Full wwPDB NMR Structure Validation Report (i)

May 7, 2024 – 03:50 pm BST

PDB ID	:	6SO9
BMRB ID	:	34428
Title	:	Mouse RBM20 RRM domain in complex with AUCUUA RNA
Authors	:	Mackereth, C.D.; Upadhyay, S.K.
Deposited on	:	2019-08-29

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
Mogul	:	1.8.4, CSD as541be (2020)
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)	:	2.36.2

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 90%.

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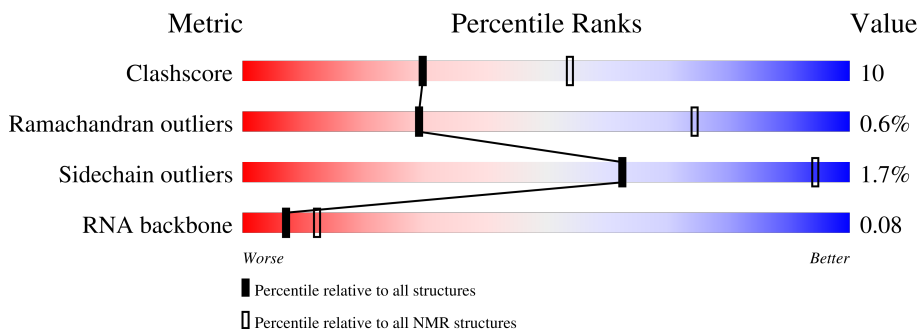
Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428
RNA backbone	4643	676

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	B	6	33% 67%
2	A	112	64% 21% 15%

2 Ensemble composition and analysis (i)

This entry contains 25 models. Model 16 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average.*

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:520-A:603, A:607-A:617 (95)	0.34	16

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 2 single-model clusters were found.

Cluster number	Models
1	1, 2, 4, 9, 10, 14, 16, 17, 18, 20, 23
2	3, 5, 6, 7, 8, 11, 19, 22
3	12, 13
4	21, 25
Single-model clusters	15; 24

3 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 1971 atoms, of which 972 are hydrogens and 0 are deuteriums.

- Molecule 1 is a RNA chain called AUCUUA.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	B	6	186	56	65	19	41	5	0

- Molecule 2 is a protein called RNA-binding protein 20.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	A	112	1785	551	907	158	162	7	0

There are 3 discrepancies between the modelled and reference sequences:

| Chain	Residue	Modelled	Actual	Comment	Reference
A	510	GLY	-	expression tag	UNP Q3UQS8
A	511	ALA	-	expression tag	UNP Q3UQS8
A	512	MET	-	expression tag	UNP Q3UQS8

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: AUCUUA

Chain B:

Chain A:
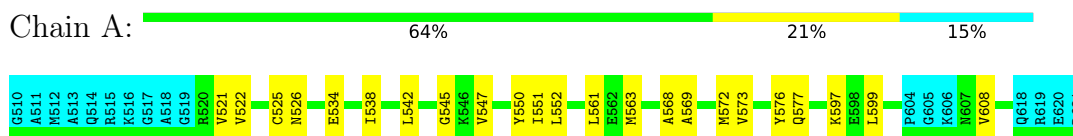
4.2 Scores per residue for each member of the ensemble
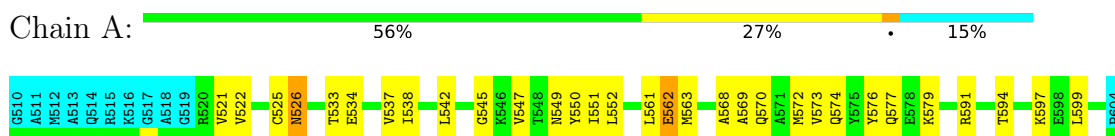
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: AUCUUA
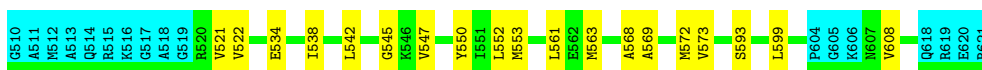
Chain B:

- Molecule 2: RNA-binding protein 20

Chain A:

G605 K606 I612 Q618 R619 E620 R621

4.2.2 Score per residue for model 2

• Molecule 1: AUCUUA

Chain B: 33% 50% 17%

ADN1 U2 C3 U4 U5 A6

• Molecule 2: RNA-binding protein 20

Chain A: 68% 17% 15%

G510 A511 M512 A513 Q514 R515 K516 G517 A518 G519 R520 V521 V522 E534 I538 L542 G545 K546 V547 Y550 I551 L552 M553 L561 E562 M563 A568 A569 M572 V573 S593 L599 P604 G605 K606 N607 V608 Q618 R619 E620 R621

4.2.3 Score per residue for model 3

• Molecule 1: AUCUUA

Chain B: 50% 33% 17%

ADN1 U2 C3 U4 U5 A6

• Molecule 2: RNA-binding protein 20

Chain A: 66% 19% 15%

G510 A511 M512 A513 Q514 R515 K516 G517 A518 G519 R520 V521 C525 E534 I538 G541 L542 G545 Y550 I551 L552 M553 M563 A568 A569 M572 V573 Y576 Q577 A581 S593 L599 P604 G605 K606 I611 Q618 R619 E620 R621
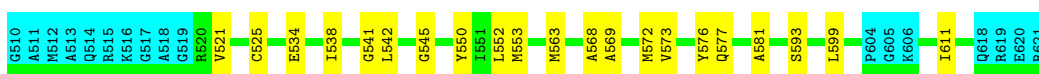
4.2.4 Score per residue for model 4

• Molecule 1: AUCUUA

Chain B: 67% 33%

ADN1 U2 C3 U4 U5 A6

• Molecule 2: RNA-binding protein 20

Chain A: 62% 22% • 15%

G510 A511 M512 A513 Q514 R515 K516 G517 A518 G519 R520 V521 V522 N526 T533 E534 V537 I538 V547 Y550 I551 L552 M553 L561 E562 M563 A568 A569 M572 V573 Y576 Q577 I590 S593 K597 E598 L599 P604 G605 K606 N607 V608 Q618 R619 E620 R621
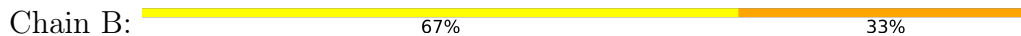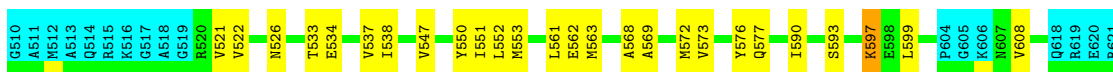
4.2.5 Score per residue for model 5

• Molecule 1: AUCUUA

Chain B: 33% 50% 17%

ADN1 U2 C3 U4 U5 A6

• Molecule 2: RNA-binding protein 20

Chain A: 65% 20% 15%

G510 A511 M512 A513 Q514 R515 K516 G517 A518 G519 R520 V521 V522 C525 N526 E534 I538 V547 Y550 I551 L552 L561 E562 M563 A568 A569 M572 V573 Y576 Q577 I590 S593 L599 P604 G605 K606 Q618 R619 E620 R621

4.2.6 Score per residue for model 6

• Molecule 1: AUCUUA

Chain B: 50% 50%

ADN1 U2 C3 U4 U5 A6

• Molecule 2: RNA-binding protein 20

Chain A: 70% 15% 15%

G510 A511 M512 A513 Q514 R515 K516 G517 A518 G519 R520 V521 N526 I538 V547 Y550 I551 M563 A568 A569 M572 V573 Y576 Q577 A581 L588 L589 I590 L599 P604 G605 K606 Q618 R619 E620 R621

4.2.7 Score per residue for model 7

• Molecule 1: AUCUUA

Chain B: 33% 67%

ADN1 U2 C3 U4 U5 A6

• Molecule 2: RNA-binding protein 20

Chain A: 60% 25% 15%

G510 A511 M512 A513 Q514 R515 K516 G517 A518 G519 R520 V521 V522 C525 T533 E534 V537 I538 L542 G545 K546 V547 Y550 I551 L552 L561 E562 M563 A568 A569 M572 V573 Y576 Q577 I590 R591 M592 S593 T594 R595 L599 P604 G605 K606 I615

Q618 R619 E620 R621
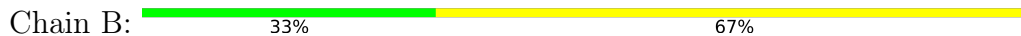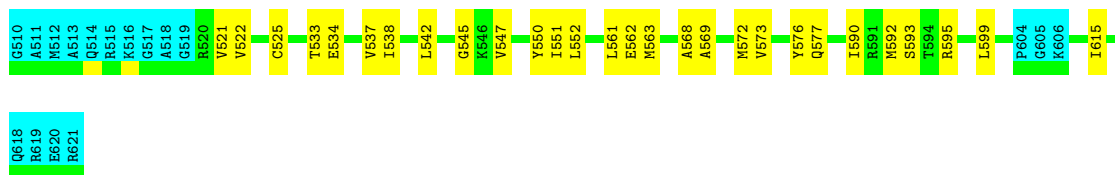
4.2.8 Score per residue for model 8

• Molecule 1: AUCUUA

Chain B: 17% 83%

ADN1 U2 C3 U4 U5 A6

• Molecule 2: RNA-binding protein 20

Chain A: 54% 30% · 15%

G510 A511 M512 A513 Q514 R515 K516 G517 A518 G519 R520 V521 C525 N526 T533 E534 V537 I538 G541 L542 P543 F544 G545 K546 V547 T548 N549 Y550 I551 L552 M553 F560 L561 E562 M563 A568 A569 M572 Y576 A581 L589 I590 K597 E598 L599 Q600 L601 K602 K603

P604 G605 K606 N607 V608 Q618 R619 E620 R621
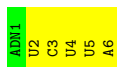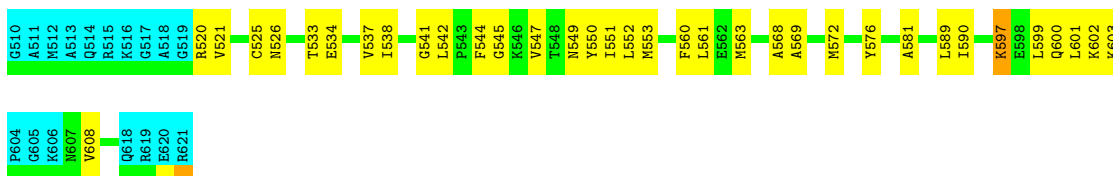
4.2.9 Score per residue for model 9

• Molecule 1: AUCUUA

Chain B: 100%

ADN1 U2 C3 U4 U5 A6

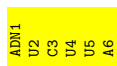• Molecule 2: RNA-binding protein 20

Chain A: 63% 21% · 15%

G510 A511 M512 A513 Q514 R515 K516 G517 A518 G519 R520 V521 V522 C525 N526 I538 G541 L542 G545 K546 V547 Y550 M553 K554 N557 L561 E562 M563 A568 A569 M572 V573 Q577 L589 L599 P604 G605 K606 N607 V608 Q618 R619 E620 R621

4.2.10 Score per residue for model 10

• Molecule 1: AUCUUA

Chain B: 33% 33% 33%

ADN1 U2 C3 U4 U5 A6

• Molecule 2: RNA-binding protein 20

Chain A: 61% 21% • 15%

G510 A511 M512 A513 Q514 R515 K516 G517 A518 G519 R520 V521 C525 N526 L527 T533 E534 V537 I538 L542 G545 K546 V547 Y550 I551 L552 N557 M563 T566 E567 A568 A569 M572 V573 Y576 N584 I590 K597 E598 L599 P604 G605 K606 N607 V608

Q618 R619 E620 R621

4.2.11 Score per residue for model 11

• Molecule 1: AUCUUA

Chain B: 33% 50% 17%

ADN1 U2 C3 U4 U5 A6

• Molecule 2: RNA-binding protein 20

Chain A: 62% 22% 15%

G510 A511 M512 A513 Q514 R515 K516 G517 A518 G519 R520 V521 V522 C525 N526 E534 I538 V547 Y550 I551 L552 M553 L561 E562 M563 A568 A569 M572 V573 Y576 Q577 A581 I590 K597 E598 L599 P604 G605 K606 N607 V608 S617 Q618 R619 E620 R621

4.2.12 Score per residue for model 12

• Molecule 1: AUCUUA

Chain B: 17% 83%

ADN1 U2 C3 U4 U5 A6

• Molecule 2: RNA-binding protein 20

Chain A: 61% 24% 15%

G510 A511 M512 A513 Q514 R515 K516 G517 A518 G519 R520 V521 V522 C525 N526 L527 C532 T533 V537 I538 V547 Y550 I551 L561 E562 M563 A568 M572 V573 Y576 Q577 L589 I590 L599 Q600 L601 P604 G605 K606 N607 V608 I611 D614 Q618 R619

E620 R621

4.2.13 Score per residue for model 13

• Molecule 1: AUCUUA

Chain B:

• Molecule 2: RNA-binding protein 20

Chain A:

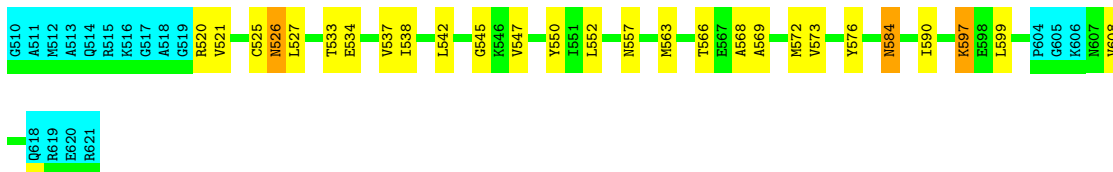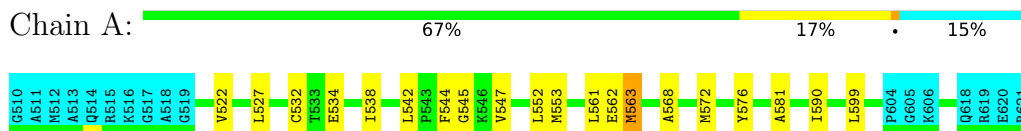### 4.2.14 Score per residue for model 14

• Molecule 1: AUCUUA
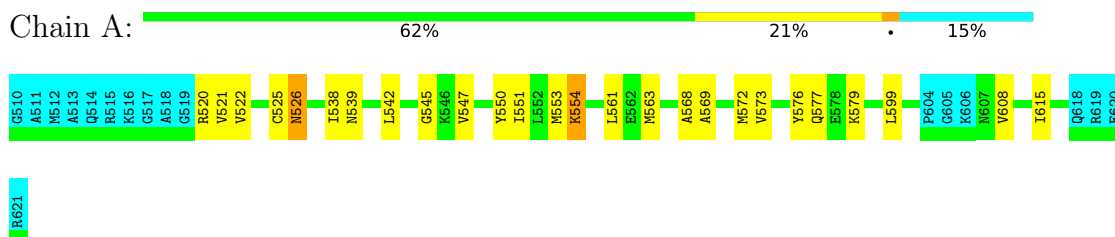
Chain B:

• Molecule 2: RNA-binding protein 20

Chain A:

4.2.15 Score per residue for model 15

• Molecule 1: AUCUUA

Chain B:

• Molecule 2: RNA-binding protein 20

Chain A:

4.2.16 Score per residue for model 16 (medoid)

• Molecule 1: AUCUUA

Chain B:

• Molecule 2: RNA-binding protein 20

Chain A:

4.2.17 Score per residue for model 17

• Molecule 1: AUCUUA

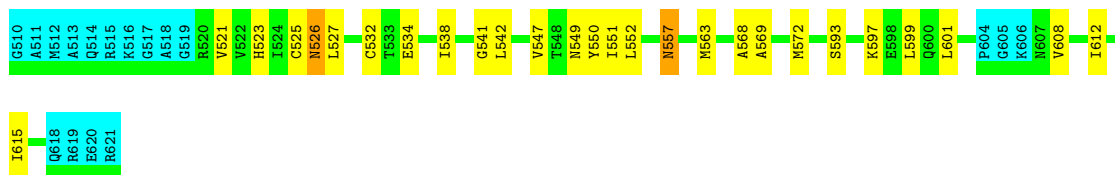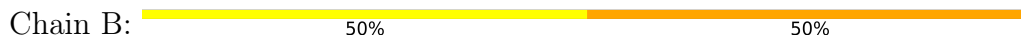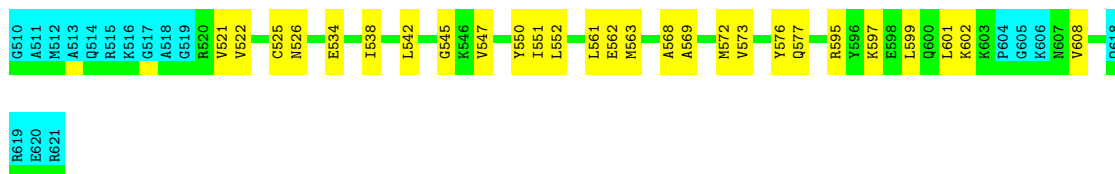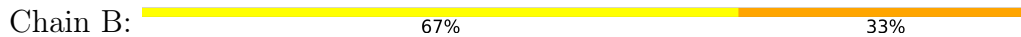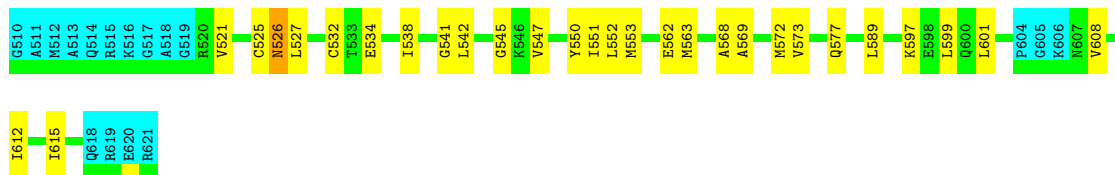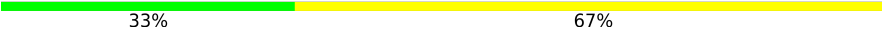Chain B:

• Molecule 2: RNA-binding protein 20

Chain A:

4.2.18 Score per residue for model 18

• Molecule 1: AUCUUA

Chain B: 33%   67%

ADN1 U2 C3 U4 U5 A6

• Molecule 2: RNA-binding protein 20

Chain A: 59%   25%   •   15%

G510 A511 M512 A513 Q514 R515 K516 G517 A518 G519 R520 V521 V522 C525 N526 T533 E534 V537 I538 L542 G545 Y550 I551 L552 M553 L561 E562 T566 E567 A568 A569 M572 V573 Y576 Q577 E578 K579 N584 K597 E598 L599 P604 G605 K606 I615

Q618 R619 E620 R621
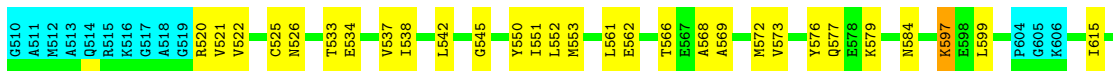
4.2.19 Score per residue for model 19

• Molecule 1: AUCUUA

Chain B: 67%   33%

ADN1 U2 C3 A6

• Molecule 2: RNA-binding protein 20

Chain A: 69%   16%   15%

G510 A511 M512 A513 Q514 R515 K516 G517 A518 G519 R520 V521 C525 N526 E534 I538 G541 V547 Y550 I551 L552 M563 A568 A569 M572 V573 Q577 K597 E598 L599 P604 G605 K606 N607 V608 Q618 R619 E620 R621

4.2.20 Score per residue for model 20

• Molecule 1: AUCUUA

Chain B: 50%   50%

ADN1 U2 C3 U4 U5 A6

• Molecule 2: RNA-binding protein 20

Chain A: 68%   17%   15%

G510 A511 M512 A513 Q514 R515 K516 G517 A518 G519 R520 V521 C525 N526 I538 L542 V547 Y550 L561 E562 M563 A568 A569 M572 V573 Q577 S593 K597 E598 L599 P604 G605 K606 N607 V608 I615 Q618 R619 E620 R621
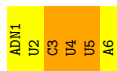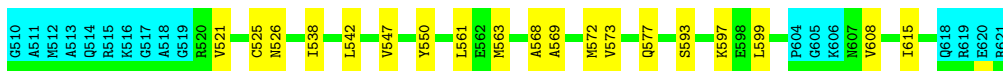
4.2.21 Score per residue for model 21

• Molecule 1: AUCUUA

Chain B:

33% 33% 33%

ADN1 U2 C3 U4 U5 A6

• Molecule 2: RNA-binding protein 20

Chain A:

65% 20% 15%

G510 A511 M512 A513 Q514 R515 K516 G517 A518 G519 R520 V521 V522 C525 N526 T533 E534 V537 I538 V547 Y550 I551 L552 L561 E562 M563 A568 A569 M572 V573 Q577 M592 L599 P604 G605 K606 N607 V608 Q618 R619 E620 R621
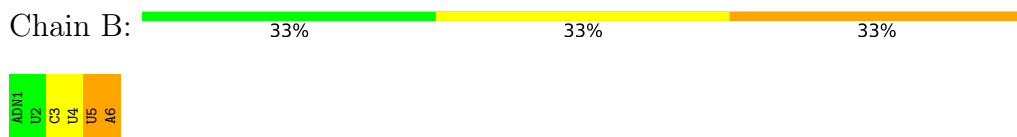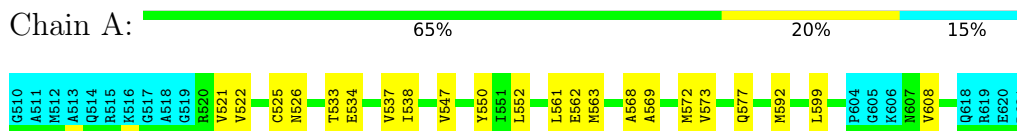
4.2.22 Score per residue for model 22

• Molecule 1: AUCUUA

Chain B:

50% 33% 17%

ADN1 U2 C3 U4 U5 A6

• Molecule 2: RNA-binding protein 20

Chain A:

59% 26% 15%

G510 A511 M512 A513 Q514 R515 K516 G517 A518 G519 R520 V521 V522 C525 T533 E534 V537 I538 G541 L542 G545 K546 V547 Y550 I551 L552 L561 E562 M563 A568 A569 M572 V573 Y576 Q577 E578 K579 P580 A581 I590 S593 L599 P604 G605 K606

Q613 D614 I615 Q618 R619 E620 R621
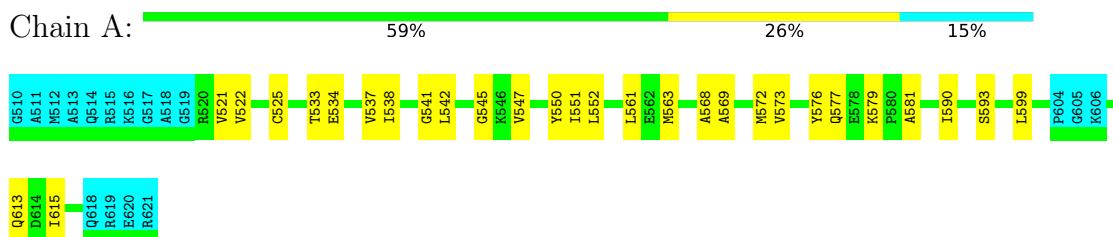
4.2.23 Score per residue for model 23

• Molecule 1: AUCUUA

Chain B:

33% 67%

ADN1 U2 C3 U4 U5 A6

• Molecule 2: RNA-binding protein 20

Chain A:

65% 19% • 15%

G510 A511 M512 A513 Q514 R515 K516 G517 A518 G519 R520 V521 V522 C525 N526 I538 V547 Y550 I551 L561 E562 M563 A568 A569 M572 V573 Y576 K579 P580 A581 L589 I590 K597 E598 L599 P604 G605 K606 I611 Q618 R619 E620 R621

4.2.24 Score per residue for model 24

● Molecule 1: AUCUUA

Chain B:

50%   50%

ADN1 U2 C3 U4 U5 A6

● Molecule 2: RNA-binding protein 20

Chain A:

62%   23%   15%

G510 A511 M512 A513 Q514 R515 K516 G517 A518 G519 R520 V521 N526 T533 E534 V537 I538 N539 L540 V547 Y550 I551 L552 M553 M563 T566 E567 A568 A569 M572 V573 Y576 Q577 E578 K579 P580 A581 I590 S593 T594 R595 P604 G605 K606 I611 Q618

R619 E620 R621

4.2.25 Score per residue for model 25

● Molecule 1: AUCUUA

Chain B:

33%   67%

ADN1 U2 C3 U4 U5 A6

● Molecule 2: RNA-binding protein 20

Chain A:

62%   22%   •   15%

G510 A511 M512 A513 Q514 R515 K516 G517 A518 G519 R520 V521 V522 C525 N526 T533 E534 V537 I538 N539 L540 G541 L542 V547 Y550 I551 L552 N557 L561 E562 M563 A568 A569 M572 V573 I583 L588 L599 P604 G605 K606 N607 V608 I612 Q618

R619 E620 R621
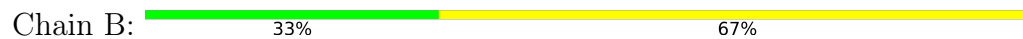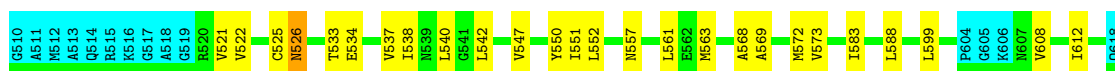
5 Refinement protocol and experimental data overview ⓘ

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

Of the 100 calculated structures, 25 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
ARIA	structure calculation	
CNS	structure calculation	

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	1478
Number of shifts mapped to atoms	1478
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	90%

6 Model quality ⓘ

6.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: ADN

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 ⓘ

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	B	121	65	65	3±2
2	A	753	775	773	17±3
All	All	21850	21000	20950	433

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 10.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Models Total
1:B:1:ADN:H4'	1:B:2:U:O5'	0.75	1.82	24	2
2:A:563:MET:SD	2:A:569:ALA:HA	0.70	2.27	23	21
2:A:521:VAL:HG13	2:A:599:LEU:HD21	0.69	1.64	6	23
1:B:1:ADN:O2'	1:B:2:U:H5"	0.67	1.89	20	1
2:A:522:VAL:HG21	2:A:563:MET:SD	0.67	2.29	12	1
1:B:5:U:H4'	2:A:553:MET:HA	0.66	1.68	4	10
2:A:568:ALA:O	2:A:572:MET:HG2	0.64	1.93	8	2
2:A:534:GLU:HA	2:A:552:LEU:HD11	0.63	1.69	15	9
2:A:538:ILE:HD11	2:A:550:TYR:CG	0.63	2.29	3	23
2:A:520:ARG:NH1	2:A:566:THR:HA	0.62	2.09	18	2
2:A:576:TYR:CE2	2:A:590:ILE:HD12	0.61	2.29	24	8

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:B:3:C:H2'	2:A:595:ARG:HD2	0.61	1.73	24	1
2:A:562:GLU:HB2	2:A:599:LEU:HD12	0.60	1.72	13	1
2:A:568:ALA:O	2:A:572:MET:HG3	0.60	1.96	18	23
2:A:521:VAL:HB	2:A:593:SER:OG	0.59	1.96	2	2
2:A:573:VAL:O	2:A:577:GLN:HG2	0.59	1.97	17	9
2:A:544:PHE:HB2	2:A:572:MET:HE1	0.58	1.75	13	2
1:B:5:U:N3	2:A:551:ILE:HG23	0.56	2.15	5	10
1:B:5:U:C2	2:A:551:ILE:HG23	0.55	2.36	17	15
1:B:5:U:O4	2:A:601:LEU:HA	0.55	2.02	16	3
1:B:5:U:C4	2:A:551:ILE:HG23	0.54	2.38	8	4
1:B:3:C:N3	2:A:593:SER:HA	0.54	2.17	24	2
2:A:538:ILE:HG23	2:A:547:VAL:HG21	0.53	1.79	24	10
2:A:522:VAL:HB	2:A:561:LEU:O	0.53	2.03	12	16
2:A:597:LYS:HD2	2:A:597:LYS:N	0.53	2.19	19	6
2:A:533:THR:O	2:A:537:VAL:HG23	0.53	2.04	24	11
1:B:3:C:O2	2:A:593:SER:HB3	0.52	2.03	5	6
1:B:1:ADN:HO2'	1:B:2:U:H6	0.52	1.45	9	1
2:A:576:TYR:CE1	2:A:581:ALA:HB2	0.51	2.40	3	7
2:A:573:VAL:O	2:A:577:GLN:HG3	0.51	2.05	1	8
2:A:538:ILE:HD11	2:A:550:TYR:CD2	0.50	2.41	3	5
1:B:2:U:C2	2:A:525:CYS:SG	0.50	3.04	14	1
1:B:1:ADN:C6	2:A:594:THR:HG21	0.50	2.37	1	1
2:A:570:GLN:O	2:A:574:GLN:HG2	0.50	2.07	1	1
1:B:4:U:O2	2:A:602:LYS:HE2	0.49	2.07	16	1
2:A:541:GLY:HA2	2:A:572:MET:SD	0.49	2.48	3	5
1:B:4:U:O2	1:B:5:U:H5	0.49	1.90	20	1
2:A:542:LEU:HB2	2:A:615:ILE:CD1	0.49	2.37	14	7
2:A:569:ALA:O	2:A:573:VAL:HG23	0.48	2.08	7	15
2:A:597:LYS:N	2:A:597:LYS:HD2	0.48	2.23	1	3
1:B:3:C:C2	2:A:593:SER:HB3	0.48	2.43	15	2
2:A:527:LEU:HB3	2:A:532:CYS:SG	0.48	2.48	15	4
2:A:525:CYS:SG	2:A:589:LEU:HB3	0.48	2.49	9	5
1:B:5:U:C6	2:A:551:ILE:HG23	0.48	2.44	12	1
2:A:538:ILE:HD11	2:A:550:TYR:CD1	0.48	2.44	21	16
2:A:538:ILE:HD12	2:A:611:ILE:CD1	0.48	2.37	24	4
2:A:576:TYR:HA	2:A:579:LYS:O	0.48	2.08	14	6
2:A:576:TYR:CD2	2:A:590:ILE:HD12	0.47	2.44	4	4
1:B:1:ADN:H3'	1:B:2:U:O4'	0.47	2.09	17	1
2:A:534:GLU:N	2:A:552:LEU:HD11	0.47	2.24	8	10
2:A:601:LEU:HB3	2:A:603:LYS:O	0.47	2.09	8	1
1:B:5:U:C4'	2:A:553:MET:HA	0.47	2.38	2	3

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:A:520:ARG:NH2	2:A:563:MET:HB2	0.47	2.24	10	2
2:A:541:GLY:HA3	2:A:561:LEU:HD21	0.47	1.86	22	1
1:B:5:U:O2'	1:B:6:A:H5'	0.46	2.10	11	1
2:A:542:LEU:HA	2:A:545:GLY:O	0.46	2.10	16	13
2:A:540:LEU:HB3	2:A:576:TYR:OH	0.46	2.10	24	1
2:A:520:ARG:CZ	2:A:566:THR:HA	0.46	2.40	10	1
2:A:539:ASN:HA	2:A:615:ILE:HD11	0.46	1.87	14	1
1:B:1:ADN:O2'	1:B:2:U:H6	0.45	1.93	9	1
2:A:547:VAL:HB	2:A:608:VAL:HG11	0.45	1.89	20	15
2:A:540:LEU:HD11	2:A:588:LEU:HD12	0.45	1.88	25	1
2:A:525:CYS:O	2:A:526:ASN:HB2	0.45	2.12	25	14
1:B:1:ADN:H2'	1:B:2:U:C6	0.45	2.47	4	1
2:A:541:GLY:CA	2:A:561:LEU:HD21	0.44	2.43	8	1
1:B:3:C:H2'	2:A:595:ARG:HD3	0.44	1.89	16	1
2:A:523:HIS:HB2	2:A:593:SER:OG	0.44	2.13	15	1
2:A:563:MET:HE1	2:A:572:MET:HG3	0.43	1.91	13	1
2:A:557:ASN:N	2:A:557:ASN:HD22	0.42	2.11	15	1
2:A:583:ILE:HG13	2:A:588:LEU:HD11	0.42	1.91	25	1
2:A:542:LEU:HD13	2:A:612:ILE:CG1	0.42	2.44	17	4
2:A:550:TYR:N	2:A:550:TYR:CD2	0.42	2.87	15	1
1:B:4:U:H4'	1:B:5:U:C5'	0.42	2.45	17	1
1:B:5:U:O5'	1:B:6:A:H2	0.42	1.98	10	1
2:A:549:ASN:OD1	2:A:601:LEU:HD11	0.42	2.15	15	2
1:B:1:ADN:H2'	1:B:2:U:C2	0.42	2.49	4	1
2:A:538:ILE:HG21	2:A:608:VAL:HG13	0.42	1.91	17	2
2:A:577:GLN:HA	2:A:577:GLN:OE1	0.42	2.15	19	2
1:B:5:U:H2'	1:B:6:A:C2	0.42	2.49	21	1
1:B:5:U:O5'	1:B:6:A:C2	0.41	2.73	10	1
1:B:5:U:C4	2:A:602:LYS:HB2	0.41	2.51	8	1
2:A:544:PHE:HB2	2:A:572:MET:CE	0.41	2.44	13	1
2:A:550:TYR:CD2	2:A:550:TYR:N	0.41	2.88	20	4
1:B:4:U:O2	1:B:5:U:C5	0.41	2.73	20	1
2:A:554:LYS:CD	2:A:554:LYS:H	0.41	2.28	14	1
1:B:6:A:OP2	1:B:6:A:C8	0.41	2.73	4	1
2:A:534:GLU:CA	2:A:552:LEU:HD11	0.41	2.46	24	1
2:A:581:ALA:O	2:A:588:LEU:HG	0.41	2.15	6	1
2:A:553:MET:SD	2:A:560:PHE:CE2	0.41	3.14	8	1
2:A:584:ASN:HD22	2:A:584:ASN:N	0.41	2.12	10	1
1:B:1:ADN:O3'	1:B:2:U:H6	0.41	1.97	12	1
2:A:572:MET:O	2:A:576:TYR:HD2	0.41	1.99	16	1
1:B:1:ADN:O5'	2:A:591:ARG:HG3	0.41	2.16	1	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:B:3:C:O2'	2:A:595:ARG:HD2	0.41	2.16	7	1
2:A:527:LEU:O	2:A:557:ASN:HB3	0.40	2.16	10	1
2:A:549:ASN:HB2	2:A:562:GLU:CB	0.40	2.45	1	1
2:A:563:MET:CE	2:A:572:MET:HG3	0.40	2.47	8	1
1:B:5:U:N1	2:A:551:ILE:HG23	0.40	2.31	12	1
2:A:561:LEU:HD23	2:A:563:MET:HE2	0.40	1.93	20	1
2:A:613:GLN:HA	2:A:613:GLN:OE1	0.40	2.17	22	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	
2	A	95/112 (85%)	85±25 (89±26%)	2±1 (2±1%)	1±0 (1±1%)	32	76
All	All	2185/2800 (78%)	2121 (97%)	51 (2%)	13 (1%)	29	74

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
2	A	526	ASN	13

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	
2	A	83/93 (89%)	78±16 (94±19%)	1±1 (2±1%)	64	94
All	All	1992/2325 (86%)	1958 (98%)	34 (2%)	62	94

All 10 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)
2	A	562	GLU	9
2	A	597	LYS	8
2	A	525	CYS	4
2	A	520	ARG	3
2	A	557	ASN	3
2	A	554	LYS	2
2	A	584	ASN	2
2	A	600	GLN	1
2	A	617	SER	1
2	A	563	MET	1

6.3.3   RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
1	B	4/6 (67%)	3±1 (78±19%)	0±0 (2±7%)	0.10±0.14
All	All	102/150 (68%)	78 (76%)	2 (2%)	0.10

The overall RNA backbone suiteness is 0.08.

All unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	B	3	C	22
1	B	4	U	22
1	B	6	A	22
1	B	5	U	12

All unique RNA pucker outliers are listed below:

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

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 ⓘ

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	7-B	1
1	22-B	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
7	B	1:ADN	O2'	2:U	P	3.00
22	B	1:ADN	O2'	2:U	P	2.89

7 Chemical shift validation ⓘ

The completeness of assignment taking into account all chemical shift lists is 90% for the well-defined parts and 88% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *starch_output*

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	1478
Number of shifts mapped to atoms	1478
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	111	-0.43 ± 0.16	None needed (< 0.5 ppm)
$^{13}C_\beta$	104	0.25 ± 0.19	None needed (< 0.5 ppm)
$^{13}C'$	108	-0.07 ± 0.14	None needed (< 0.5 ppm)
^{15}N	107	0.60 ± 0.30	None needed (imprecise)

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 90%, i.e. 1274 atoms were assigned a chemical shift out of a possible 1422. 0 out of 15 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	470/473 (99%)	191/191 (100%)	187/190 (98%)	92/92 (100%)
Sidechain	692/779 (89%)	472/509 (93%)	206/241 (85%)	14/29 (48%)

Continued on next page...

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	64/81 (79%)	34/38 (89%)	30/39 (77%)	0/4 (0%)
Sugar	38/55 (69%)	30/30 (100%)	8/25 (32%)	0/0 (—%)
Base	10/34 (29%)	9/19 (47%)	1/10 (10%)	0/5 (0%)
Overall	1274/1422 (90%)	736/787 (94%)	432/505 (86%)	106/130 (82%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 88%, i.e. 1457 atoms were assigned a chemical shift out of a possible 1647. 0 out of 15 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	551/560 (98%)	225/228 (99%)	219/224 (98%)	107/108 (99%)
Sidechain	794/917 (87%)	542/596 (91%)	236/279 (85%)	16/42 (38%)
Aromatic	64/81 (79%)	34/38 (89%)	30/39 (77%)	0/4 (0%)
Sugar	38/55 (69%)	30/30 (100%)	8/25 (32%)	0/0 (—%)
Base	10/34 (29%)	9/19 (47%)	1/10 (10%)	0/5 (0%)
Overall	1457/1647 (88%)	840/911 (92%)	494/577 (86%)	123/159 (77%)

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	B	4	U	H5	3.08	3.92 − 6.99	-7.8
1	A	543	PRO	HB2	0.10	0.37 − 3.78	-5.8
1	A	572	MET	HA	2.06	2.11 − 6.67	-5.1

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: