wwPDB NMR Structure Validation Summary Report (i)

Feb 27, 2022 – 02:17 PM EST

PDB ID	:	2D9O
Title	:	Solution structure of RNA binding domain in Hypothetical protein FLJ10634
Authors	:	Tsuda, K.; Muto, Y.; Inoue, M.; Kigawa, T.; Terada, T.; Shirouzu, M.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2005-12-12

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 following versions of software and data (see references (i)) were used in the production of this report:

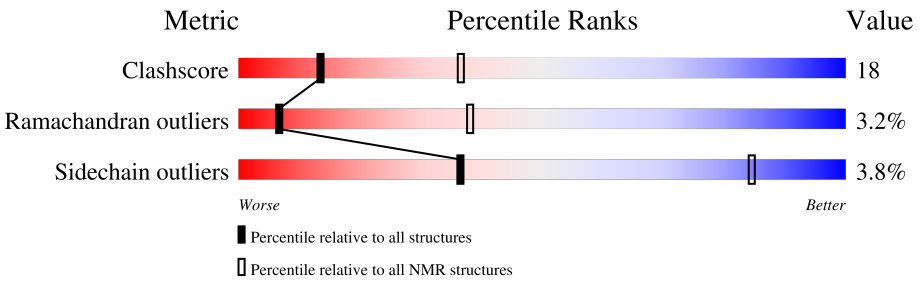
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
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	100	39% 22% • 38%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 9 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:170-A:177, A:189-A:209, A:216-A:248 (62)	0.19	9

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 4 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1509 atoms, of which 767 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called DnaJ (Hsp40) homolog, subfamily C, member 17.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	100	1509	468	767	126	147	1	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	161	GLY	-	cloning artifact	GB 8922563
A	162	SER	-	cloning artifact	GB 8922563
A	163	SER	-	cloning artifact	GB 8922563
A	164	GLY	-	cloning artifact	GB 8922563
A	165	SER	-	cloning artifact	GB 8922563
A	166	SER	-	cloning artifact	GB 8922563
A	167	GLY	-	cloning artifact	GB 8922563
A	255	SER	-	cloning artifact	GB 8922563
A	256	GLY	-	cloning artifact	GB 8922563
A	257	PRO	-	cloning artifact	GB 8922563
A	258	SER	-	cloning artifact	GB 8922563
A	259	SER	-	cloning artifact	GB 8922563
A	260	GLY	-	cloning artifact	GB 8922563

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: DnaJ (Hsp40) homolog, subfamily C, member 17

Chain A: 39% 22% • 38%

G161 S162 S163 G164 S165 S166 G167 Q168 G169 L173 K178 C179 K180 K181 E182 D183 E184 S185 K186 G187 G188 Y189 S190 K191 D192 V193 L198 Q199 K200 Y201 S210 S211 K212 K213 P214 G215 V219 E220 F221 K225 A226 A227 V231 Q232 N233 E234 L237 V238 D239 N240 P241 L242

K243 I244 G249 Q250 P251 Q252 D253 A254 S255 G256 P257 S258 S259 G260

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

The representative model is number 9. Colouring as in section 4.1 above.

- Molecule 1: DnaJ (Hsp40) homolog, subfamily C, member 17

Chain A: 35% 25% • 38%

G161 S162 S163 G164 S165 S166 G167 Q168 G169 L173 K178 C179 K180 K181 E182 D183 E184 S185 K186 G187 G188 Y189 S190 K191 L198 Q199 K200 Y201 N206 L207 V208 L209 S210 S211 K212 K213 P214 G215 V219 E220 F221 V224 K225 L229 A230 V231 Q232 N233 E234 L237 V238 D239

N240 P241 L242 K243 G249 Q250 P251 Q252 D253 A254 S255 G256 P257 S258 S259 G260

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *torsion angle dynamics, restrainted molecular dynamics.*

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
CYANA	refinement	2.0

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	495	534	534	18±4
All	All	9900	10680	10680	370

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 18.

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:198:LEU:HD13	1:A:219:VAL:HG21	0.88	1.45	15	9
1:A:173:LEU:HD12	1:A:221:PHE:CE1	0.83	2.09	8	15
1:A:176:LYS:HD2	1:A:216:THR:HG22	0.82	1.49	2	2
1:A:176:LYS:CD	1:A:216:THR:HG22	0.79	2.08	12	3
1:A:176:LYS:HD3	1:A:216:THR:HG22	0.79	1.52	12	1

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	62/100 (62%)	47±3 (76±4%)	13±3 (21±5%)	2±1 (3±2%)	7 38
All	All	1240/2000 (62%)	937 (76%)	263 (21%)	40 (3%)	7 38

5 of 7 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	LEU	15
1	A	233	ASN	8
1	A	193	VAL	6
1	A	204	VAL	4
1	A	200	LYS	3

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	55/82 (67%)	53±1 (96±2%)	2±1 (4±2%)	36 84
All	All	1100/1640 (67%)	1058 (96%)	42 (4%)	36 84

5 of 15 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	234	GLU	9
1	A	243	LYS	7
1	A	225	LYS	4
1	A	176	LYS	3
1	A	206	ASN	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 (i)

No chemical shift data were provided