wwPDB NMR Structure Validation Summary Report

Mar 16, 2022 – 04:13 PM EDT

PDB ID	:	6C8U
Title	:	Solution structure of Musashi2 RRM1
Authors	:	Xing, M.; Lan, L.; Douglas, J.T.; Gao, P.; Hanzlik, R.P.; Xu, L.
Deposited on	:	2018-01-25

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 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)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

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 was not calculated.

There are no overall percentile quality scores available for this entry.

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	1	115	100%

2 Ensemble composition and analysis

This entry contains 10 models. The authors have identified model 1 as representative, based on the following criterion: *lowest energy*. No medoid model was calculated.

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**

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

NmrClust was unable to cluster the ensemble.

Error message: Wrapper check: not enough residues in core to run NmrClust

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1834 atoms, of which 910 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	1	115	1834	580	910	171	167	6	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
1	-3	MET	-	expression tag	UNP Q96DH6
1	-2	HIS	-	expression tag	UNP Q96DH6
1	-1	HIS	-	expression tag	UNP Q96DH6
1	0	HIS	-	expression tag	UNP Q96DH6
1	1	HIS	-	expression tag	UNP Q96DH6
1	2	HIS	-	expression tag	UNP Q96DH6
1	3	HIS	-	expression tag	UNP Q96DH6
1	4	SER	-	expression tag	UNP Q96DH6
1	5	THR	-	expression tag	UNP Q96DH6
1	6	SER	-	expression tag	UNP Q96DH6
1	7	VAL	-	expression tag	UNP Q96DH6
1	8	ASP	-	expression tag	UNP Q96DH6
1	9	LEU	-	expression tag	UNP Q96DH6
1	10	GLY	-	expression tag	UNP Q96DH6
1	11	THR	-	expression tag	UNP Q96DH6
1	12	GLU	-	expression tag	UNP Q96DH6
1	13	ASN	-	expression tag	UNP Q96DH6
1	14	LEU	-	expression tag	UNP Q96DH6
1	15	TYR	-	expression tag	UNP Q96DH6
1	16	PHE	-	expression tag	UNP Q96DH6
1	17	GLN	-	expression tag	UNP Q96DH6
1	18	SER	-	expression tag	UNP Q96DH6
1	19	ASN	-	expression tag	UNP Q96DH6
1	20	ALA	-	expression tag	UNP Q96DH6

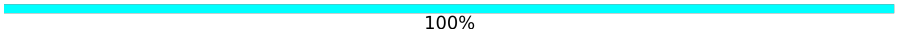
4 Residue-property plots

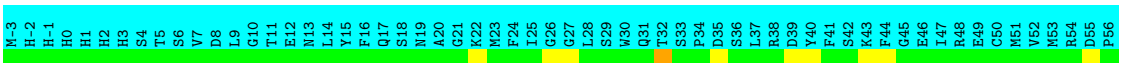
4.1 Average score per residue in the NMR ensemble

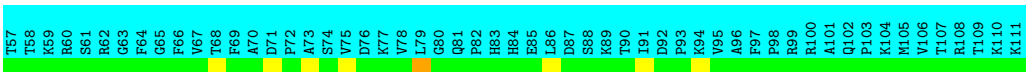
These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for 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: RNA-binding protein Musashi homolog 2

Chain 1:





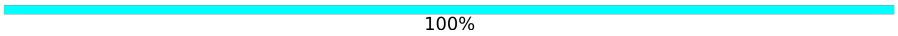


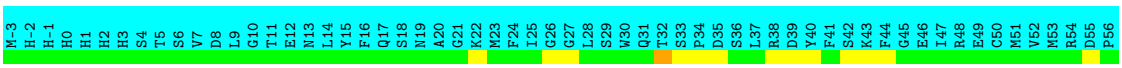
4.2 Residue scores for the representative (author defined) model from the NMR ensemble

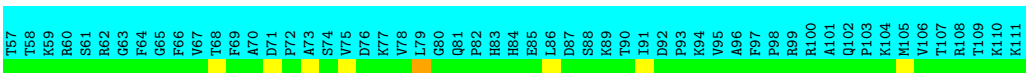
The representative model is number 1. Colouring as in section 4.1 above.

- Molecule 1: RNA-binding protein Musashi homolog 2

Chain 1:







5 Refinement protocol and experimental data overview

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

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
CNS	refinement	1.3
CNS	structure calculation	1.3

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	1326
Number of shifts mapped to atoms	1326
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	—

6 Model quality

6.1 Standard geometry

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

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
1	1	0	0	0	0±0
All	All	0	0	0	-

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

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	1	0	-	-	-	-
All	All	0	-	-	-	-

There are no Ramachandran outliers.

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
1	1	0	-	-	-
All	All	0	-	-	-

There are no protein residues with a non-rotameric sidechain to report.

6.3.3 RNA

There are no RNA molecules in this entry.

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 monosaccharides in this entry.

6.6 Ligand geometry

There are no ligands in this entry.

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is *undefined* for the well-defined parts and 82% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *MSI2-RRM1_assignments_unassigned_deleted_180125.txt*

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	1326
Number of shifts mapped to atoms	1326
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 ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	108	-0.40 ± 0.10	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	100	0.18 ± 0.05	None needed (< 0.5 ppm)
$^{13}C'$	108	0.05 ± 0.15	None needed (< 0.5 ppm)
^{15}N	98	0.13 ± 0.42	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 —%, i.e. 0 atoms were assigned a chemical shift out of a possible 0. 0 out of 0 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	0/0 (—%)	0/0 (—%)	0/0 (—%)	0/0 (—%)
Sidechain	0/0 (—%)	0/0 (—%)	0/0 (—%)	0/0 (—%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	0/0 (—%)	0/0 (—%)	0/0 (—%)	0/0 (—%)
Overall	0/0 (—%)	0/0 (—%)	0/0 (—%)	0/0 (—%)

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	74	SER	HB3	2.39	5.25 – 2.45	-5.2

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 1: