wwPDB NMR Structure Validation Summary Report

May 29, 2020 – 01:40 am BST

PDB ID	:	2XEB
Title	:	NMR STRUCTURE OF THE PROTEIN-UNBOUND SPLICEOSOMAL U4 SNRNA 5' STEM LOOP
Authors	:	Falb, M.; Amata, I.; Gabel, F.; Simon, B.; Carlomagno, T.
Deposited on	:	2010-05-12

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 (1)) were used in the production of this report:

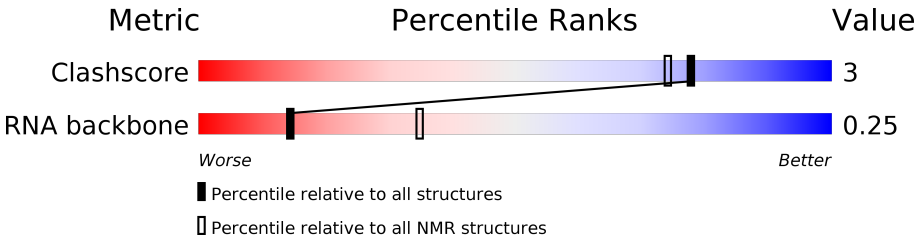
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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
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	19	21% 68% 11%
2	B	14	79% 21%

2 Ensemble composition and analysis (i)

This entry contains 10 models. This entry does not contain polypeptide chains, therefore identification of well-defined residues and clustering analysis are not possible. All residues are included in the validation scores.

3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1073 atoms, of which 361 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	A	19	615	182	206	74	134	19	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	19	G	U	engineered mutation	GB 36174

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	14	458	134	155	57	98	14	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	40	G	U	engineered mutation	GB 36174
B	53	C	U	engineered mutation	GB 36174

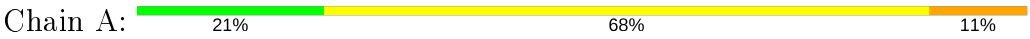
4 Residue-property plots (i)

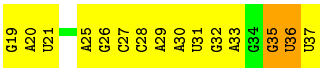
4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

• Molecule 1: 5'-R(P*GP*AP*UP*CP*GP*UP*AP*GP*CP*CP*AP*AP*UP*GP*AP*GP*GP*UP*U)-3'

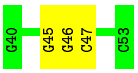
Chain A:





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

Chain B:

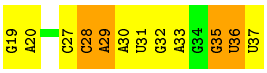


4.2 Residue scores for the representative (author defined) model from the NMR ensemble

The representative model is number 1. Colouring as in section 4.1 above.

• Molecule 1: 5'-R(P*GP*AP*UP*CP*GP*UP*AP*GP*CP*CP*AP*AP*UP*GP*AP*GP*GP*UP*U)-3'

Chain A:



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

Chain B:

5 Refinement protocol and experimental data overview

The models were refined using the following method: *RDC REFINEMENT, WATER REFINEMENT*.

Of the 100 calculated structures, 10 were deposited, based on the following criterion: *LOWEST ENERGY STRUCTURES*.

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

Software name	Classification	Version
ARIA-CNS	refinement	
ARIA CNS 1.2	structure solution	1.1

No chemical shift data was provided. No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

5.1 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	409	206	205	2±1
2	B	303	155	154	2±1
All	All	7120	3610	3590	31

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 3.

5 of 9 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:30:A:C2	2:B:44:A:C2	0.51	2.99	8	3
1:A:27:C:C4	1:A:28:C:N4	0.50	2.79	1	9
2:B:46:G:H2'	2:B:47:C:C6	0.49	2.43	5	9
1:A:35:G:H2'	1:A:36:U:C5	0.46	2.45	1	4
1:A:35:G:O2'	1:A:36:U:C6	0.45	2.66	6	1

5.2 Torsion angles

5.2.1 Protein backbone

There are no protein molecules in this entry.

5.2.2 Protein sidechains

There are no protein molecules in this entry.

5.2.3 RNA

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
1	A	19/19 (100%)	12±1 (63±6%)	1±0 (5±2%)	0.16±0.05
2	B	13/14 (93%)	3±1 (25±9%)	0±0 (0±0%)	0.33±0.06
All	All	312/330 (95%)	153 (49%)	10 (3%)	0.25

The overall RNA backbone suiteness is 0.25.

5 of 25 unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	A	32	G	10
1	A	37	U	10
1	A	31	U	10
1	A	35	G	10
1	A	20	A	10

All unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	A	31	U	8
1	A	19	G	2

5.3 Non-standard residues in protein, DNA, RNA chains

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

5.4 Carbohydrates

There are no carbohydrates in this entry.

5.5 Ligand geometry

There are no ligands in this entry.

5.6 Other polymers

There are no such molecules in this entry.

5.7 Polymer linkage issues

There are no chain breaks in this entry.

6 Chemical shift validation

No chemical shift data were provided