Full wwPDB NMR Structure Validation Report (i)

Feb 19, 2022 – 05:30 PM EST

PDB ID	:	1RGO
Title	:	Structural Basis for Recognition of the mRNA Class II AU-Rich Element by the Tandem Zinc Finger Domain of TIS11d
Authors	:	Hudson, B.P.; Martinez-Yamout, M.A.; Dyson, H.J.; Wright, P.E.
Deposited on	:	2003-11-12

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 following versions of software and data (see references (i)) were used in the production of this report:

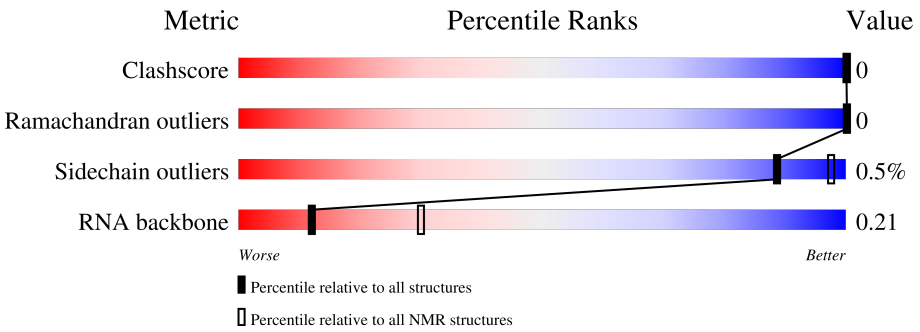
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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
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	D	9	33% 44% 22%
2	A	70	63% • 33%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 3 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:155-A:179 (25)	0.08	3
2	A:195-A:216 (22)	0.07	10

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 5 clusters and 2 single-model clusters were found.

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

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 1402 atoms, of which 637 are hydrogens and 0 are deuteriums.

- Molecule 1 is a RNA chain called RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3').

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	D	9	274	83	93	24	66	8	0

- Molecule 2 is a protein called Butyrate response factor 2.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	A	70	1126	367	544	107	102	6	0

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
3	A	2	2	2

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

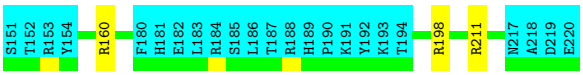
These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for 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 (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D: 33% 44% 22%

• Molecule 2: Butyrate response factor 2

Chain A: 63% • 33%



4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

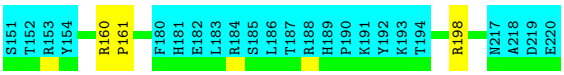
4.2.1 Score per residue for model 1

• Molecule 1: RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D: 22% 67% 11%

• Molecule 2: Butyrate response factor 2

Chain A: 63% • 33%



4.2.2 Score per residue for model 2

- Molecule 1: RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D: 44% 33% 22%

U1 U2 A3 U4 U5 U6 A7 U8 U9

- Molecule 2: Butyrate response factor 2

Chain A: 63% 33%

S151 T152 R153 Y154 R160 F180 H181 E182 L183 R184 S185 L186 T187 R188 H189 P190 K191 Y192 K193 T194 R198 R211 N217 A218 D219 E220

4.2.3 Score per residue for model 3 (medoid)

- Molecule 1: RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D: 56% 44%

U1 U2 A3 U6 A7 U8 U9

- Molecule 2: Butyrate response factor 2

Chain A: 64% 33%

S151 T152 R153 Y154 R160 F180 H181 E182 L183 R184 S185 L186 T187 R188 H189 P190 K191 Y192 K193 T194 R198 N217 A218 D219 E220

4.2.4 Score per residue for model 4

- Molecule 1: RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D: 33% 44% 22%

U1 U2 A3 U4 U5 U6 A7 U8 U9

- Molecule 2: Butyrate response factor 2

Chain A: 63% 33%

S151 T152 R153 Y154 R160 F180 H181 E182 L183 R184 S185 L186 T187 R188 H189 P190 K191 Y192 K193 T194 R198 R211 N217 A218 D219 E220

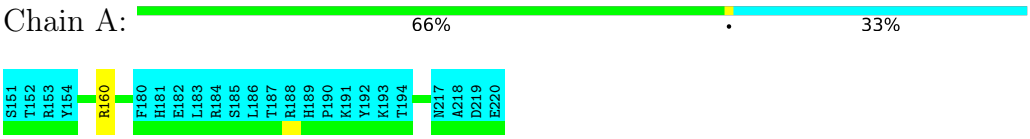
4.2.5 Score per residue for model 5

- Molecule 1: RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D:

- Molecule 2: Butyrate response factor 2

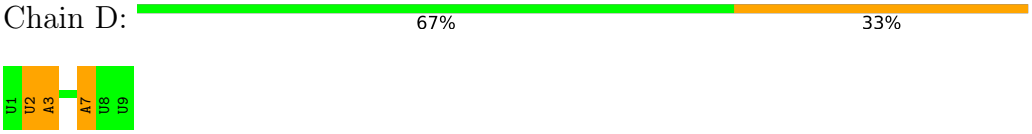
Chain A:



4.2.6 Score per residue for model 6

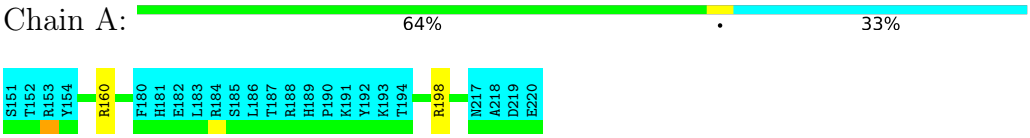
- Molecule 1: RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D:



- Molecule 2: Butyrate response factor 2

Chain A:



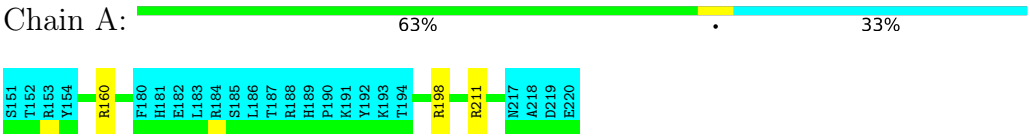
4.2.7 Score per residue for model 7

- Molecule 1: RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D:

- Molecule 2: Butyrate response factor 2

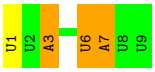
Chain A:



4.2.8 Score per residue for model 8

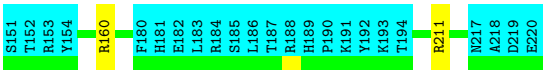
- Molecule 1: RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D: 56% 11% 33%



- Molecule 2: Butyrate response factor 2

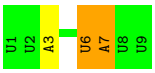
Chain A: 64% • 33%



4.2.9 Score per residue for model 9

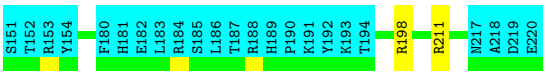
- Molecule 1: RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D: 67% 11% 22%



- Molecule 2: Butyrate response factor 2

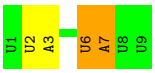
Chain A: 64% • 33%



4.2.10 Score per residue for model 10

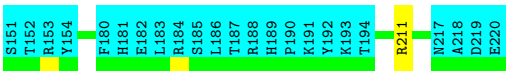
- Molecule 1: RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D: 56% 22% 22%



- Molecule 2: Butyrate response factor 2

Chain A: 66% • 33%



4.2.11 Score per residue for model 11

- Molecule 1: RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D: 33% 33% 11% 22%

U1 U2 A3 U4 U5 U6 A7 U8 U9

- Molecule 2: Butyrate response factor 2

Chain A: 63% • 33%

S151 T152 R153 Y154 R160 F180 H181 E182 L183 R184 S185 L186 T187 R188 H189 P190 K191 Y192 K193 T194 R198 R211 N217 A218 D219 E220

4.2.12 Score per residue for model 12

- Molecule 1: RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D: 44% 22% 22% 11%

U1 U2 A3 U4 U5 U6 A7 U8 U9

- Molecule 2: Butyrate response factor 2

Chain A: 63% • • 33%

S151 T152 R153 Y154 R160 F180 H181 E182 L183 R184 S185 L186 T187 R188 H189 P190 K191 Y192 K193 T194 R198 R211 N217 A218 D219 E220

4.2.13 Score per residue for model 13

- Molecule 1: RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D: 44% 33% 22%

U1 U2 A3 U6 A7 U8 U9

- Molecule 2: Butyrate response factor 2

Chain A: 63% • 33%

S151 T152 R153 Y154 R160 Y170 F180 H181 E182 L183 R184 S185 L186 T187 R188 H189 P190 K191 Y192 K193 T194 R198 N217 A218 D219 E220

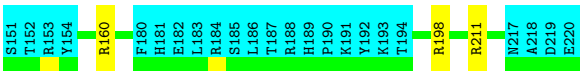
4.2.14 Score per residue for model 14

- Molecule 1: RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D: 22% 22% 56%

- Molecule 2: Butyrate response factor 2

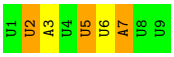
Chain A: 63% • 33%



4.2.15 Score per residue for model 15

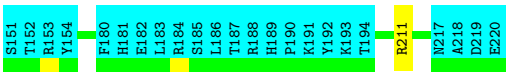
- Molecule 1: RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D: 44% 22% 33%



- Molecule 2: Butyrate response factor 2

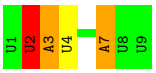
Chain A: 66% • 33%



4.2.16 Score per residue for model 16

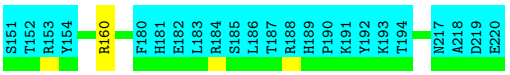
- Molecule 1: RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D: 56% 11% 22% 11%



- Molecule 2: Butyrate response factor 2

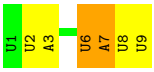
Chain A: 66% • 33%



4.2.17 Score per residue for model 17

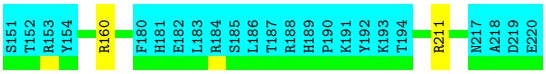
- Molecule 1: RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D: 33% 44% 22%



- Molecule 2: Butyrate response factor 2

Chain A: 64% 33%



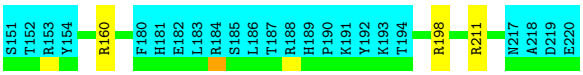
4.2.18 Score per residue for model 18

- Molecule 1: RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D: 11% 44% 44%

- Molecule 2: Butyrate response factor 2

Chain A: 63% 33%



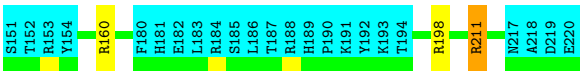
4.2.19 Score per residue for model 19

- Molecule 1: RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D: 33% 11% 56%

- Molecule 2: Butyrate response factor 2

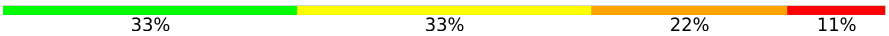
Chain A: 63% 33%



4.2.20 Score per residue for model 20

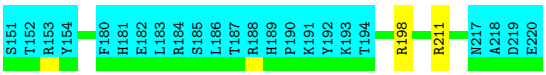
- Molecule 1: RNA (5'-R(*UP*UP*AP*UP*UP*UP*AP*UP*U)-3')

Chain D:



- Molecule 2: Butyrate response factor 2

Chain A:



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *Individual zinc finger domains were created in DYANA and docked to UAUU (finger I) and UUAUU (finger II) using five rounds of simulated annealing in AMBER using distance and torsional restraints. Finger I/UAUU and Finger II/UUAUU complexes were connected to form TZF/UUAUUUAUU complexes and refined in AMBER with residual dipolar coupling restraints added. Calculations were then switched from in vacuo to Generalized Born continuum solvent model and one additional round of simulated annealing was performed..*

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	1.5
Amber	structure solution	8
Amber	refinement	8

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

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

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	D	1.33±0.01	0±0/200 (0.0± 0.0%)	2.11±0.04	8±1/308 (2.5± 0.3%)
2	A	0.85±0.01	0±0/395 (0.0± 0.0%)	1.09±0.01	2±1/531 (0.4± 0.1%)
All	All	1.04	0/11900 (0.0%)	1.55	199/16780 (1.2%)

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
1	D	0.0±0.0	2.0±1.0
2	A	0.0±0.0	0.1±0.2
All	All	0	42

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
1	D	3	A	N1-C6-N6	-9.65	112.81	118.60	15	20
1	D	2	U	O4'-C1'-N1	8.56	115.05	108.20	15	8
1	D	7	A	N1-C6-N6	-8.35	113.59	118.60	14	20
1	D	7	A	C5-C6-N1	8.17	121.78	117.70	13	20
2	A	198	ARG	NE-CZ-NH1	8.09	124.34	120.30	9	14
2	A	211	ARG	NE-CZ-NH1	7.47	124.04	120.30	12	13
1	D	3	A	C5-C6-N1	7.42	121.41	117.70	9	20
1	D	6	U	O4'-C1'-N1	7.17	113.93	108.20	4	4
1	D	2	U	N3-C2-O2	-6.61	117.57	122.20	15	1
1	D	3	A	C4-C5-C6	-6.61	113.70	117.00	4	20

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	D	2	U	C1'-O4'-C4'	-6.46	104.73	109.90	6	1
1	D	4	U	O4'-C1'-N1	6.43	113.35	108.20	4	4
1	D	7	A	C4-C5-C6	-6.43	113.78	117.00	13	20
1	D	8	U	O4'-C1'-N1	6.16	113.13	108.20	18	7
2	A	160	ARG	NE-CZ-NH1	6.02	123.31	120.30	13	16
1	D	6	U	C1'-O4'-C4'	-5.61	105.41	109.90	4	2
1	D	1	U	O4'-C1'-N1	5.45	112.56	108.20	12	4
1	D	5	U	O4'-C1'-N1	5.30	112.44	108.20	15	2
1	D	9	U	O4'-C1'-N1	5.29	112.43	108.20	17	1
1	D	1	U	N3-C2-O2	-5.08	118.65	122.20	1	1
2	A	161	PRO	N-CA-CB	5.03	109.33	103.30	1	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)
1	D	7	A	Sidechain	17
1	D	6	U	Sidechain	10
1	D	3	A	Sidechain	5
1	D	2	U	Sidechain	5
1	D	5	U	Sidechain	3
1	D	4	U	Sidechain	1
2	A	170	TYR	Sidechain	1

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
All	All	11280	8840	8860	-

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.

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	
2	A	47/70 (67%)	45±0 (96±1%)	2±0 (4±1%)	0±0 (0±0%)	100	100
All	All	940/1400 (67%)	907 (96%)	33 (4%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	
2	A	41/63 (65%)	41±0 (100±1%)	0±0 (0±1%)	89	97
All	All	820/1260 (65%)	816 (100%)	4 (0%)	89	97

All 1 unique residues with a non-rotameric sidechain are listed below.

Mol	Chain	Res	Type	Models (Total)
2	A	211	ARG	4

6.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
1	D	8/9 (89%)	2±1 (31±17%)	0±1 (5±7%)	0.21±0.07
All	All	160/180 (89%)	50 (31%)	8 (5%)	0.21

The overall RNA backbone suiteness is 0.21.

All unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	D	2	U	13
1	D	6	U	12
1	D	3	A	8
1	D	5	U	7
1	D	4	U	6
1	D	9	U	2
1	D	7	A	1
1	D	8	U	1

All unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	D	3	A	3
1	D	2	U	3
1	D	8	U	1
1	D	5	U	1

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)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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