wwPDB NMR Structure Validation Summary Report

Jun 5, 2023 – 01:38 PM JST

PDB ID	:	7FBR
BMRB ID	:	36430
Title	:	Solution structure of The first RNA binding domain of Matrin-3
Authors	:	Muto, Y.; Kobayashi, N.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2021-07-12

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 (1)) were used in the production of this report:

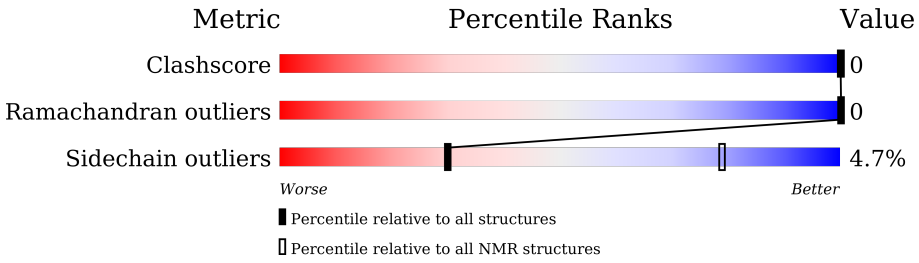
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.33

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 93%.

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

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	102	67% · 29%

2 Ensemble composition and analysis

This entry contains 20 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	A:399-A:470 (72)	0.19	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 3 clusters and 4 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1615 atoms, of which 816 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Matrin-3.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	102	1615	503	816	146	148	2	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	383	GLY	-	expression tag	UNP Q8K310
A	384	SER	-	expression tag	UNP Q8K310
A	385	SER	-	expression tag	UNP Q8K310
A	386	GLY	-	expression tag	UNP Q8K310
A	387	SER	-	expression tag	UNP Q8K310
A	388	SER	-	expression tag	UNP Q8K310
A	389	GLY	-	expression tag	UNP Q8K310
A	397	ARG	SER	engineered mutation	UNP Q8K310
A	479	SER	-	expression tag	UNP Q8K310
A	480	GLY	-	expression tag	UNP Q8K310
A	481	PRO	-	expression tag	UNP Q8K310
A	482	SER	-	expression tag	UNP Q8K310
A	483	SER	-	expression tag	UNP Q8K310
A	484	GLY	-	expression tag	UNP Q8K310

4 Residue-property plots (i)

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: Matrin-3

Chain A:

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: Matrin-3

Chain A:

5 Refinement protocol and experimental data overview

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

Of the 200 calculated structures, 20 were deposited, based on the following criterion: *structures with favorable non-bond energy*.

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

Software name	Classification	Version
Amber	refinement	12
CYANA	structure calculation	
TALOS	geometry optimization	2007

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	1
Total number of shifts	1210
Number of shifts mapped to atoms	1210
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	93%

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	0.70±0.00	0±0/591 (0.0± 0.0%)	1.00±0.01	2±1/805 (0.2± 0.1%)
All	All	0.70	0/11820 (0.0%)	1.00	33/16100 (0.2%)

There are no bond-length outliers.

All 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	A	407	ARG	NE-CZ-NH1	8.42	124.51	120.30	2	18
1	A	467	ARG	NE-CZ-NH1	6.65	123.62	120.30	4	12
1	A	467	ARG	NE-CZ-NH2	-5.48	117.56	120.30	7	1
1	A	412	ARG	NE-CZ-NH1	5.12	122.86	120.30	18	2

There are no chirality outliers.

There are no planarity outliers.

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
All	All	11560	11740	11740	-

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 -.

There are no clashes.

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	
1	A	72/102 (71%)	71±0 (98±1%)	1±0 (2±1%)	0±0 (0±0%)	100	100
All	All	1440/2040 (71%)	1416 (98%)	24 (2%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	63/87 (72%)	60±1 (95±1%)	3±1 (5±1%)	30	79
All	All	1260/1740 (72%)	1201 (95%)	59 (5%)	30	79

5 of 8 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	434	ILE	20
1	A	443	THR	20
1	A	417	GLN	8
1	A	427	ASN	4
1	A	426	SER	3

6.3.3 RNA (i)

There are no RNA molecules in this entry.

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 93% for the well-defined parts and 86% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *MATRIN3-RRM1*

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	1210
Number of shifts mapped to atoms	1210
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

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	92	-0.36 ± 0.17	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	87	-0.01 ± 0.13	None needed (< 0.5 ppm)
$^{13}C'$	85	0.04 ± 0.15	None needed (< 0.5 ppm)
^{15}N	79	0.77 ± 0.36	Should be applied

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 93%, i.e. 955 atoms were assigned a chemical shift out of a possible 1029. 0 out of 16 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	347/357 (97%)	140/144 (97%)	143/144 (99%)	64/69 (93%)
Sidechain	539/584 (92%)	369/384 (96%)	162/180 (90%)	8/20 (40%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	69/88 (78%)	36/44 (82%)	33/41 (80%)	0/3 (0%)
Overall	955/1029 (93%)	545/572 (95%)	338/365 (93%)	72/92 (78%)

7.1.4 Statistically unusual chemical shifts (i)

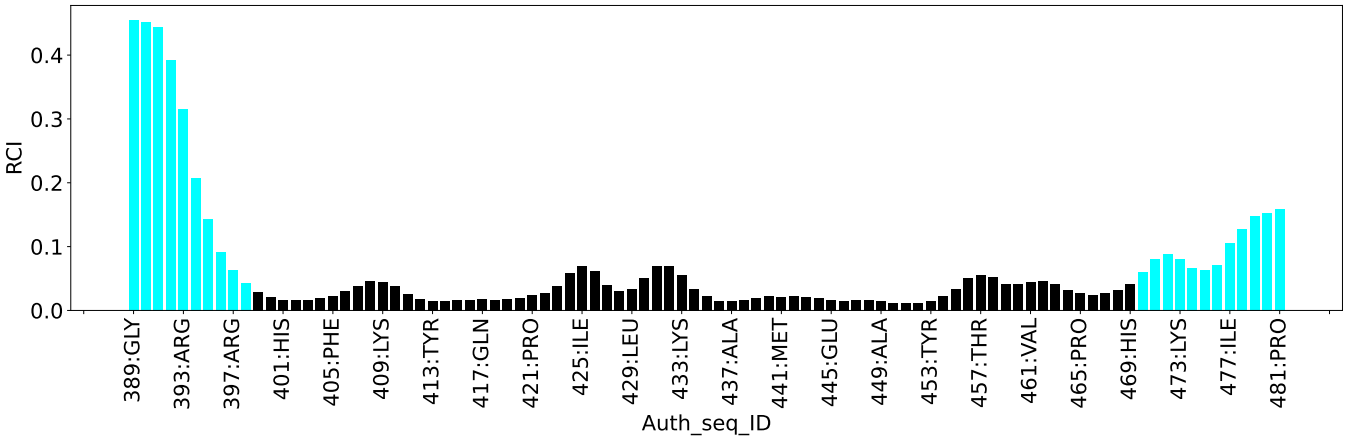
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	455	THR	HG1	5.70	0.08 – 2.19	21.6
1	A	444	THR	HG1	5.62	0.08 – 2.19	21.2
1	A	457	THR	HG1	5.38	0.08 – 2.19	20.1

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.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:



8 NMR restraints analysis

8.1 Conformationally restricting restraints

The following table provides the summary of experimentally observed NMR restraints in different categories. Restraints are classified into different categories based on the sequence separation of the atoms involved.

Description	Value
Total distance restraints	1768
Intra-residue (|i-j|=0)	318
Sequential (|i-j|=1)	465
Medium range (|i-j|>1 and |i-j|<5)	307
Long range (|i-j|≥5)	678
Inter-chain	0
Hydrogen bond restraints	0
Disulfide bond restraints	0
Total dihedral-angle restraints	0
Number of unmapped restraints	0
Number of restraints per residue	17.3
Number of long range restraints per residue[1]	6.6

[1]Long range hydrogen bonds and disulfide bonds are counted as long range restraints while calculating the number of long range restraints per residue

8.2 Residual restraint violations

This section provides the overview of the restraint violations analysis. The violations are binned as small, medium and large violations based on its absolute value. Average number of violations per model is calculated by dividing the total number of violations in each bin by the size of the ensemble.

8.2.1 Average number of distance violations per model

Distance violations less than 0.1 Å are not included in the calculation.There are no distance violations

8.2.2 Average number of dihedral-angle violations per model

Dihedral-angle violations less than 1° are not included in the calculation.There are no dihedral-angle violations

9 Distance violation analysis (i)

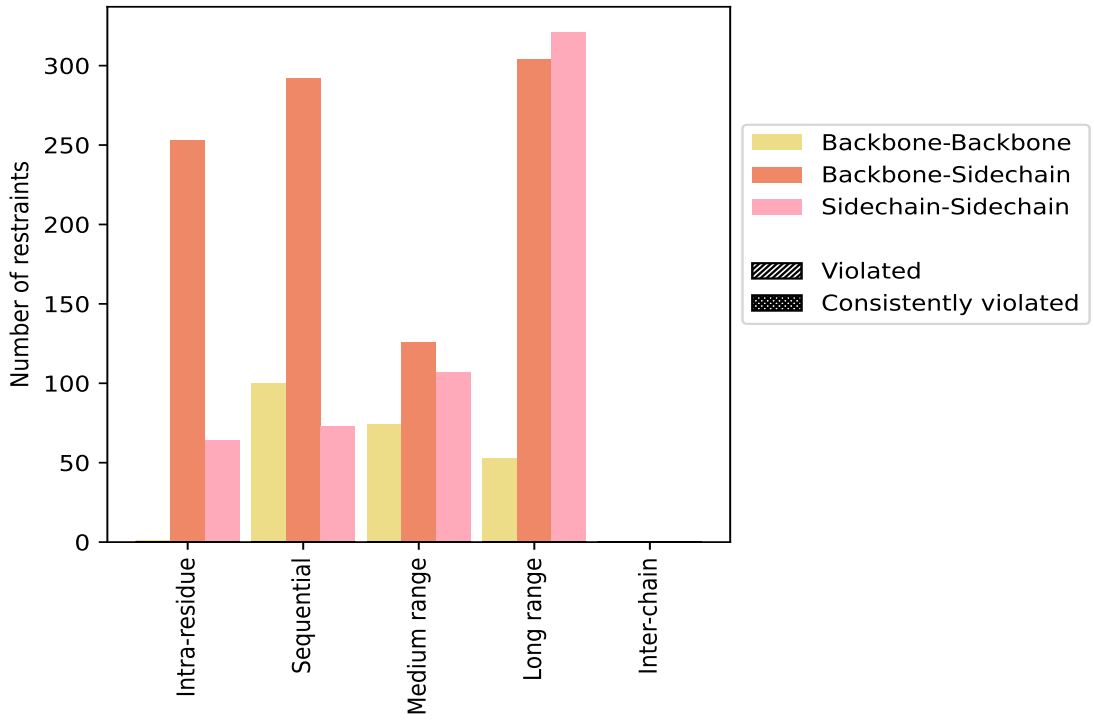
9.1 Summary of distance violations (i)

The following table shows the summary of distance violations in different restraint categories based on the sequence separation of the atoms involved. Each category is further sub-divided into three sub-categories based on the atoms involved. Violations less than 0.1 Å are not included in the statistics.

Restraints type	Count	%[1]	Violated[3]			Consistently Violated[4]		
			Count	%[2]	%[1]	Count	%[2]	%[1]
Intra-residue (|i-j|=0)	318	18.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	1	0.1	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	253	14.3	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	64	3.6	0	0.0	0.0	0	0.0	0.0
Sequential (|i-j|=1)	465	26.3	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	100	5.7	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	292	16.5	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	73	4.1	0	0.0	0.0	0	0.0	0.0
Medium range (|i-j|>1 & |i-j|<5)	307	17.4	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	74	4.2	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	126	7.1	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	107	6.1	0	0.0	0.0	0	0.0	0.0
Long range (|i-j|≥5)	678	38.3	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	53	3.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	304	17.2	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	321	18.2	0	0.0	0.0	0	0.0	0.0
Inter-chain	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Hydrogen bond	0	0.0	0	0.0	0.0	0	0.0	0.0
Disulfide bond	0	0.0	0	0.0	0.0	0	0.0	0.0
Total	1768	100.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	228	12.9	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	975	55.1	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	565	32.0	0	0.0	0.0	0	0.0	0.0

[1] percentage calculated with respect to the total number of distance restraints, [2] percentage calculated with respect to the number of restraints in a particular restraint category, [3] violated in at least one model, [4] violated in all the models

9.1.1 Bar chart : Distribution of distance restraints and violations



Violated and consistently violated restraints are shown using different hatch patterns in their respective categories. The hydrogen bonds and disulfied bonds are counted in their appropriate category on the x-axis

9.2 Distance violation statistics for each model

No violations found

9.3 Distance violation statistics for the ensemble

No violations found

9.4 Most violated distance restraints in the ensemble

No violations found

9.5 All violated distance restraints

No violations found

10 Dihedral-angle violation analysis (i)

Dihedral angle analysis failed due to data error in the dihedral angle restraints, possibly missing target value