



wwPDB/EMDatabank EM Map/Model Validation Summary Report ⓘ

Dec 11, 2019 – 03:32 AM EST

PDB ID : 6DBQ
EMDB ID: : EMD-7847
Title : Cryo-EM structure of RAG in complex with 12-RSS and 23-RSS substrate DNAs
Authors : Wu, H.; Liao, M.; Ru, H.; Mi, W.
Deposited on : 2018-05-03
Resolution : 4.22 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

<https://www.wwpdb.org/validation/2017/EMValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

MolProbity : 4.02b-467
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : 2.4

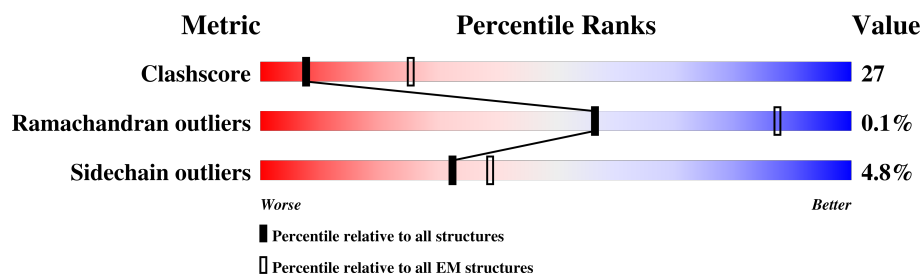
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 4.22 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	136327	1886
Ramachandran outliers	132723	1663
Sidechain outliers	132532	1531

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	1159	29% 23% . 46%
1	C	1159	29% 22% . 47%
2	B	533	34% 30% . 34%
2	D	533	33% 32% . 34%
3	E	50	28% 70% .
4	F	50	24% 76%
5	G	61	28% 70% .
6	H	61	38% 61% .

2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 19956 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 Recombination activating gene 1 - MBP chimera.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	621	Total	C	N	O	S	0	0
			4999	3128	899	935	37		
1	C	616	Total	C	N	O	S	0	0
			4972	3113	893	929	37		

There are 68 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-127	MET	-	initiating methionine	UNP P0AEX9
A	-126	GLY	-	expression tag	UNP P0AEX9
A	-125	SER	-	expression tag	UNP P0AEX9
A	-124	SER	-	expression tag	UNP P0AEX9
A	-123	HIS	-	expression tag	UNP P0AEX9
A	-122	HIS	-	expression tag	UNP P0AEX9
A	-121	HIS	-	expression tag	UNP P0AEX9
A	-120	HIS	-	expression tag	UNP P0AEX9
A	-119	HIS	-	expression tag	UNP P0AEX9
A	-118	HIS	-	expression tag	UNP P0AEX9
A	-117	GLY	-	expression tag	UNP P0AEX9
A	-116	THR	-	expression tag	UNP P0AEX9
A	-115	LYS	-	expression tag	UNP P0AEX9
A	-114	THR	-	expression tag	UNP P0AEX9
A	251	GLY	-	linker	UNP P0AEX9
A	252	THR	-	linker	UNP P0AEX9
A	253	ASP	-	linker	UNP P0AEX9
A	254	TYR	-	linker	UNP P0AEX9
A	255	ASP	-	linker	UNP P0AEX9
A	256	ILE	-	linker	UNP P0AEX9
A	257	PRO	-	linker	UNP P0AEX9
A	258	THR	-	linker	UNP P0AEX9
A	259	THR	-	linker	UNP P0AEX9
A	260	LEU	-	linker	UNP P0AEX9
A	261	GLU	-	linker	UNP P0AEX9
A	262	VAL	-	linker	UNP P0AEX9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	263	LEU	-	linker	UNP P0AEX9
A	264	PHE	-	linker	UNP P0AEX9
A	265	GLN	-	linker	UNP P0AEX9
A	266	GLY	-	linker	UNP P0AEX9
A	267	PRO	-	linker	UNP P0AEX9
A	268	LEU	-	linker	UNP P0AEX9
A	269	GLY	-	linker	UNP P0AEX9
A	270	SER	-	linker	UNP P0AEX9
C	-127	MET	-	initiating methionine	UNP P0AEX9
C	-126	GLY	-	expression tag	UNP P0AEX9
C	-125	SER	-	expression tag	UNP P0AEX9
C	-124	SER	-	expression tag	UNP P0AEX9
C	-123	HIS	-	expression tag	UNP P0AEX9
C	-122	HIS	-	expression tag	UNP P0AEX9
C	-121	HIS	-	expression tag	UNP P0AEX9
C	-120	HIS	-	expression tag	UNP P0AEX9
C	-119	HIS	-	expression tag	UNP P0AEX9
C	-118	HIS	-	expression tag	UNP P0AEX9
C	-117	GLY	-	expression tag	UNP P0AEX9
C	-116	THR	-	expression tag	UNP P0AEX9
C	-115	LYS	-	expression tag	UNP P0AEX9
C	-114	THR	-	expression tag	UNP P0AEX9
C	251	GLY	-	linker	UNP P0AEX9
C	252	THR	-	linker	UNP P0AEX9
C	253	ASP	-	linker	UNP P0AEX9
C	254	TYR	-	linker	UNP P0AEX9
C	255	ASP	-	linker	UNP P0AEX9
C	256	ILE	-	linker	UNP P0AEX9
C	257	PRO	-	linker	UNP P0AEX9
C	258	THR	-	linker	UNP P0AEX9
C	259	THR	-	linker	UNP P0AEX9
C	260	LEU	-	linker	UNP P0AEX9
C	261	GLU	-	linker	UNP P0AEX9
C	262	VAL	-	linker	UNP P0AEX9
C	263	LEU	-	linker	UNP P0AEX9
C	264	PHE	-	linker	UNP P0AEX9
C	265	GLN	-	linker	UNP P0AEX9
C	266	GLY	-	linker	UNP P0AEX9
C	267	PRO	-	linker	UNP P0AEX9
C	268	LEU	-	linker	UNP P0AEX9
C	269	GLY	-	linker	UNP P0AEX9
C	270	SER	-	linker	UNP P0AEX9

- Molecule 2 is a protein called Recombination activating gene 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	351	Total	C	N	O	S	0	0
			2714	1716	470	509	19		
2	D	351	Total	C	N	O	S	0	0
			2714	1716	470	509	19		

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-2	GLY	-	expression tag	UNP Q1RLW7
B	-1	GLY	-	expression tag	UNP Q1RLW7
B	0	SER	-	expression tag	UNP Q1RLW7
D	-2	GLY	-	expression tag	UNP Q1RLW7
D	-1	GLY	-	expression tag	UNP Q1RLW7
D	0	SER	-	expression tag	UNP Q1RLW7

- Molecule 3 is a DNA chain called Molecule name: Forward strand of 12-RSS substrate DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	E	50	Total	C	N	O	P	0	0
			1023	486	192	295	50		

- Molecule 4 is a DNA chain called Molecule name: Reverse strand of 12-RSS substrate DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	F	50	Total	C	N	O	P	0	0
			1027	489	183	305	50		

- Molecule 5 is a DNA chain called Molecule name: Forward strand of 23-RSS substrate DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	G	61	Total	C	N	O	P	0	0
			1245	593	223	368	61		

- Molecule 6 is a DNA chain called Reverse strand of 23-RSS substrate DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	H	61	Total	C	N	O	P	0	0
			1256	596	235	364	61		

- Molecule 7 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		AltConf
7	A	1	Total 1	Zn 1	0
7	C	1	Total 1	Zn 1	0

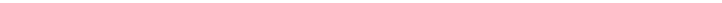
- Molecule 8 is CALCIUM ION (three-letter code: CA) (formula: Ca).

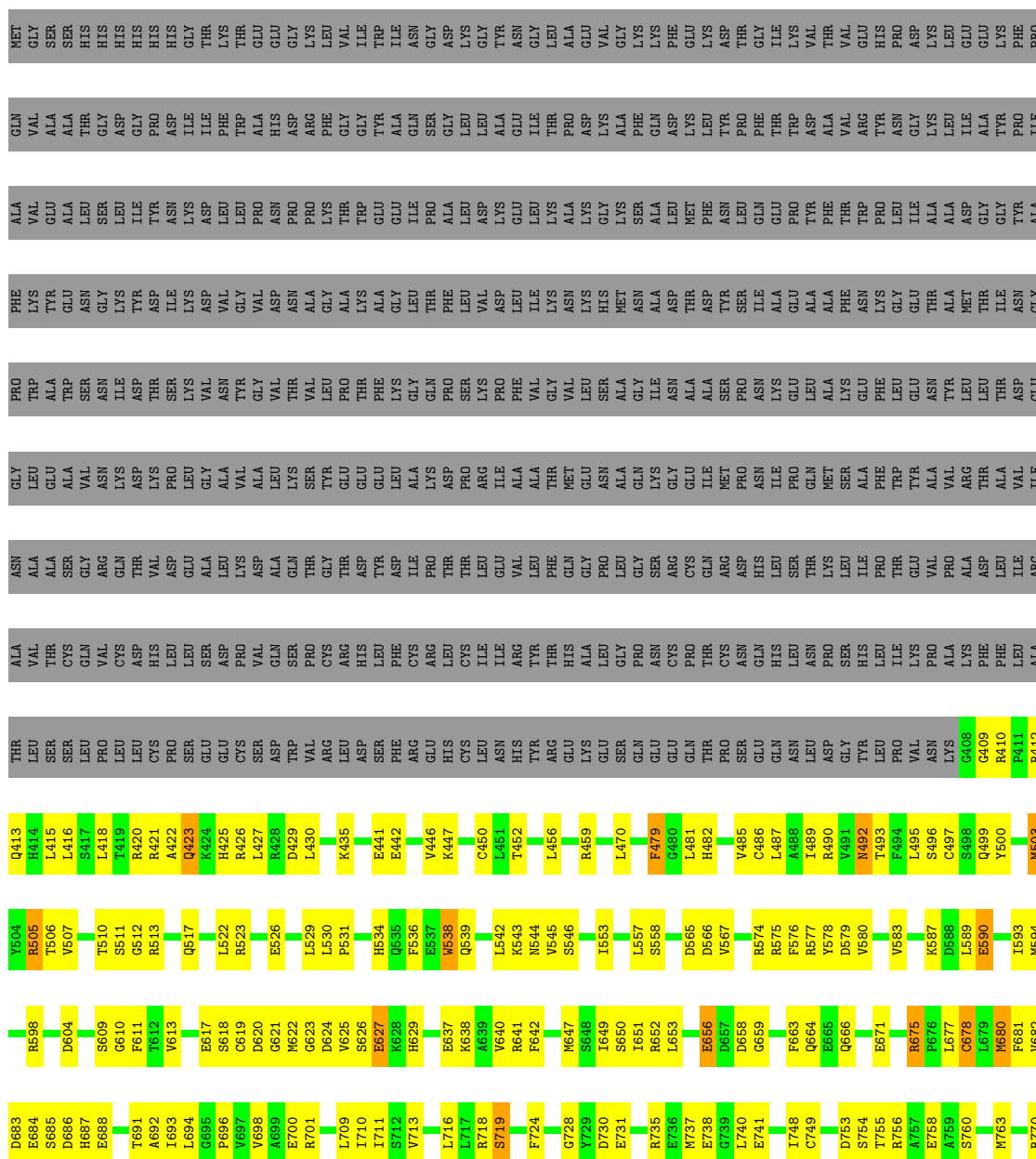
Mol	Chain	Residues	Atoms		AltConf
8	A	2	Total 2	Ca 2	0
8	C	2	Total 2	Ca 2	0

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: Recombination activating gene 1 - MBP chimera

Chain A:  29% 23% . 46%



4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	48002	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 ($e^-/\text{\AA}^2$)	47	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, CA

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	0.32	0/5097	0.59	4/6856 (0.1%)
1	C	0.33	0/5067	0.66	7/6812 (0.1%)
2	B	0.34	0/2784	0.63	5/3784 (0.1%)
2	D	0.35	1/2784 (0.0%)	0.61	3/3784 (0.1%)
3	E	0.61	0/1148	0.95	1/1768 (0.1%)
4	F	0.62	0/1150	0.99	0/1774
5	G	0.62	1/1394 (0.1%)	0.91	0/2148
6	H	0.65	0/1410	1.00	1/2175 (0.0%)
All	All	0.42	2/20834 (0.0%)	0.73	21/29101 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
1	C	0	1
2	B	0	3
2	D	0	1
All	All	0	6

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	83	GLN	CG-CD	-5.47	1.38	1.51
5	G	35	DC	C1'-N1	5.12	1.55	1.49

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	432	ASN	CB-CG-OD1	24.80	171.21	121.60
1	C	432	ASN	CB-CG-ND2	-20.88	66.60	116.70
1	A	423	GLN	CG-CD-OE1	17.99	157.58	121.60
1	A	423	GLN	CG-CD-NE2	-15.65	79.13	116.70
2	B	344	HIS	N-CA-CB	-11.29	90.27	110.60

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	423	GLN	Sidechain
2	B	10	ASN	Sidechain,Mainchain
2	B	344	HIS	Sidechain
1	C	817	HIS	Sidechain
2	D	83	GLN	Sidechain

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	4999	0	4954	263	0
1	C	4972	0	4937	245	0
2	B	2714	0	2665	160	0
2	D	2714	0	2665	173	0
3	E	1023	0	561	69	0
4	F	1027	0	566	67	0
5	G	1245	0	687	99	0
6	H	1256	0	685	50	0
7	A	1	0	0	0	0
7	C	1	0	0	0	0
8	A	2	0	0	0	0
8	C	2	0	0	0	0
All	All	19956	0	17720	1020	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 27.

The worst 5 of 1020 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)
5:G:12:DT:O4	6:H:50:DA:N6	1.60	1.34
2:D:123:ARG:NH2	2:D:125:GLU:OE2	1.70	1.25
2:D:232:ARG:HH11	2:D:234:ILE:CG1	1.49	1.23
1:C:435:LYS:NZ	1:C:439:GLU:OE2	1.81	1.12
1:A:789:GLU:OE2	1:A:797:ARG:NH1	1.84	1.10

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	619/1159 (53%)	558 (90%)	59 (10%)	2 (0%)	43	80
1	C	612/1159 (53%)	551 (90%)	61 (10%)	0	100	100
2	B	349/533 (66%)	306 (88%)	43 (12%)	0	100	100
2	D	349/533 (66%)	312 (89%)	37 (11%)	0	100	100
All	All	1929/3384 (57%)	1727 (90%)	200 (10%)	2 (0%)	56	87

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	818	CYS
1	A	512	GLY

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	547/1000 (55%)	521 (95%)	26 (5%)	28	60
1	C	545/1000 (54%)	515 (94%)	30 (6%)	24	57
2	B	303/465 (65%)	290 (96%)	13 (4%)	32	63
2	D	303/465 (65%)	291 (96%)	12 (4%)	34	64
All	All	1698/2930 (58%)	1617 (95%)	81 (5%)	32	60

5 of 81 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	289	TYR
1	C	573	SER
2	D	178	CYS
2	B	292	LEU
1	C	490	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	10	ASN
2	D	33	GLN
1	C	775	ASN
1	A	956	ASN
1	C	1012	HIS

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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 carbohydrates in this entry.

5.6 Ligand geometry

Of 6 ligands modelled in this entry, 6 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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.