wwPDB NMR Structure Validation Summary Report (i)

May 28, 2020 – 11:06 pm BST

PDB ID	:	2LEB
Title	:	Solution structure of human SRSF2 (SC35) RRM in complex with 5'-UCCAGU-3'
Authors	:	Daubner, G.M.; Clery, A.; Jayne, S.; Stevenin, J.; Allain, F.H.-T.
Deposited on	:	2011-06-15

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:

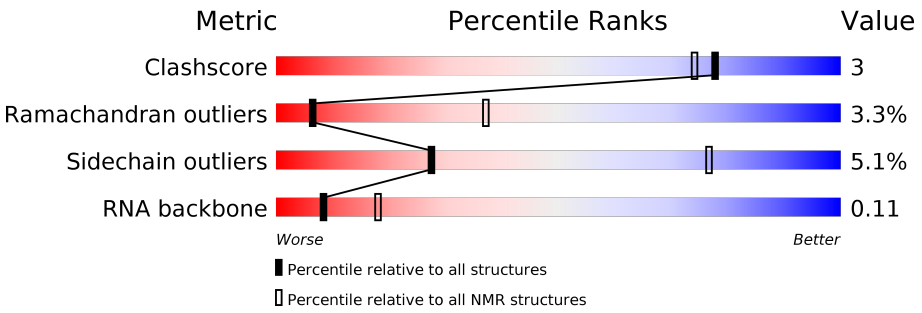
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	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 is 68%.

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	135	66% • • • 25%
2	B	6	17% 50% 33%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 11 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:3-A:98 (96)	0.49	11

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 2 clusters and 2 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1792 atoms, of which 849 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Serine/arginine-rich splicing factor 2.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	101	1604	506	783	155	155	5	0

There are 34 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-33	MET	-	EXPRESSION TAG	UNP Q01130
A	-32	GLY	-	EXPRESSION TAG	UNP Q01130
A	-31	SER	-	EXPRESSION TAG	UNP Q01130
A	-30	SER	-	EXPRESSION TAG	UNP Q01130
A	-29	HIS	-	EXPRESSION TAG	UNP Q01130
A	-28	HIS	-	EXPRESSION TAG	UNP Q01130
A	-27	HIS	-	EXPRESSION TAG	UNP Q01130
A	-26	HIS	-	EXPRESSION TAG	UNP Q01130
A	-25	HIS	-	EXPRESSION TAG	UNP Q01130
A	-24	HIS	-	EXPRESSION TAG	UNP Q01130
A	-23	SER	-	EXPRESSION TAG	UNP Q01130
A	-22	SER	-	EXPRESSION TAG	UNP Q01130
A	-21	GLY	-	EXPRESSION TAG	UNP Q01130
A	-20	LEU	-	EXPRESSION TAG	UNP Q01130
A	-19	VAL	-	EXPRESSION TAG	UNP Q01130
A	-18	PRO	-	EXPRESSION TAG	UNP Q01130
A	-17	ARG	-	EXPRESSION TAG	UNP Q01130
A	-16	GLY	-	EXPRESSION TAG	UNP Q01130
A	-15	SER	-	EXPRESSION TAG	UNP Q01130
A	-14	HIS	-	EXPRESSION TAG	UNP Q01130
A	-13	MET	-	EXPRESSION TAG	UNP Q01130
A	-12	ALA	-	EXPRESSION TAG	UNP Q01130
A	-11	SER	-	EXPRESSION TAG	UNP Q01130
A	-10	MET	-	EXPRESSION TAG	UNP Q01130
A	-9	THR	-	EXPRESSION TAG	UNP Q01130
A	-8	GLY	-	EXPRESSION TAG	UNP Q01130
A	-7	GLY	-	EXPRESSION TAG	UNP Q01130
A	-6	GLN	-	EXPRESSION TAG	UNP Q01130
A	-5	GLN	-	EXPRESSION TAG	UNP Q01130
A	-4	MET	-	EXPRESSION TAG	UNP Q01130
A	-3	GLY	-	EXPRESSION TAG	UNP Q01130

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	ARG	-	EXPRESSION TAG	UNP Q01130
A	-1	GLY	-	EXPRESSION TAG	UNP Q01130
A	0	SER	-	EXPRESSION TAG	UNP Q01130

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	6	188	56	66	20	41	5	0

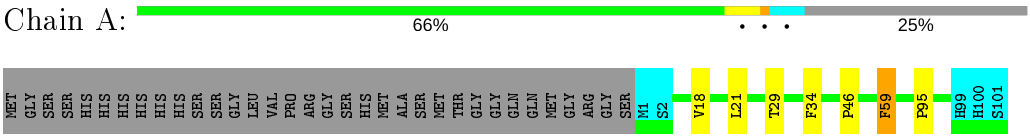
4 Residue-property plots (i)

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.

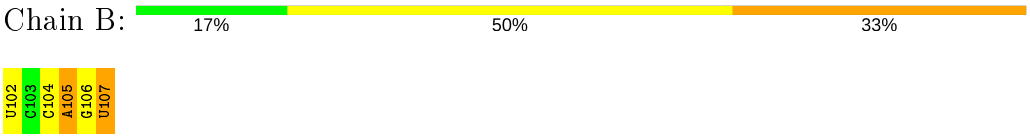
- Molecule 1: Serine/arginine-rich splicing factor 2

Chain A:



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

Chain B:

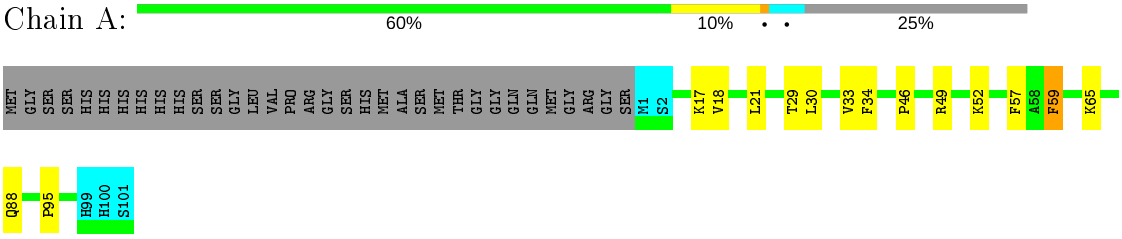


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 11. Colouring as in section 4.1 above.

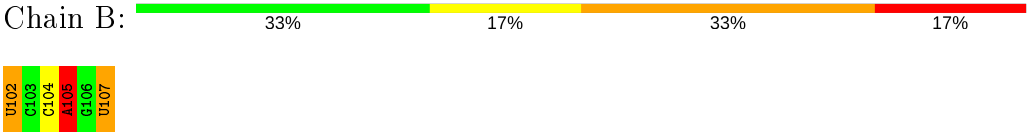
- Molecule 1: Serine/arginine-rich splicing factor 2

Chain A:



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

Chain B:



5 Refinement protocol and experimental data overview (i)

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

Of the 50 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
CYANA	structure solution	3.0
AMBER	refinement	9

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	1028
Number of shifts mapped to atoms	1028
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	68%

No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

5.1 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	750	755	5±2
2	B	122	66	66	3±1
All	All	18040	16320	16420	101

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:21:LEU:H	1:A:21:LEU:HD13	0.71	1.44	13	2
1:A:21:LEU:HD13	1:A:21:LEU:H	0.71	1.45	5	3
1:A:21:LEU:CD1	1:A:21:LEU:H	0.56	2.14	3	4
1:A:95:PRO:CD	2:B:104:C:H3'	0.54	2.32	18	18
1:A:21:LEU:H	1:A:21:LEU:CD1	0.53	2.16	13	1

5.2 Torsion angles (i)

5.2.1 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	96/135 (71%)	76±3 (79±4%)	17±3 (17±3%)	3±1 (3±1%)	6	37
All	All	1920/2700 (71%)	1523 (79%)	334 (17%)	63 (3%)	6	37

5 of 18 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	46	PRO	11
1	A	94	ARG	8
1	A	55	ARG	7
1	A	40	VAL	7
1	A	47	ARG	6

5.2.2 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	82/113 (73%)	78±1 (95±2%)	4±1 (5±2%)	27	77
All	All	1640/2260 (73%)	1557 (95%)	83 (5%)	27	77

5 of 19 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	59	PHE	20
1	A	29	THR	19
1	A	21	LEU	14
1	A	89	MET	8
1	A	80	LEU	3

5.2.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
2	B	5/6 (83%)	2±1 (47±18%)	1±0 (17±10%)	0.11±0.09
All	All	100/120 (83%)	47 (47%)	17 (17%)	0.11

The overall RNA backbone suiteness is 0.11.

All unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	B	105	A	20
2	B	107	U	14
2	B	106	G	10
2	B	103	C	2
2	B	104	C	1

All unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	B	105	A	16
2	B	106	G	1

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

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

5.4 Carbohydrates (i)

There are no carbohydrates in this entry.

5.5 Ligand geometry (i)

There are no ligands in this entry.

5.6 Other polymers (i)

There are no such molecules in this entry.

5.7 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Chemical shift validation (i)

The completeness of assignment taking into account all chemical shift lists is 68% for the well-defined parts and 67% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

6.1.1 Bookkeeping (i)

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	1028
Number of shifts mapped to atoms	1028
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	2

6.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	67	2.83 ± 0.06	Should be applied
$^{13}C_{\beta}$	76	2.95 ± 0.18	Should be applied
$^{13}C'$	0	—	None (insufficient data)
^{15}N	90	0.19 ± 0.49	None needed (< 0.5 ppm)

6.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 68%, i.e. 913 atoms were assigned a chemical shift out of a possible 1346. 0 out of 13 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	332/466 (71%)	179/185 (97%)	65/192 (34%)	88/89 (99%)
Sidechain	462/678 (68%)	307/402 (76%)	145/228 (64%)	10/48 (21%)

Continued on next page...

Continued from previous page...

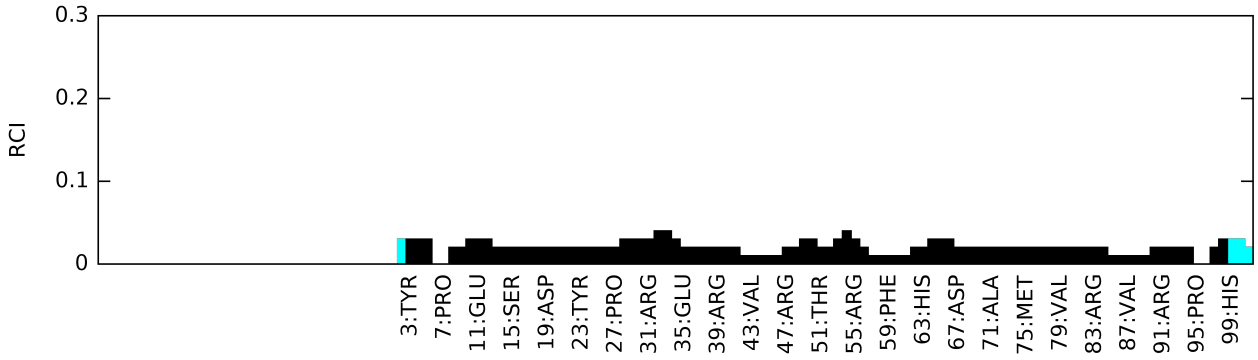
	Total	^{1}H	^{13}C	^{15}N
Aromatic	83/91 (91%)	45/48 (94%)	38/42 (90%)	0/1 (0%)
Overall	913/1346 (68%)	567/698 (81%)	248/503 (49%)	98/145 (68%)

6.1.4 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.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	17	LYS	HE2	1.84	3.87 – 1.97	-5.7
2	B	103	C	H1'	6.42	6.40 – 4.70	5.1

6.1.5 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.

Random coil index (RCI) for chain A: