wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 09:11 PM BST

PDB ID	:	2IHX
Title	:	Solution Structure of the Rous Sarcoma Virus Nucleocapsid Protein:uPsi RNA Packaging Signal Complex
Authors	:	Zhou, J.; Summers, M.
Deposited on	:	2006-09-27

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
http://wwpdb.org/validation/2016/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

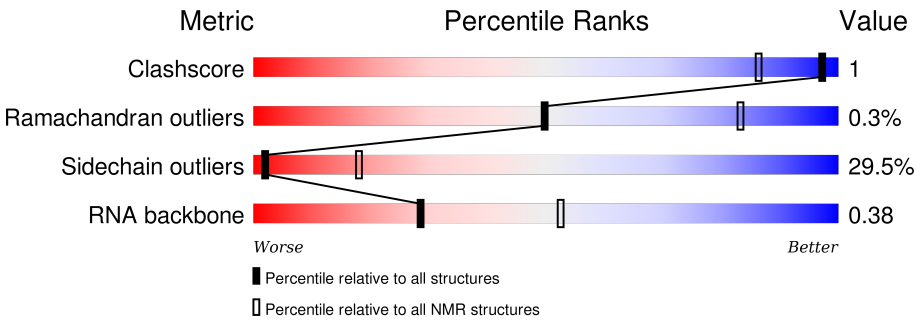
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

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 31%.

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958
RNA backbone	3027	600

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	75	65% 33% •
2	A	61	46% 11% • 23% 18%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 13 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:17-A:37 (21)	0.16	13
2	A:46-A:60 (15)	0.04	15

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 4 clusters and 5 single-model clusters were found.

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

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 3180 atoms, of which 1185 are hydrogens and 0 are deuteriums.

- Molecule 1 is a RNA chain called uPsi RNA.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	B	75	2418	715	810	290	528	75	0

- Molecule 2 is a protein called Nucleocapsid (NC) Protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	A	50	760	224	375	87	67	7	0

- 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 (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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: uPsi RNA

Chain B: 65% 33%

C160 U161 G162 C163 C164 C165 U166 C167 A168 U169 C170 C171 G172 U173 C174 U175 C176 G177 C178 U179 U180 A181 U182 U183 C184 G185 G186 G187 G188 A189 G190 C191 G192 G193 A194 C195 G196 A197 U198 G199 A200 C201 C202 C203 U204 A205 G206 U207 A208 G209 A210 G211 G212 G213 G214 G215 C216 U217 G218 C219

G220 G221 C222 U223 U224 A225 G226 G227 A228 G229 G230 G231 C232 A233 G234

- Molecule 2: Nucleocapsid (NC) Protein

Chain A: 46% 11% 23% 18%

G15 R16 A17 R18 G19 L20 Y30 K37 R38 K39 S40 G41 N42 S43 R44 E45 Q48 L49 A57 K58 R61 K62 R63 D64 GLY ASN GLN GLY GLN ARG PRO GLY LYS GLY LEU

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

The representative model is number 13. Colouring as in section 4.1 above.

- Molecule 1: uPsi RNA

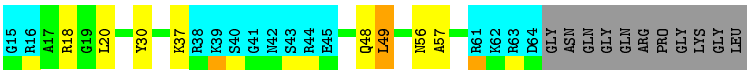
Chain B: 71% 29%

C160 U161 G162 C163 C164 C165 U166 C167 A168 U169 C170 C171 G172 U173 C174 U175 C176 G177 C178 U179 U180 A181 U182 U183 C184 G185 G186 G187 G188 A189 G190 C191 G192 G193 A194 C195 G196 A197 U198 G199 A200 C201 C202 C203 U204 A205 G206 U207 A208 G209 A210 G211 G212 G213 G214 G215 C216 U217 G218 C219

G220 G221 C222 U223 U224 A225 G226 G227 A228 G229 G230 G231 C232 A233 G234

- Molecule 2: Nucleocapsid (NC) Protein

Chain A: 46% 11% 23% 18%


G15
R16
A17
R18
G19
L20
Y30
K37
R38
K39
S40
G41
N42
S43
R44
E45
Q48
L49
N56
A57
R61
K62
R63
D64
GLY
ASN
GLN
GLY
GLN
ARG
PRO
GLY
LYS
GLY
LEU

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *1H-1H distance restraints, Hydrogen-bond restraints, Torsion angle restraints, Inter-phosphate restraints.*

Of the 800 calculated structures, 20 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	refinement	2.1

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)	BMRB entry 15113
Number of chemical shift lists	1
Total number of shifts	655
Number of shifts mapped to atoms	655
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	31%

No validations of the models with respect to experimental NMR restraints is performed at this time.

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

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	1.52±0.01	1±0/1797 (0.0±0.0%)	3.42±0.01	317±2/2802 (11.3±0.1%)
2	A	0.23±0.00	0±0/273 (0.0±0.0%)	0.49±0.00	0±0/365 (0.0±0.0%)
All	All	1.42	15/41400 (0.0%)	3.22	6331/63340 (10.0%)

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	B	219	C	C4'-O4'	-10.92	1.31	1.45	8	15

5 of 345 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	220	G	O4'-C4'-C3'	-14.35	89.65	104.00	18	20
1	B	198	U	O4'-C4'-C3'	-13.92	90.08	104.00	18	16
1	B	183	U	O4'-C4'-C3'	-13.89	90.11	104.00	14	19
1	B	168	A	O4'-C4'-C3'	-13.84	90.16	104.00	13	16
1	B	201	C	O4'-C4'-C3'	-13.83	90.17	104.00	9	20

There are no chirality outliers.

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	1608	810	811	2±2
2	A	267	252	252	1±0
All	All	37540	21240	21260	55

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 1.

5 of 12 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:A:49:LEU:HD12	2:A:57:ALA:HB1	0.64	1.70	3	20
1:B:219:C:O2'	1:B:220:G:OP2	0.61	2.19	11	1
1:B:223:U:O2'	1:B:224:U:O5'	0.59	2.19	10	5
1:B:223:U:C5'	1:B:224:U:OP1	0.49	2.61	10	5
1:B:218:G:O2'	1:B:219:C:OP1	0.47	2.32	18	6

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	36/61 (59%)	31±2 (87±5%)	5±2 (12±5%)	0±0 (0±1%)	50	83
All	All	720/1220 (59%)	628 (87%)	90 (12%)	2 (0%)	50	83

All 2 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	28	GLY	1
2	A	18	ARG	1

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	28/47 (60%)	20±2 (71±6%)	8±2 (29±6%)	2 18
All	All	560/940 (60%)	395 (71%)	165 (29%)	2 18

5 of 16 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	49	LEU	20
2	A	20	LEU	20
2	A	30	TYR	20
2	A	18	ARG	15
2	A	48	GLN	15

6.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
1	B	74/75 (99%)	24±2 (32±2%)	2±1 (3±1%)	0.37±0.03
All	All	1480/1500 (99%)	479 (32%)	39 (3%)	0.37

The overall RNA backbone suiteness is 0.38.

5 of 34 unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	B	176	C	20
1	B	207	U	20
1	B	220	G	20
1	B	208	A	20
1	B	206	G	20

5 of 9 unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	B	218	G	15
1	B	219	C	5
1	B	223	U	5
1	B	198	U	4
1	B	182	U	3

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

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

6.5 Carbohydrates (i)

There are no carbohydrates 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)

The completeness of assignment taking into account all chemical shift lists is 31% for the well-defined parts and 30% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 15113

Chemical shift list name: *assigned_ chem_ shift_ list_ 1*

7.1.1 Bookkeeping (i)

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	655
Number of shifts mapped to atoms	655
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	28	1.46 ± 0.94	None needed (imprecise)
$^{13}C_{\beta}$	25	2.33 ± 0.51	Should be applied
$^{13}C'$	0	—	—
^{15}N	0	—	—

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 31%, i.e. 581 atoms were assigned a chemical shift out of a possible 1845. 2 out of 2 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	53/176 (30%)	28/70 (40%)	25/72 (35%)	0/34 (0%)
Sidechain	118/216 (55%)	76/132 (58%)	42/69 (61%)	0/15 (0%)

Continued on next page...

Continued from previous page...

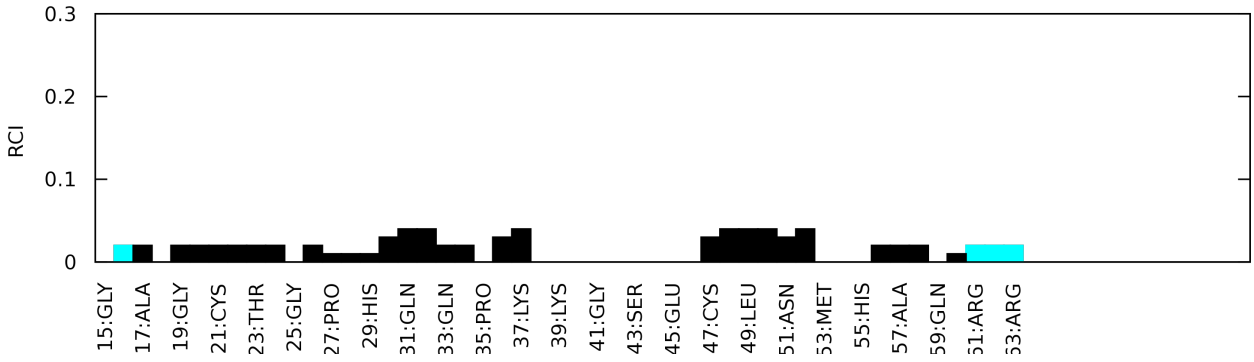
	Total	^{1}H	^{13}C	^{15}N
Aromatic	15/30 (50%)	8/16 (50%)	7/12 (58%)	0/2 (0%)
Overall	581/1845 (31%)	507/1041 (49%)	74/650 (11%)	0/154 (0%)

7.1.4 Statistically unusual chemical shifts (i)

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.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	36	LYS	CD	52.64	34.86 – 23.06	20.1
2	B	220	G	H3'	5.78	5.68 – 3.38	5.4
2	B	219	C	H5'	2.86	5.59 – 2.89	-5.1

7.1.5 Random Coil Index (RCI) plots (i)

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.

Random coil index (RCI) for chain A: