wwPDB NMR Structure Validation Summary Report

Apr 21, 2024 – 05:18 AM EDT

PDB ID	:	2LI8
BMRB ID	:	17883
Title	:	The solution structure of the Lin28-ZnF domains bound to AGGAGAU of pre-let-7 miRNA
Authors	:	Allain, F.H.-T.; Loughlin, F.E.
Deposited on	:	2011-08-25

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:

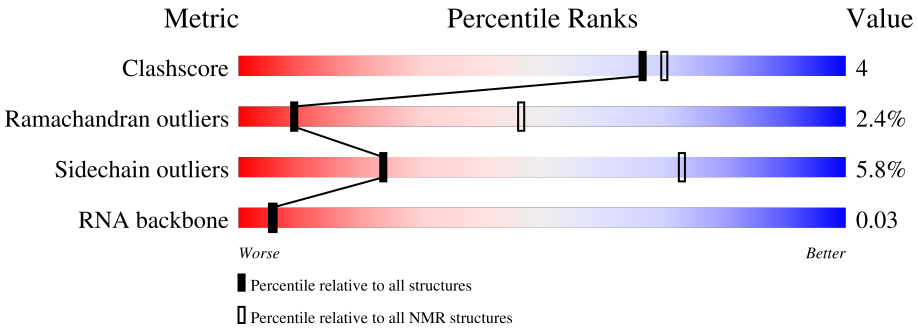
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson 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 is 72%.

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	74	38% 11% 36% 15%
2	B	7	86% 14%

2 Ensemble composition and analysis

This entry contains 20 models. Model 9 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

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:139-A:152 (14)	0.11	15
2	A:160-A:181 (22)	0.16	9

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 6 single-model clusters were found.

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

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 1185 atoms, of which 553 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Protein lin-28 homolog A.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	63	953	290	475	96	84	8	0

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	113	SER	-	expression tag	UNP Q9H9Z2
A	114	SER	-	expression tag	UNP Q9H9Z2
A	115	GLY	-	expression tag	UNP Q9H9Z2
A	116	LEU	-	expression tag	UNP Q9H9Z2
A	117	VAL	-	expression tag	UNP Q9H9Z2
A	118	PRO	-	expression tag	UNP Q9H9Z2
A	119	ARG	-	expression tag	UNP Q9H9Z2
A	120	GLY	-	expression tag	UNP Q9H9Z2
A	121	SER	-	expression tag	UNP Q9H9Z2
A	122	HIS	-	expression tag	UNP Q9H9Z2
A	123	MET	-	expression tag	UNP Q9H9Z2

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	7	230	69	78	32	45	6	0

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

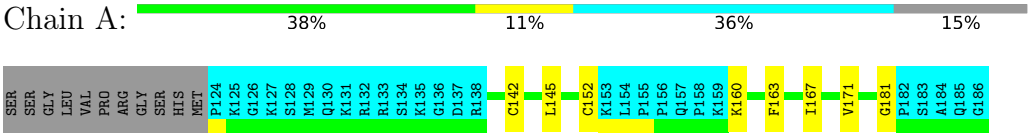
Mol	Chain	Residues	Atoms	
			Total	Zn
3	A	2	2	2

4 Residue-property plots (i)

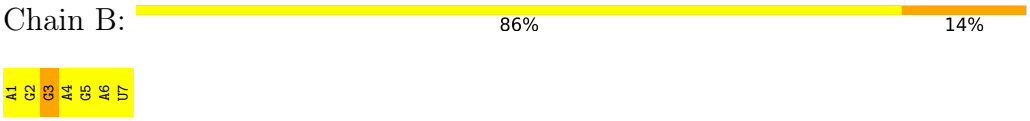
4.1 Average score per residue in the NMR ensemble

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: Protein lin-28 homolog A



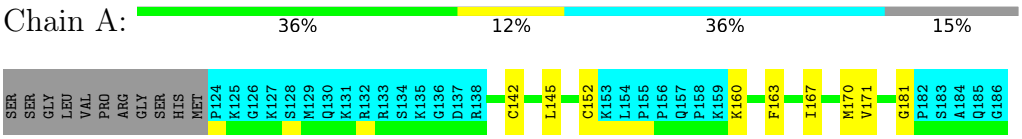
• Molecule 2: RNA (5'-R(*AP*GP*GP*AP*GP*AP*U)-3')



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

The representative model is number 9. Colouring as in section 4.1 above.

• Molecule 1: Protein lin-28 homolog A



• Molecule 2: RNA (5'-R(*AP*GP*GP*AP*GP*AP*U)-3')

5 Refinement protocol and experimental data overview

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

Of the 200 calculated structures, 20 were deposited, based on the following criterion: *lowest energy/fewest violation*.

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

Software name	Classification	Version
CYANA	structure solution	3.0
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	706
Number of shifts mapped to atoms	706
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	72%

6 Model quality

6.1 Standard geometry

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	A	0.48±0.01	0±0/276 (0.0± 0.0%)	0.80±0.02	0±0/371 (0.0± 0.1%)
2	B	1.05±0.02	0±0/171 (0.0± 0.0%)	1.62±0.08	3±1/266 (1.2± 0.5%)
All	All	0.75	0/8940 (0.0%)	1.21	65/12740 (0.5%)

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	0.1±0.2
All	All	0	1

There are no bond-length outliers.

5 of 13 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	1	A	O4'-C1'-N9	14.87	120.10	108.20	14	20
2	B	3	G	O4'-C1'-N9	6.57	113.46	108.20	18	13
2	B	4	A	O4'-C1'-N9	6.56	113.45	108.20	14	4
2	B	1	A	C3'-C2'-C1'	-6.29	96.47	101.50	14	1
2	B	7	U	O4'-C1'-N1	6.14	113.11	108.20	9	6

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
2	B	5	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	269	246	246	3±1
2	B	152	78	78	1±1
All	All	8460	6480	6480	60

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 4.

5 of 10 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:163:PHE:HB2	1:A:171:VAL:HG23	0.58	1.75	4	20
2:B:6:A:H1'	2:B:7:U:C6	0.56	2.36	5	3
1:A:142:CYS:HB3	1:A:152:CYS:SG	0.46	2.49	20	13
2:B:1:A:O2'	2:B:2:G:C8	0.46	2.68	17	2
1:A:160:LYS:HG2	1:A:167:ILE:HG23	0.45	1.87	13	11

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	36/74 (49%)	29±1 (79±3%)	7±1 (18±3%)	1±0 (2±1%)	9	46
All	All	720/1480 (49%)	572 (79%)	131 (18%)	17 (2%)	9	46

All 2 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	181	GLY	16
1	A	160	LYS	1

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	30/62 (48%)	28±1 (94±2%)	2±1 (6±2%)	24	73
All	All	600/1240 (48%)	565 (94%)	35 (6%)	24	73

All 5 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	145	LEU	20
1	A	170	MET	6
1	A	173	SER	5
1	A	180	GLN	3
1	A	150	LYS	1

6.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
2	B	6/7 (86%)	5±1 (84±11%)	1±1 (15±12%)	0.03±0.04
All	All	124/140 (89%)	101 (81%)	18 (15%)	0.03

The overall RNA backbone suiteness is 0.03.

5 of 6 unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	B	3	G	20
2	B	4	A	19
2	B	7	U	19
2	B	2	G	18
2	B	6	A	16

All unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	B	2	G	12
2	B	1	A	4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
2	B	5	G	1
2	B	4	A	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

The completeness of assignment taking into account all chemical shift lists is 72% for the well-defined parts and 74% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

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	706
Number of shifts mapped to atoms	706
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	74

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	61	0.00 ± 0.00	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	55	0.35 ± 0.10	None needed (< 0.5 ppm)
$^{13}C'$	6	—	None (insufficient data)
^{15}N	50	-0.07 ± 0.47	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 72%, i.e. 415 atoms were assigned a chemical shift out of a possible 574. 0 out of 3 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	133/181 (73%)	62/74 (84%)	42/72 (58%)	29/35 (83%)
Sidechain	203/212 (96%)	138/138 (100%)	61/67 (91%)	4/7 (57%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	32/47 (68%)	18/25 (72%)	12/18 (67%)	2/4 (50%)
Sugar	37/77 (48%)	37/42 (88%)	0/35 (0%)	0/0 (—%)
Base	10/57 (18%)	10/36 (28%)	0/11 (0%)	0/10 (0%)
Overall	415/574 (72%)	265/315 (84%)	115/203 (57%)	35/56 (62%)

7.1.4 Statistically unusual chemical shifts (i)

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	142	CYS	C	8.90	164.77 – 184.97	-82.2
1	A	182	PRO	CD	11.92	45.11 – 55.58	-36.7
1	A	155	PRO	CD	12.21	45.11 – 55.58	-36.4
1	A	156	PRO	CD	12.30	45.11 – 55.58	-36.3
1	A	158	PRO	CD	12.39	45.11 – 55.58	-36.2
1	A	175	PRO	CD	12.51	45.11 – 55.58	-36.1
1	A	124	PRO	CD	12.75	45.11 – 55.58	-35.9
1	A	164	CYS	C	118.61	164.77 – 184.97	-27.9
1	A	155	PRO	CA	23.59	55.85 – 70.84	-26.5
1	A	158	PRO	CA	24.74	55.85 – 70.84	-25.8
1	A	182	PRO	CA	25.55	55.85 – 70.84	-25.2
1	A	183	SER	CB	25.92	56.28 – 71.32	-25.2
1	A	128	SER	CB	25.93	56.28 – 71.32	-25.2
1	A	173	SER	CB	25.94	56.28 – 71.32	-25.2
1	A	134	SER	CB	25.96	56.28 – 71.32	-25.2
1	A	156	PRO	CA	25.87	55.85 – 70.84	-25.0
1	A	175	PRO	CA	26.04	55.85 – 70.84	-24.9
1	A	168	SER	CB	26.56	56.28 – 71.32	-24.8
1	A	152	CYS	C	124.89	164.77 – 184.97	-24.7
1	A	166	SER	CB	27.05	56.28 – 71.32	-24.4
1	A	174	CYS	C	125.85	164.77 – 184.97	-24.3
1	A	139	CYS	C	127.46	164.77 – 184.97	-23.5
1	A	161	CYS	C	128.76	164.77 – 184.97	-22.8
1	A	157	GLN	CA	14.24	46.17 – 66.97	-20.4
1	A	184	ALA	CA	14.75	43.52 – 62.81	-19.9
1	A	173	SER	CA	18.22	48.46 – 68.96	-19.8
1	A	178	ALA	CA	15.33	43.52 – 62.81	-19.6
1	A	151	GLU	CA	17.29	47.03 – 67.62	-19.4
1	A	137	ASP	CA	16.07	44.71 – 64.67	-19.4
1	A	154	LEU	CA	15.58	45.17 – 66.21	-19.1

Continued on next page...

Continued from previous page...

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	141	ASN	CA	18.43	44.28 – 62.79	-19.0
1	A	160	LYS	CA	16.20	46.18 – 67.77	-18.9
1	A	172	ALA	CA	16.76	43.52 – 62.81	-18.9
1	A	146	ASP	CA	17.03	44.71 – 64.67	-18.9
1	A	148	HIS	CA	13.37	45.04 – 67.94	-18.8
1	A	183	SER	CA	20.43	48.46 – 68.96	-18.7
1	A	128	SER	CA	20.48	48.46 – 68.96	-18.6
1	A	134	SER	CA	20.48	48.46 – 68.96	-18.6
1	A	185	GLN	CA	17.99	46.17 – 66.97	-18.6
1	A	130	GLN	CA	18.06	46.17 – 66.97	-18.5
1	A	170	MET	CA	15.84	45.26 – 67.07	-18.5
1	A	180	GLN	CA	18.11	46.17 – 66.97	-18.5
1	A	166	SER	CA	20.88	48.46 – 68.96	-18.5
1	A	179	GLN	CA	18.27	46.17 – 66.97	-18.4
1	A	168	SER	CA	21.23	48.46 – 68.96	-18.3
1	A	145	LEU	CA	17.54	45.17 – 66.21	-18.1
1	A	131	LYS	CA	18.27	46.18 – 67.77	-17.9
1	A	127	LYS	CA	18.29	46.18 – 67.77	-17.9
1	A	153	LYS	CA	18.32	46.18 – 67.77	-17.9
1	A	149	ALA	CA	18.67	43.52 – 62.81	-17.9
1	A	138	ARG	CA	16.55	45.44 – 68.13	-17.7
1	A	129	MET	CA	17.53	45.26 – 67.07	-17.7
1	A	176	LEU	CA	18.58	45.17 – 66.21	-17.6
1	A	135	LYS	CA	19.02	46.18 – 67.77	-17.6
1	A	165	GLN	CA	20.21	46.17 – 66.97	-17.5
1	A	177	LYS	CA	19.46	46.18 – 67.77	-17.4
1	A	159	LYS	CA	19.79	46.18 – 67.77	-17.2
1	A	133	ARG	CA	18.14	45.44 – 68.13	-17.0
1	A	132	ARG	CA	18.22	45.44 – 68.13	-17.0
1	A	169	HIS	CA	17.94	45.04 – 67.94	-16.8
1	A	150	LYS	CA	20.70	46.18 – 67.77	-16.8
1	A	147	HIS	CA	18.28	45.04 – 67.94	-16.7
1	A	138	ARG	NE	111.13	76.53 – 92.65	16.5
1	A	162	HIS	CA	20.55	45.04 – 67.94	-15.7
1	A	140	TYR	CA	19.26	45.75 – 70.63	-15.7
1	A	167	ILE	CA	23.64	48.30 – 75.08	-14.2
1	A	163	PHE	CA	22.95	45.38 – 70.89	-13.8
1	A	171	VAL	CA	28.09	48.38 – 76.73	-12.2
1	A	174	CYS	CA	19.87	40.80 – 75.33	-11.1
1	A	142	CYS	CA	20.88	40.80 – 75.33	-10.8
1	A	164	CYS	CA	21.02	40.80 – 75.33	-10.7
1	A	139	CYS	CA	21.71	40.80 – 75.33	-10.5

Continued on next page...

Continued from previous page...

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	161	CYS	CA	22.26	40.80 – 75.33	-10.4
1	A	152	CYS	CA	24.86	40.80 – 75.33	-9.6

7.1.5 Random Coil Index (RCI) plots (i)

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: