wwPDB EM Map/Model Validation Report (i)

Jun 16, 2016 – 06:18 PM EDT

PDB ID	:	5G2X
EMDB ID:	:	EMD-3331
Title	:	Structure a of Group II Intron Complexed with its Reverse Transcriptase
Authors	:	Qu, G.; Kaushal, P.S.; Wang, J.; Shigematsu, H.; Piazza, C.L.; Agrawal, R.K.; Belfort, M.; Wang, H.W.
Deposited on	:	2016-04-16
Resolution	:	3.80 Å(reported)

This is a wwPDB EM Map/Model Validation Report for a publicly released PDB/EMDB entry.
For rigid body fitted models, validation errors reported here could
stem from errors in the original structure(s) used in the fitting.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/EMValidationReportHelp

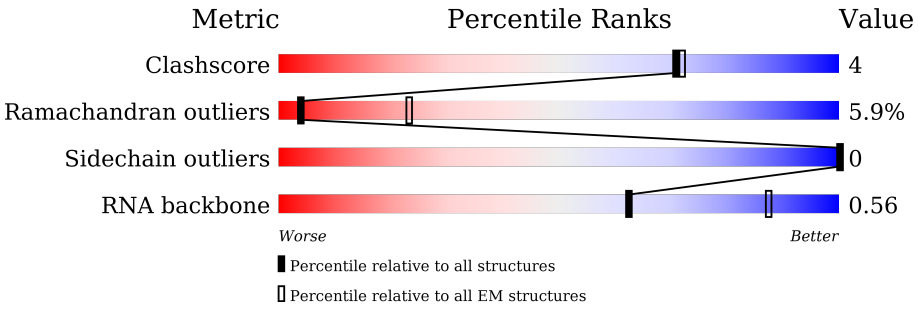
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027790

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 3.80 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	114402	924
Ramachandran outliers	111179	726
Sidechain outliers	111093	686
RNA backbone	3027	244

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	704	67% 28%
2	B	12	67% 33%
3	C	599	62% 18% 19%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 19072 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 GROUP II INTRON.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	692	Total	C	N	O	P	0	0
			14825	6637	2735	4762	691		

- Molecule 2 is a RNA chain called 5'-R(*CP*AP*CP*AP*UP*CP*CP*AP*UP*AP*AP*CP)-3'.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	12	Total	C	N	O	P	0	0
			250	113	44	81	12		

- Molecule 3 is a protein called GROUP II INTRON-ENCODED PROTEIN LTRA.

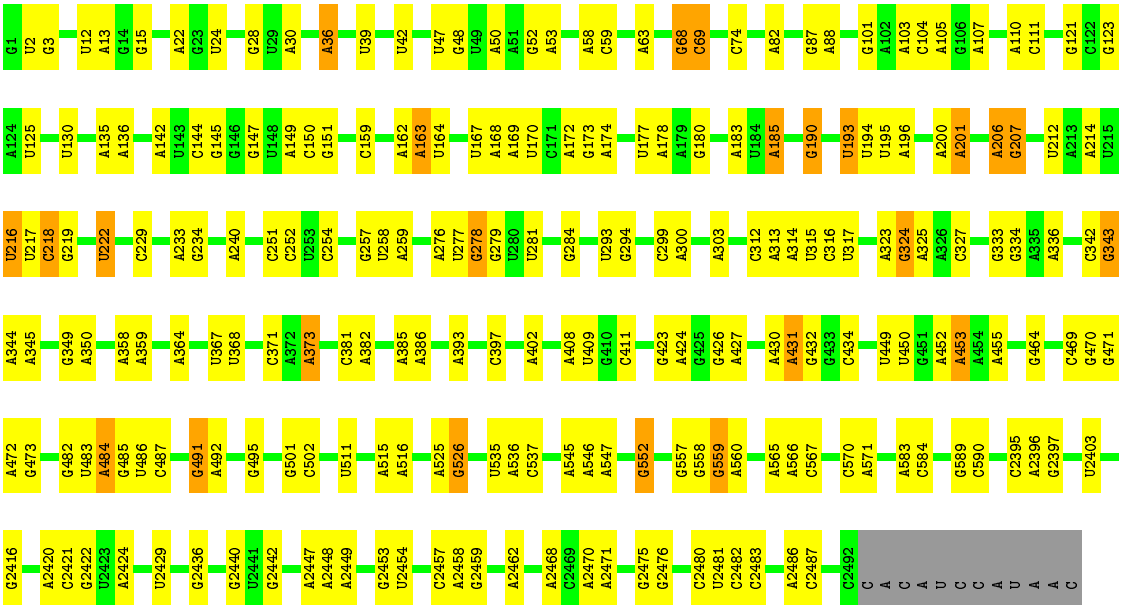
Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	487	Total	C	N	O	S	0	1
			3997	2570	688	720	19		

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 errors 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.

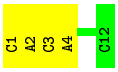
- Molecule 1: GROUP II INTRON

Chain A:



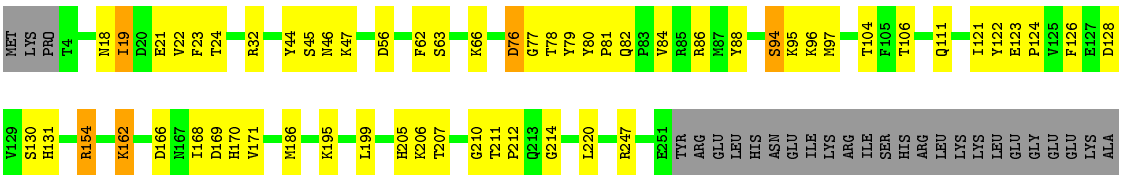
- Molecule 2: 5'-R(*CP*AP*CP*AP*UP*CP*CP*AP*UP*AP*AP*CP)-3'

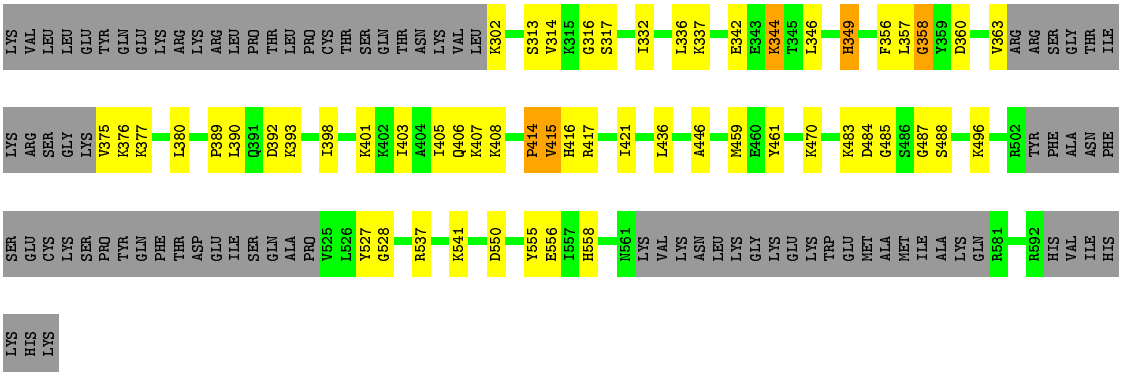
Chain B:



- Molecule 3: GROUP II INTRON-ENCODED PROTEIN LTRA

Chain C:




LYS VAL LEU LEU GLU TYR GLN GLU LYS ARG LYS ARG LEU PRO THR LEU PRO CYS THR SER GLN THR ASN LYS VAL LEU K302 S313 V314 K315 G316 S317 I332 L336 K337 E342 E343 K344 T345 L346 H349 F356 L357 G358 Y359 D360 V363 ARG ARG SER GLY THR ILE
LYS ARG SER GLY LYS V375 K376 K377 L380 P389 L390 Q391 D392 K393 I398 K401 K402 I403 A404 I405 Q406 K407 K408 P414 V415 H416 R417 I421 L436 A446 M459 E460 Y461 K470 K483 D484 G485 S486 G487 S488 K496 R502 TYR PHE ALA ASN PHE
SER GLU CYS LYS SER PRO TYR GLN PHE THR ASP GLU ILE SER GLN ALA PRO V525 L526 Y527 G528 R537 K541 D550 Y555 E556 I557 H558 N561 LYS VAL LYS ASN LEU LYS GLY LYS GLU LYS TRP GLU MET ALA MET ILE ALA LYS GLN R581 R592 HIS VAL ILE HIS
LYS HIS LYS

4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of particles used	Not provided	Depositor
Resolution determination method	Not provided	Depositor
CTF correction method	INDIVIDUAL PARTICLES	Depositor
Microscope	OTHER	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	Not provided	Depositor
Minimum defocus (nm)	1200	Depositor
Maximum defocus (nm)	3000	Depositor
Magnification	22500	Depositor
Image detector	GATAN K2 SUMMIT (4K 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	#$\|Z\|>2$	RMSZ	#$\|Z\|>2$
1	A	0.11	0/16614	0.65	0/25893
2	B	0.11	0/278	0.65	0/429
3	C	0.25	0/4080	0.42	0/5482
All	All	0.15	0/20972	0.62	0/31804

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	14825	0	7458	58	0
2	B	250	0	131	10	0
3	C	3997	0	4065	65	0
All	All	19072	0	11654	124	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 4.

The worst 5 of 124 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:B:1:C:H2'	2:B:2:A:H8	1.19	1.03

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:1:C:H2'	2:B:2:A:C8	2.01	0.95
2:B:1:C:P	3:C:380:LEU:HD22	2.11	0.90
2:B:1:C:O2'	2:B:2:A:H5'	1.89	0.72
3:C:18:ASN:HB2	3:C:19:ILE:HB	1.73	0.69

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
3	C	477/599 (80%)	355 (74%)	94 (20%)	28 (6%)	2 27

5 of 28 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	22	VAL
3	C	81	PRO
3	C	95	LYS
3	C	349	HIS
3	C	357	LEU

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.

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
3	C	441/545 (81%)	441 (100%)	0	100 100

There are no protein residues with a non-rotameric sidechain to report.

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

Mol	Chain	Res	Type
3	C	111	GLN
3	C	140	HIS
3	C	213	GLN
3	C	453	ASN

5.3.3 RNA

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	688/704 (97%)	146 (21%)	7 (1%)
2	B	11/12 (91%)	0	0
All	All	699/716 (97%)	146 (20%)	7 (1%)

5 of 146 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	2	U
1	A	3	G
1	A	12	U
1	A	13	A
1	A	28	G

5 of 7 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	A	177	U
1	A	2395	C
1	A	546	A
1	A	103	A
1	A	559	G

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

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.