wwPDB NMR Structure Validation Summary Report

Apr 21, 2024 – 10:30 AM EDT

PDB ID	:	2MQQ
BMRB ID	:	25043
Title	:	Structural Investigation of hnRNP L bound to RNA
Authors	:	Blatter, M.; Allain, F.
Deposited on	:	2014-06-24

This is a wwPDB NMR Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

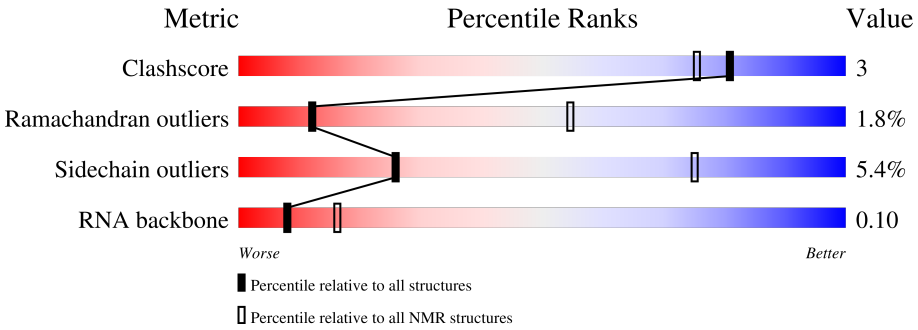
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 76%.

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	A	215	90% 5% • •
2	B	5	20% 40% 40%
2	C	5	40% 40% 20%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 2 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:374-A:565, A:571-A:586 (208)	0.65	2

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 2 clusters and 2 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 3631 atoms, of which 1735 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Heterogenous nuclear ribonucleoprotein L.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	215	3315	1065	1625	288	322	15	0

- Molecule 2 is a RNA chain called RNA (5'-R(*AP*CP*AP*CP*A)-3').

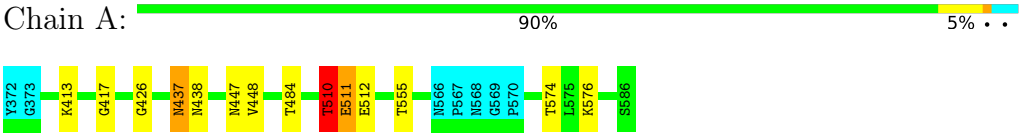
Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	5	158	48	55	21	30	4	0
			Total	C	H	N	O	P	
2	C	5	158	48	55	21	30	4	0

4 Residue-property plots

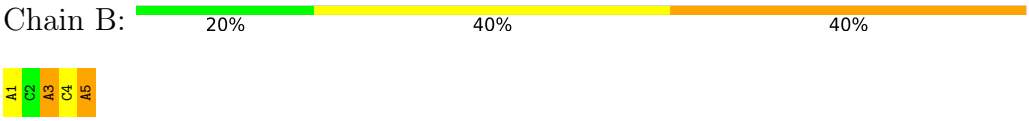
4.1 Average score per residue in the NMR ensemble

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: Heterogenous nuclear ribonucleoprotein L



• Molecule 2: RNA (5'-R(*AP*CP*AP*CP*A)-3')

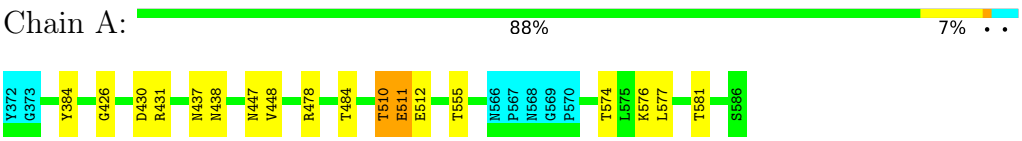


• Molecule 2: RNA (5'-R(*AP*CP*AP*CP*A)-3')

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

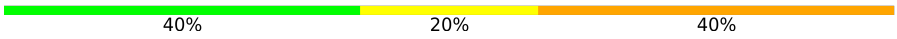
The representative model is number 2. Colouring as in section 4.1 above.

• Molecule 1: Heterogenous nuclear ribonucleoprotein L



• Molecule 2: RNA (5'-R(*AP*CP*AP*CP*A)-3')

Chain B:



• Molecule 2: RNA (5'-R(*AP*CP*AP*CP*A)-3')

Chain C:

5 Refinement protocol and experimental data overview

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

Of the 250 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	

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	3
Total number of shifts	2389
Number of shifts mapped to atoms	2389
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	76%

6 Model quality

6.1 Standard geometry

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	0.53±0.00	0±0/1679 (0.0± 0.0%)	0.78±0.01	0±0/2262 (0.0± 0.0%)
2	B	1.01±0.01	0±0/115 (0.0± 0.0%)	1.54±0.04	2±1/177 (1.1± 0.5%)
2	C	1.02±0.01	0±0/115 (0.0± 0.0%)	1.47±0.03	0±1/177 (0.2± 0.3%)
All	All	0.61	0/38180 (0.0%)	0.92	46/52320 (0.1%)

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	C	0.0±0.0	0.6±0.5
All	All	0	11

There are no bond-length outliers.

5 of 10 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	5	A	O4'-C1'-N9	7.26	114.01	108.20	6	20
2	B	3	A	C5'-C4'-C3'	-6.37	105.81	116.00	19	2
2	B	3	A	O4'-C1'-N9	6.03	113.02	108.20	1	3
2	C	5	A	O4'-C1'-N9	5.87	112.90	108.20	14	3
2	B	1	A	O4'-C1'-N9	5.78	112.82	108.20	15	11

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
2	C	4	C	Sidechain	11

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	1640	1582	1582	8±2
2	B	103	55	57	1±1
2	C	103	55	57	1±1
All	All	36920	33840	33920	186

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 32 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:397:PHE:CE1	1:A:406:VAL:HG13	0.60	2.32	6	1
1:A:437:ASN:CB	1:A:448:VAL:HB	0.56	2.31	4	20
1:A:510:THR:O	1:A:512:GLU:N	0.54	2.41	9	20
1:A:438:ASN:HA	1:A:447:ASN:HA	0.53	1.80	10	17
1:A:438:ASN:CA	1:A:447:ASN:HA	0.53	2.33	10	18

6.3 Torsion angles

6.3.1 Protein backbone

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	
1	A	207/215 (96%)	195±2 (94±1%)	9±2 (4±1%)	4±1 (2±0%)	12	54
All	All	4140/4300 (96%)	3891 (94%)	175 (4%)	74 (2%)	12	54

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)
1	A	426	GLY	20

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	510	THR	20
1	A	511	GLU	20
1	A	417	GLY	11
1	A	445	LYS	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
1	A	183/188 (97%)	173±2 (95±1%)	10±2 (5±1%)	26 75
All	All	3660/3760 (97%)	3464 (95%)	196 (5%)	26 75

5 of 40 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)
1	A	574	THR	20
1	A	484	THR	18
1	A	510	THR	17
1	A	576	LYS	16
1	A	555	THR	13

6.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
2	B	4/5 (80%)	3±0 (66±12%)	0±1 (10±17%)	0.14±0.03
2	C	5/5 (100%)	1±0 (28±10%)	1±0 (19±4%)	0.07±0.02
All	All	179/200 (90%)	81 (45%)	27 (15%)	0.10

The overall RNA backbone suiteness is 0.10.

5 of 6 unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	B	5	A	20
2	C	5	A	20

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
2	B	4	C	18
2	B	3	A	15
2	C	3	A	7

All unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	C	1	A	19
2	B	3	A	5
2	B	4	C	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 monosaccharides 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 76% for the well-defined parts and 76% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_2*

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	74
Number of shifts mapped to atoms	74
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

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 3%, i.e. 74 atoms were assigned a chemical shift out of a possible 2957. 0 out of 30 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	0/1032 (0%)	0/418 (0%)	0/416 (0%)	0/198 (0%)
Sidechain	0/1469 (0%)	0/951 (0%)	0/460 (0%)	0/58 (0%)
Aromatic	0/266 (0%)	0/132 (0%)	0/127 (0%)	0/7 (0%)
Sugar	54/110 (49%)	30/60 (50%)	24/50 (48%)	0/0 (—%)
Base	20/80 (25%)	10/50 (20%)	10/20 (50%)	0/10 (0%)
Overall	74/2957 (3%)	40/1611 (2%)	34/1073 (3%)	0/273 (0%)

7.1.4 Statistically unusual chemical shifts

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots

No *random coil index*(RCI) plot could be generated from the current chemical shift list. RCI is only applicable to proteins

7.2 Chemical shift list 2

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_3*

7.2.1 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	72
Number of shifts mapped to atoms	72
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.2.2 Chemical shift referencing

No chemical shift referencing corrections were calculated (not enough data).

7.2.3 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 2%, i.e. 72 atoms were assigned a chemical shift out of a possible 2957. 0 out of 30 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	0/1032 (0%)	0/418 (0%)	0/416 (0%)	0/198 (0%)
Sidechain	0/1469 (0%)	0/951 (0%)	0/460 (0%)	0/58 (0%)
Aromatic	0/266 (0%)	0/132 (0%)	0/127 (0%)	0/7 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Sugar	52/110 (47%)	28/60 (47%)	24/50 (48%)	0/0 (—%)
Base	20/80 (25%)	10/50 (20%)	10/20 (50%)	0/10 (0%)
Overall	72/2957 (2%)	38/1611 (2%)	34/1073 (3%)	0/273 (0%)

7.2.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.2.5 Random Coil Index (RCI) plots (i)

No *random coil index*(RCI) plot could be generated from the current chemical shift list. RCI is only applicable to proteins

7.3 Chemical shift list 3

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

7.3.1 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	2243
Number of shifts mapped to atoms	2243
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	21

7.3.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	196	2.46 ± 0.23	Should be applied
$^{13}C_{\beta}$	186	2.71 ± 0.09	Should be applied
$^{13}C'$	0	—	None (insufficient data)
^{15}N	200	0.71 ± 0.33	Should be applied

7.3.3 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 74%, i.e. 2184 atoms were assigned a chemical shift out of a possible 2957. 0 out of 30 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	787/1032 (76%)	402/418 (96%)	190/416 (46%)	195/198 (98%)
Sidechain	1235/1469 (84%)	846/951 (89%)	368/460 (80%)	21/58 (36%)
Aromatic	162/266 (61%)	103/132 (78%)	58/127 (46%)	1/7 (14%)
Sugar	0/110 (0%)	0/60 (0%)	0/50 (0%)	0/0 (—%)
Base	0/80 (0%)	0/50 (0%)	0/20 (0%)	0/10 (0%)
Overall	2184/2957 (74%)	1351/1611 (84%)	616/1073 (57%)	217/273 (79%)

7.3.4 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.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
3	A	481	ARG	NE	123.00	76.53 – 92.65	23.8
3	A	395	ARG	NE	119.22	76.53 – 92.65	21.5
3	A	492	ARG	NE	118.58	76.53 – 92.65	21.1
3	A	538	ARG	NE	118.54	76.53 – 92.65	21.1
3	A	525	ARG	NE	117.63	76.53 – 92.65	20.5
3	A	431	ARG	NE	114.64	76.53 – 92.65	18.6
3	A	496	PRO	HA	1.44	2.78 – 6.00	-9.2
3	A	410	LYS	HD2	3.33	0.58 – 2.64	8.3
3	A	410	LYS	HD3	3.33	0.54 – 2.65	8.2
3	A	576	LYS	HB3	-0.25	0.46 – 3.04	-7.8
3	A	553	LEU	CG	33.60	21.37 – 32.19	6.3
3	A	525	ARG	NH1	104.70	49.05 – 99.42	6.0
3	A	404	GLY	HA3	1.78	2.08 – 5.71	-5.8
3	A	542	GLY	HA3	1.83	2.08 – 5.71	-5.7
3	A	472	LYS	HG3	-0.12	0.04 – 2.67	-5.6
3	A	526	PRO	HA	2.71	2.78 – 6.00	-5.2
3	A	470	SER	HB3	2.46	2.49 – 5.20	-5.1
3	A	576	LYS	HB2	0.56	0.58 – 2.97	-5.1
3	A	419	ALA	HB1	0.13	0.14 – 2.58	-5.0
3	A	419	ALA	HB2	0.13	0.14 – 2.58	-5.0
3	A	419	ALA	HB3	0.13	0.14 – 2.58	-5.0

7.3.5 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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A: