WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB NMR Structure Validation Summary Report (i)

May 29, 2020 – 01:16 am BST

PDB ID : 2RRA

Title : Solution structure of RNA binding domain in human Tra2 beta protein in complex with RNA (GAAGAA)

Authors : Tsuda, K.; Kuwasako, K.; Takahashi, M.; Someya, T.; Inoue, M.; Kigawa, T.; Terada, T.; Shirouzu, M.; Sugano, S.; Muto, Y.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)

Deposited on : 2010-06-17

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:

Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
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	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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	Percentile Ranks	Value
Clashscore		0
Ramachandran outliers		0
Sidechain outliers		1.9%
RNA backbone		0.09

Worse ← → *Better*

∎ Percentile relative to all structures
⊐ Percentile relative to all NMR structures
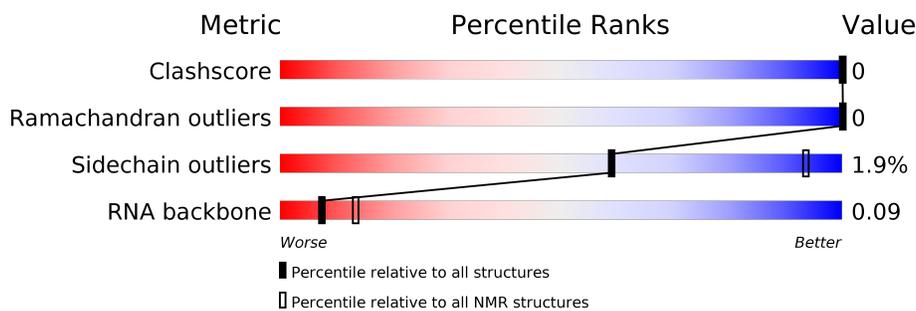
Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428
RNA backbone	4643	676

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	99	66% • 32%
2	B	6	50% 50%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 8 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

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:119-A:150, A:160-A:194 (67)	0.12	8

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 2 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1732 atoms, of which 820 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called cDNA FLJ40872 fis, clone TUTER2000283, highly similar to Homo sapiens transformer-2-beta (SFRS10) gene.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	99	1533	480	752	145	153	3	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	103	GLY	-	EXPRESSION TAG	UNP Q8N1H4
A	104	SER	-	EXPRESSION TAG	UNP Q8N1H4
A	105	SER	-	EXPRESSION TAG	UNP Q8N1H4
A	106	GLY	-	EXPRESSION TAG	UNP Q8N1H4
A	107	SER	-	EXPRESSION TAG	UNP Q8N1H4
A	108	SER	-	EXPRESSION TAG	UNP Q8N1H4

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	6	199	60	68	30	36	5	0

4 Residue-property plots ⓘ

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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: cDNA FLJ40872 fis, clone TUTER2000283, highly similar to Homo sapiens transformer-2-beta (SFRS10) gene
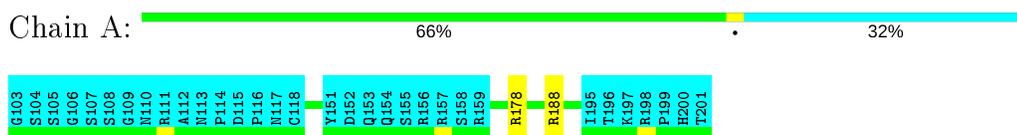
• Molecule 2: 5'-R(*GP*AP*AP*GP*AP*A)-3'

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

The representative model is number 8. Colouring as in section 4.1 above.

• Molecule 1: cDNA FLJ40872 fis, clone TUTER2000283, highly similar to Homo sapiens transformer-2-beta (SFRS10) gene
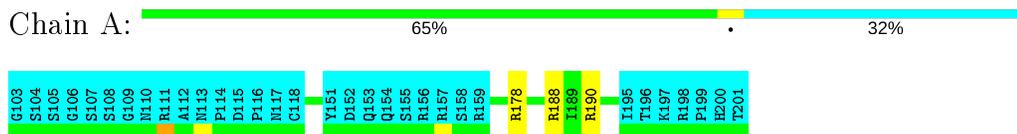
• Molecule 2: 5'-R(*GP*AP*AP*GP*AP*A)-3'

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DGSA-distance geometry 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	9
CYANA	structure solution	2.1

No chemical shift data was provided. No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

5.1 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
All	All	13300	11740	11740	-

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.

5.2 Torsion angles (i)

5.2.1 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	67/99 (68%)	66±0 (99±1%)	1±0 (1±1%)	0±0 (0±0%)	100	100

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1340/1980 (68%)	1322 (99%)	18 (1%)	0 (0%)	100 100

There are no Ramachandran outliers.

5.2.2 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	56/84 (67%)	55±1 (98±1%)	1±1 (2±1%)	59 93
All	All	1120/1680 (67%)	1099 (98%)	21 (2%)	59 93

All 4 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	188	ARG	16
1	A	187	ARG	3
1	A	135	ARG	1
1	A	177	GLU	1

5.2.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
2	B	5/6 (83%)	3±1 (59±17%)	0±0 (1±4%)	0.09±0.06
All	All	100/120 (83%)	59 (59%)	1 (1%)	0.08

The overall RNA backbone suiteness is 0.09.

All unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	B	5	A	20
2	B	6	A	13
2	B	4	G	13
2	B	2	A	10
2	B	3	A	3

All unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	B	2	A	1

5.3 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.4 Carbohydrates (i)

There are no carbohydrates in this entry.

5.5 Ligand geometry (i)

There are no ligands in this entry.

5.6 Other polymers (i)

There are no such molecules in this entry.

5.7 Polymer linkage issues (i)

There are no chain breaks in this entry.

6   Chemical shift validation (i)

No chemical shift data were provided