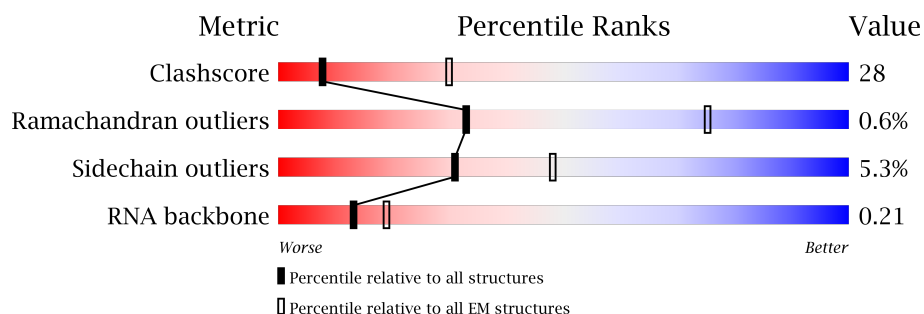
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 3.30 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	125131	1336
Ramachandran outliers	121729	1120
Sidechain outliers	121581	1026
RNA backbone	3398	335

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 ≥ 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	232	30% 42% 9% . 17%
2	C	549	41% 42% 8% . 8%
3	D	373	39% 28% . 30%
3	E	373	64% 33% . .
3	F	373	66% 32% . .
3	G	373	66% 31% . .
3	H	373	67% 30% . .
3	I	373	61% 26% . 12%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	J	244	<div><div></div><div>48%21%•29%</div></div>
4	L	244	<div><div></div><div>45%21%•32%</div></div>
5	K	61	<div><div></div><div>13%49%31%••</div></div>
6	M	50	<div><div></div><div>38%58%•</div></div>
7	N	254	<div><div></div><div>49%40%6%5%</div></div>
8	O	39	<div><div></div><div>38%46%5%10%</div></div>

2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 28923 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	192	Total	C	N	O	S	0	0
			1439	886	290	261	2		

- Molecule 2 is a protein called CRISPR-associated protein, Cse1 family.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	C	503	Total	C	N	O	S	0	0
			3980	2523	732	716	9		

- Molecule 3 is a protein called CRISPR-associated protein, Cse4 family.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	D	260	Total	C	N	O	S	0	0
			2017	1266	356	390	5		
3	E	367	Total	C	N	O	S	0	0
			2840	1781	509	545	5		
3	F	367	Total	C	N	O	S	0	0
			2840	1781	509	545	5		
3	G	366	Total	C	N	O	S	0	0
			2829	1772	508	544	5		
3	H	366	Total	C	N	O	S	0	0
			2829	1772	508	544	5		
3	I	329	Total	C	N	O	S	0	0
			2543	1596	460	484	3		

- Molecule 4 is a protein called Cse2.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	J	173	Total	C	N	O	S	0	0
			1388	873	278	235	2		
4	L	165	Total	C	N	O	S	0	0
			1322	832	264	224	2		

- 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	50	Total	C	N	O	P	0	0
			1013	481	179	303	50		

- 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
			1896	1204	346	343	3		

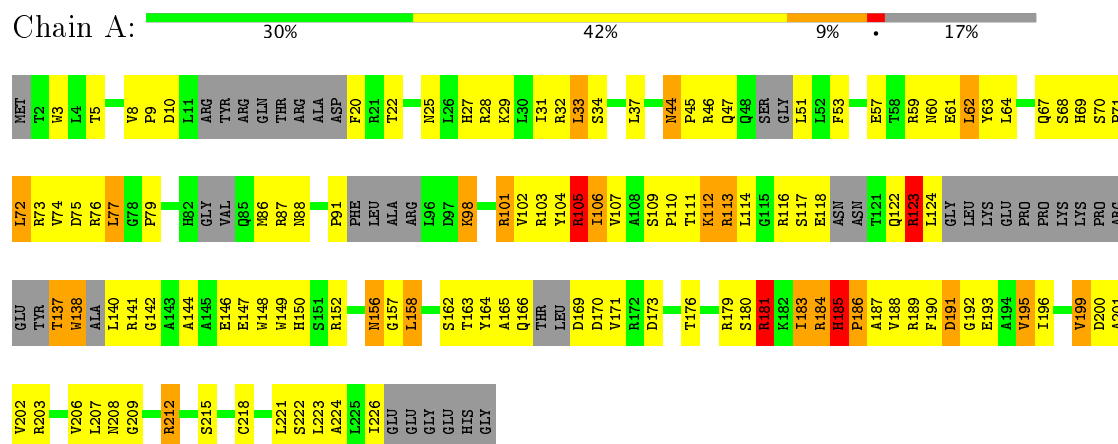
- Molecule 8 is a DNA chain called Nontarget Strand.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	O	35	Total	C	N	O	P	0	0
			720	338	139	208	35		

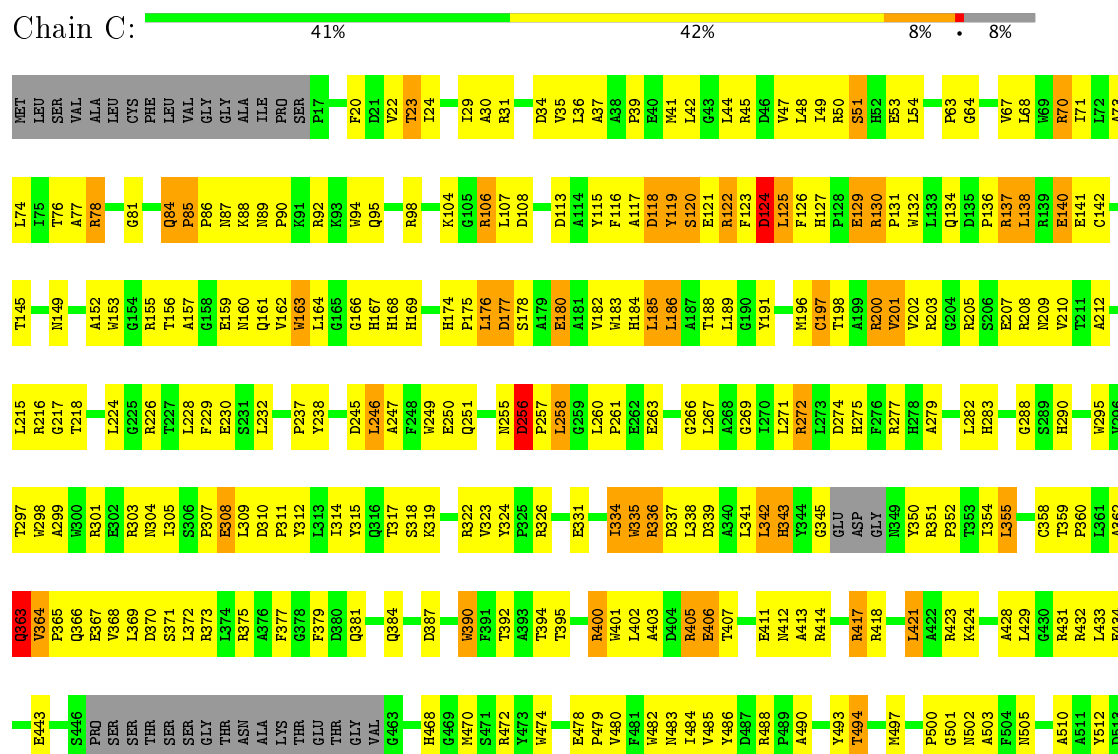
3 Residue-property plots

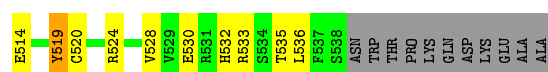
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according 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-associated protein, Cse3 family

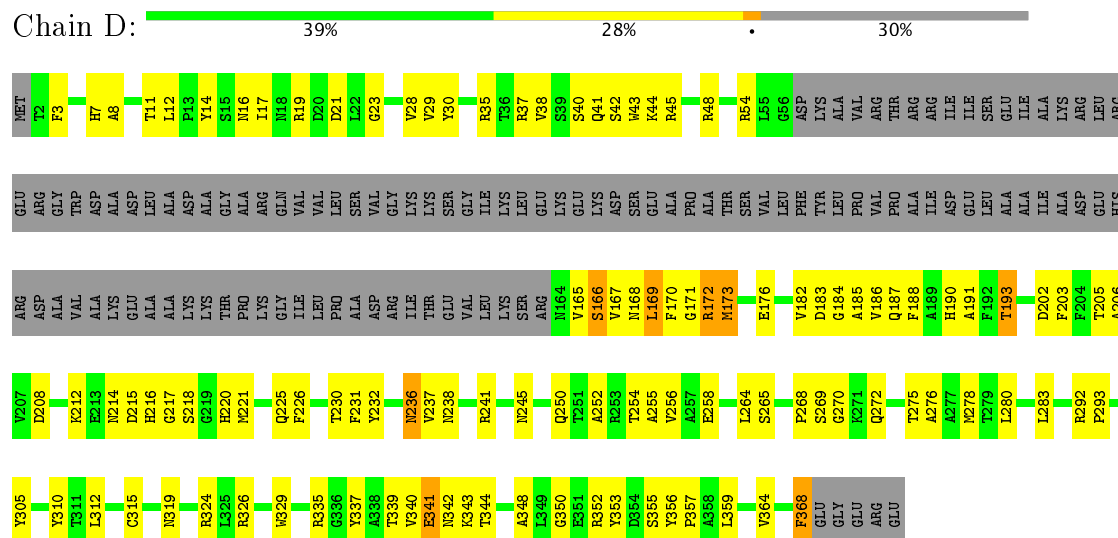


- Molecule 2: CRISPR-associated protein, Cse1 family

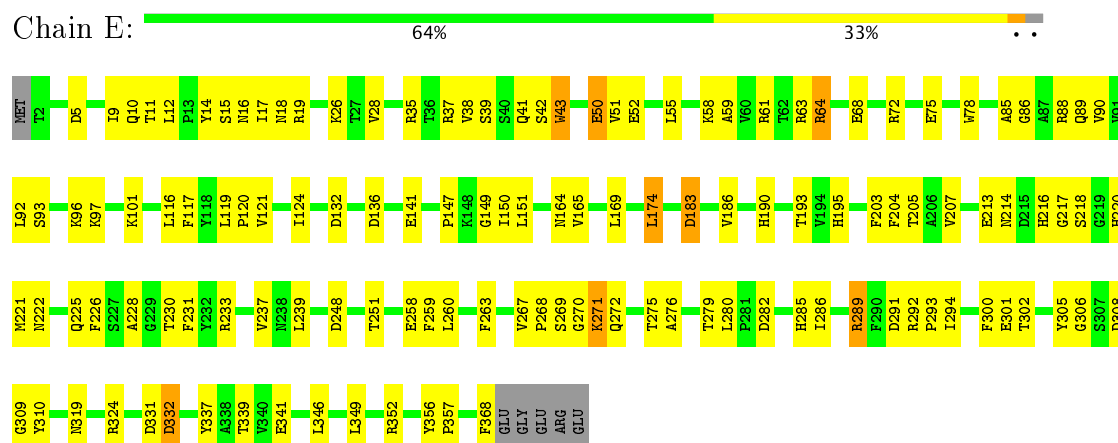




- Molecule 3: CRISPR-associated protein, Cse4 family



- Molecule 3: CRISPR-associated protein, Cse4 family

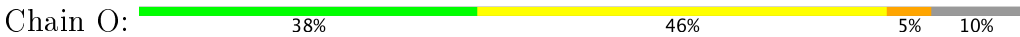


- Molecule 3: CRISPR-associated protein, Cse4 family



CYS
ALA
GLU
GLU
ALA
ALA

● Molecule 8: Nontarget Strand



4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	131292	Depositor
Resolution determination method	FSC 0.143 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)	21000	Depositor
Magnification	31000	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	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 >2	RMSZ	# Z >2
1	A	0.42	0/1456	1.06	10/1960 (0.5%)
2	C	0.54	0/4097	0.84	7/5592 (0.1%)
3	D	0.72	0/2060	0.81	2/2802 (0.1%)
3	E	0.64	0/2894	0.66	0/3927
3	F	0.64	0/2894	0.63	0/3927
3	G	0.65	0/2882	0.64	0/3911
3	H	0.65	0/2882	0.64	0/3911
3	I	0.65	0/2587	0.63	0/3509
4	J	0.56	0/1417	0.78	3/1924 (0.2%)
4	L	0.53	0/1348	0.75	1/1828 (0.1%)
5	K	1.04	1/1417 (0.1%)	1.14	7/2208 (0.3%)
6	M	1.15	0/1132	0.99	3/1742 (0.2%)
7	N	0.64	0/1939	0.80	1/2638 (0.0%)
8	O	0.90	3/807 (0.4%)	0.92	1/1241 (0.1%)
All	All	0.68	4/29812 (0.0%)	0.79	35/41120 (0.1%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
8	O	43	DC	C1'-N1	5.94	1.56	1.49
5	K	55	C	C1'-N1	5.59	1.57	1.48
8	O	35	DT	C1'-N1	5.37	1.56	1.49
8	O	31	DC	C1'-N1	5.32	1.56	1.49

The worst 5 of 35 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	185	HIS	C-N-CD	-20.38	75.77	120.60
1	A	158	LEU	CA-CB-CG	8.54	134.94	115.30
6	M	50	DA	O4'-C1'-N9	7.84	113.48	108.00
4	J	13	LEU	CA-CB-CG	7.76	133.14	115.30
1	A	72	LEU	CA-CB-CG	7.15	131.74	115.30

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	1439	0	1372	223	0
2	C	3980	0	3851	416	0
3	D	2017	0	1937	143	0
3	E	2840	0	2809	141	0
3	F	2840	0	2809	108	0
3	G	2829	0	2800	105	0
3	H	2829	0	2800	101	0
3	I	2543	0	2541	107	0
4	J	1388	0	1409	61	0
4	L	1322	0	1337	57	0
5	K	1267	0	642	132	0
6	M	1013	0	562	46	0
7	N	1896	0	1925	150	0
8	O	720	0	391	28	0
All	All	28923	0	27185	1580	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 1580 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:C:142:CYS:SG	2:C:200:ARG:HB3	1.46	1.55
3:D:169:LEU:HD11	3:D:170:PHE:CE2	1.54	1.42
1:A:122:GLN:NE2	5:K:44:A:N1	1.67	1.40
2:C:429:LEU:HA	2:C:432:ARG:CG	1.53	1.35
3:D:169:LEU:CD1	3:D:170:PHE:CE2	2.07	1.35

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	174/232 (75%)	152 (87%)	13 (8%)	9 (5%)	2	16
2	C	497/549 (90%)	462 (93%)	29 (6%)	6 (1%)	15	50
3	D	256/373 (69%)	236 (92%)	19 (7%)	1 (0%)	38	71
3	E	365/373 (98%)	352 (96%)	13 (4%)	0	100	100
3	F	365/373 (98%)	350 (96%)	15 (4%)	0	100	100
3	G	364/373 (98%)	350 (96%)	13 (4%)	1 (0%)	44	76
3	H	364/373 (98%)	349 (96%)	15 (4%)	0	100	100
3	I	323/373 (87%)	314 (97%)	9 (3%)	0	100	100
4	J	167/244 (68%)	162 (97%)	4 (2%)	1 (1%)	28	63
4	L	157/244 (64%)	152 (97%)	5 (3%)	0	100	100
7	N	237/254 (93%)	220 (93%)	17 (7%)	0	100	100
All	All	3269/3761 (87%)	3099 (95%)	152 (5%)	18 (1%)	33	63

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	44	ASN
1	A	62	LEU
1	A	106	ILE
1	A	156	ASN
1	A	186	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	137/193 (71%)	125 (91%)	12 (9%)	12	40
2	C	405/452 (90%)	354 (87%)	51 (13%)	5	23
3	D	209/299 (70%)	201 (96%)	8 (4%)	38	70
3	E	294/299 (98%)	283 (96%)	11 (4%)	39	71
3	F	294/299 (98%)	286 (97%)	8 (3%)	50	77
3	G	293/299 (98%)	286 (98%)	7 (2%)	54	79
3	H	293/299 (98%)	285 (97%)	8 (3%)	50	77
3	I	263/299 (88%)	254 (97%)	9 (3%)	42	73
4	J	141/200 (70%)	136 (96%)	5 (4%)	41	72
4	L	134/200 (67%)	132 (98%)	2 (2%)	70	85
7	N	202/211 (96%)	183 (91%)	19 (9%)	10	36
All	All	2665/3050 (87%)	2525 (95%)	140 (5%)	31	62

5 of 140 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	166	SER
3	E	271	LYS
7	N	69	ARG
3	D	172	ARG
3	E	43	TRP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 44 such sidechains are listed below:

Mol	Chain	Res	Type
3	E	272	GLN
3	G	18	ASN
4	L	208	HIS
3	E	342	ASN
3	F	89	GLN

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
5	K	57/61 (93%)	27 (47%)	1 (1%)

5 of 27 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

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
5	K	44	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](#)

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.