wwPDB NMR Structure Validation Summary Report (i)

May 7, 2024 – 02:33 pm BST

PDB ID	:	1DZ5
Title	:	The NMR structure of the 38KDa U1A protein-PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein
Authors	:	Varani, L.; Gunderson, S.I.; Mattaj, I.W.; Kay, L.E.; Neuhaus, D.; Varani, G.
Deposited on	:	2000-02-16

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 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:

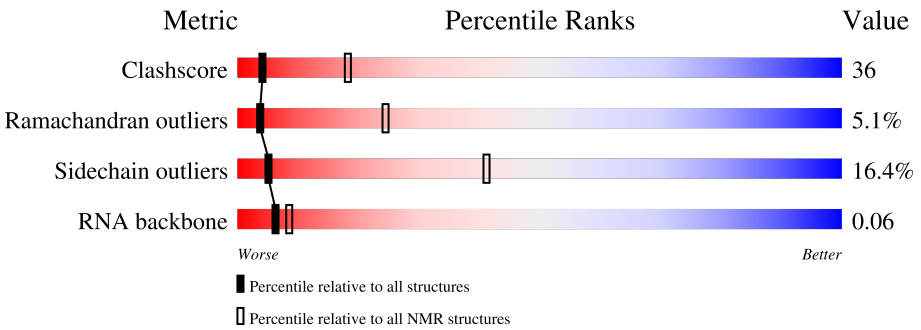
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 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	A	101	42% 49% 5% 5%
1	B	101	40% 50% 5% 5%
2	C	22	14% 59% 27%
2	D	22	14% 59% 27%

2 Ensemble composition and analysis (i)

This entry contains 13 models. The atoms present in the NMR models are not consistent. Some calculations may have failed as a result. All residues are included in the validation scores. Model 1 is the overall representative, medoid model (most similar to other models).

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:6-A:101, B:6-B:101 (192)	1.36	1

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

NmrClust was unable to cluster the ensemble.

Error message: Inconsistent models in file

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 4742 atoms, of which 2182 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called U1 SMALL NUCLEAR RIBONUCLEOPROTEIN A.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	101	1666	524	851	143	144	4	0
1	B	101	1666	524	851	143	144	4	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	30	HIS	TYR	engineered mutation	UNP P09012
A	35	ARG	GLN	engineered mutation	UNP P09012
B	30	HIS	TYR	engineered mutation	UNP P09012
B	35	ARG	GLN	engineered mutation	UNP P09012

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	C	22	705	209	240	84	151	21	0
2	D	22	705	209	240	84	151	21	0

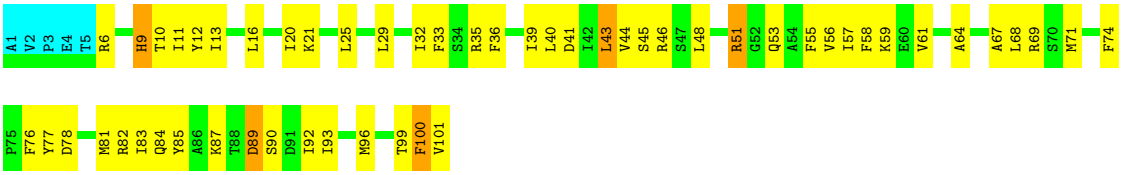
4 Residue-property plots (i)

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.

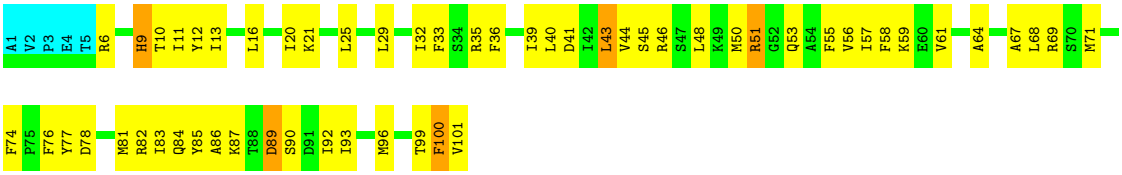
• Molecule 1: U1 SMALL NUCLEAR RIBONUCLEOPROTEIN A

Chain A: 42% 49% 5% 5%



• Molecule 1: U1 SMALL NUCLEAR RIBONUCLEOPROTEIN A

Chain B: 40% 50% 5% 5%



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

Chain C: 14% 59% 27%

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

Chain D: 14% 59% 27%

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

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

• Molecule 1: U1 SMALL NUCLEAR RIBONUCLEOPROTEIN A

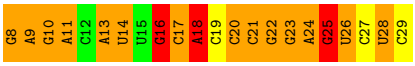
Chain A: 44% 43% 8% • 5%

• Molecule 1: U1 SMALL NUCLEAR RIBONUCLEOPROTEIN A

Chain B: 44% 43% 8% • 5%

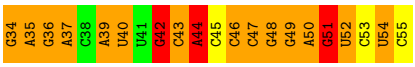
• Molecule 2: PIE, RNA (5'-R(*GP*AP*GP*AP*CP*AP*UP*UP*GP*CP*AP*CP*CP* CP*GP*GP*AP*GP*UP*CP*UP*C)-3')

Chain C: 9% 14% 64% 14%



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

Chain D: 9% 14% 64% 14%



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *RESTRAINED MOLECULAR DYNAMICS*.

Of the 50 calculated structures, 13 were deposited, based on the following criterion: *AGREEMENT WITH EXPERIMENTAL DATA*.

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

Software name	Classification	Version
X-PLOR	refinement	3.8.1
X-PLOR	structure solution	3.8.1

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

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	0.97±0.00	0±0/794 (0.0± 0.0%)	0.79±0.00	0±0/1061 (0.0± 0.0%)
1	B	0.97±0.00	0±0/794 (0.0± 0.0%)	0.79±0.00	0±0/1061 (0.0± 0.0%)
2	C	1.13±0.01	0±0/519 (0.0± 0.0%)	1.94±0.01	27±1/807 (3.4± 0.1%)
2	D	1.13±0.00	0±0/519 (0.0± 0.0%)	1.94±0.01	27±1/807 (3.3± 0.2%)
All	All	1.04	0/34138 (0.0%)	1.41	701/48568 (1.4%)

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
1	A	0.0±0.0	6.0±0.0
1	B	0.0±0.0	6.0±0.0
All	All	0	156

There are no bond-length outliers.

5 of 58 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	C	10	G	N7-C8-N9	9.28	117.74	113.10	7	13
2	C	16	G	N7-C8-N9	9.28	117.74	113.10	11	13
2	D	42	G	N7-C8-N9	9.27	117.73	113.10	5	13
2	C	25	G	N7-C8-N9	9.26	117.73	113.10	10	13
2	D	48	G	N7-C8-N9	9.23	117.72	113.10	13	13

There are no chirality outliers.

5 of 12 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	A	6	ARG	Sidechain	13
1	A	35	ARG	Sidechain	13
1	A	46	ARG	Sidechain	13
1	A	51	ARG	Sidechain	13
1	A	69	ARG	Sidechain	13

6.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in each chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	780	815	815	59±11
1	B	780	815	815	59±12
2	C	465	240	240	42±7
2	D	465	240	240	41±7
All	All	32370	27430	27430	2175

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 36.

5 of 1017 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:99:THR:HG21	1:B:99:THR:HG21	1.04	1.20	9	2
1:A:101:VAL:HG11	1:B:101:VAL:HG11	1.03	1.30	8	1
1:A:101:VAL:HG13	1:B:101:VAL:HG13	0.98	1.29	2	5
1:A:11:ILE:HD12	1:A:83:ILE:HG21	0.97	1.37	8	6
1:B:11:ILE:HD12	1:B:83:ILE:HG21	0.93	1.39	4	6

6.3 Torsion angles (i)

6.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the backbone conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	95/101 (94%)	74±3 (78±3%)	16±4 (17±4%)	5±1 (5±2%)	4 24
1	B	95/101 (94%)	74±3 (78±3%)	16±4 (17±4%)	5±1 (5±2%)	4 24
All	All	2470/2626 (94%)	1918 (78%)	426 (17%)	126 (5%)	4 24

5 of 40 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	89	ASP	12
1	B	89	ASP	12
1	A	100	PHE	10
1	B	100	PHE	10
1	A	59	LYS	7

6.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the sidechain conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	86/90 (96%)	72±3 (84±4%)	14±3 (16±4%)	5 41
1	B	86/90 (96%)	72±3 (84±4%)	14±3 (16±4%)	5 41
All	All	2236/2340 (96%)	1870 (84%)	366 (16%)	5 41

5 of 104 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	43	LEU	13
1	A	74	PHE	13
1	B	43	LEU	13
1	B	74	PHE	13
1	A	36	PHE	10

6.3.3 RNA (i)

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
2	C	21/22 (95%)	14±1 (67±6%)	4±1 (19±5%)	0.06±0.02
2	D	21/22 (95%)	14±1 (67±6%)	4±1 (19±5%)	0.06±0.02
All	All	566/572 (99%)	368 (65%)	104 (18%)	0.06

The overall RNA backbone suiteness is 0.06.

5 of 38 unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	C	11	A	13
2	C	16	G	13
2	C	17	C	13
2	C	18	A	13
2	D	37	A	13

5 of 16 unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	C	17	C	13
2	D	43	C	13
2	C	16	G	11
2	D	42	G	11
2	C	8	G	10

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

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

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

There are no ligands in this entry.

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation (i)

No chemical shift data were provided