wwPDB NMR Structure Validation Summary Report (i)

Feb 28, 2022 – 12:07 AM EST

PDB ID	:	2DNZ
Title	:	Solution structure of the second RNA binding domain of RNA binding motif protein 23
Authors	:	Suzuki, S.; Muto, Y.; Inoue, M.; Kigawa, T.; Terada, T.; Shirouzu, M.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2006-04-27

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:

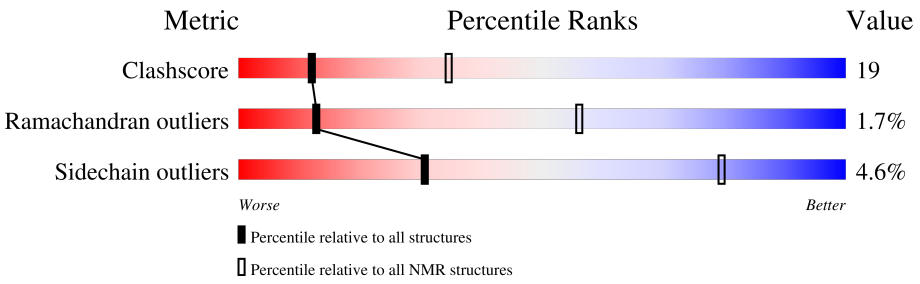
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
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.

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	95	48% 22% 29%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 17 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:230-A:261, A:271-A:305 (67)	0.15	17

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 3 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1396 atoms, of which 685 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Probable RNA-binding protein 23.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	95	1396	439	685	127	141	4	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	224	GLY	-	cloning artifact	UNP Q86U06
A	225	SER	-	cloning artifact	UNP Q86U06
A	226	SER	-	cloning artifact	UNP Q86U06
A	227	GLY	-	cloning artifact	UNP Q86U06
A	228	SER	-	cloning artifact	UNP Q86U06
A	229	SER	-	cloning artifact	UNP Q86U06
A	230	GLY	-	cloning artifact	UNP Q86U06
A	313	SER	-	cloning artifact	UNP Q86U06
A	314	GLY	-	cloning artifact	UNP Q86U06
A	315	PRO	-	cloning artifact	UNP Q86U06
A	316	SER	-	cloning artifact	UNP Q86U06
A	317	SER	-	cloning artifact	UNP Q86U06
A	318	GLY	-	cloning artifact	UNP Q86U06

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: Probable RNA-binding protein 23

Chain A: 48% 22% 29%

G224 S225 S226 G227 S228 S229 G230 L231 Y232 H237 F238 N239 I240 R246 G247 I248 F249 E250 P251 F252 M261 K262 D263 S264 D265 T266 G267 R268 S269 K270 G271 Y272 A286 L287 L290 N291 V302 V305 T306 E307 R308 L309 D310 G311 G312 S313 G314 P315 S316 S317 G318

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

The representative model is number 17. Colouring as in section 4.1 above.

• Molecule 1: Probable RNA-binding protein 23

Chain A: 45% 21% • 29%

G224 S225 S226 G227 S228 S229 G230 L231 Y232 H237 F238 N239 I240 L245 R246 G247 I248 F249 E250 P251 F252 L260 M261 K262 D263 S264 D265 T266 G267 R268 S269 K270 D279 A283 A286 L287 L290 N291 M300 R301 V302 G303 H304 V305 T306 E307 R308 L309 D310 G311 G312 S313

G314 P315 S316 S317 G318

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *target function, structures with the lowest energy, structures with the least restraint violations.*

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	527	517	517	20±4
All	All	10540	10340	10340	402

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 19.

5 of 94 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:231:LEU:HD12	1:A:277:PHE:CE1	0.85	2.06	14	5
1:A:290:LEU:HB3	1:A:302:VAL:HG21	0.78	1.54	15	20
1:A:232:TYR:HB2	1:A:305:VAL:HG12	0.75	1.56	20	3
1:A:249:PHE:CZ	1:A:290:LEU:HD13	0.72	2.19	1	18
1:A:287:LEU:HD12	1:A:302:VAL:HG12	0.71	1.63	19	10

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	67/95 (71%)	60±1 (90±2%)	6±2 (8±2%)	1±1 (2±1%)	13 56
All	All	1340/1900 (71%)	1204 (90%)	113 (8%)	23 (2%)	13 56

5 of 6 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	294	GLU	7
1	A	305	VAL	5
1	A	240	ILE	4
1	A	291	ASN	4
1	A	271	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	55/76 (72%)	52±1 (95±3%)	3±1 (5±3%)	31 79
All	All	1100/1520 (72%)	1049 (95%)	51 (5%)	31 79

5 of 14 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	232	TYR	9
1	A	250	GLU	9
1	A	272	TYR	6
1	A	291	ASN	6
1	A	246	ARG	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