



wwPDB EM Validation Summary Report ⓘ

May 4, 2026 – 10:34 PM JST

PDB ID : 9VJ9 / pdb_00009vj9
EMDB ID : EMD-65107
Title : Type I-A CRISPR integrase prespacer catching complex, State I
Authors : Li, Z.X.; Li, Y.T.; Lu, M.L.; Xiao, Y.B.
Deposited on : 2025-06-19
Resolution : 2.96 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 : **FAILED**
MolProbity : 4-5-2 with Phenix2.0
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
EM percentile statistics : **NOT EXECUTED**
MapQ : **FAILED**
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.49

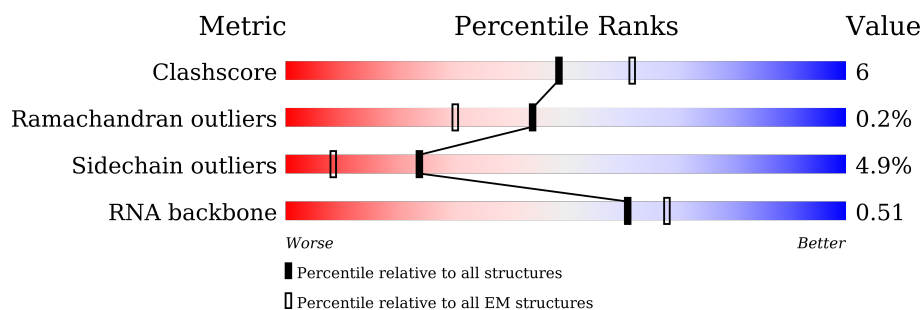
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 2.96 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	229148	23984
Ramachandran outliers	224038	23583
Sidechain outliers	223484	23102
RNA backbone	8273	3508

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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	219	79% (green) 21% (yellow)
1	B	219	81% (green) 17% (yellow) . (orange)
1	C	219	87% (green) 12% (yellow) . (orange)
1	D	219	81% (green) 18% (yellow) . (orange)
2	K	288	67% (green) 32% (yellow) .. (grey)
2	L	288	75% (green) 23% (yellow) . (orange)
3	N	51	55% (green) 33% (yellow) 12% (grey)
4	T	51	65% (green) 24% (yellow) 12% (grey)

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	g	55	 33% 15% • 51%
6	t	83	 42% 30% • 27%
7	S	1337	 82% 16% ••

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 26415 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 Type II-A CRISPR-associated protein Csn2.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	B	219	Total	C	N	O	S	0	0
			1790	1158	276	351	5		
1	C	219	Total	C	N	O	S	0	0
			1790	1158	276	351	5		
1	D	219	Total	C	N	O	S	0	0
			1790	1158	276	351	5		
1	A	219	Total	C	N	O	S	0	0
			1790	1158	276	351	5		

- Molecule 2 is a protein called CRISPR-associated endonuclease Cas1.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	K	286	Total	C	N	O	S	0	0
			2347	1508	407	423	9		
2	L	287	Total	C	N	O	S	0	0
			2351	1510	408	424	9		

- Molecule 3 is a DNA chain called DNA (51-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
3	N	45	Total	C	N	O	P	0	0
			928	442	179	262	45		

- Molecule 4 is a DNA chain called DNA (51-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
4	T	45	Total	C	N	O	P	0	0
			917	441	153	278	45		

- Molecule 5 is a RNA chain called RNA (55-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
5	g	27	Total	C	N	O	P	0	0
			566	253	88	198	27		

- Molecule 6 is a RNA chain called RNA (83-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
6	t	61	Total	C	N	O	P	0	0
			1300	582	232	425	61		

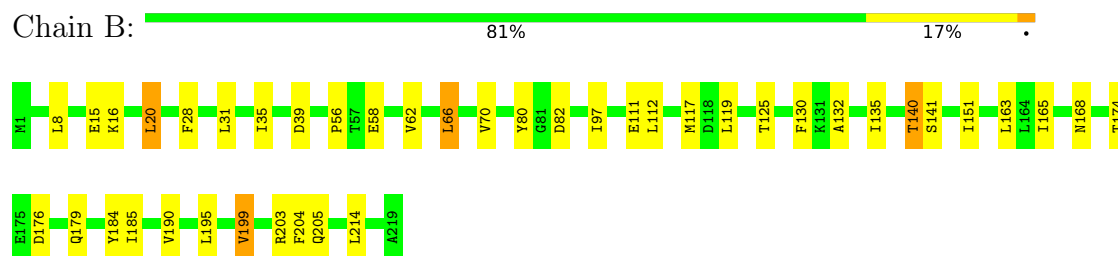
- Molecule 7 is a protein called CRISPR-associated endonuclease Cas9.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	S	1322	Total	C	N	O	S	0	0
			10846	6953	1851	2015	27		

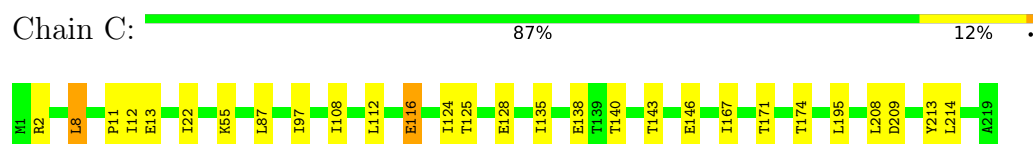
3 Residue-property plots

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. 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.

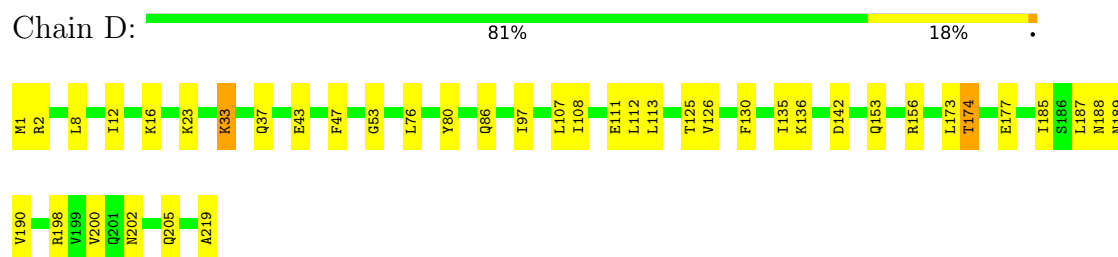
- Molecule 1: Type II-A CRISPR-associated protein Csn2



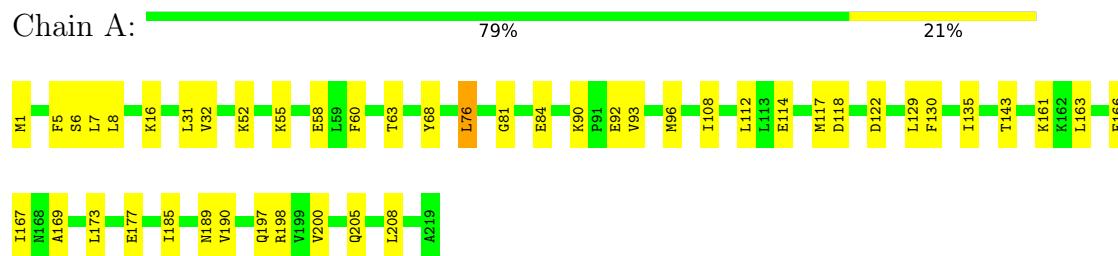
- Molecule 1: Type II-A CRISPR-associated protein Csn2



- Molecule 1: Type II-A CRISPR-associated protein Csn2

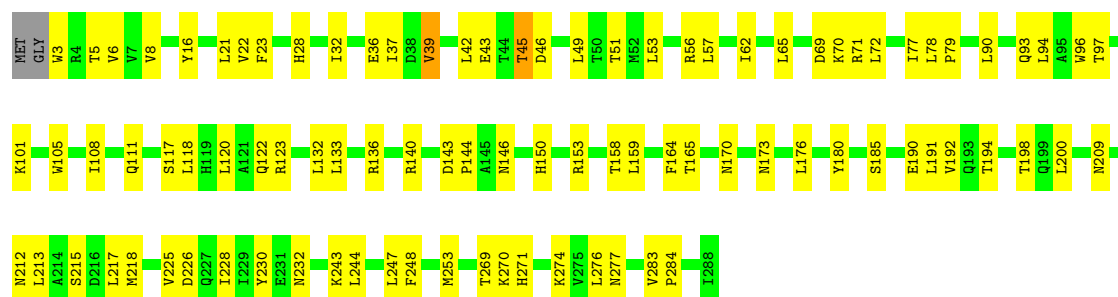


- Molecule 1: Type II-A CRISPR-associated protein Csn2




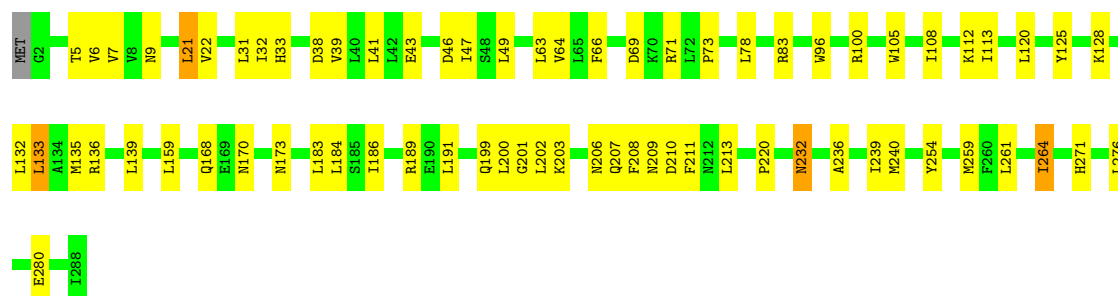
- Molecule 2: CRISPR-associated endonuclease Cas1

Chain K:  67% 32% ..



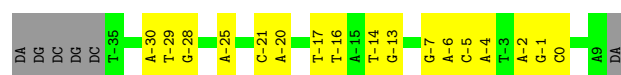
• Molecule 2: CRISPR-associated endonuclease Cas1

Chain L:  75% 23% .



• Molecule 3: DNA (51-MER)

Chain N:  55% 33% 12%




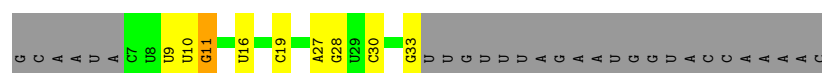
• Molecule 4: DNA (51-MER)

Chain T:  65% 24% 12%



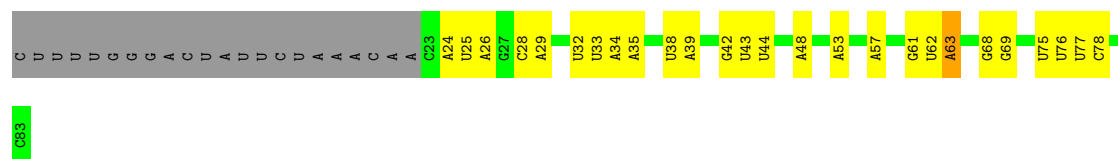
• Molecule 5: RNA (55-MER)

Chain g:  33% 15% . 51%



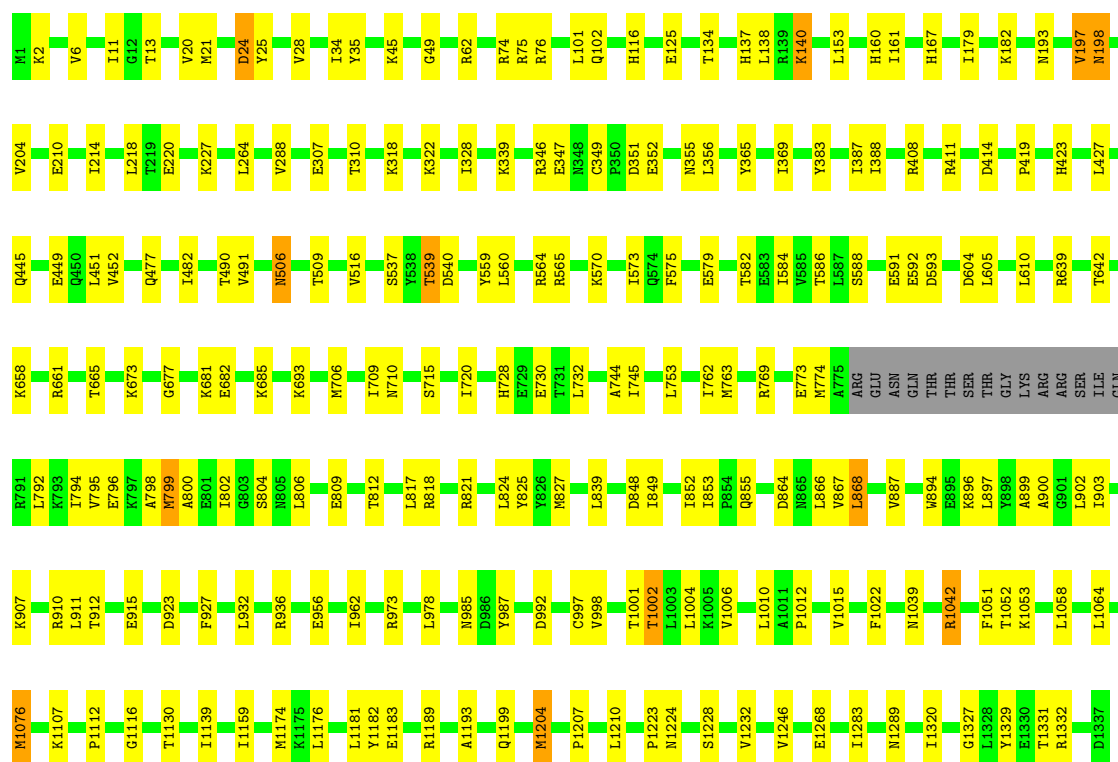
• Molecule 6: RNA (83-MER)

Chain t:  42% 30% . 27%



● Molecule 7: CRISPR-associated endonuclease Cas9

Chain S: 82% 16% ..



4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	136878	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	TFS KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	45	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	2400	Depositor
Magnification	130000	Depositor
Image detector	FEI FALCON IV (4k x 4k)	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	0.22	0/1817	0.34	0/2456
1	B	0.24	0/1817	0.35	0/2456
1	C	0.21	0/1817	0.33	0/2456
1	D	0.21	0/1817	0.31	0/2456
2	K	0.18	0/2395	0.41	0/3232
2	L	0.17	0/2399	0.39	0/3237
3	N	0.21	0/1044	0.39	0/1609
4	T	0.20	0/1024	0.40	0/1578
5	g	0.21	0/629	0.30	0/976
6	t	0.19	0/1454	0.27	0/2263
7	S	0.21	0/11064	0.33	1/14881 (0.0%)
All	All	0.20	0/27277	0.34	1/37600 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	S	197	VAL	N-CA-C	-5.86	108.14	113.71

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	1790	0	1811	26	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	1790	0	1811	20	0
1	C	1790	0	1811	14	0
1	D	1790	0	1811	24	0
2	K	2347	0	2371	50	0
2	L	2351	0	2374	37	0
3	N	928	0	506	12	0
4	T	917	0	514	8	0
5	g	566	0	285	4	0
6	t	1300	0	655	14	0
7	S	10846	0	10959	123	0
All	All	26415	0	24908	299	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 299 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)
6:t:33:U:H4'	7:S:134:THR:HG23	1.66	0.77
7:S:806:LEU:HB3	7:S:809:GLU:HA	1.67	0.76
2:K:140:ARG:HH12	2:K:144:PRO:HD2	1.52	0.73
1:A:16:LYS:NZ	1:A:189:ASN:OD1	2.26	0.69
7:S:116:HIS:HB3	7:S:125:GLU:HG3	1.75	0.68

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	217/219 (99%)	210 (97%)	6 (3%)	1 (0%)	24 49
1	B	217/219 (99%)	204 (94%)	11 (5%)	2 (1%)	14 34

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	217/219 (99%)	208 (96%)	9 (4%)	0	100	100
1	D	217/219 (99%)	204 (94%)	13 (6%)	0	100	100
2	K	284/288 (99%)	263 (93%)	21 (7%)	0	100	100
2	L	285/288 (99%)	255 (90%)	29 (10%)	1 (0%)	30	53
7	S	1318/1337 (99%)	1245 (94%)	71 (5%)	2 (0%)	43	66
All	All	2755/2789 (99%)	2589 (94%)	160 (6%)	6 (0%)	44	66

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	6	SER
7	S	1052	THR
1	B	140	THR
2	L	207	GLN
1	B	163	LEU

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	202/202 (100%)	192 (95%)	10 (5%)	22	47
1	B	202/202 (100%)	189 (94%)	13 (6%)	16	38
1	C	202/202 (100%)	193 (96%)	9 (4%)	24	50
1	D	202/202 (100%)	190 (94%)	12 (6%)	18	41
2	K	254/255 (100%)	235 (92%)	19 (8%)	12	31
2	L	254/255 (100%)	235 (92%)	19 (8%)	12	31
7	S	1178/1192 (99%)	1138 (97%)	40 (3%)	32	58
All	All	2494/2510 (99%)	2372 (95%)	122 (5%)	24	48

5 of 122 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	K	243	LYS
7	S	923	ASP
2	L	64	VAL
7	S	868	LEU
7	S	1159	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 29 such sidechains are listed below:

Mol	Chain	Res	Type
2	L	111	GLN
7	S	1289	ASN
7	S	184	GLN
7	S	1131	HIS
2	L	263	ASN

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
5	g	26/55 (47%)	6 (23%)	0
6	t	60/83 (72%)	13 (21%)	0
All	All	86/138 (62%)	19 (22%)	0

5 of 19 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
5	g	9	U
5	g	10	U
5	g	11	G
5	g	27	A
5	g	28	G

There are no RNA pucker outliers to report.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

5.5 Carbohydrates ⓘ

There are no oligosaccharides 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.