wwPDB NMR Structure Validation Summary Report (i)

Apr 22, 2024 – 10:43 AM JST

PDB ID	:	5X3Z
BMRB ID	:	36059
Title	:	Solution structure of musashi1 RBD2 in complex with RNA
Authors	:	Iwaoka, R.; Nagata, T.; Tsuda, K.; Imai, T.; Okano, H.; Kobayashi, N.; Katahira, M.
Deposited on	:	2017-02-09

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:

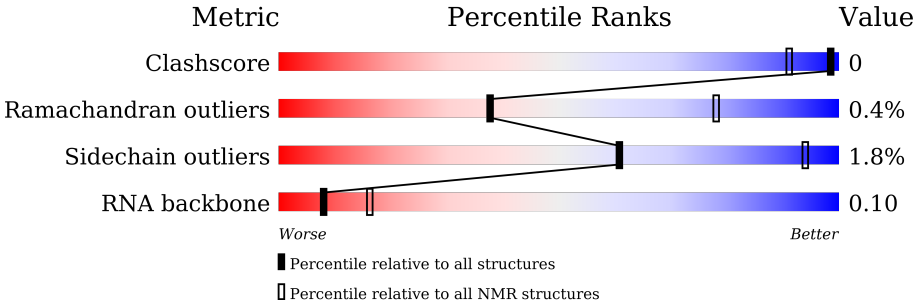
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson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

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 85%.

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	96	71% . 28%
2	B	5	20% 60% 20%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 16 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	A:109-A:141, A:149-A:184 (69)	0.29	16

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 1 single-model cluster was found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1675 atoms, of which 805 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	96	1515	480	750	134	144	7	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	105	GLY	-	expression tag	UNP Q61474
A	106	SER	-	expression tag	UNP Q61474
A	107	HIS	-	expression tag	UNP Q61474
A	108	MET	-	expression tag	UNP Q61474

- Molecule 2 is a RNA chain called RNA (5'-R(*GP*UP*AP*GP*U)-3').

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	5	160	48	55	19	34	4	0

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: RNA-binding protein Musashi homolog 1

Chain A: 71% 28%

G105 S106 H107 M108 F142 D143 K144 T145 T146 N147 R148 H149 R150 Q185 P186 K187 E188 V189 M190 S191 P192 T193 G194 S195 A196 R197 G198 R199 S200

• Molecule 2: RNA (5'-R(*GP*UP*AP*GP*U)-3')

Chain B: 20% 60% 20%

G201 U202 A203 G204 U205

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

The representative model is number 16. Colouring as in section 4.1 above.

• Molecule 1: RNA-binding protein Musashi homolog 1

Chain A: 71% 28%

G105 S106 H107 M108 F142 D143 K144 T145 T146 N147 R148 H149 R150 Q185 P186 K187 E188 V189 M190 S191 P192 T193 G194 S195 A196 R197 G198 R199 S200

• Molecule 2: RNA (5'-R(*GP*UP*AP*GP*U)-3')

Chain B: 20% 80%

G201 U202 A203 G204 U205

5 Refinement protocol and experimental data overview (i)

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 the least restraint violations*.

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

Software name	Classification	Version
Amber	refinement	
CYANA	structure calculation	

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	1128
Number of shifts mapped to atoms	1128
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	85%

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	0.70±0.00	0±0/571 (0.0± 0.0%)	0.93±0.03	1±0/764 (0.1± 0.1%)
2	B	1.39±0.02	0±0/117 (0.0± 0.0%)	2.23±0.07	5±1/181 (2.8± 0.5%)
All	All	0.86	0/13760 (0.0%)	1.29	121/18900 (0.6%)

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±0.2
2	B	0.0±0.0	0.5±0.6
All	All	0	10

There are no bond-length outliers.

5 of 15 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	203	A	N1-C6-N6	-11.85	111.49	118.60	3	20
1	A	150	ARG	NE-CZ-NH1	10.24	125.42	120.30	16	17
2	B	204	G	O4'-C1'-N9	9.07	115.45	108.20	19	10
2	B	201	G	O4'-C1'-N9	8.05	114.64	108.20	9	6
2	B	203	A	C5-C6-N1	7.82	121.61	117.70	4	20

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	205	U	Sidechain	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
2	B	203	A	Sidechain	2
2	B	204	G	Sidechain	2
2	B	202	U	Sidechain	1
2	B	201	G	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	559	547	547	0±0
All	All	13280	12040	12040	2

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 0.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:110:LYS:HE2	1:A:154:PHE:CD2	0.48	2.44	14	2

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	69/96 (72%)	66±1 (96±1%)	3±1 (4±1%)	0±1 (0±1%)	38	78
All	All	1380/1920 (72%)	1320 (96%)	55 (4%)	5 (0%)	38	78

All 2 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	133	GLY	3
1	A	109	LYS	2

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	62/85 (73%)	61±1 (98±1%)	1±1 (2±1%)	61 94
All	All	1240/1700 (73%)	1218 (98%)	22 (2%)	61 94

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	156	THR	6
1	A	176	ASN	4
1	A	121	THR	4
1	A	178	MET	3
1	A	150	ARG	2

6.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
2	B	4/5 (80%)	2±1 (52±27%)	0±0 (5±10%)	0.10±0.09
All	All	81/100 (81%)	42 (52%)	4 (5%)	0.11

The overall RNA backbone suiteness is 0.10.

All unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	B	204	G	18
2	B	205	U	11
2	B	203	A	7
2	B	202	U	6

All unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	B	203	A	2
2	B	201	G	1
2	B	202	U	1

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 (i)

The completeness of assignment taking into account all chemical shift lists is 85% for the well-defined parts and 81% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *cplx170207_deposit.str*

7.1.1 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	1128
Number of shifts mapped to atoms	1128
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 (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	93	-0.15 ± 0.08	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	86	0.03 ± 0.17	None needed (< 0.5 ppm)
$^{13}C'$	91	0.09 ± 0.09	None needed (< 0.5 ppm)
^{15}N	90	0.16 ± 0.34	None needed (< 0.5 ppm)

7.1.3 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 85%, i.e. 886 atoms were assigned a chemical shift out of a possible 1047. 0 out of 11 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	350/350 (100%)	143/143 (100%)	138/138 (100%)	69/69 (100%)
Sidechain	460/497 (93%)	311/321 (97%)	144/161 (89%)	5/15 (33%)

Continued on next page...

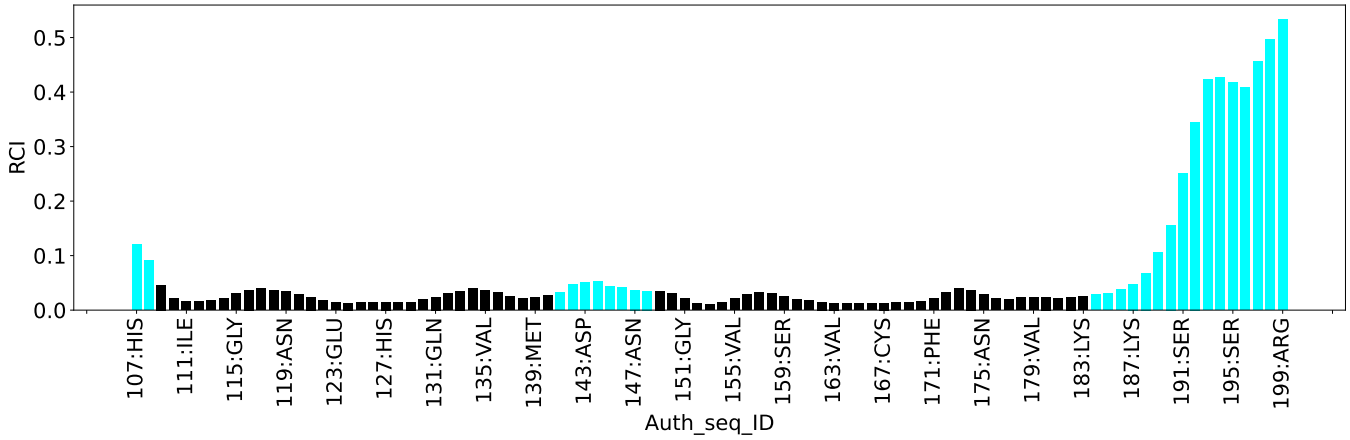
Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	76/107 (71%)	38/55 (69%)	38/48 (79%)	0/4 (0%)
Sugar	0/55 (0%)	0/30 (0%)	0/25 (0%)	0/0 (—%)
Base	0/38 (0%)	0/23 (0%)	0/8 (0%)	0/7 (0%)
Overall	886/1047 (85%)	492/572 (86%)	320/380 (84%)	74/95 (78%)

7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

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: