wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 04:43 PM BST

PDB ID	:	1RFL
Title	:	NMR data driven structural model of G-domain of MnmE protein
Authors	:	Monleon, D.; Esteve, V.; Martinez-Vicente, M.; Yim, L.; Armengod, M.E.; Celda, B.
Deposited on	:	2003-11-10

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
http://wwpdb.org/validation/2016/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:

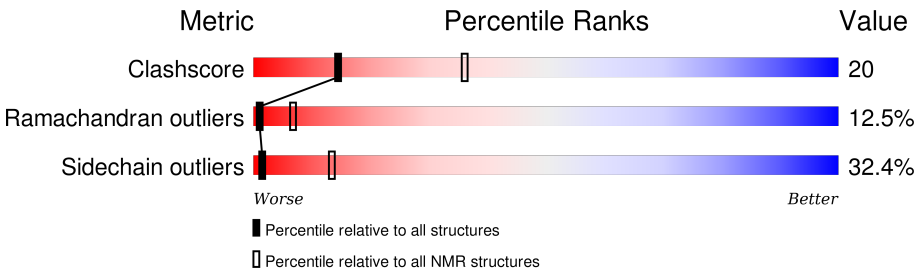
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

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 46%.

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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	172	28% 65% . .

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 12 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:4-A:170 (167)	0.65	12

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 2 clusters and 2 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2675 atoms, of which 1364 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Probable tRNA modification GTPase trmE.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	172	2675	812	1364	244	250	5	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	169	ILE	-	CLONING ARTIFACT	UNP P25522
A	170	HIS	-	CLONING ARTIFACT	UNP P25522
A	171	ARG	-	CLONING ARTIFACT	UNP P25522
A	172	ASP	-	CLONING ARTIFACT	UNP P25522

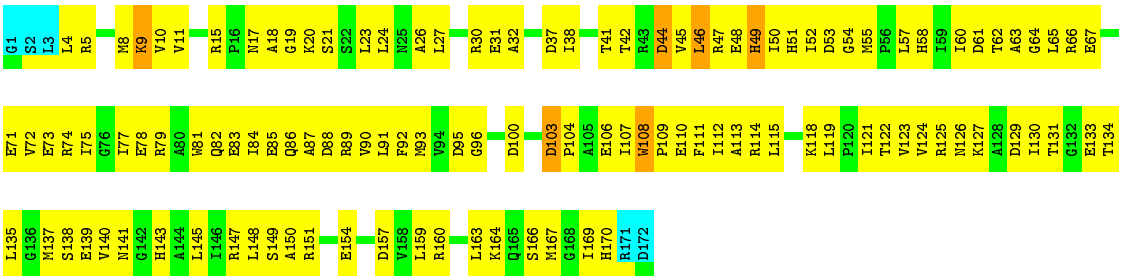
4 Residue-property plots (i)

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: Probable tRNA modification GTPase trmE

Chain A: 28% 65%

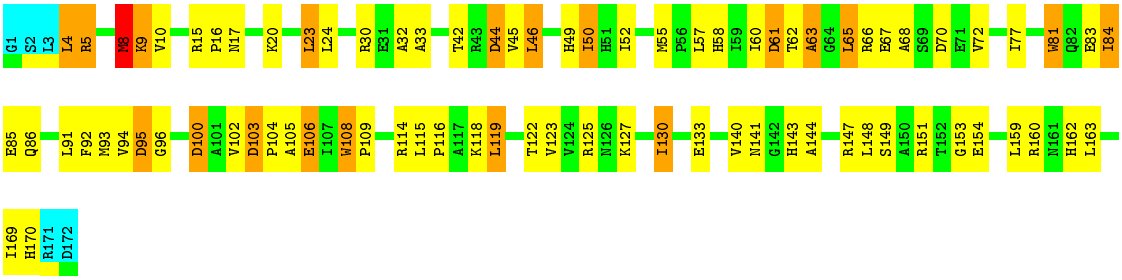


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 12. Colouring as in section 4.1 above.

- Molecule 1: Probable tRNA modification GTPase trmE

Chain A: 50% 35% 11%



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *NMR restraints were obtained from the backbone assignment (TALOS analysis for dihedral angle constraints) and from the 15N-edited 3D NOESY. Additional structural information was obtained from the homologous GTPAse protein structures..*

Of the 200 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy.*

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

Software name	Classification	Version
DYANA	structure solution	2000
DYANA	refinement	2000

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)	BMRB entry 5861
Number of chemical shift lists	1
Total number of shifts	950
Number of shifts mapped to atoms	950
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	46%

No validations of the models with respect to experimental NMR restraints is performed at this time.

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	1274	1327	1296	52±8
All	All	25480	26540	25920	1035

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 20.

5 of 725 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:124:VAL:HG12	1:A:130:ILE:HD11	0.86	1.47	1	1
1:A:107:ILE:O	1:A:107:ILE:HD13	0.86	1.71	1	3
1:A:130:ILE:HG23	1:A:134:THR:HG21	0.86	1.45	3	1
1:A:92:PHE:CE2	1:A:119:LEU:HD23	0.86	2.06	13	1
1:A:92:PHE:CE2	1:A:107:ILE:HG22	0.84	2.07	2	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	167/172 (97%)	100±6 (60±4%)	46±5 (27±3%)	21±4 (12±2%)	1 7
All	All	3340/3440 (97%)	2008 (60%)	916 (27%)	416 (12%)	1 7

5 of 124 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	49	HIS	17
1	A	63	ALA	12
1	A	77	ILE	12
1	A	84	ILE	11
1	A	19	GLY	10

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	134/138 (97%)	91±5 (68±3%)	43±5 (32±3%)	1 13
All	All	2680/2760 (97%)	1812 (68%)	868 (32%)	1 13

5 of 116 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	103	ASP	20
1	A	9	LYS	19
1	A	108	TRP	19
1	A	57	LEU	17
1	A	55	MET	16

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 carbohydrates 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 46% for the well-defined parts and 46% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 5861

Chemical shift list name: *assigned_ chem_ shift_ list_ 1*

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	950
Number of shifts mapped to atoms	950
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
$^{13}C_{\alpha}$	168	0.02 ± 0.25	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	137	0.10 ± 0.16	None needed (< 0.5 ppm)
$^{13}C'$	157	-0.45 ± 0.14	None needed (< 0.5 ppm)
^{15}N	159	-0.03 ± 0.53	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 46%, i.e. 930 atoms were assigned a chemical shift out of a possible 2016. 0 out of 28 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	796/823 (97%)	321/328 (98%)	319/334 (96%)	156/161 (97%)
Sidechain	134/1103 (12%)	0/637 (0%)	134/411 (33%)	0/55 (0%)

Continued on next page...

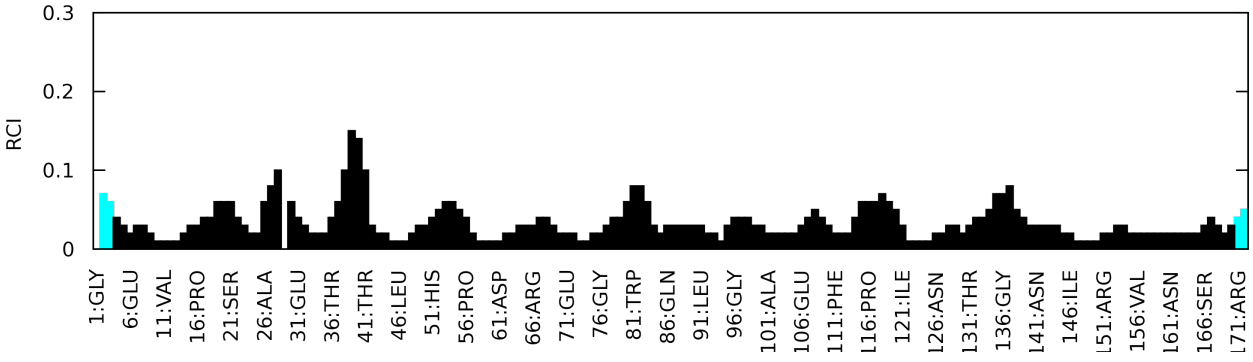
Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	0/90 (0%)	0/46 (0%)	0/30 (0%)	0/14 (0%)
Overall	930/2016 (46%)	321/1011 (32%)	453/775 (58%)	156/230 (68%)

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.

Random coil index (RCI) for chain A: