wwPDB/EMDataBank EM Map/Model Validation Summary Report (i)

Aug 5, 2019 – 01:33 PM EDT

PDB ID	:	6ME0
EMDB ID:	:	EMD-9105
Title	:	Structure of a group II intron retroelement prior to DNA integration
Authors	:	Haack, D.; Yan, X.; Zhang, C.; Hingey, J.; Lyumkis, D.; Baker, T.S.; Toor, N.
Deposited on	:	2018-09-05
Resolution	:	3.60 Å(reported)
Based on PDB ID	:	4R0D

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 (i) symbol.

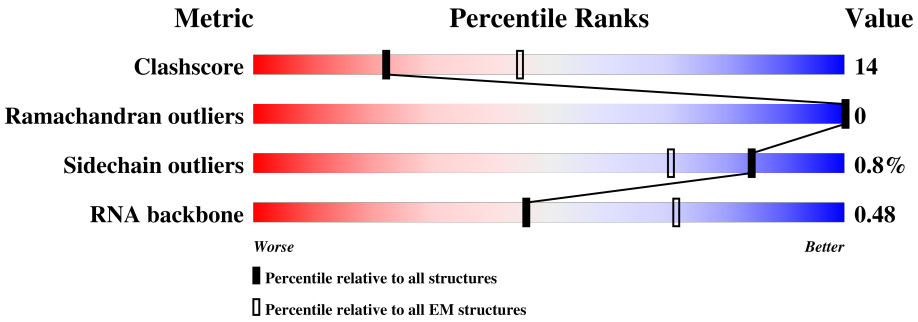
MolProbity	:	4.02b-467
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.4

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 3.60 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
RNA backbone	3747	458

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 <=5%

Mol	Chain	Length	Quality of chain
1	A	866	45% 38% 11%
2	B	45	22% 11% 64%
3	C	562	39% 34% 26%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 21539 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 RNA chain called T.el4h RNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
1	A	829	17776	7930	3275	5742	829	0	0

- Molecule 2 is a DNA chain called Sense Target DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
2	B	16	337	157	68	96	16	0	0

- Molecule 3 is a protein called Maturase reverse transcriptase.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	C	414	3393	2169	633	583	8	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	275	ASP	GLY	conflict	UNP Q8DMK2

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		AltConf
			Total	Mg	
4	A	32	32	32	0

- Molecule 5 is SODIUM ION (three-letter code: NA) (formula: Na).

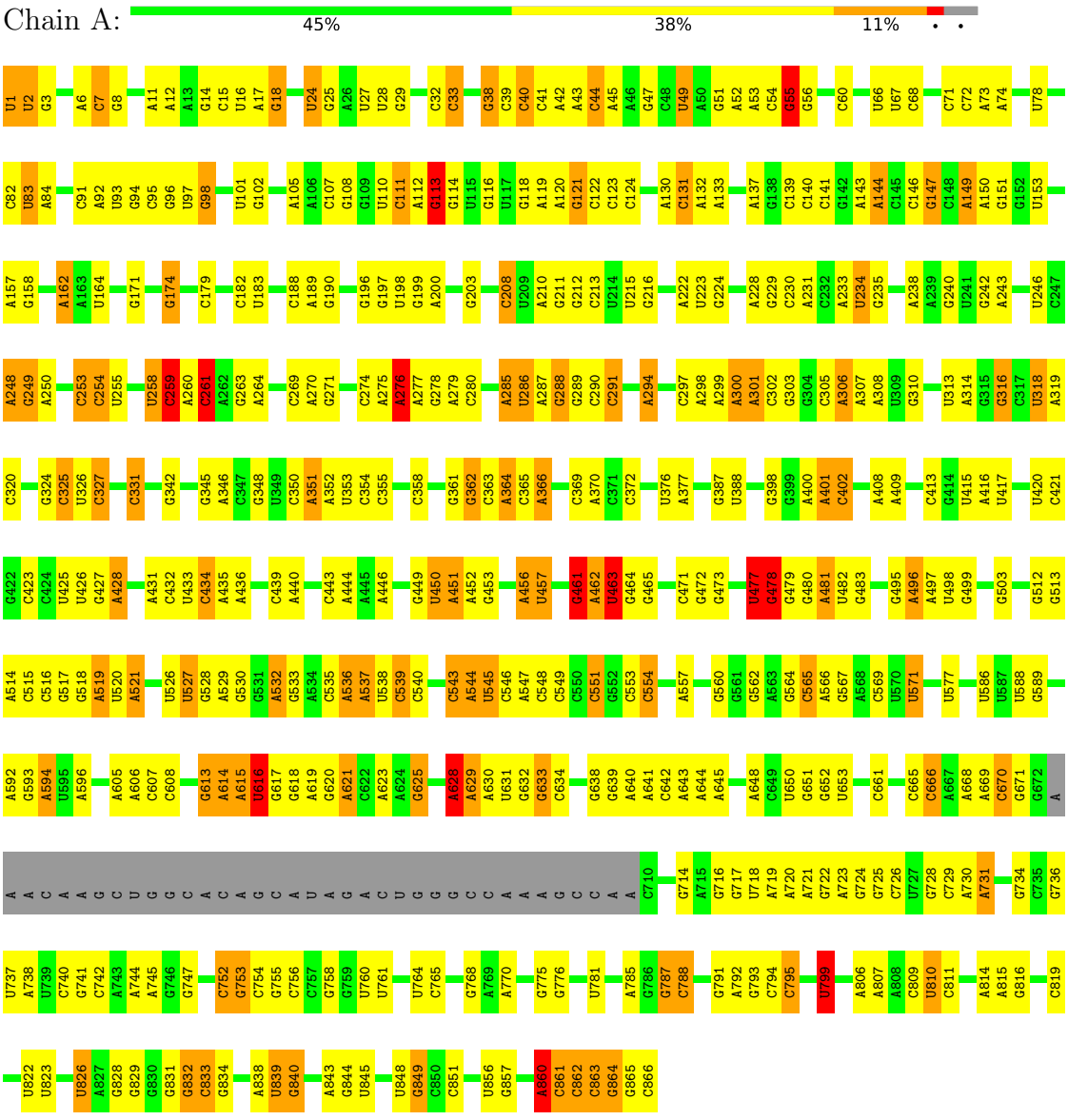
Mol	Chain	Residues	Atoms		AltConf
			Total	Na	
5	A	1	1	1	0

3 Residue-property plots (i)

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.

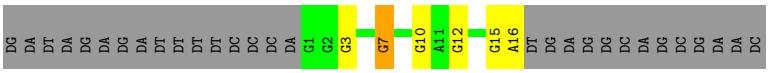
- Molecule 1: T.el4h RNA

Chain A:



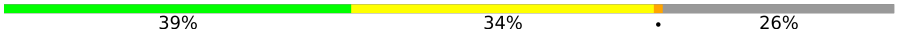
- Molecule 2: Sense Target DNA

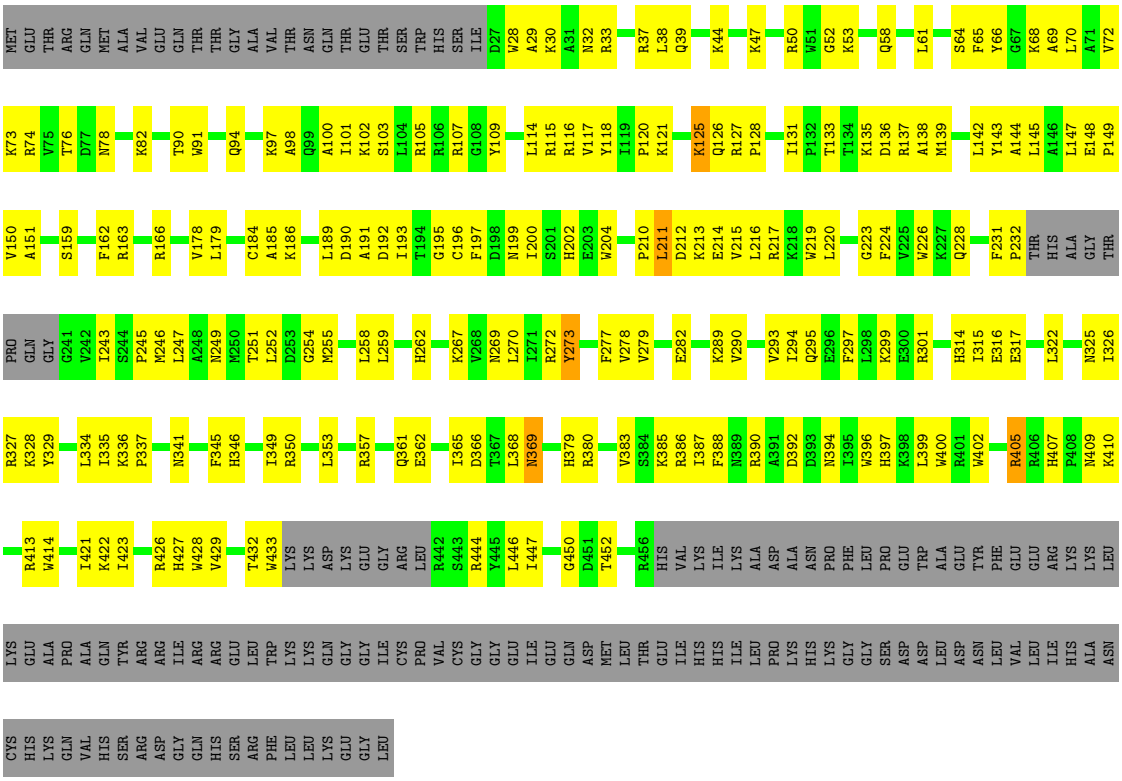
Chain B:



• Molecule 3: Maturase reverse transcriptase

Chain C:





4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	73144	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 ($e^-/Å^2$)	58.0	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: NA, MG

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	1.02	5/19901 (0.0%)	1.30	164/31039 (0.5%)
2	B	1.61	1/379 (0.3%)	1.09	2/585 (0.3%)
3	C	0.60	0/3468	0.72	1/4668 (0.0%)
All	All	0.98	6/23748 (0.0%)	1.24	167/36292 (0.5%)

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
3	C	0	5

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	1	U	O3'-P	-12.29	1.46	1.61
1	A	428	A	N9-C4	6.26	1.41	1.37
1	A	481	A	N9-C4	-5.96	1.34	1.37
2	B	7	DG	C3'-O3'	-5.42	1.36	1.44
1	A	276	A	N9-C4	-5.38	1.34	1.37

The worst 5 of 167 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	33	C	N3-C2-O2	-11.29	114.00	121.90
1	A	543	C	N1-C2-O2	11.15	125.59	118.90
1	A	543	C	C2-N1-C1'	10.01	129.81	118.80
1	A	33	C	C6-N1-C2	-9.37	116.55	120.30
1	A	540	C	N1-C2-O2	8.82	124.19	118.90

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	C	125	LYS	Peptide
3	C	185	ALA	Peptide
3	C	223	GLY	Peptide
3	C	273	TYR	Peptide
3	C	452	THR	Peptide

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	17776	0	8967	181	0
2	B	337	0	179	10	0
3	C	3393	0	3496	288	0
4	A	32	0	0	0	0
5	A	1	0	0	0	0
All	All	21539	0	12642	443	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 14.

The worst 5 of 443 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)
3:C:116:ARG:NE	3:C:226:TRP:CZ3	1.77	1.49
3:C:116:ARG:CZ	3:C:226:TRP:CZ3	1.98	1.43
3:C:163:ARG:NH1	3:C:166:ARG:HH12	1.20	1.38
1:A:1:U:O2	1:A:865:G:N1	1.70	1.25
3:C:366:ASP:OD1	3:C:444:ARG:NH2	1.69	1.24

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
3	C	408/562 (73%)	360 (88%)	48 (12%)	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
3	C	355/482 (74%)	352 (99%)	3 (1%)	83 93

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	369	ASN
3	C	380	ARG
3	C	405	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	269	ASN
3	C	427	HIS
3	C	346	HIS
3	C	262	HIS
3	C	341	ASN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	827/866 (95%)	250 (30%)	12 (1%)

5 of 250 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	2	U
1	A	3	G
1	A	6	A
1	A	11	A
1	A	12	A

5 of 12 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	A	545	U
1	A	628	A
1	A	832	G
1	A	543	C
1	A	764	U

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 carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 33 ligands modelled in this entry, 33 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 (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.