W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

Full wwPDB NMR Structure Validation Report (i)

Sep 6, 2023 – 10:11 am BST

PDB ID	:	7Q33
BMRB ID	:	34673
Title	:	Solution structure of RBM39 RRM2 bound to 5'-AGCUUUG-3
Authors	:	Campagne, S.; Allain, F.H.
Deposited on	:	2021-10-26

This is a Full wwPDB NMR Structure Validation 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

1   Overall quality at a glance   ⓘ

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 84%.

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.
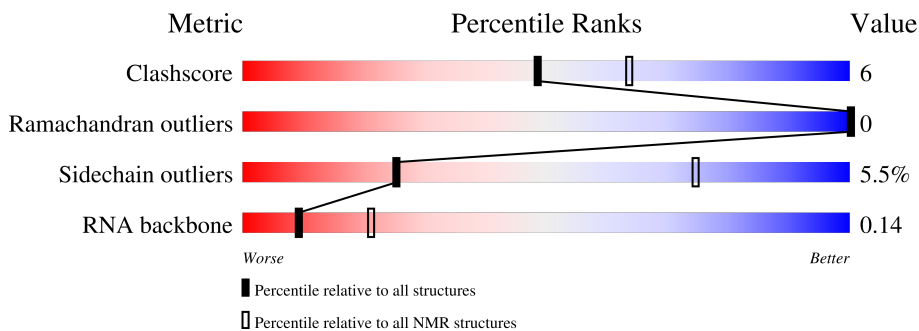
Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428
RNA backbone	4643	676

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	93	66% 10% 25%
2	B	7	57% 29% 14%

2 Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 7 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:249-A:284, A:293-A:326 (70)	0.12	7

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1671 atoms, of which 798 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called RNA-binding protein 39.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	93	1452	459	724	126	136	7	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	246	MET	-	initiating methionine	UNP Q14498
A	333	LEU	-	expression tag	UNP Q14498
A	334	GLU	-	expression tag	UNP Q14498
A	335	LEU	-	expression tag	UNP Q14498
A	336	VAL	-	expression tag	UNP Q14498
A	337	PRO	-	expression tag	UNP Q14498
A	338	ARG	-	expression tag	UNP Q14498

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	7	219	66	74	24	49	6	0

4 Residue-property plots ⓘ

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 protein 39
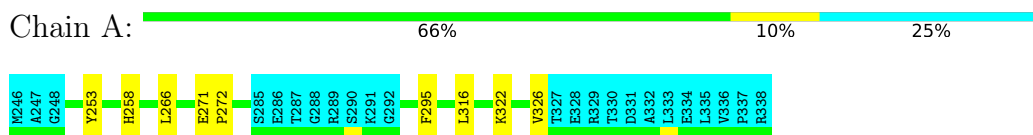
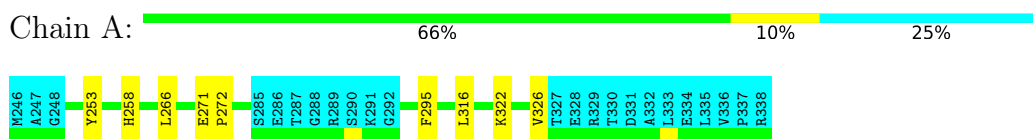- Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1
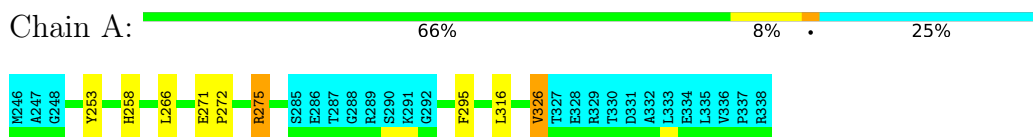
- Molecule 1: RNA-binding protein 39
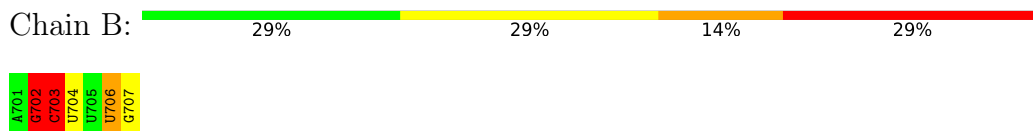

- Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

4.2.2 Score per residue for model 2
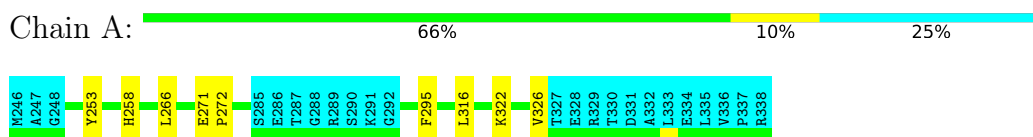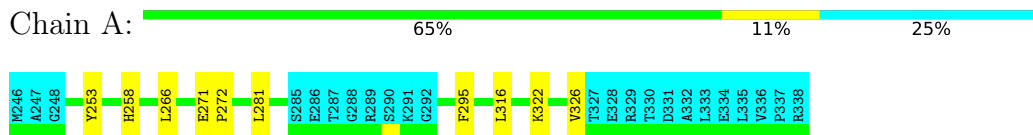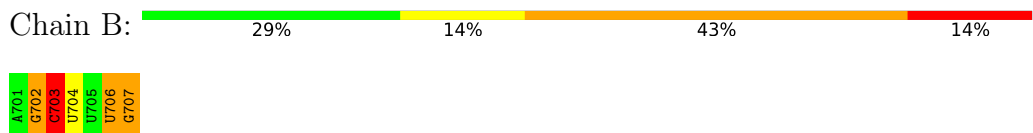
• Molecule 1: RNA-binding protein 39

Chain A:

• Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

Chain B:

4.2.3 Score per residue for model 3

• Molecule 1: RNA-binding protein 39

Chain A:

• Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

Chain B:

4.2.4 Score per residue for model 4
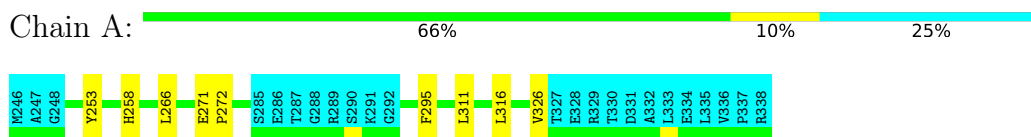
• Molecule 1: RNA-binding protein 39

Chain A:

• Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

Chain B:

4.2.5 Score per residue for model 5

• Molecule 1: RNA-binding protein 39

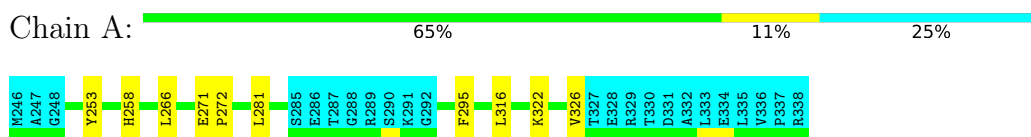Chain A: 66% 10% 25%

M246 A247 G248 Y253 H258 L266 E271 P272 S285 E286 T287 G288 R289 S290 K291 G292 F295 L311 L316 V326 T327 E328 R329 T330 D331 A332 L333 E334 L335 V336 P337 R338

• Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

Chain B: 43% 43% 14%

A701 G702 C703 U704 U705 U706 G707

4.2.6 Score per residue for model 6

• Molecule 1: RNA-binding protein 39

Chain A: 65% 11% 25%

M246 A247 G248 Y253 H258 L266 E271 P272 L281 S285 E286 T287 G288 R289 S290 K291 G292 F295 L316 K322 V326 T327 E328 R329 T330 D331 A332 L333 E334 L335 V336 P337 R338

• Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

Chain B: 14% 29% 43% 14%

A701 G702 C703 U704 U705 U706 G707

4.2.7 Score per residue for model 7 (medoid)

• Molecule 1: RNA-binding protein 39

Chain A: 66% 10% 25%

M246 A247 G248 Y253 H258 L266 E271 P272 L281 S285 E286 T287 G288 R289 S290 K291 G292 F295 L316 V326 T327 E328 R329 T330 D331 A332 L333 E334 L335 V336 P337 R338

• Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

Chain B: 57% 29% 14%

A701 G702 C703 U704 U705 U706 G707
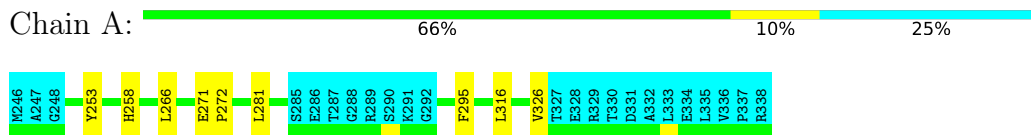
4.2.8 Score per residue for model 8

• Molecule 1: RNA-binding protein 39

Chain A:

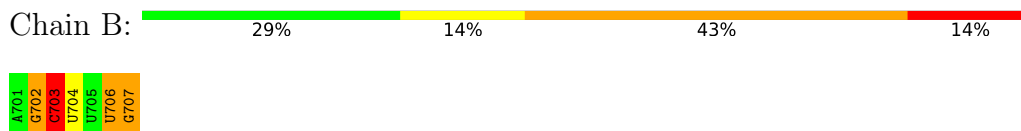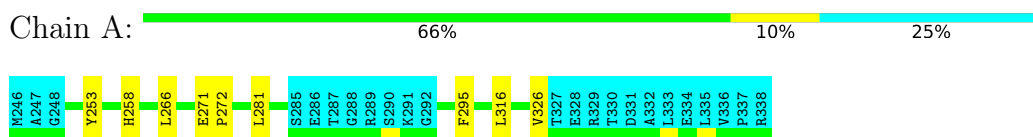
• Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

Chain B:
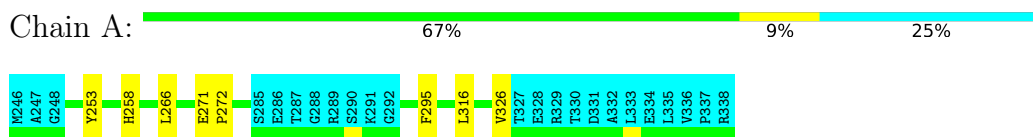
4.2.9 Score per residue for model 9

• Molecule 1: RNA-binding protein 39

Chain A:

• Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

Chain B:
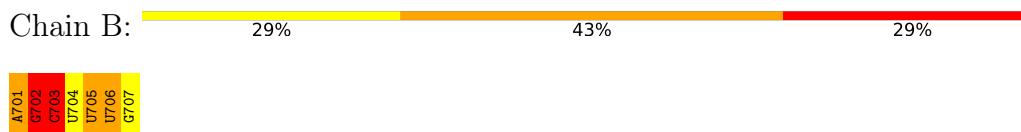
4.2.10 Score per residue for model 10

• Molecule 1: RNA-binding protein 39

Chain A:

• Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

Chain B:
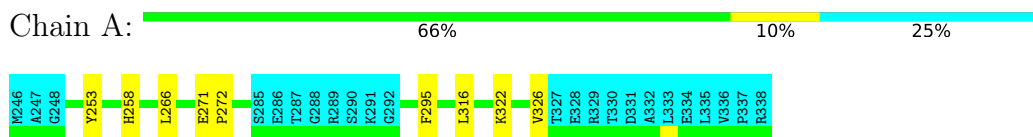
4.2.11   Score per residue for model 11

• Molecule 1: RNA-binding protein 39

Chain A:

• Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

Chain B:

4.2.12   Score per residue for model 12

• Molecule 1: RNA-binding protein 39

Chain A:

• Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

Chain B:

4.2.13   Score per residue for model 13

• Molecule 1: RNA-binding protein 39

Chain A:

• Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

Chain B:

4.2.14 Score per residue for model 14

• Molecule 1: RNA-binding protein 39

Chain A:

• Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

Chain B:
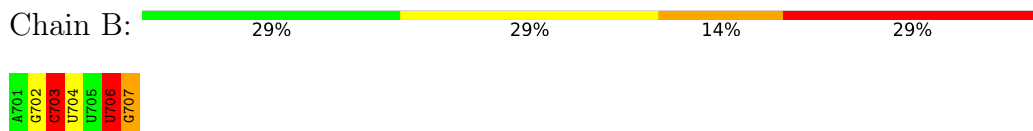
4.2.15 Score per residue for model 15

• Molecule 1: RNA-binding protein 39
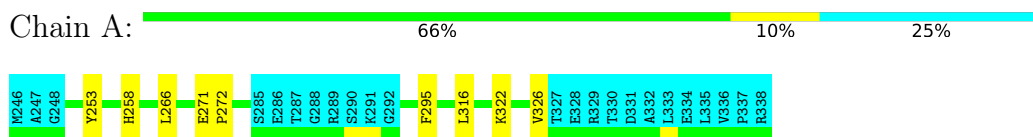
Chain A:

• Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

Chain B:

4.2.16 Score per residue for model 16

• Molecule 1: RNA-binding protein 39

Chain A:

• Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

Chain B:
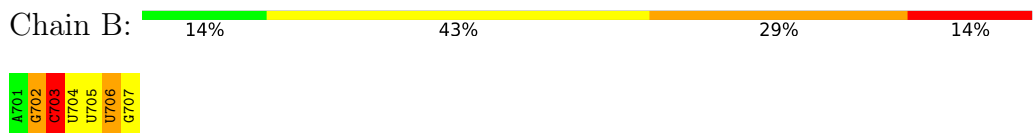
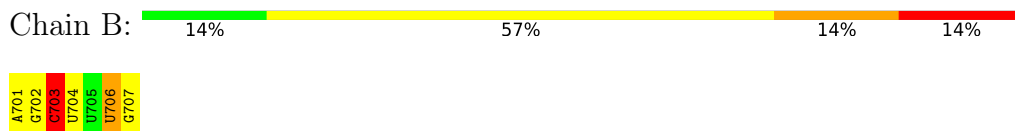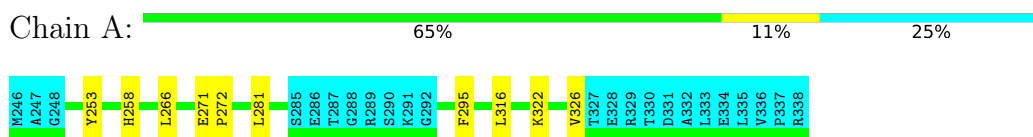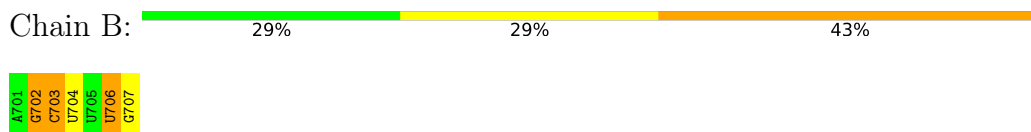### 4.2.17 Score per residue for model 17

• Molecule 1: RNA-binding protein 39

Chain A:

• Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

Chain B:
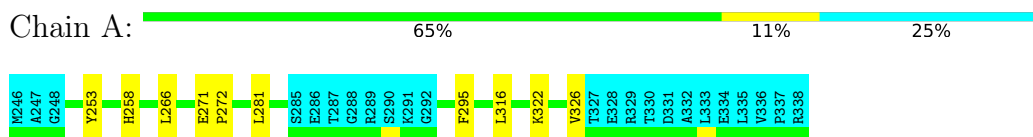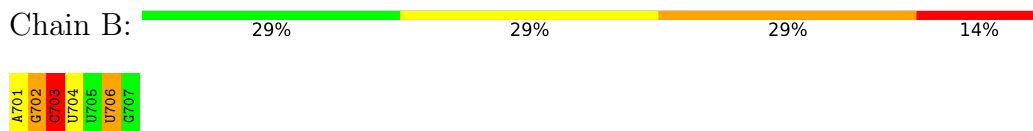
4.2.18 Score per residue for model 18

• Molecule 1: RNA-binding protein 39

Chain A:

• Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

Chain B:
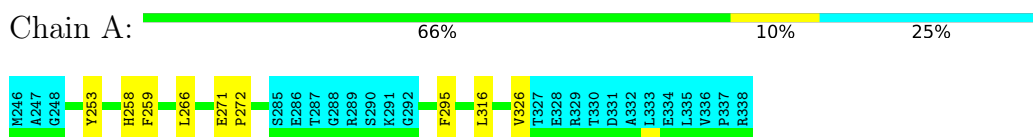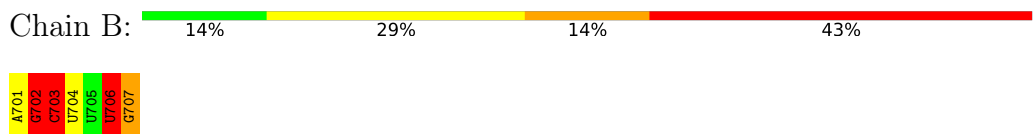
4.2.19 Score per residue for model 19

• Molecule 1: RNA-binding protein 39

Chain A:

• Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

Chain B:

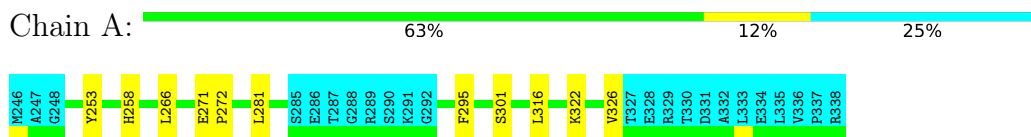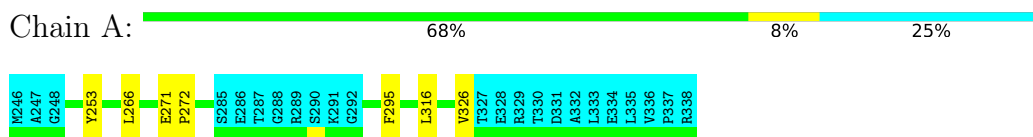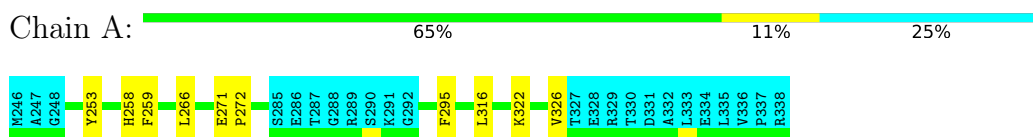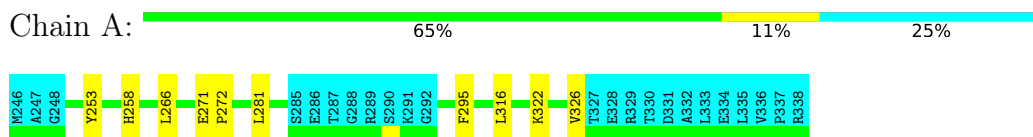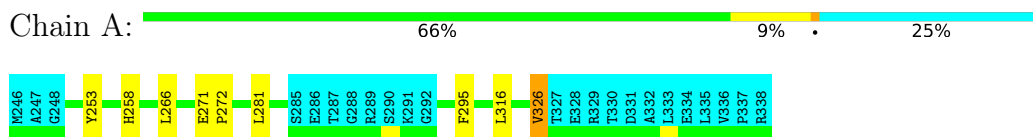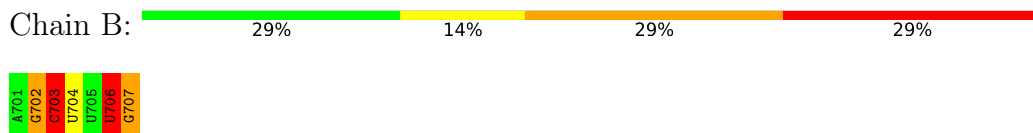### 4.2.20 Score per residue for model 20

• Molecule 1: RNA-binding protein 39

Chain A:

• Molecule 2: RNA (5'-R(*AP*GP*CP*UP*UP*UP*G)-3')

Chain B:
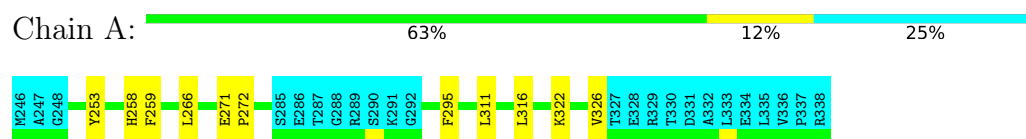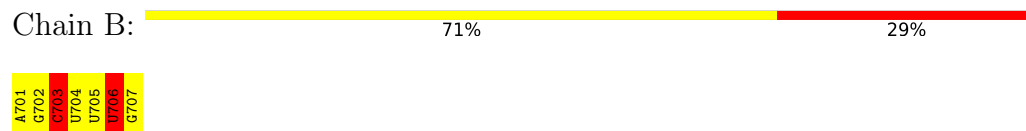
5 Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *simulated annealing.*

Of the 50 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
CYANA	structure calculation	
Amber	refinement	

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)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	1085
Number of shifts mapped to atoms	1085
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	84%

6   Model quality ⓘ

6.1   Standard geometry ⓘ

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the (average) root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#Z>5	RMSZ	#Z>5
1	A	0.52±0.00	0±0/571 (0.0± 0.0%)	0.75±0.01	0±0/765 (0.0± 0.0%)
2	B	1.06±0.02	0±0/161 (0.0± 0.0%)	1.77±0.10	4±1/249 (1.5± 0.5%)
All	All	0.67	0/14640 (0.0%)	1.10	77/20280 (0.4%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
2	B	0.0±0.0	0.7±0.6
All	All	0	13

There are no bond-length outliers.

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	B	703	C	O4'-C1'-N1	12.57	118.26	108.20	1	19
2	B	703	C	C1'-O4'-C4'	-10.12	101.81	109.90	18	16
2	B	702	G	C1'-O4'-C4'	-7.41	103.97	109.90	13	15
2	B	702	G	O4'-C1'-N9	7.00	113.80	108.20	17	12
2	B	703	C	C5'-C4'-O4'	6.93	117.42	109.10	9	2
2	B	706	U	O4'-C1'-N1	6.26	113.21	108.20	13	5
2	B	705	U	O4'-C1'-N1	5.50	112.60	108.20	9	2
2	B	705	U	P-O3'-C3'	5.28	126.04	119.70	3	3
2	B	706	U	C1'-O4'-C4'	-5.21	105.73	109.90	19	2
2	B	703	C	C4'-C3'-C2'	-5.04	97.56	102.60	15	1

There are no chirality outliers.

All unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
2	B	703	C	Sidechain	12
2	B	707	G	Sidechain	1

6.2 Too-close contacts ⓘ

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	558	553	553	8±1
2	B	145	74	76	6±1
All	All	14060	12540	12580	164

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 6.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Total
1:A:258:HIS:CD2	2:B:703:C:H41	0.70	2.04	16	19
1:A:295:PHE:CE2	2:B:707:G:C8	0.65	2.85	18	10
1:A:295:PHE:CE1	2:B:707:G:C8	0.64	2.85	6	8
1:A:295:PHE:CZ	2:B:707:G:C8	0.59	2.90	15	18
1:A:253:TYR:CD2	2:B:706:U:C6	0.55	2.95	1	9
1:A:253:TYR:CD1	2:B:706:U:C6	0.54	2.96	7	11
1:A:295:PHE:CE1	1:A:326:VAL:CG2	0.52	2.93	13	3
1:A:295:PHE:CE1	1:A:326:VAL:HG21	0.51	2.41	1	9
1:A:295:PHE:CE2	1:A:326:VAL:HG21	0.51	2.41	7	11
1:A:295:PHE:CE2	1:A:326:VAL:CG2	0.50	2.94	17	3
1:A:253:TYR:CG	2:B:706:U:C5	0.50	2.99	13	20
2:B:701:A:C8	2:B:701:A:H5'	0.49	2.43	13	2
1:A:271:GLU:N	1:A:272:PRO:HD2	0.46	2.26	16	20
2:B:702:G:C8	2:B:703:C:C5	0.46	3.03	20	3
2:B:702:G:H1'	2:B:703:C:C6	0.45	2.47	9	3
1:A:259:PHE:CG	2:B:701:A:C2	0.45	3.05	15	3
1:A:259:PHE:CD1	2:B:701:A:C2	0.44	3.06	20	1
1:A:258:HIS:CD2	2:B:703:C:N4	0.43	2.85	1	5
1:A:275:ARG:H	1:A:275:ARG:HD2	0.43	1.72	2	1
1:A:258:HIS:HA	2:B:702:G:N7	0.41	2.30	3	1
2:B:701:A:H5"	2:B:701:A:C8	0.41	2.51	5	2

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:701:A:C8	2:B:701:A:H5"	0.40	2.52	9	1
1:A:253:TYR:HA	1:A:295:PHE:CE2	0.40	2.51	17	1

6.3 Torsion angles ⓘ

6.3.1 Protein backbone ⓘ

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	70/93 (75%)	65±1 (93±1%)	5±1 (7±1%)	0±0 (0±0%)	100	100
All	All	1400/1860 (75%)	1305 (93%)	95 (7%)	0 (0%)	100	100

There are no Ramachandran outliers.

6.3.2 Protein sidechains ⓘ

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	60/78 (77%)	57±1 (94±1%)	3±1 (6±1%)	25	74
All	All	1200/1560 (77%)	1134 (94%)	66 (6%)	25	74

All 7 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	266	LEU	20
1	A	316	LEU	20
1	A	322	LYS	12
1	A	281	LEU	9
1	A	326	VAL	2
1	A	311	LEU	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	275	ARG	1

6.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
2	B	6/7 (86%)	4±1 (65±14%)	3±1 (48±21%)	0.15±0.09
All	All	130/140 (93%)	78 (60%)	57 (44%)	0.14

The overall RNA backbone suiteness is 0.14.

All unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	B	704	U	20
2	B	706	U	20
2	B	703	C	17
2	B	702	G	10
2	B	707	G	6
2	B	705	U	5

All unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	B	702	G	19
2	B	701	A	10
2	B	704	U	8
2	B	705	U	8
2	B	703	C	6
2	B	706	U	6

6.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

6.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.6 Ligand geometry ⓘ

There are no ligands in this entry.

6.7 Other polymers ⓘ

There are no such molecules in this entry.

6.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

7 Chemical shift validation ⓘ

The completeness of assignment taking into account all chemical shift lists is 84% for the well-defined parts and 77% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *RBM39_2-AGCUUUG.str*

7.1.1 Bookkeeping ⓘ

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	1085
Number of shifts mapped to atoms	1085
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	85	-0.09 ± 0.10	None needed (< 0.5 ppm)
$^{13}C_\beta$	76	-0.17 ± 0.11	None needed (< 0.5 ppm)
$^{13}C'$	73	0.04 ± 0.17	None needed (< 0.5 ppm)
^{15}N	83	0.25 ± 0.37	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments ⓘ

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 84%, i.e. 930 atoms were assigned a chemical shift out of a possible 1103. 0 out of 10 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	344/351 (98%)	144/144 (100%)	133/140 (95%)	67/67 (100%)
Sidechain	472/531 (89%)	320/347 (92%)	148/165 (90%)	4/19 (21%)

Continued on next page...

Continued from previous page...

	Total	**¹H**	**¹³C**	**¹⁵N**
Aromatic	71/92 (77%)	42/46 (91%)	29/44 (66%)	0/2 (0%)
Sugar	34/77 (44%)	34/42 (81%)	0/35 (0%)	0/0 (—%)
Base	9/52 (17%)	9/31 (29%)	0/12 (0%)	0/9 (0%)
Overall	930/1103 (84%)	549/610 (90%)	310/396 (78%)	71/97 (73%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 77%, i.e. 1082 atoms were assigned a chemical shift out of a possible 1398. 0 out of 13 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	**¹H**	**¹³C**	**¹⁵N**
Backbone	419/467 (90%)	178/192 (93%)	158/186 (85%)	83/89 (93%)
Sidechain	549/710 (77%)	372/462 (81%)	173/219 (79%)	4/29 (14%)
Aromatic	71/92 (77%)	42/46 (91%)	29/44 (66%)	0/2 (0%)
Sugar	34/77 (44%)	34/42 (81%)	0/35 (0%)	0/0 (—%)
Base	9/52 (17%)	9/31 (29%)	0/12 (0%)	0/9 (0%)
Overall	1082/1398 (77%)	635/773 (82%)	360/496 (73%)	87/129 (67%)

7.1.4   Statistically unusual chemical shifts ⓘ

The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	291	LYS	CG	33.54	19.35 − 30.45	7.8

7.1.5   Random Coil Index (RCI) plots ⓘ

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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A: