



# wwPDB NMR Structure Validation Summary Report ⓘ

Apr 26, 2016 – 11:39 PM BST

PDB ID : 2KM8  
Title : Interdomain RRM packing contributes to RNA recognition in the rna15, hrp1, anchor RNA 3' processing ternary complex  
Authors : Leeper, T.C.; Varani, G.  
Deposited on : 2009-07-24

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-20027457  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027457

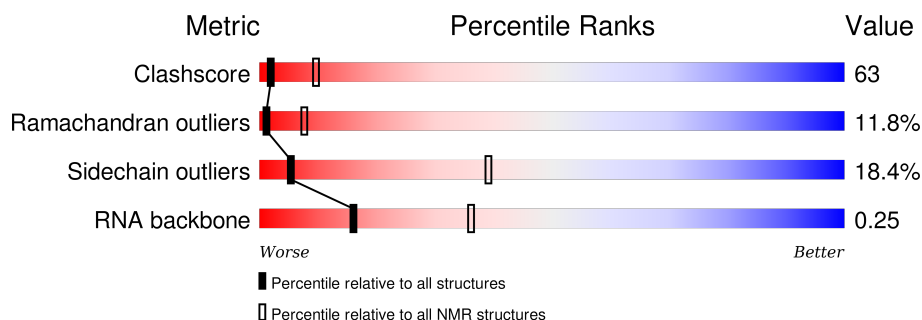
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 25%.

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	13	 77% 23%
2	B	84	 14% 65% 17% ..
3	C	167	 23% 62% 12% .

## 2 Ensemble composition and analysis

This entry contains 10 models. Model 1 is the overall representative, medoid model (most similar to other models).

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:22-B:103, C:158-C:319 (244)	1.31	1

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. No single-model clusters were found.

Cluster number	Models
1	1, 3, 4, 6
2	7, 8
3	2, 5
4	9, 10

### 3 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4375 atoms, of which 2102 are hydrogens and 0 are deuteriums.

- Molecule 1 is a RNA chain called 5'-R(P\*UP\*AP\*UP\*AP\*UP\*AP\*UP\*AP\*AP\*UP\*AP\*AP\*U)-3'.

Mol	Chain	Residues	Atoms						Trace
1	A	13	Total	C	H	N	O	P	0
			414	124	139	47	91	13	

- Molecule 2 is a protein called mRNA 3'-end-processing protein RNA15.

Mol	Chain	Residues	Atoms						Trace
2	B	84	Total	C	H	N	O	S	0
			1296	411	640	112	129	4	

- Molecule 3 is a protein called Nuclear polyadenylated RNA-binding protein 4.


Mol	Chain	Residues	Atoms						Trace
3	C	167	Total	C	H	N	O	S	0
			2665	849	1323	232	256	5	

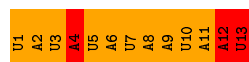
## 4 Residue-property plots

### 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.

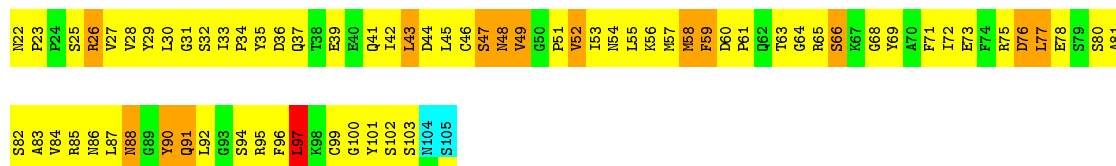
- Molecule 1: 5'-R(P\*UP\*AP\*UP\*AP\*UP\*AP\*UP\*AP\*AP\*UP\*AP\*AP\*U)-3'

Chain A: 



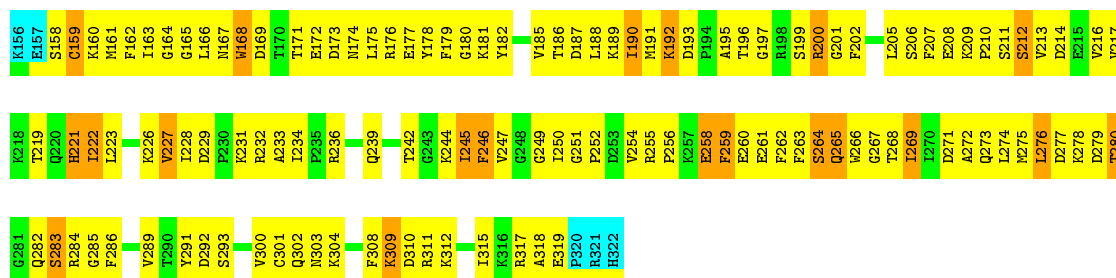
- Molecule 2: mRNA 3'-end-processing protein RNA15

Chain B: 



- Molecule 3: Nuclear polyadenylated RNA-binding protein 4


Chain C: 



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

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

- Molecule 1: 5'-R(P\*UP\*AP\*UP\*AP\*UP\*AP\*UP\*AP\*AP\*UP\*AP\*AP\*U)-3'

Chain A:  15% 54% 31%

U1  
A2  
U3  
A4  
U5  
A6  
U7  
A8  
A9  
U10  
A11  
A12  
U13

- Molecule 2: mRNA 3'-end-processing protein RNA15

Chain B:  26% 49% 21% ..

R22  
S25  
R26  
V27  
Y29  
L30  
G31  
S32  
I33  
P34  
Y35  
D36  
Q37  
T38  
E39  
E40  
Q41  
I42  
L43  
D44  
L45  
C46  
S47  
R48  
V49  
V52  
I53  
N54  
L55  
K56  
M57  
M58  
F59  
D60  
P61  
R65  
S66  
K67  
G68  
Y69  
A70  
F71  
I72  
E73  
F74  
R75  
D76  
L77  
E78  
A81  
S82  
A83  
V84  
R85  
N86

L87  
M88  
G89  
Y90  
Q91  
L92  
G93  
S94  
F96  
L97  
G100  
Y101  
S102  
S103  
N104  
S105

- Molecule 3: Nuclear polyadenylated RNA-binding protein 4

Chain C:  37% 46% 12% ..

K156  
E157  
S158  
C159  
K160  
M161  
F162  
G165  
L166  
M167  
V168  
D169  
T170  
T171  
M174  
L175  
R176  
E177  
Y178  
F179  
G180  
K181  
Y182  
D187  
F188  
K189  
Q190  
M191  
K192  
D193  
T196  
S199  
R200  
G201  
F202  
L205  
S206  
K209  
P210  
S211  
Q212  
S213  
V213  
D214  
E215  
V216  
V217  
H221  
I222  
L223  
D224  
G225  
K226

V227  
I228  
D229  
R232  
A233  
D237  
E238  
Q239  
T242  
G243  
K244  
T245  
F246  
V247  
G248  
G249  
V254  
E258  
F259  
E260  
E261  
F262  
F263  
S264  
Q265  
V266  
G267  
T268  
I269  
D271  
A272  
Q273  
L274  
M275  
L276  
D277  
K278  
D279  
T280  
G281  
Q282  
S283  
R284  
G285  
F286  
G287  
F288  
V289  
T290  
Y291  
D292  
R299  
V300

C301  
Q302  
M303  
K304  
F308  
K309  
D310  
R311  
E314  
I315  
K316  
R317  
A318  
E319  
P320  
R321  
H322

## 5 Refinement protocol and experimental data overview

The models were refined using the following method: *torsion angle dynamics*.

Of the 100 calculated structures, 10 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
X-PLOR NIH	refinement	
CYANA	structure solution	

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 16425
Number of chemical shift lists	1
Total number of shifts	922
Number of shifts mapped to atoms	922
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	25%

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	1.12±0.10	0±1/307 (0.1±0.4%)	1.65±0.05	11±2/473 (2.2±0.4%)
2	B	1.05±0.02	0±0/655 (0.0±0.0%)	0.90±0.02	0±0/886 (0.0±0.0%)
3	C	1.05±0.00	0±0/1322 (0.0±0.0%)	0.89±0.00	0±0/1776 (0.0±0.0%)
All	All	1.06	5/22840 (0.0%)	1.04	106/31350 (0.3%)

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	0.1±0.3
All	All	0	1

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	11	A	O3'-P	8.23	1.71	1.61	7	1
1	A	12	A	P-O5'	6.47	1.66	1.59	7	1
1	A	11	A	N9-C4	-6.31	1.34	1.37	7	1
1	A	11	A	C3'-O3'	5.52	1.49	1.42	7	1
2	B	68	GLY	N-CA	5.20	1.53	1.46	7	1

5 of 24 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	9	A	O4'-C1'-N9	10.28	116.43	108.20	5	7
1	A	12	A	O4'-C1'-N9	7.37	114.10	108.20	7	1
1	A	9	A	C1'-O4'-C4'	-6.79	104.46	109.90	5	2

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	8	A	C5'-C4'-C3'	-6.77	105.17	116.00	10	6
1	A	3	U	O4'-C1'-N1	6.53	113.42	108.20	2	1

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	4	A	Sidechain	1

## 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	275	139	138	98±23
2	B	641	629	626	99±25
3	C	1295	1274	1273	134±14
All	All	22110	20420	20370	2660

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 63.

5 of 1466 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:11:A:H3'	2:B:69:TYR:N	1.31	1.37	7	1
1:A:6:A:H4'	1:A:7:U:H5'	1.13	1.17	9	7
1:A:12:A:H5'	2:B:67:LYS:O	1.09	1.47	7	1
1:A:12:A:C5'	2:B:67:LYS:O	1.07	2.00	7	1
1:A:10:U:H5'	2:B:71:PHE:CE1	1.03	1.86	5	4

## 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	81/84 (96%)	56±2 (69±3%)	13±2 (16±3%)	12±2 (15±2%)	<b>1</b> <b>5</b>
3	C	162/167 (97%)	119±3 (74±2%)	26±3 (16±2%)	17±2 (10±1%)	<b>1</b> <b>10</b>
All	All	2430/2510 (97%)	1752 (72%)	392 (16%)	286 (12%)	<b>1</b> <b>8</b>

5 of 83 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	77	LEU	10
2	B	26	ARG	10
2	B	91	GLN	10
2	B	76	ASP	10
2	B	47	SER	10

### 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	72/74 (97%)	58±2 (80±3%)	14±2 (20±3%)	<b>5</b> <b>36</b>
3	C	140/145 (97%)	115±3 (82±2%)	25±3 (18±2%)	<b>6</b> <b>41</b>
All	All	2120/2190 (97%)	1730 (82%)	390 (18%)	<b>5</b> <b>39</b>

5 of 129 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	90	TYR	10
3	C	276	LEU	9

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
3	C	171	THR	9
2	B	37	GLN	9
3	C	222	ILE	9

### 6.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
1	A	12/13 (92%)	6±1 (46±11%)	1±1 (8±8%)	0.25±0.07
All	All	120/130 (92%)	55 (46%)	9 (8%)	0.25

The overall RNA backbone suiteness is 0.25.

5 of 11 unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	A	3	U	8
1	A	13	U	7
1	A	4	A	6
1	A	10	U	6
1	A	12	A	6

All unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	A	5	U	3
1	A	6	A	3
1	A	10	U	3

## 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 carbohydrates in this entry.

## 6.6 Ligand geometry ⓘ

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 25% for the well-defined parts and 25% for the entire structure.

### 7.1 Chemical shift list 1

File name: BMRB entry 16425

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	922
Number of shifts mapped to atoms	922
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

#### 7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	80	$-0.14 \pm 0.15$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}_\beta$	65	$0.11 \pm 0.13$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}'$	77	$-0.38 \pm 0.19$	None needed ( $< 0.5$ ppm)
$^{15}\text{N}$	78	$0.61 \pm 0.72$	None needed (imprecise)

#### 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 25%, i.e. 812 atoms were assigned a chemical shift out of a possible 3265. 9 out of 30 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	387/1198 (32%)	158/477 (33%)	153/488 (31%)	76/233 (33%)
Sidechain	392/1583 (25%)	264/933 (28%)	119/568 (21%)	9/82 (11%)

*Continued on next page...*

Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	33/249 (13%)	33/133 (25%)	0/112 (0%)	0/4 (0%)
Overall	812/3265 (25%)	455/1674 (27%)	272/1259 (22%)	85/332 (26%)

#### 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
2	B	23	PRO	HA	1.69	6.05 – 2.75	-8.2

#### 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 B:

