wwPDB NMR Structure Validation Summary Report (i)

Feb 26, 2022 – 11:54 AM EST

PDB ID	:	2CPH
Title	:	Solution structure of the C-terminal RNA recognition motif of hypothetical RNA-binding protein RBM19
Authors	:	Nagata, T.; Muto, Y.; Inoue, M.; Kigawa, T.; Terada, T.; Shirouzu, M.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2005-05-19

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:

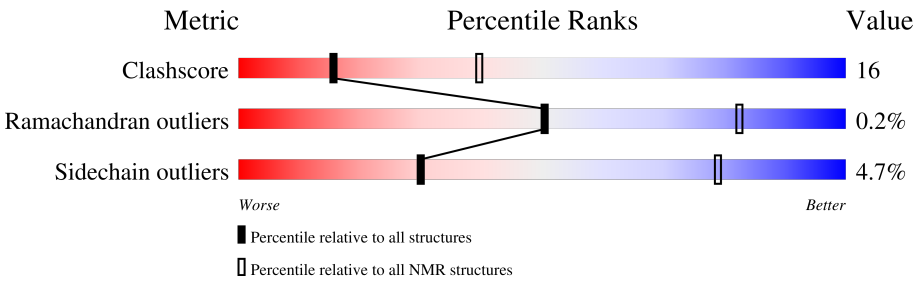
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
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	107	46% 18% 36%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 11 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:462-A:493, A:504-A:539 (68)	0.20	11

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. No single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1665 atoms, of which 837 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called RNA binding motif protein 19.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	107	1665	520	837	153	153	2	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	447	GLY	-	cloning artifact	UNP Q8R3C6
A	448	SER	-	cloning artifact	UNP Q8R3C6
A	449	SER	-	cloning artifact	UNP Q8R3C6
A	450	GLY	-	cloning artifact	UNP Q8R3C6
A	451	SER	-	cloning artifact	UNP Q8R3C6
A	452	SER	-	cloning artifact	UNP Q8R3C6
A	453	GLY	-	cloning artifact	UNP Q8R3C6
A	548	SER	-	cloning artifact	UNP Q8R3C6
A	549	GLY	-	cloning artifact	UNP Q8R3C6
A	550	PRO	-	cloning artifact	UNP Q8R3C6
A	551	SER	-	cloning artifact	UNP Q8R3C6
A	552	SER	-	cloning artifact	UNP Q8R3C6
A	553	GLY	-	cloning artifact	UNP Q8R3C6

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: RNA binding motif protein 19

Chain A: 46% 18% 36%

G447 S448 S449 G450 S451 S452 G453 Q454 V455 P456 K457 K458 Q459 T460 T461 V466 I469 Q475 R479 F482 S483 T484 F485 L488 K489 T490 V491 R492 L493 P494 K495 K496 M497 T498 G499 T500 G501 A502 H503 F506 F511 I512 F521 L524 C525 H526 W539 A540 D541

S542 E543 V544 T545 V546 Q547 S548 G549 P550 S551 S552 G553

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

The representative model is number 11. Colouring as in section 4.1 above.

• Molecule 1: RNA binding motif protein 19

Chain A: 39% 23% • 36%

G447 S448 S449 G450 S451 S452 G453 Q454 V455 P456 K457 K458 Q459 T460 T461 I464 L465 V466 I469 Q475 S483 T484 F485 G486 K489 T490 V491 R492 L493 P494 K495 K496 M497 T498 G499 T500 G501 A502 H503 R504 G505 F506 G507 F508 F511 I512 T513 K514 K519 A520 F521 N522 A523

L524 C525 H526 W539 A540 D541 S542 E543 V544 T545 V546 Q547 S548 G549 P550 S551 S552 G553

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	567	581	581	19±4
All	All	11340	11620	11620	375

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 16.

5 of 141 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:466:VAL:HG13	1:A:469:ILE:HD11	0.91	1.38	3	17
1:A:482:PHE:CE2	1:A:524:LEU:HD12	0.88	2.03	14	3
1:A:489:LYS:HG2	1:A:512:ILE:HD13	0.77	1.54	8	13
1:A:479:ARG:HA	1:A:491:VAL:HG21	0.73	1.59	1	1
1:A:464:ILE:HD12	1:A:511:PHE:CE1	0.73	2.18	8	5

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	68/107 (64%)	59±2 (87±4%)	8±2 (12±4%)	0±0 (0±1%)	50 82
All	All	1360/2140 (64%)	1189 (87%)	168 (12%)	3 (0%)	50 82

All 3 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	525	CYS	1
1	A	484	THR	1
1	A	483	SER	1

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	60/90 (67%)	57±1 (95±2%)	3±1 (5±2%)	30 79
All	All	1200/1800 (67%)	1144 (95%)	56 (5%)	30 79

5 of 19 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	506	PHE	11
1	A	516	ASP	5
1	A	475	GLN	5
1	A	504	ARG	4
1	A	518	LYS	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