W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB NMR Structure Validation Summary Report ⓘ

Apr 20, 2024 – 08:44 AM EDT

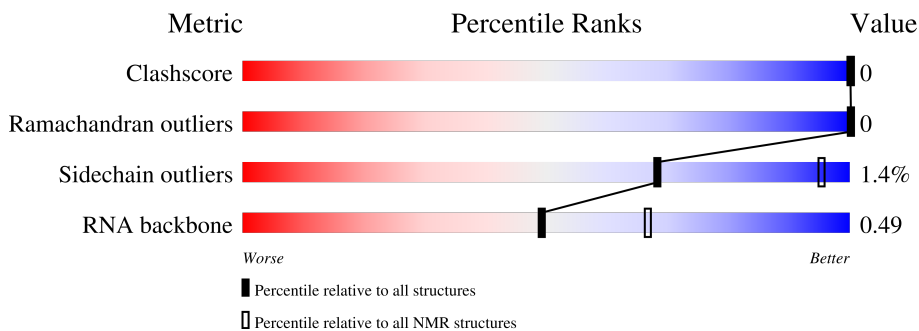|		
PDB ID	:	6MCF
BMRB ID	:	30511
Title	:	Solution structure of 7SK stem-loop 1 with HIV-1 Tat RNA Binding Domain
Authors	:	Pham, V.V.; D'Souza, V.M.
Deposited on	:	2018-08-31

1 Overall quality at a glance ⓘ

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

The overall completeness of chemical shifts assignment is 36%.

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	Percentile Ranks	Value
Clashscore		0
Ramachandran outliers		0
Sidechain outliers		1.4%
RNA backbone		0.49

Worse   *Better*

▮ Percentile relative to all structures

▯ Percentile relative to all NMR structures

Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428
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	57	16%	60%	23%	•
2	B	17	82%		18%	

2 Ensemble composition and analysis (i)

This entry contains 10 models.

Cyrange was unable to find well-defined residues.

Error message: Only domains with < 8 residues could be identified.

NmrClust was unable to cluster the ensemble.

Error message: Wrapper check: not enough residues in core to run NmrClust

3 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 2149 atoms, of which 784 are hydrogens and 0 are deuteriums.

- Molecule 1 is a RNA chain called 7SK Stem-loop 1 RNA.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	A	57	1832	541	619	215	401	56	0

- Molecule 2 is a protein called Protein Tat.

| Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
2	B	17	317	89	165	41	22	0

4   Residue-property plots ⓘ

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.
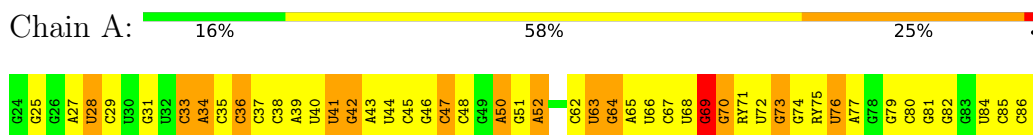
• Molecule 1: 7SK Stem-loop 1 RNA
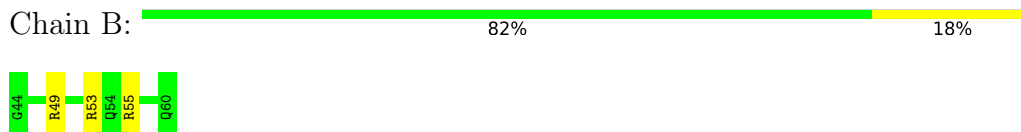
Chain A:

• Molecule 2: Protein Tat

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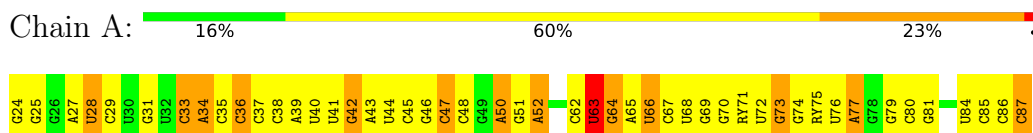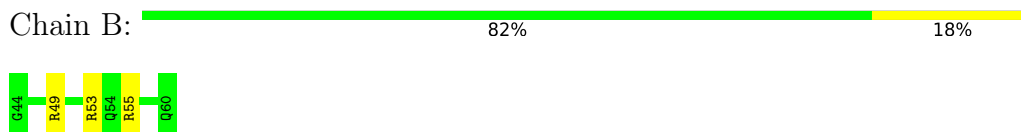
4.2   Residue scores for the representative (author defined) model from the NMR ensemble

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

• Molecule 1: 7SK Stem-loop 1 RNA

Chain A:

• Molecule 2: Protein Tat

Chain B:

5 Refinement protocol and experimental data overview ⓘ

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

Of the 20 calculated structures, 10 were deposited, based on the following criterion: *structures with the least restraint violations.*

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

Software name	Classification	Version
CYANA	structure calculation	
Amber	refinement	

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	517
Number of shifts mapped to atoms	517
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	36%

6 Model quality ⓘ

6.1 Standard geometry ⓘ

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

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	1.40±0.00	0±0/1308 (0.0± 0.0%)	2.28±0.02	80±4/2034 (3.9± 0.2%)
2	B	0.93±0.01	0±0/153 (0.0± 0.0%)	1.79±0.07	4±1/196 (1.9± 0.6%)
All	All	1.36	0/14610 (0.0%)	2.24	840/22300 (3.8%)

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	Planarity
1	A	0.0±0.0	12.2±1.2
All	All	0	122

There are no bond-length outliers.

5 of 124 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
2	B	53	ARG	NE-CZ-NH1	14.73	127.67	120.30	7	10
1	A	77	A	N1-C6-N6	-10.64	112.22	118.60	4	10
1	A	34	A	N1-C6-N6	-10.18	112.50	118.60	3	10
1	A	41	U	O4'-C1'-N1	10.05	116.24	108.20	4	3
1	A	43	A	N1-C6-N6	-9.57	112.86	118.60	9	10

There are no chirality outliers.

5 of 23 unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	28	U	Sidechain	10
1	A	46	G	Sidechain	10
1	A	50	A	Sidechain	10
1	A	63	U	Sidechain	10
1	A	25	G	Sidechain	9

6.2 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 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
All	All	13650	7840	7570	-

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 -.

There are no clashes.

6.3 Torsion angles (i)

6.3.1 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 NMR 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	
2	B	15/17 (88%)	11±1 (75±8%)	4±1 (25±8%)	0±0 (0±0%)	100	100
All	All	150/170 (88%)	113 (75%)	37 (25%)	0 (0%)	100	100

There are no Ramachandran outliers.

6.3.2 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 NMR 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	
2	B	14/14 (100%)	14±0 (99±3%)	0±0 (1±3%)	68	95
All	All	140/140 (100%)	138 (99%)	2 (1%)	68	95

All 2 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	B	53	ARG	1
2	B	56	ARG	1

6.3.3   RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
1	A	54/57 (95%)	8±1 (14±2%)	2±1 (4±3%)	0.49±0.02
All	All	540/570 (95%)	76 (14%)	24 (4%)	0.49

The overall RNA backbone suiteness is 0.49.

5 of 12 unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	A	63	U	10
1	A	64	G	10
1	A	72	U	10
1	A	41	U	9
1	A	76	U	8

5 of 9 unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	A	64	G	8
1	A	40	U	4
1	A	76	U	4
1	A	67	C	2
1	A	72	U	2

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

2 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
1	RY	A	71	1	17,21,22	2.05±0.02	5±0 (29±1%)
1	RY	A	75	1	17,21,22	2.01±0.01	5±0 (29±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
1	RY	A	71	1	20,30,33	1.61±0.09	6±1 (28±4%)
1	RY	A	75	1	20,30,33	1.57±0.09	4±1 (19±3%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
1	RY	A	71	1	-	0±0,7,38,39	0±0,2,2,2
1	RY	A	75	1	-	0±0,7,38,39	0±0,2,2,2

5 of 11 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	75	RY	C4-N3	5.42	1.29	1.37	6	10
1	A	71	RY	C4-N3	5.36	1.29	1.37	8	10
1	A	71	RY	C2-N1	3.79	1.41	1.35	7	10
1	A	71	RY	C6-C5	3.51	1.36	1.48	9	10

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	75	RY	C6-C5	3.37	1.36	1.48	9	10

5 of 15 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	75	RY	O4'-C1'-N1	4.49	115.42	109.30	2	10
1	A	71	RY	O4'-C1'-N1	3.28	113.77	109.30	8	9
1	A	75	RY	N3-C2-N1	3.26	120.10	116.65	2	10
1	A	71	RY	N3-C2-N1	3.23	120.07	116.65	7	10
1	A	75	RY	C4'-O4'-C1'	3.17	102.48	109.47	7	6

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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 36% for the well-defined parts and 36% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *starch_output*

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	517
Number of shifts mapped to atoms	517
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	2

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 36%, i.e. 480 atoms were assigned a chemical shift out of a possible 1324. 0 out of 0 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	53/87 (61%)	33/36 (92%)	10/34 (29%)	10/17 (59%)
Sidechain	121/180 (67%)	83/112 (74%)	29/46 (63%)	9/22 (41%)
Aromatic	6/16 (38%)	6/8 (75%)	0/7 (0%)	0/1 (0%)
Sugar	96/605 (16%)	56/330 (17%)	40/275 (15%)	0/0 (—%)
Base	204/436 (47%)	133/271 (49%)	43/90 (48%)	28/75 (37%)
Overall	480/1324 (36%)	311/757 (41%)	122/452 (27%)	47/115 (41%)

7.1.4 Statistically unusual chemical shifts ⓘ

The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	B	50	LYS	NZ	47.34	$19.79 - 46.09$	5.5
1	B	51	LYS	NZ	47.20	$19.79 - 46.09$	5.4

7.1.5 Random Coil Index (RCI) plots ⓘ

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain B: