wwPDB EM Validation Summary Report (i)

May 12, 2024 – 12:07 AM JST

PDB ID	:	8JQB
EMDB ID	:	EMD-36563
Title	:	Structure of Gabija GajA-GajB 4:4 Complex
Authors	:	Li, J.; Wang, Z.; Wang, L.
Deposited on	:	2023-06-13
Resolution	:	3.20 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 (i) 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 (i)) were used in the production of this report:

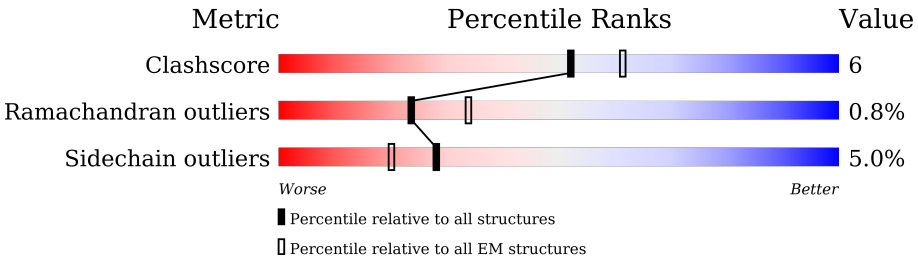
EMDB validation analysis	:	FAILED
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	FAILED
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 3.20 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	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

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	578	70% 10% • 19%
1	B	578	69% 12% • 19%
1	C	578	69% 11% • 19%
1	D	578	69% 11% • 19%
2	E	499	75% 22% •
2	F	499	75% 23% •
2	G	499	76% 22% •
2	H	499	74% 23% •

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 31608 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 Endonuclease GajA.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	C	468	Total 3835	C 2463	N 631	O 729	S 12	0	0
1	A	468	Total 3835	C 2463	N 631	O 729	S 12	0	0
1	B	468	Total 3835	C 2463	N 631	O 729	S 12	0	0
1	D	468	Total 3835	C 2463	N 631	O 729	S 12	0	0

- Molecule 2 is a protein called Gabija protein GajB.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	E	499	Total 4067	C 2604	N 670	O 781	S 12	0	0
2	F	499	Total 4067	C 2604	N 670	O 781	S 12	0	0
2	G	499	Total 4067	C 2604	N 670	O 781	S 12	0	0
2	H	499	Total 4067	C 2604	N 670	O 781	S 12	0	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	1	MET	-	initiating methionine	UNP J8HQ06
E	2	ILE	-	expression tag	UNP J8HQ06
E	3	GLU	-	expression tag	UNP J8HQ06
E	4	ASP	-	expression tag	UNP J8HQ06
E	5	GLU	-	expression tag	UNP J8HQ06
F	1	MET	-	initiating methionine	UNP J8HQ06
F	2	ILE	-	expression tag	UNP J8HQ06
F	3	GLU	-	expression tag	UNP J8HQ06
F	4	ASP	-	expression tag	UNP J8HQ06

Continued on next page...

Continued from previous page...

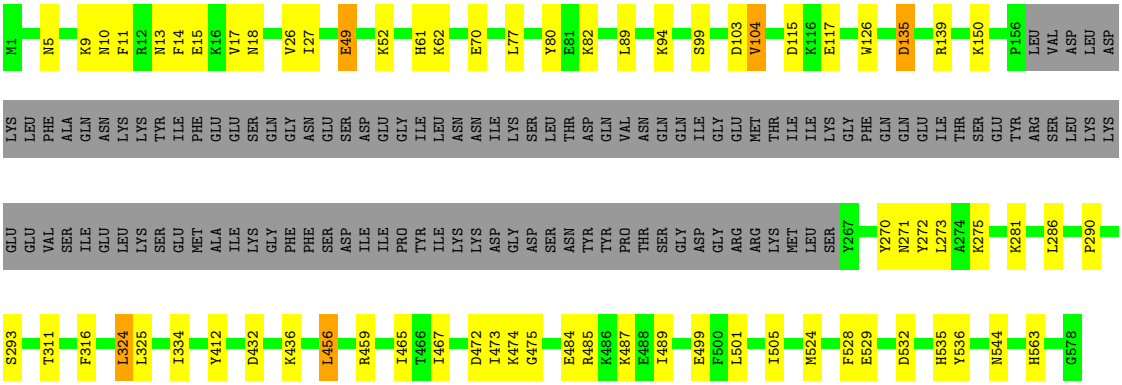
Chain	Residue	Modelled	Actual	Comment	Reference
F	5	GLU	-	expression tag	UNP J8HQ06
G	1	MET	-	initiating methionine	UNP J8HQ06
G	2	ILE	-	expression tag	UNP J8HQ06
G	3	GLU	-	expression tag	UNP J8HQ06
G	4	ASP	-	expression tag	UNP J8HQ06
G	5	GLU	-	expression tag	UNP J8HQ06
H	1	MET	-	initiating methionine	UNP J8HQ06
H	2	ILE	-	expression tag	UNP J8HQ06
H	3	GLU	-	expression tag	UNP J8HQ06
H	4	ASP	-	expression tag	UNP J8HQ06
H	5	GLU	-	expression tag	UNP J8HQ06

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.

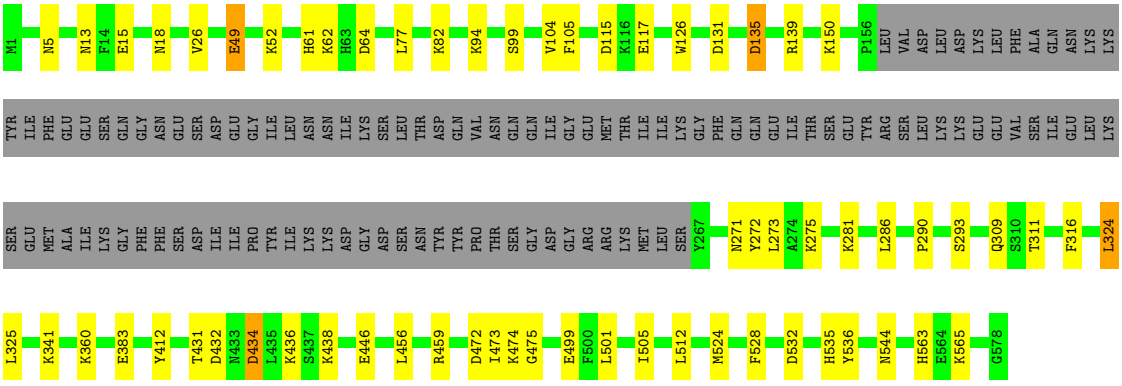
• Molecule 1: Endonuclease GajA

Chain C:



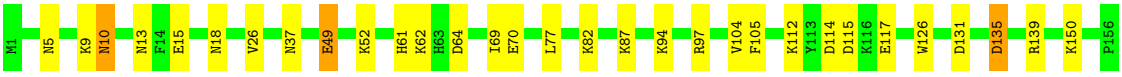
• Molecule 1: Endonuclease GajA

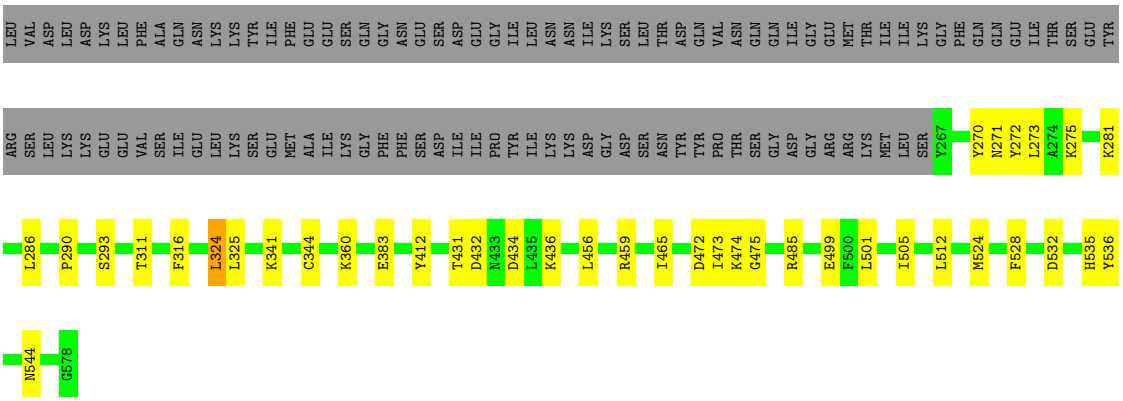
Chain A:



• Molecule 1: Endonuclease GajA

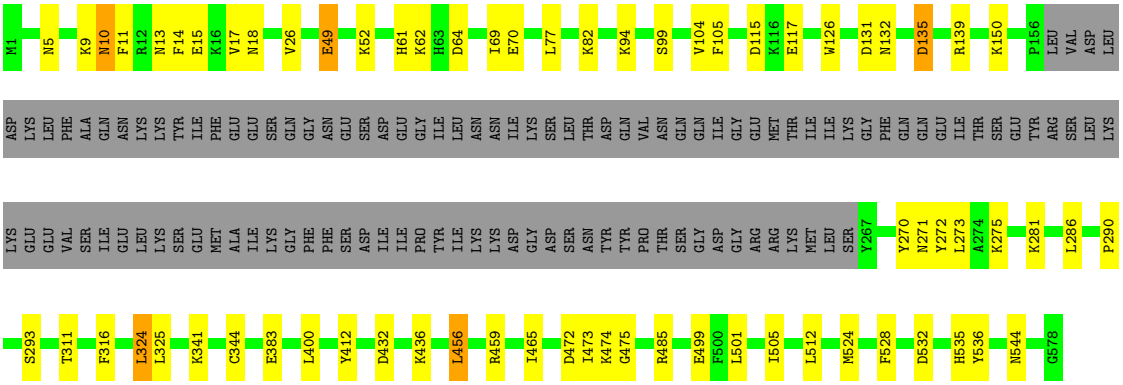
Chain B:





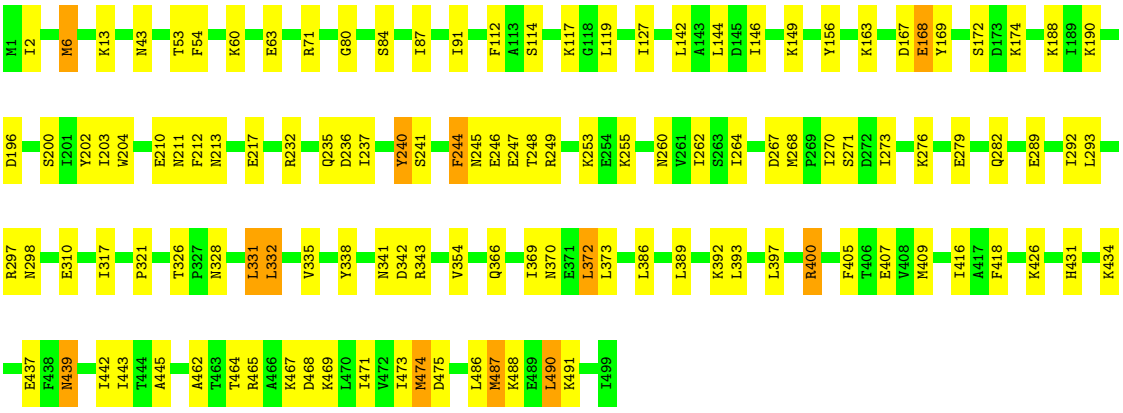
• Molecule 1: Endonuclease GajA

Chain D:



• Molecule 2: Gabija protein GajB

Chain E:

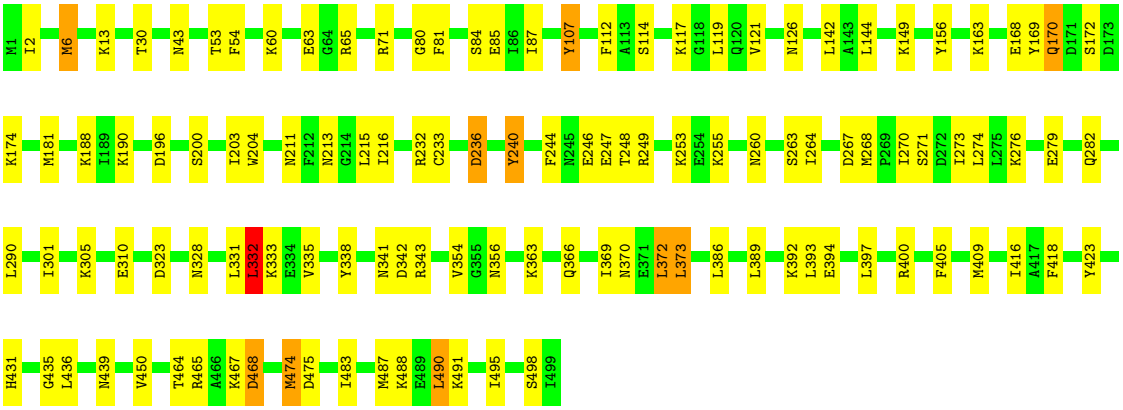


• Molecule 2: Gabija protein GajB

Chain F:

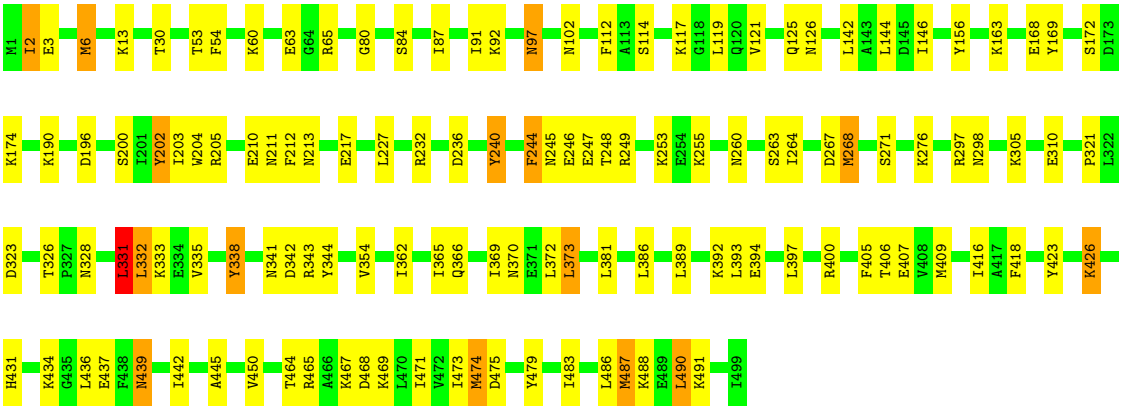
• Molecule 2: Gabija protein GajB

Chain G:



• Molecule 2: Gabija protein GajB

Chain H:



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	2245766	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 ($e^-/Å^2$)	50	Depositor
Minimum defocus (nm)	1500	Depositor
Maximum defocus (nm)	1800	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	0.30	0/3902	0.60	2/5240 (0.0%)
1	B	0.30	0/3902	0.61	2/5240 (0.0%)
1	C	0.29	0/3902	0.61	5/5240 (0.1%)
1	D	0.31	0/3902	0.62	4/5240 (0.1%)
2	E	0.32	0/4136	0.73	11/5579 (0.2%)
2	F	0.31	0/4136	0.71	10/5579 (0.2%)
2	G	0.31	0/4136	0.71	11/5579 (0.2%)
2	H	0.31	0/4136	0.73	13/5579 (0.2%)
All	All	0.31	0/32152	0.67	58/43276 (0.1%)

There are no bond length outliers.

The worst 5 of 58 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	196	ASP	CB-CG-OD2	9.31	126.68	118.30
2	H	196	ASP	CB-CG-OD2	9.11	126.50	118.30
2	G	196	ASP	CB-CG-OD2	8.79	126.21	118.30
2	G	332	LEU	CA-CB-CG	7.43	132.39	115.30
2	F	119	LEU	CA-CB-CG	7.16	131.76	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	3835	0	3872	34	0
1	B	3835	0	3872	38	0
1	C	3835	0	3872	38	0
1	D	3835	0	3872	36	0
2	E	4067	0	4089	62	0
2	F	4067	0	4089	61	0
2	G	4067	0	4089	52	0
2	H	4067	0	4089	66	0
All	All	31608	0	31844	363	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 363 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:F:232:ARG:HE	2:F:464:THR:HA	1.48	0.78
2:E:232:ARG:HE	2:E:464:THR:HA	1.52	0.74
2:E:434:LYS:HA	2:E:465:ARG:HH21	1.54	0.71
2:G:232:ARG:HE	2:G:464:THR:HA	1.58	0.69
2:H:434:LYS:HA	2:H:465:ARG:HH21	1.61	0.65

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	464/578 (80%)	436 (94%)	26 (6%)	2 (0%)	34 69
1	B	464/578 (80%)	438 (94%)	24 (5%)	2 (0%)	34 69
1	C	464/578 (80%)	433 (93%)	29 (6%)	2 (0%)	34 69
1	D	464/578 (80%)	434 (94%)	28 (6%)	2 (0%)	34 69

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	E	497/499 (100%)	443 (89%)	48 (10%)	6 (1%)	13	49
2	F	497/499 (100%)	449 (90%)	42 (8%)	6 (1%)	13	49
2	G	497/499 (100%)	443 (89%)	48 (10%)	6 (1%)	13	49
2	H	497/499 (100%)	450 (90%)	41 (8%)	6 (1%)	13	49
All	All	3844/4308 (89%)	3526 (92%)	286 (7%)	32 (1%)	24	58

5 of 32 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	E	247	GLU
2	F	247	GLU
2	G	247	GLU
2	H	247	GLU
1	C	104	VAL

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	429/529 (81%)	413 (96%)	16 (4%)	34	68
1	B	429/529 (81%)	411 (96%)	18 (4%)	30	65
1	C	429/529 (81%)	415 (97%)	14 (3%)	38	71
1	D	429/529 (81%)	413 (96%)	16 (4%)	34	68
2	E	453/453 (100%)	426 (94%)	27 (6%)	19	54
2	F	453/453 (100%)	428 (94%)	25 (6%)	21	57
2	G	453/453 (100%)	424 (94%)	29 (6%)	17	52
2	H	453/453 (100%)	422 (93%)	31 (7%)	16	49
All	All	3528/3928 (90%)	3352 (95%)	176 (5%)	28	60

5 of 176 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	F	487	MET
2	G	474	MET
2	G	149	LYS
2	G	323	ASP
2	H	174	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
2	E	260	ASN
2	F	260	ASN
2	H	97	ASN
2	H	260	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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 monosaccharides 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.