



# wwPDB NMR Structure Validation Summary Report ⓘ

May 23, 2016 – 06:45 PM EDT

PDB ID : 2N82  
Title : solution structure of the complex of microRNA 20b pre-element with Rbfox RRM  
Authors : Yang, F.; Chen, Y.; Varani, G.  
Deposited on : 2015-09-30

This is a wwPDB NMR Structure Validation Summary Report for a publicly released PDB entry.  
We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/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:

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-20027674  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027674

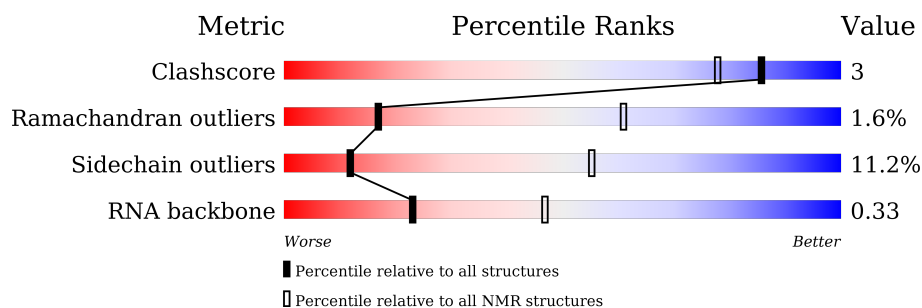
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

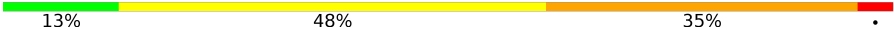

The overall completeness of chemical shifts assignment is 79%.

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  $\geq 3$ , 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	23	
2	B	104	

## 2 Ensemble composition and analysis

This entry contains 20 models. Model 14 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

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	B:116-B:192 (77)	0.25	14

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 3 single-model clusters were found.

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

### 3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2385 atoms, of which 1070 are hydrogens and 0 are deuteriums.

- Molecule 1 is a RNA chain called RNA (5'-R(\*GP\*GP\*UP\*AP\*GP\*UP\*UP\*UP\*UP\*GP\*GP\*CP\*AP\*UP\*GP\*AP\*CP\*UP\*CP\*UP\*AP\*CP\*C)-3').

Mol	Chain	Residues	Atoms						Trace
1	A	23	Total	C	H	N	O	P	0
			730	217	247	81	163	22	

- Molecule 2 is a protein called RNA binding protein fox-1 homolog 1.

Mol	Chain	Residues	Atoms						Trace
2	B	104	Total	C	H	N	O	S	0
			1655	516	823	158	155	3	

There are 4 discrepancies between the modelled and reference sequences:

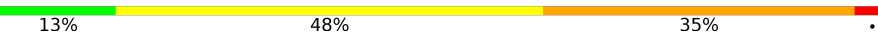
Chain	Residue	Modelled	Actual	Comment	Reference
B	105	GLY	-	EXPRESSION TAG	UNP Q9NWB1
B	106	SER	-	EXPRESSION TAG	UNP Q9NWB1
B	107	HIS	-	EXPRESSION TAG	UNP Q9NWB1
B	108	MET	-	EXPRESSION TAG	UNP Q9NWB1

## 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: RNA (5'-R(\*GP\*GP\*UP\*AP\*GP\*UP\*UP\*UP\*UP\*GP\*GP\*CP\*AP\*UP\*GP\*AP\*CP\*UP\*CP\*UP\*AP\*CP\*C)-3')

Chain A: 



- Molecule 2: RNA binding protein fox-1 homolog 1

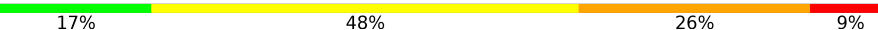
Chain B: 



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

The representative model is number 14. Colouring as in section 4.1 above.

- Molecule 1: RNA (5'-R(\*GP\*GP\*UP\*AP\*GP\*UP\*UP\*UP\*UP\*GP\*GP\*CP\*AP\*UP\*GP\*AP\*CP\*UP\*CP\*UP\*AP\*CP\*C)-3')

Chain A: 



- Molecule 2: RNA binding protein fox-1 homolog 1

Chain B: 



## 5 Refinement protocol and experimental data overview

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

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy*.

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

Software name	Classification	Version
CYANA	structure solution	
AMBER	refinement	
TALOS	geometry optimization	

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)	2n82_cs.cif
Number of chemical shift lists	1
Total number of shifts	1520
Number of shifts mapped to atoms	1520
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	79%

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)

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	1.48±0.02	1±1/538 (0.3±0.2%)	2.26±0.05	31±4/836 (3.7±0.5%)
2	B	0.71±0.01	0±0/636 (0.0±0.0%)	1.21±0.04	5±2/852 (0.6±0.2%)
All	All	1.13	29/23480 (0.1%)	1.81	718/33760 (2.1%)

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
2	B	0.0±0.0	0.7±0.7
All	All	0	13

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	A	26	U	N1-C2	7.05	1.44	1.38	17	13
1	A	30	C	C4-N4	-6.07	1.28	1.33	8	7
1	A	33	G	C2-N2	-5.62	1.28	1.34	8	8
1	A	30	C	N3-C4	-5.18	1.30	1.33	16	1

5 of 82 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(°)	Ideal(°)	Models	
								Worst	Total
1	A	26	U	N3-C2-O2	-15.51	111.34	122.20	17	20
1	A	26	U	N1-C2-O2	14.35	132.84	122.80	17	17
1	A	34	A	N1-C6-N6	-12.82	110.91	118.60	1	20
1	A	31	A	C5-C6-N1	10.59	123.00	117.70	5	20
1	A	28	G	N3-C4-C5	-10.52	123.34	128.60	19	17

There are no chirality outliers.

5 of 6 unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
2	B	150	PHE	Sidechain	5
2	B	160	PHE	Sidechain	4
2	B	129	ARG	Sidechain	1
2	B	134	ARG	Sidechain	1
2	B	171	ARG	Sidechain	1

## 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	A	483	247	245	3±1
2	B	623	617	617	4±2
All	All	22120	17280	17250	106

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 3.

5 of 40 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:149:ILE:HD11	2:B:160:PHE:CD2	0.63	2.29	9	8
1:A:33:G:C4	2:B:160:PHE:CE2	0.62	2.88	17	14
2:B:149:ILE:HD11	2:B:160:PHE:CD1	0.58	2.34	13	2
1:A:33:G:C8	2:B:160:PHE:CE1	0.57	2.92	1	3
1:A:30:C:C5	2:B:156:LYS:HA	0.57	2.35	17	9

## 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	77/104 (74%)	71±2 (92±2%)	5±2 (6±3%)	1±1 (2±1%)	17	61
All	All	1540/2080 (74%)	1422 (92%)	93 (6%)	25 (2%)	17	61

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	B	152	GLU	10
2	B	153	ARG	4
2	B	156	LYS	3
2	B	182	GLU	3
2	B	123	ASN	3

### 6.3.2 Protein sidechains ⓘ

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	66/90 (73%)	59±2 (89±3%)	7±2 (11±3%)	12	55
All	All	1320/1800 (73%)	1172 (89%)	148 (11%)	12	55

5 of 27 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	192	THR	19
2	B	176	LEU	17
2	B	180	VAL	13
2	B	128	PHE	13
2	B	179	THR	11

### 6.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
1	A	22/23 (96%)	9±1 (41±6%)	3±1 (12±5%)	0.33±0.06
All	All	443/460 (96%)	180 (41%)	51 (12%)	0.33

The overall RNA backbone suiteness is 0.33.

5 of 18 unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	A	26	U	20
1	A	30	C	20
1	A	27	U	18
1	A	28	G	18
1	A	34	A	17

5 of 10 unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	A	26	U	12
1	A	30	C	9
1	A	33	G	8
1	A	32	U	6
1	A	34	A	4

## 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](#)

There are no ligands in this entry.

## 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

The completeness of assignment taking into account all chemical shift lists is 79% for the well-defined parts and 79% for the entire structure.

### 7.1 Chemical shift list 1

File name: 2n82\_cs.cif

Chemical shift list name: *assigned\_chem\_shift\_list\_1*

#### 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	1520
Number of shifts mapped to atoms	1520
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	11

#### 7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	103	$-0.27 \pm 0.09$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}_\beta$	95	$0.11 \pm 0.17$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}'$	101	$-0.18 \pm 0.10$	None needed ( $< 0.5$ ppm)
$^{15}\text{N}$	97	$-0.01 \pm 0.46$	None needed ( $< 0.5$ ppm)

#### 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 79%, i.e. 1123 atoms were assigned a chemical shift out of a possible 1429. 9 out of 10 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	379/379 (100%)	151/151 (100%)	154/154 (100%)	74/74 (100%)
Sidechain	433/537 (81%)	265/316 (84%)	155/185 (84%)	13/36 (36%)

*Continued on next page...*

Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	60/86 (70%)	41/48 (85%)	19/36 (53%)	0/2 (0%)
Overall	1123/1429 (79%)	591/758 (78%)	445/530 (84%)	87/141 (62%)

#### 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.

Mol	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
2	B	129	ARG	NE	115.04	92.63 – 76.73	19.1
2	B	127	ARG	NE	115.01	92.63 – 76.73	19.1
2	B	173	ARG	NE	114.26	92.63 – 76.73	18.6
2	B	194	ARG	NE	114.02	92.63 – 76.73	18.5
2	B	118	ARG	NE	113.61	92.63 – 76.73	18.2
2	B	153	ARG	NE	113.09	92.63 – 76.73	17.9
2	B	184	ARG	NE	109.06	92.63 – 76.73	15.3
1	A	30	C	H3'	3.21	5.20 – 3.60	-7.4
2	B	132	ASP	HB2	0.87	4.07 – 1.37	-6.9
2	B	125	PRO	HD2	1.67	5.45 – 1.85	-5.5
1	A	33	G	H2'	5.75	5.66 – 3.46	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.

Random coil index (RCI) for chain B:

