wwPDB EM Validation Summary Report

Oct 20, 2024 – 12:08 AM JST

PDB ID	:	8Z0L
EMDB ID	:	EMD-39707
Title	:	Cryo-EM structure of Cas8-HNH system at partial R-loop state
Authors	:	Zhang, H.; Zhu, H.; Li, X.; Liu, Y.
Deposited on	:	2024-04-09
Resolution	:	2.57 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:

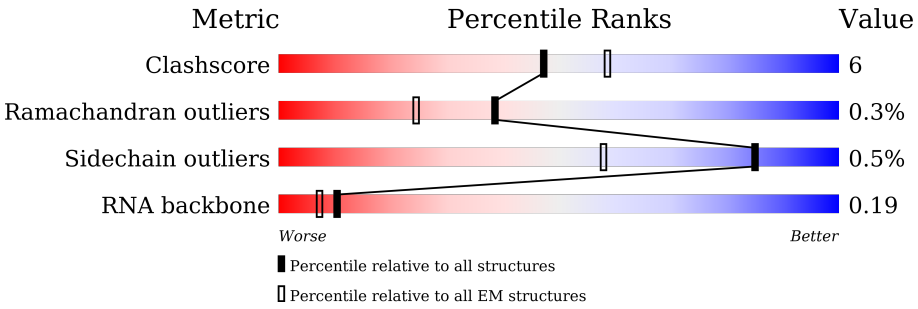
EMDB validation analysis	:	FAILED
MolProbity	:	4.02b-467
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
MapQ	:	FAILED
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

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 2.57 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	210492	15764
Ramachandran outliers	207382	16835
Sidechain outliers	206894	16415
RNA backbone	6643	2191

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 <=5%

Mol	Chain	Length	Quality of chain
1	A	325	94% 5%
1	B	325	94% 6% •
1	C	325	92% 6% •
1	D	325	88% 11% •
1	G	325	86% 6% 7%
1	H	325	76% 13% • 10%
2	E	255	88% 12%
3	F	344	80% 13% • 6%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	I	32	69% 31%
5	J	7	43% 57%
6	L	69	26% 32% 26% 16%
7	M	181	67% 27%

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 22858 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 I-F CRISPR-associated protein Csy3.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	325	Total	C	N	O	S	0	0
			2632	1682	434	501	15		
1	B	323	Total	C	N	O	S	0	0
			2604	1663	428	498	15		
1	C	320	Total	C	N	O	S	0	0
			2576	1648	424	489	15		
1	D	321	Total	C	N	O	S	0	0
			2581	1649	425	492	15		
1	G	302	Total	C	N	O	S	0	0
			2430	1557	398	462	13		
1	H	292	Total	C	N	O	S	0	0
			2339	1500	380	446	13		

- Molecule 2 is a protein called hypothetical protein J6N51_11000.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	E	255	Total	C	N	O	S	0	0
			1968	1250	329	377	12		

- Molecule 3 is a protein called HNH endonuclease.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	F	322	Total	C	N	O	S	0	0
			2422	1534	410	469	9		

- Molecule 4 is a DNA chain called DNA (32-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
4	I	32	Total	C	N	O	P	0	0
			648	308	118	190	32		

- Molecule 5 is a DNA chain called DNA (5'-D(P*GP*TP*GP*CP*GP*GP*A)-3').

Mol	Chain	Residues	Atoms					AltConf	Trace
5	J	7	Total	C	N	O	P	0	0
			148	69	30	42	7		

- Molecule 6 is a RNA chain called RNA (69-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
6	L	58	Total	C	N	O	P	0	0
			1249	556	229	406	58		

- Molecule 7 is a protein called type I-F CRISPR-associated endoribonuclease Cas6/Csy4.

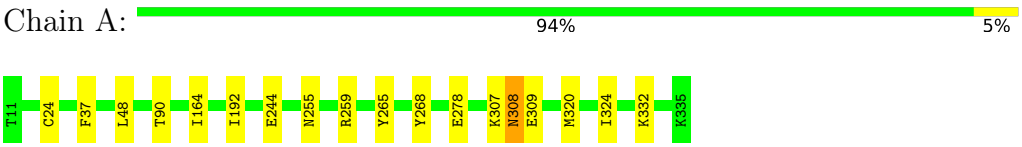
Mol	Chain	Residues	Atoms					AltConf	Trace
7	M	173	Total	C	N	O	S	0	0
			1261	798	207	253	3		

3 Residue-property plots (i)

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.

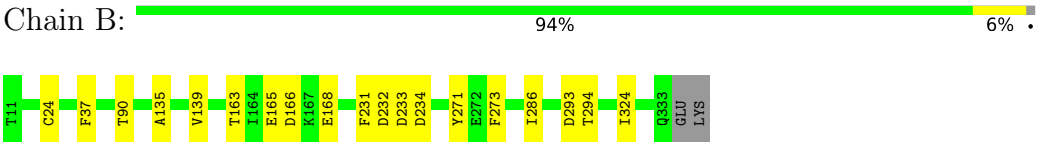
• Molecule 1: type I-F CRISPR-associated protein Csy3

Chain A:



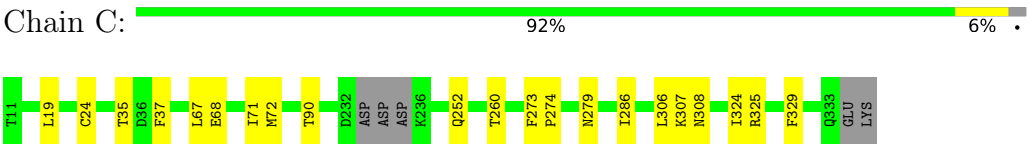
• Molecule 1: type I-F CRISPR-associated protein Csy3

Chain B:



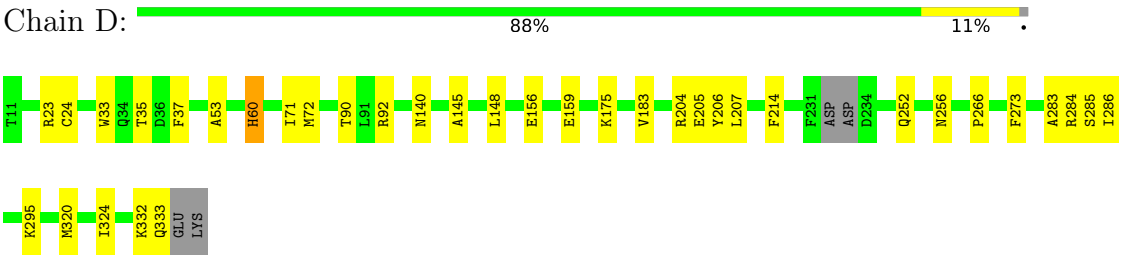
• Molecule 1: type I-F CRISPR-associated protein Csy3

Chain C:



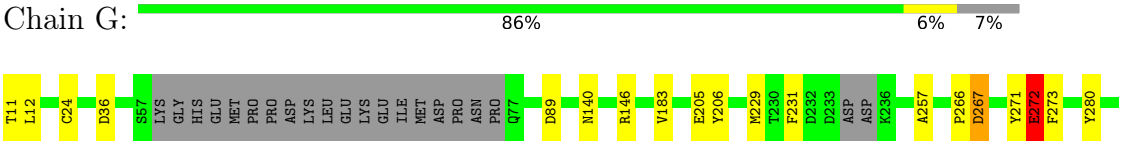
• Molecule 1: type I-F CRISPR-associated protein Csy3

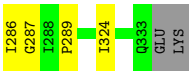
Chain D:



• Molecule 1: type I-F CRISPR-associated protein Csy3

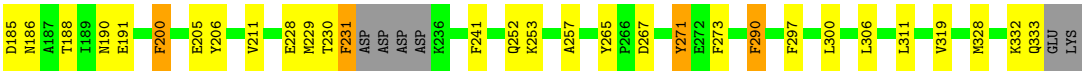
Chain G:





• Molecule 1: type I-F CRISPR-associated protein Csy3

Chain H:

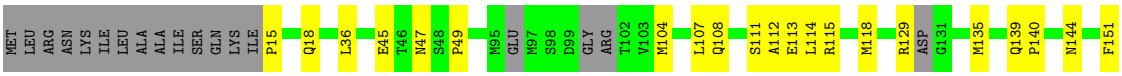


• Molecule 2: hypothetical protein J6N51_11000

Chain E:

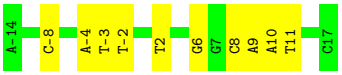
• Molecule 3: HNH endonuclease

Chain F:



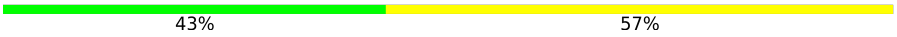
• Molecule 4: DNA (32-MER)

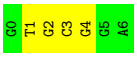
Chain I:



• Molecule 5: DNA (5'-D(P*GP*TP*GP*CP*GP*GP*A)-3')

Chain J:



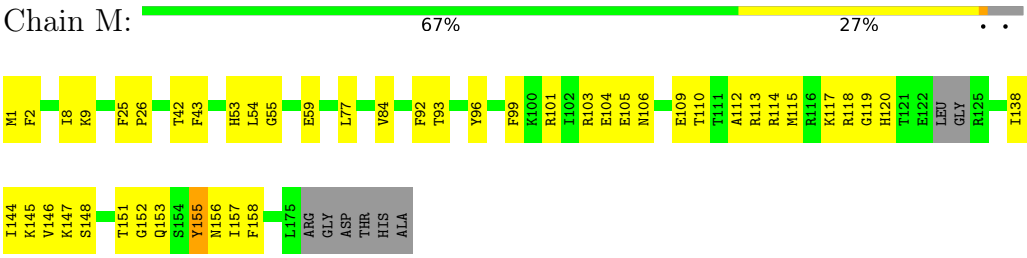


• Molecule 6: RNA (69-MER)

Chain L:

• Molecule 7: type I-F CRISPR-associated endoribonuclease Cas6/Csy4

Chain M:



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	140489	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING ONLY	Depositor
Microscope	JEOL CRYO ARM 300	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	60	Depositor
Minimum defocus (nm)	1200	Depositor
Maximum defocus (nm)	1600	Depositor
Magnification	Not provided	
Image detector	GATAN K3 (6k 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	#$\|Z\|>5$	RMSZ	#$\|Z\|>5$
1	A	0.34	0/2691	0.60	0/3636
1	B	0.36	0/2663	0.62	0/3604
1	C	0.36	0/2634	0.61	0/3564
1	D	0.37	0/2639	0.63	0/3571
1	G	0.42	0/2482	0.64	0/3357
1	H	0.44	0/2391	0.69	0/3239
2	E	0.38	0/2003	0.62	0/2711
3	F	0.40	0/2470	0.66	0/3361
4	I	0.74	0/725	1.07	0/1114
5	J	0.62	0/166	0.86	0/255
6	L	0.40	0/1397	0.82	0/2176
7	M	0.42	0/1283	0.77	0/1750
All	All	0.40	0/23544	0.68	0/32338

There are no bond length outliers.

There are no bond angle outliers.

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	2632	0	2584	12	0
1	B	2604	0	2537	13	0
1	C	2576	0	2513	15	0
1	D	2581	0	2508	26	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	G	2430	0	2367	18	0
1	H	2339	0	2248	37	0
2	E	1968	0	1921	21	0
3	F	2422	0	2274	38	0
4	I	648	0	359	11	0
5	J	148	0	79	2	0
6	L	1249	0	627	44	0
7	M	1261	0	1120	48	0
All	All	22858	0	21137	255	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 255 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:L:70:C:H42	6:L:76:G:H1	1.10	0.99
3:F:112:ALA:HA	3:F:115:ARG:HB2	1.51	0.91
2:E:248:LEU:HG	2:E:250:LEU:HD13	1.54	0.90
6:L:71:G:H3'	7:M:118:ARG:NH2	1.93	0.83
6:L:70:C:H3'	6:L:71:G:C8	2.15	0.82

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	323/325 (99%)	312 (97%)	10 (3%)	1 (0%)	37 57
1	B	321/325 (99%)	305 (95%)	15 (5%)	1 (0%)	37 57
1	C	316/325 (97%)	305 (96%)	11 (4%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	D	317/325 (98%)	303 (96%)	13 (4%)	1 (0%)	37 57
1	G	296/325 (91%)	275 (93%)	19 (6%)	2 (1%)	19 36
1	H	286/325 (88%)	260 (91%)	26 (9%)	0	100 100
2	E	253/255 (99%)	235 (93%)	18 (7%)	0	100 100
3	F	314/344 (91%)	291 (93%)	21 (7%)	2 (1%)	22 41
7	M	169/181 (93%)	141 (83%)	27 (16%)	1 (1%)	22 41
All	All	2595/2730 (95%)	2427 (94%)	160 (6%)	8 (0%)	38 57

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	285	SER
1	G	267	ASP
1	G	272	GLU
1	B	233	ASP
1	A	308	ASN

5.3.2 Protein sidechains (i)

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	288/292 (99%)	288 (100%)	0	100 100
1	B	284/292 (97%)	284 (100%)	0	100 100
1	C	280/292 (96%)	279 (100%)	1 (0%)	89 96
1	D	280/292 (96%)	278 (99%)	2 (1%)	81 92
1	G	263/292 (90%)	262 (100%)	1 (0%)	89 96
1	H	251/292 (86%)	247 (98%)	4 (2%)	58 78
2	E	208/219 (95%)	208 (100%)	0	100 100
3	F	248/304 (82%)	248 (100%)	0	100 100
7	M	121/160 (76%)	118 (98%)	3 (2%)	42 66
All	All	2223/2435 (91%)	2212 (100%)	11 (0%)	85 95

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	H	290	PHE
7	M	96	TYR
7	M	155	TYR
7	M	143	TYR
1	H	200	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
3	F	193	ASN
3	F	275	HIS
7	M	141	HIS
1	G	140	ASN
2	E	59	HIS

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
6	L	56/69 (81%)	33 (58%)	5 (8%)

5 of 33 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
6	L	22	U
6	L	23	U
6	L	29	G
6	L	30	A
6	L	34	C

All (5) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
6	L	54	C
6	L	55	U
6	L	59	C
6	L	61	G
6	L	78	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 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.