wwPDB NMR Structure Validation Summary Report (i)

Feb 22, 2022 – 01:35 PM EST

PDB ID	:	1WI5
Title	:	Solution structure of the S1 RNA binding domain from human hypothetical protein BAA11502
Authors	:	Suzuki, S.; Muto, Y.; Inoue, M.; Kigawa, T.; Terada, T.; Shirouzu, M.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2004-05-28

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:

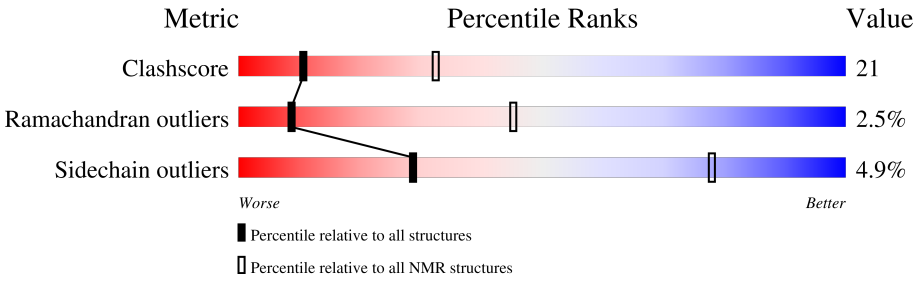
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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.

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	119	45% 18% • 37%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 20 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:197-A:271 (75)	0.35	20

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 5 clusters and 5 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1757 atoms, of which 888 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called RRP5 protein homolog.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	119	1757	539	888	155	173	2	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	179	GLY	-	cloning artifact	UNP Q14690
A	180	SER	-	cloning artifact	UNP Q14690
A	181	SER	-	cloning artifact	UNP Q14690
A	182	GLY	-	cloning artifact	UNP Q14690
A	183	SER	-	cloning artifact	UNP Q14690
A	184	SER	-	cloning artifact	UNP Q14690
A	185	GLY	-	cloning artifact	UNP Q14690
A	292	SER	-	cloning artifact	UNP Q14690
A	293	GLY	-	cloning artifact	UNP Q14690
A	294	PRO	-	cloning artifact	UNP Q14690
A	295	SER	-	cloning artifact	UNP Q14690
A	296	SER	-	cloning artifact	UNP Q14690
A	297	GLY	-	cloning artifact	UNP Q14690

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: RRP5 protein homolog

Chain A: 45% 18% • 37%

G179 S180 S181 G182 S183 S184 G185 K186 N187 V188 N189 R190 V191 L192 S193 A194 E195 A196 L202 L203 D212 H213 G214 Y215 F227 L228 P229 L230 L231 K232 Y236 I237 K240 N241 K242 G243 A244 K245 L246 Q250 L268 S269 V270 G271 H272 S273 E274 V275 S276 T277 A278 I279 A280

T281 E282 Q283 Q284 S285 W286 N287 L288 N289 N290 L291 S292 G293 P294 S295 S296 G297

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

The representative model is number 20. Colouring as in section 4.1 above.

- Molecule 1: RRP5 protein homolog

Chain A: 46% 16% • 37%

G179 S180 S181 G182 S183 S184 G185 K186 N187 V188 N189 R190 V191 L192 S193 A194 E195 A196 M201 L202 L203 L210 G214 Y215 D222 F227 K232 I237 N241 K242 G243 L246 K247 V248 I255 K260 G264 S269 V270 G271 H272 S273 E274 V275 S276 T277 A278

I279 A280 T281 E282 Q283 Q284 S285 W286 N287 L288 N289 N290 L291 S292 G293 P294 S295 S296 G297

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *target function.*

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

Software name	Classification	Version
CYANA	structure solution	2.0.17
CYANA	refinement	2.0.17

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

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 (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	560	598	597	24±5
All	All	11200	11960	11940	484

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 21.

5 of 158 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:215:TYR:CE2	1:A:246:LEU:HD12	0.98	1.92	14	5
1:A:237:ILE:HD11	1:A:246:LEU:HD11	0.84	1.49	15	2
1:A:215:TYR:CE1	1:A:246:LEU:HD12	0.81	2.11	18	8
1:A:215:TYR:CD2	1:A:246:LEU:HD12	0.79	2.11	14	5
1:A:228:LEU:HD12	1:A:229:PRO:HD2	0.77	1.56	7	9

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	75/119 (63%)	66±1 (88±2%)	7±1 (10±2%)	2±1 (2±1%)	9 45
All	All	1500/2380 (63%)	1315 (88%)	148 (10%)	37 (2%)	9 45

5 of 9 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	242	LYS	19
1	A	197	LEU	4
1	A	261	GLY	4
1	A	264	GLY	4
1	A	263	GLY	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	61/96 (64%)	58±2 (95±3%)	3±2 (5±3%)	29 78
All	All	1220/1920 (64%)	1160 (95%)	60 (5%)	29 78

5 of 21 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	236	TYR	8
1	A	245	LYS	8
1	A	232	LYS	7
1	A	250	GLN	5
1	A	260	LYS	5

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 (i)

No chemical shift data were provided