wwPDB NMR Structure Validation Summary Report

Oct 16, 2023 – 11:49 AM EDT

PDB ID	:	8THV
BMRB ID	:	31099
Title	:	FARFAR-NMR ensemble of HIV-1 TAR with apical loop capturing ground and excited conformational states
Authors	:	Roy, R.; Geng, A.; Shi, H.; Merriman, D.K.; Dethoff, E.A.; Salmon, L.; Al-Hashimi, H.M.
Deposited on	:	2023-07-18

This is a wwPDB NMR Structure 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/NMRValidationReportHelp
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:

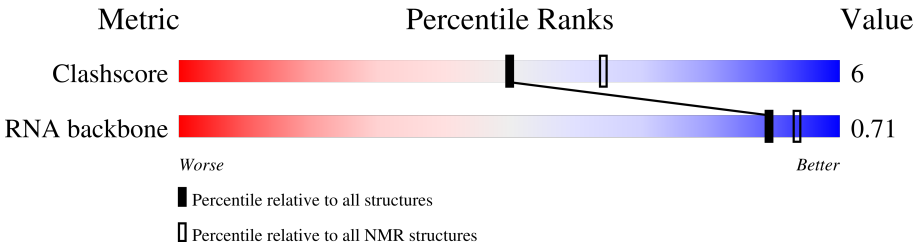
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 75%.

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)	NMR archive (#Entries)
Clashscore	158937	12864
RNA backbone	4643	676

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and 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	29	66% 31% .

2 Ensemble composition and analysis (i)

This entry contains 20 models. This entry does not contain polypeptide chains, therefore identification of well-defined residues and clustering analysis are not possible. All residues are included in the validation scores.

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 925 atoms, of which 311 are hydrogens and 0 are deuteriums.

- Molecule 1 is a RNA chain called RNA (29-MER).

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	A	29	925	275	311	109	202	28	0

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

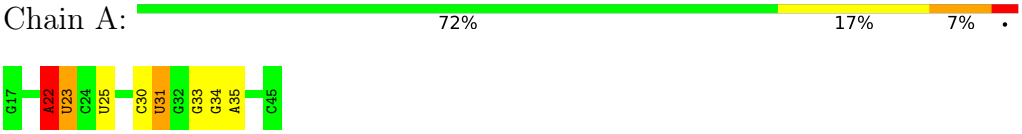
These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for 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 outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

- Molecule 1: RNA (29-MER)

4.2 Residue scores for the representative (author defined) model from the NMR ensemble

The representative model is number 1. Colouring as in section 4.1 above.

- Molecule 1: RNA (29-MER)



5 Refinement protocol and experimental data overview

The models were refined using the following method: *simulated annealing.*

Of the 20 calculated structures, 20 were deposited, based on the following criterion: *target function.*

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
Rosetta	structure calculation	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	503
Number of shifts mapped to atoms	500
Number of unparsed shifts	0
Number of shifts with mapping errors	3
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	75%

6 Model quality

6.1 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 (average) 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	0.46±0.32	0±1/685 (0.0± 0.1%)	0.93±0.31	3±14/1067 (0.3± 1.3%)
All	All	0.56	6/13700 (0.0%)	0.98	69/21340 (0.3%)

5 of 6 unique bond outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	34	G	O3'-P	-6.92	1.52	1.61	18	1
1	A	20	A	C6-N1	-6.10	1.31	1.35	9	1
1	A	35	A	C6-N1	-6.09	1.31	1.35	9	1
1	A	22	A	C6-N1	-6.05	1.31	1.35	9	1
1	A	27	A	C6-N1	-5.93	1.31	1.35	9	1

5 of 69 unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	22	A	P-O3'-C3'	9.59	131.20	119.70	1	1
1	A	24	C	OP1-P-OP2	-8.16	107.36	119.60	9	1
1	A	35	A	OP1-P-OP2	-7.52	108.32	119.60	9	1
1	A	41	C	OP1-P-OP2	-7.21	108.78	119.60	9	1
1	A	36	G	OP1-P-OP2	-6.69	109.56	119.60	9	1

There are no chirality outliers.

There are no planarity outliers.

6.2 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 each 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 averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	614	311	312	6±2
All	All	12280	6220	6240	111

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.

5 of 95 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:34:G:O2'	1:A:35:A:OP1	1.18	1.60	2	2
1:A:33:G:O4'	1:A:33:G:OP1	1.16	1.62	15	1
1:A:30:C:O2'	1:A:31:U:OP2	1.16	1.61	7	1
1:A:34:G:O2'	1:A:35:A:OP2	1.07	1.69	18	1
1:A:24:C:O2'	1:A:25:U:OP2	1.04	1.71	4	1

6.3 Torsion angles (i)

6.3.1 Protein backbone (i)

There are no protein molecules in this entry.

6.3.2 Protein sidechains (i)

There are no protein molecules in this entry.

6.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
1	A	28/29 (97%)	5±2 (18±6%)	1±1 (3±3%)	0.71±0.07
All	All	560/580 (97%)	99 (18%)	19 (3%)	0.71

The overall RNA backbone suiteness is 0.71.

5 of 14 unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	A	35	A	14
1	A	34	G	14

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	31	U	10
1	A	24	C	10
1	A	32	G	9

5 of 9 unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	A	34	G	3
1	A	24	C	3
1	A	23	U	3
1	A	22	A	2
1	A	29	C	2

6.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates

There are no monosaccharides in this entry.

6.6 Ligand geometry

There are no ligands in this entry.

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 75% for the well-defined parts and 75% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *chemical_shifts*

7.1.1 Bookkeeping

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	503
Number of shifts mapped to atoms	500
Number of unparsed shifts	0
Number of shifts with mapping errors	3
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. All 3 occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	17	G	H5'	4.283	0.002	.
1	A	17	G	H5"	4.407	0.002	.
1	A	17	G	P	-1.889	0.004	.

7.1.2 Chemical shift referencing

No chemical shift referencing corrections were calculated (not enough data).

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 75%, i.e. 425 atoms were assigned a chemical shift out of a possible 563. 0 out of 0 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Sugar	312/319 (98%)	168/174 (97%)	144/145 (99%)	0/0 (—%)
Base	113/244 (46%)	50/143 (35%)	41/48 (85%)	22/53 (42%)
Overall	425/563 (75%)	218/317 (69%)	185/193 (96%)	22/53 (42%)

7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots (i)

No *random coil index*(RCI) plot could be generated from the current chemical shift list. RCI is only applicable to proteins

8 NMR restraints analysis (i)

No restraints data found

9 Distance violation analysis (i)

No distance restraints data found

10 Dihedral-angle violation analysis (i)

No dihedral-angle restraints found