wwPDB NMR Structure Validation Summary Report

Feb 10, 2022 – 11:07 AM EST

PDB ID	:	1F6U
Title	:	NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop sl2 of the psi-RNA packaging signal. Implications for genome recognition
Authors	:	Amarasinghe, G.K.; De Guzman, R.N.; Turner, R.B.; Chancellor, K.J.; Summers, M.F.
Deposited on	:	2000-06-23

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 ⓘ symbol.

The following versions of software and data (see references ⓘ) were used in the production of this report:

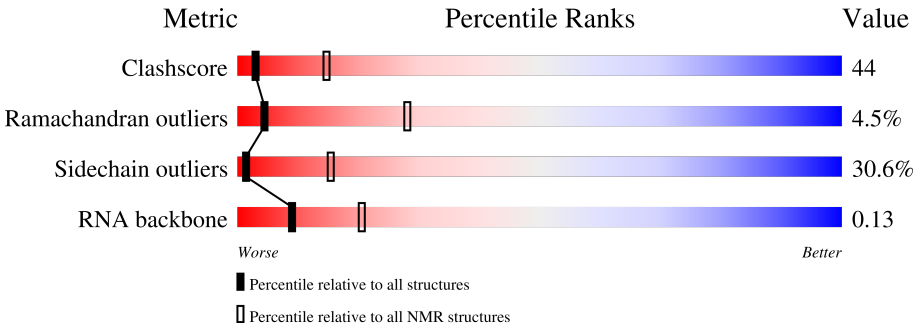
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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 was not calculated.

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
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	B	19	47% 53%
2	A	56	30% 52% 7% 11%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA and RNA chains that are outliers for geometric criteria:

Mol	Chain	Compound	Res	Total models with violations	
				Chirality	Geometry
1	B	CG1	201	5	-

2 Ensemble composition and analysis

This entry contains 20 models. Model 6 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	**Residue range (total)**	**Backbone RMSD (Å)**	**Medoid model**
1	A:2-A:51 (50)	0.76	6

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 3 clusters and 1 single-model cluster was found.

Cluster number	Models
1	1, 3, 4, 5, 6, 8, 9, 10, 11, 15, 17, 18, 19, 20
2	7, 13, 16
3	12, 14
Single-model clusters	2

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 1497 atoms, of which 643 are hydrogens and 0 are deuteriums.

- Molecule 1 is a RNA chain called HIV-1 STEM-LOOP SL2 FROM PSI-RNA PACKAGING.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	B	19	618	183	206	76	134	19	0

- Molecule 2 is a protein called HIV-1 NUCLEOCAPSID PROTEIN.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	A	56	877	263	437	95	74	8	1

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	3	LYS	ARG	conflict	UNP P35962

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
3	A	2	2	2

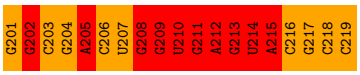
4 Residue-property plots

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.

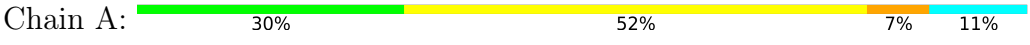
• Molecule 1: HIV-1 STEM-LOOP SL2 FROM PSI-RNA PACKAGING

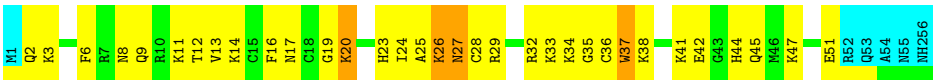
Chain B:



• Molecule 2: HIV-1 NUCLEOCAPSID PROTEIN

Chain A:





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

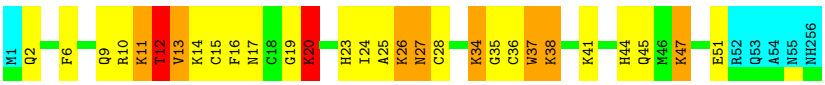
The representative model is number 6. Colouring as in section 4.1 above.

• Molecule 1: HIV-1 STEM-LOOP SL2 FROM PSI-RNA PACKAGING

Chain B:

• Molecule 2: HIV-1 NUCLEOCAPSID PROTEIN

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 400 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
DYANA	structure solution	1.5
DYANA	refinement	1.5

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, CG1, NH2

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	B	3.85±0.21	12±1/432 (2.8± 0.3%)	4.38±0.17	77±2/672 (11.4± 0.3%)
2	A	0.59±0.00	0±0/405 (0.0± 0.0%)	0.88±0.00	0±0/533 (0.0± 0.0%)
All	All	2.80	241/16740 (1.4%)	3.32	1534/24100 (6.4%)

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	B	2.6±1.1	0.0±0.0
All	All	53	0

5 of 16 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	B	211	G	C4'-O4'	-29.36	1.07	1.45	20	20
1	B	214	U	C4'-O4'	-29.11	1.07	1.45	2	20
1	B	208	G	C4'-O4'	-28.88	1.08	1.45	16	20
1	B	216	C	C4'-O4'	-28.75	1.08	1.45	16	20
1	B	204	G	C4'-O4'	-28.49	1.08	1.45	3	20

5 of 92 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	B	209	G	C5'-C4'-O4'	37.07	153.59	109.10	6	20
1	B	214	U	C5'-C4'-O4'	34.19	150.13	109.10	8	17

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	B	210	U	C5'-C4'-O4'	31.90	147.38	109.10	6	15
1	B	208	G	C5'-C4'-O4'	29.96	145.05	109.10	20	11
1	B	207	U	C5'-C4'-O4'	28.18	142.92	109.10	3	10

5 of 11 unique chiral outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Models (Total)
1	B	216	C	C4'	15
1	B	209	G	C4'	12
1	B	214	U	C4'	6
1	B	201	CG1	C4'	5
1	B	208	G	C4'	4

There are no planarity outliers.

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
1	B	412	206	209	39±7
2	A	398	396	396	34±5
All	All	16240	12040	12110	1239

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 44.

5 of 317 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:B:201:CG1:HB	1:B:201:CG1:H5'	0.91	1.41	17	3
1:B:201:CG1:C5'	1:B:201:CG1:HB	0.87	1.99	11	2
2:A:6:PHE:CE1	2:A:24:ILE:HD13	0.85	2.07	4	12
2:A:6:PHE:CZ	2:A:24:ILE:HD13	0.81	2.11	7	14
1:B:215:A:C6	1:B:216:C:C4	0.80	2.69	18	18

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	A	50/56 (89%)	40±2 (80±4%)	8±2 (16±4%)	2±1 (4±3%)	4 28
All	All	1000/1120 (89%)	795 (80%)	160 (16%)	45 (4%)	4 28

5 of 7 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	A	20	LYS	12
2	A	12	THR	10
2	A	32	ARG	6
2	A	33	LYS	6
2	A	34	LYS	5

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	A	42/46 (91%)	29±1 (69±3%)	13±1 (31±3%)	1 15
All	All	840/920 (91%)	583 (69%)	257 (31%)	1 15

5 of 31 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	A	26	LYS	20
2	A	27	ASN	20
2	A	37	TRP	15
2	A	20	LYS	14
2	A	38	LYS	14

6.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
1	B	18/19 (95%)	10±2 (56±11%)	7±2 (41±10%)	0.13±0.04
All	All	353/380 (93%)	202 (57%)	148 (42%)	0.14

The overall RNA backbone suiteness is 0.13.

5 of 17 unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	B	209	G	20
1	B	210	U	20
1	B	211	G	20
1	B	212	A	20
1	B	213	G	20

5 of 14 unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	B	212	A	20
1	B	209	G	19
1	B	211	G	19
1	B	214	U	19
1	B	210	U	15

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

1 non-standard protein/DNA/RNA residue is 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	CG1	B	201	1	23,27,27	2.08±0.78	2±1 (6±2%)

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	CG1	B	201	1	27,41,41	2.27±0.13	7±1 (27±2%)

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	CG1	B	201	1	-	0±0,9,29,29	0±0,3,3,3

All 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	B	201	CG1	O4'-C4'	13.47	1.14	1.45	2	15
1	B	201	CG1	C6-N1	2.07	1.36	1.33	20	17

5 of 8 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	B	201	CG1	O4'-C4'-C5'	7.61	134.42	109.37	13	19
1	B	201	CG1	O4'-C4'-C3'	5.88	93.47	105.11	4	10
1	B	201	CG1	C2-N3-C4	5.20	109.42	115.36	7	20
1	B	201	CG1	C5-C6-N1	4.44	117.36	123.43	14	20
1	B	201	CG1	N2-C2-N1	3.89	111.20	117.25	15	20

All unique chiral outliers are listed below.

Mol	Chain	Res	Type	Atoms	Models (Total)
1	B	201	CG1	C4'	5

There are no torsion outliers.

There are no ring outliers.

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation (i)

No chemical shift data were provided