WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB EM Validation Summary Report ⓘ

Sep 1, 2024 – 12:09 AM JST

PDB ID	:	8YH9
EMDB ID	:	EMD-39285
Title	:	Type I-FHNH Cascade complex
Authors	:	Li, Z.
Deposited on	:	2024-02-27
Resolution	:	3.35 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:
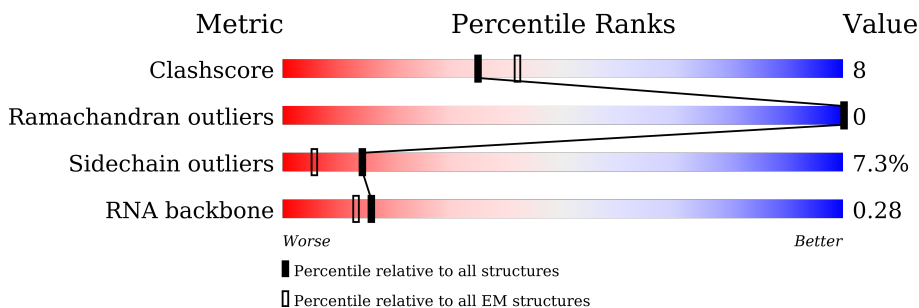
EMDB validation analysis	:	FAILED
MolProbity	:	4.02b-467
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
MapQ	:	FAILED
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.38.2

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 3.35 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	Percentile Ranks	Value
Clashscore		8
Ramachandran outliers		0
Sidechain outliers		7.3%
RNA backbone		0.28

Worse .... *Better*

▮ Percentile relative to all structures

▯ Percentile relative to all EM structures

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	255	72% 22% · ·
2	B	181	70% 25% · ·
3	C	60	42% 47% 12%
4	D	335	70% 15% · 13%
4	E	335	59% 26% · 12%
4	F	335	71% 17% · 10%
4	G	335	71% 19% · 10%
4	H	335	67% 20% · 11%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	I	335	65% 24% 10%
5	J	344	56% 14% 30%

2   Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 21153 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 Cas5f.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	245	Total	C	N	O	S	0	0
			1934	1235	324	361	14		

- Molecule 2 is a protein called Cas6f.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	175	Total	C	N	O	S	0	0
			1415	913	239	258	5		

- Molecule 3 is a RNA chain called 60-nt crRNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	60	Total	C	N	O	P	0	0
			1287	577	240	411	59		

- Molecule 4 is a protein called Cas7f.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	I	302	Total	C	N	O	S	0	0
			2467	1581	407	466	13		
4	H	297	Total	C	N	O	S	0	0
			2423	1553	399	458	13		
4	G	302	Total	C	N	O	S	0	0
			2461	1576	407	465	13		
4	F	300	Total	C	N	O	S	0	0
			2447	1570	403	461	13		
4	E	294	Total	C	N	O	S	0	0
			2397	1539	393	452	13		
4	D	293	Total	C	N	O	S	0	0
			2393	1534	393	453	13		

- Molecule 5 is a protein called Cas8f fusion with HNH.

| Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
5	J	242	1929	1229	334	360	6	0	0

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: Cas5f
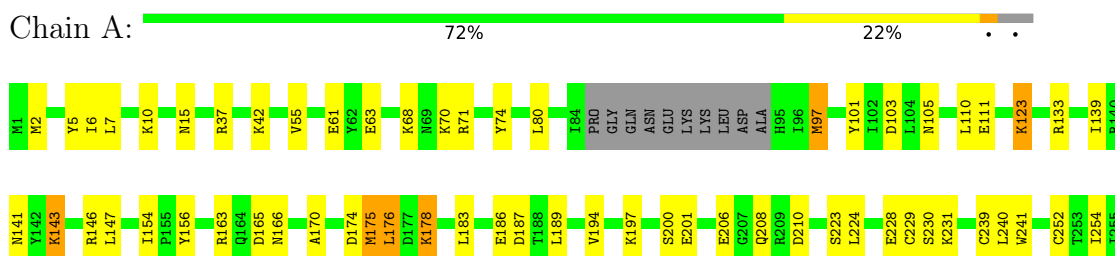
Chain A:

- Molecule 2: Cas6f
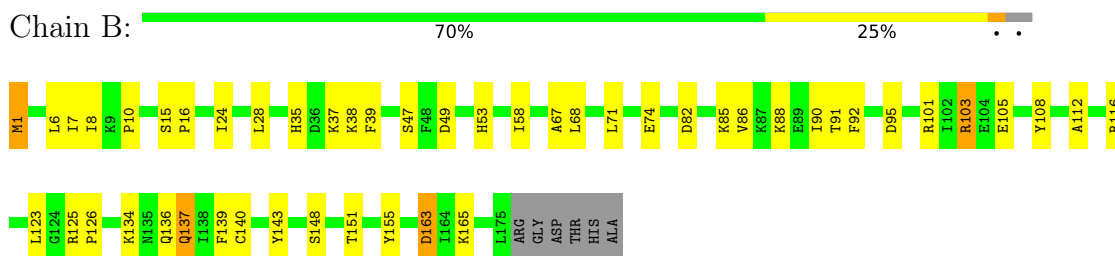
Chain B:

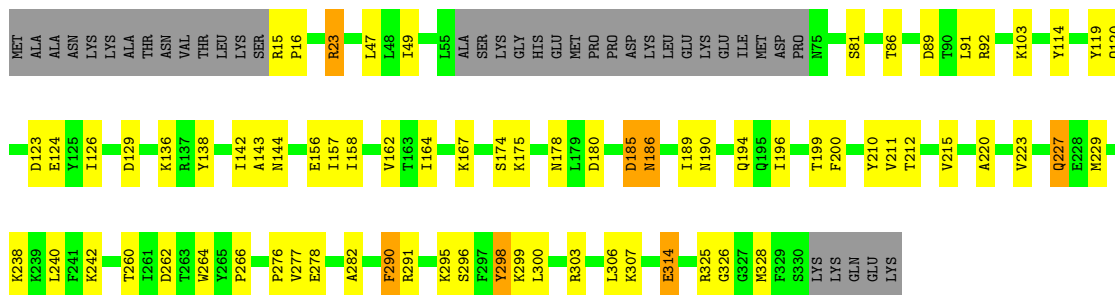- Molecule 3: 60-nt crRNA
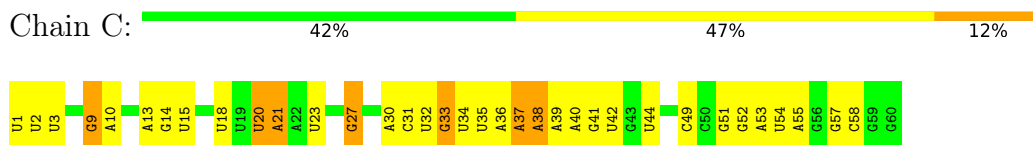
Chain C:

- Molecule 4: Cas7f
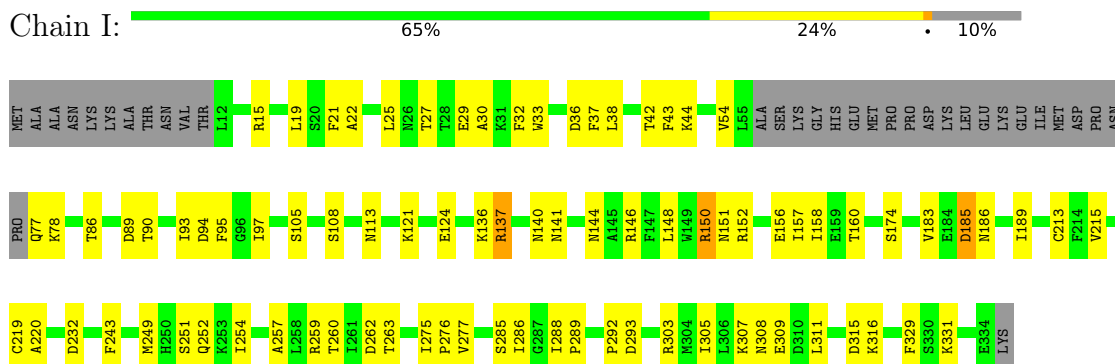
Chain I:
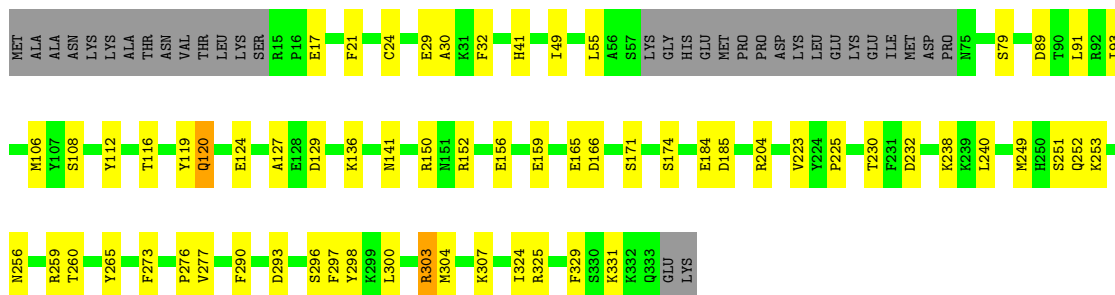
• Molecule 4: Cas7f

Chain H:

67%   20%   •   11%
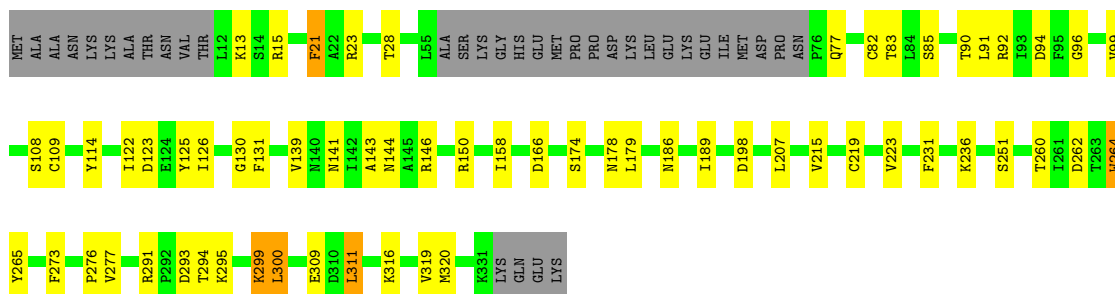
• Molecule 4: Cas7f

Chain G:

71%   19%   •   10%

• Molecule 4: Cas7f

Chain F:

71%   17%   •   10%

• Molecule 4: Cas7f

Chain E:

59%   26%   •   12%

N173 S174 K175 D180 T199 F200 R204 L207 N208 I209 T212 C213 F214 V215 K216 I217 E222 V223 S226 D233 D234 D235 K236 K239 F243 S246 A247 H250 S251 Q252 N255 L258 I261 D293 R303 M304 I305 L306 L311 D315

I324 R325 G326 S330 K331 LYS GLN GLU LYS

• Molecule 4: Cas7f

Chain D:

70% 15% • 13%

MET ALA ALA ASN LYS LYS ALA THR ASN VAL THR LEU LYS SER ARG PRO GLU N18 L19 S20 F21 A22 R23 N26 T27 T28 E29 F32 W33 Q34 F37 R40 L45 I49 A53 VAL LEU ALA SER LYS GLY HIS GLU MET PRO PRO ASP LYS LEU GLU LYS GLU ILE

MET ASP PRO ASN PRO GLN LYS S79 C82 T83 L91 R92 I93 G96 I97 K98 K103 Y114 K136 R137 L148 N151 R152 K153 E156 T163 D185 T199 F200 A201 L207 Y210 V211 T212 K216 I217 M221 E222 V223 Y224 P225 S226 Q227 E228

M229 D233 D234 D235 K236 A247 H250 S251 Q252 N255 N256 R259 T260 I261 F273 I301 R325 K331 K335

• Molecule 5: Cas8f fusion with HNH

Chain J:

56% 14% • 30%

MET LEU ARG ASN LYS ILE LEU ALA ALA ILE SER GLN ILE LYS PRO GLU GLU GLN ILE LYS ASN LYS TYR ILE GLU GLY LEU GLN SER ILE ASP LYS ASN H35 H39 F43 S48 D61 L67 D68 T69 A70 N71 A72 G73 E76 PRO ASP ILE VAL THR ASN ASP

ALA LYS TYR LYS ARG PRO GLN GLY PHE VAL ALA MET GLU MET SER ASP GLY ARG THR VAL MET GLU HIS LEU GLU ASP GLU ALA GLU SER ARG HIS GLU MET GLU LEU THR ASP LYS TYR ASP GLU ILE ARG ASP GLY ILE LEU ASN MET PRO SER MET GLN PRO Y141 F146

F151 S157 Y158 P165 L177 Y178 R179 S180 K184 I185 R186 L190 S191 S192 N193 F212 Q213 A214 L215 K218 F219 K222 K226 D229 Y232 L233 L234 E241 G244 V245 D252 I260 K266 R272 H275 Q286 L289 G300

G313 K318 N319 M320 I321 M322 R330 V331 A334 K340 TRP ARG ASN THR

4   Experimental information ⓘ

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	110316	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	NONE	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose $(e^-/\text{Å}^2)$	54	Depositor
Minimum defocus (nm)	1200	Depositor
Maximum defocus (nm)	2200	Depositor
Magnification	Not provided	
Image detector	GATAN K3 BIOQUANTUM (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	0.28	0/1967	0.58	0/2650
2	B	0.26	0/1446	0.51	0/1945
3	C	0.30	0/1443	0.89	0/2250
4	D	0.26	0/2445	0.52	0/3299
4	E	0.29	0/2450	0.55	0/3308
4	F	0.30	0/2501	0.54	0/3376
4	G	0.28	0/2515	0.56	0/3396
4	H	0.27	0/2477	0.55	0/3347
4	I	0.28	0/2520	0.54	0/3400
5	J	0.26	0/1972	0.59	0/2668
All	All	0.28	0/21736	0.58	0/29639

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	1934	0	1953	39	0
2	B	1415	0	1437	25	0
3	C	1287	0	645	23	0
4	D	2393	0	2359	33	0
4	E	2397	0	2368	59	0
4	F	2447	0	2423	34	0

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	G	2461	0	2430	36	0
4	H	2423	0	2386	35	0
4	I	2467	0	2442	42	0
5	J	1929	0	1921	29	0
All	All	21153	0	20364	322	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 8.

The worst 5 of 322 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:F:300:LEU:HD11	4:F:319:VAL:HG11	1.55	0.88
4:E:139:VAL:HG22	4:E:172:PHE:HZ	1.42	0.85
4:D:91:LEU:HD22	4:D:223:VAL:HG21	1.63	0.80
4:F:300:LEU:HD21	4:F:319:VAL:HG12	1.74	0.69
4:I:150:ARG:HH22	4:H:23:ARG:HH21	1.44	0.66

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	241/255 (94%)	206 (86%)	35 (14%)	0	100	100
2	B	173/181 (96%)	155 (90%)	18 (10%)	0	100	100
4	D	289/335 (86%)	271 (94%)	18 (6%)	0	100	100
4	E	290/335 (87%)	261 (90%)	29 (10%)	0	100	100
4	F	296/335 (88%)	278 (94%)	18 (6%)	0	100	100
4	G	298/335 (89%)	276 (93%)	22 (7%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
4	H	293/335 (88%)	275 (94%)	18 (6%)	0	100	100
4	I	298/335 (89%)	281 (94%)	17 (6%)	0	100	100
5	J	238/344 (69%)	212 (89%)	26 (11%)	0	100	100
All	All	2416/2790 (87%)	2215 (92%)	201 (8%)	0	100	100

There are no Ramachandran outliers to report.

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	211/219 (96%)	197 (93%)	14 (7%)	14	39
2	B	156/160 (98%)	142 (91%)	14 (9%)	8	28
4	D	262/299 (88%)	248 (95%)	14 (5%)	19	45
4	E	263/299 (88%)	238 (90%)	25 (10%)	7	25
4	F	269/299 (90%)	248 (92%)	21 (8%)	10	33
4	G	270/299 (90%)	253 (94%)	17 (6%)	15	40
4	H	266/299 (89%)	243 (91%)	23 (9%)	8	30
4	I	271/299 (91%)	253 (93%)	18 (7%)	14	39
5	J	211/304 (69%)	199 (94%)	12 (6%)	17	44
All	All	2179/2477 (88%)	2021 (93%)	158 (7%)	14	36

5 of 158 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	E	92	ARG
4	D	233	ASP
4	E	107	TYR
4	E	243	PHE
5	J	157	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
4	I	141	ASN
4	G	120	GLN
4	E	26	ASN
4	E	141	ASN
5	J	328	HIS

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
3	C	59/60 (98%)	26 (44%)	0

5 of 26 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
3	C	3	U
3	C	9	G
3	C	10	A
3	C	13	A
3	C	14	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 ⓘ

There are no ligands in this entry.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.