



wwPDB NMR Structure Validation Summary Report ⓘ

Aug 10, 2021 – 04:10 PM EDT

PDB ID : 2LEH
Title : Solution structure of the core SMN-Gemin2 complex
Authors : Sarachan, K.L.; Valentine, K.; Gupta, K.; Moorman, V.; Gledhill, J.; Bernens, M.; Tommos, C.; Wand, A.J.; Van Duyne, G.
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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) 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 : 2.23.1
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.23.1

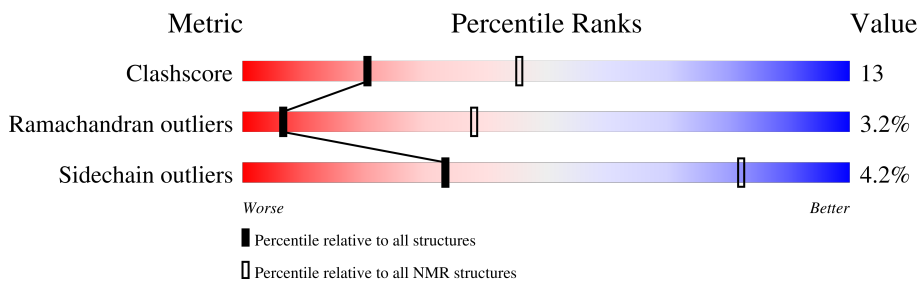
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment is 85%.

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

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	189	
2	B	26	

2 Ensemble composition and analysis

This entry contains 32 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 27 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:103-A:126, A:134-A:150, A:177-A:200, A:207-A:277, B:35-B:51 (153)	0.54	27

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 3403 atoms, of which 1681 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Survival of motor neuron protein-interacting protein 1.

Mol	Chain	Residues	Atoms					Trace	
			Total	C	H	N	O		S
1	A	189	3013	960	1493	264	289	7	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	281	LEU	-	expression tag	UNP O14893
A	282	GLU	-	expression tag	UNP O14893
A	283	TYR	-	expression tag	UNP O14893

- Molecule 2 is a protein called Survival motor neuron protein.

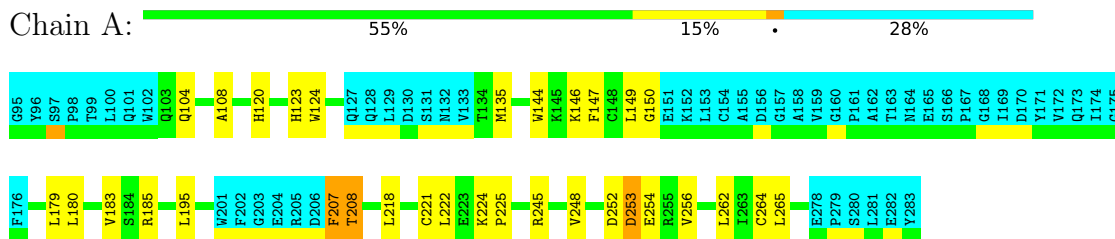
Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
2	B	26	390	126	188	31	45	0

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: Survival of motor neuron protein-interacting protein 1



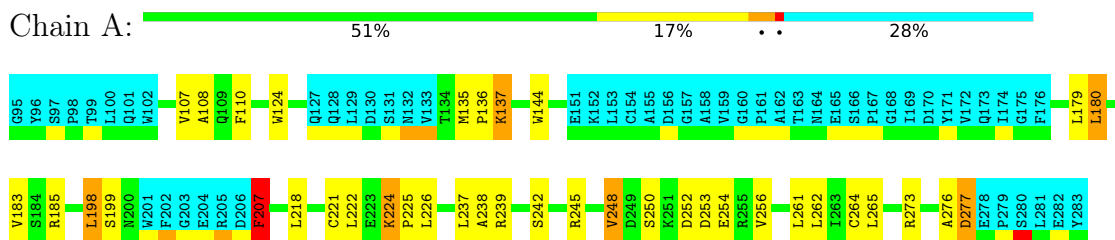
- Molecule 2: Survival motor neuron protein



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

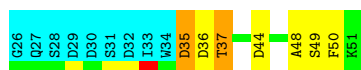
The representative model is number 27. Colouring as in section 4.1 above.

- Molecule 1: Survival of motor neuron protein-interacting protein 1



- Molecule 2: Survival motor neuron protein





5 Refinement protocol and experimental data overview

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

Of the 400 calculated structures, 32 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
X-PLOR NIH	refinement	2.26
X-PLOR NIH	structure solution	2.26

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	2544
Number of shifts mapped to atoms	2544
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	85%

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	0.67±0.00	0±0/1127 (0.0± 0.0%)	0.76±0.01	0±0/1527 (0.0± 0.0%)
2	B	0.56±0.01	0±0/132 (0.0± 0.0%)	0.69±0.02	0±0/175 (0.0± 0.0%)
All	All	0.66	0/40286 (0.0%)	0.75	2/54462 (0.0%)

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(°)	Ideal(°)	Models	
								Worst	Total
1	A	221	CYS	CB-CA-C	-5.61	99.18	110.40	5	1
1	A	270	PHE	CB-CG-CD2	-5.19	117.17	120.80	25	1

There are no chirality outliers.

There are no planarity outliers.

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	1104	1117	1117	30±6
2	B	131	132	132	4±2
All	All	39518	39968	39968	1072

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 13.

5 of 273 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:261:LEU:O	1:A:264:CYS:SG	0.98	2.21	10	4
1:A:183:VAL:CG1	1:A:221:CYS:SG	0.91	2.57	14	7
1:A:237:LEU:HD12	1:A:238:ALA:N	0.85	1.87	9	5
1:A:221:CYS:SG	1:A:222:LEU:CD1	0.82	2.66	8	11
1:A:221:CYS:SG	1:A:222:LEU:HD22	0.76	2.20	28	3

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	136/189 (72%)	126±2 (93±1%)	6±2 (4±1%)	4±1 (3±1%)	8	42
2	B	16/26 (62%)	14±1 (87±4%)	1±1 (6±5%)	1±1 (7±4%)	2	16
All	All	4864/6880 (71%)	4483 (92%)	224 (5%)	157 (3%)	7	38

5 of 16 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	B	49	SER	27
1	A	253	ASP	24
1	A	208	THR	23
1	A	207	PHE	22
1	A	254	GLU	12

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	125/169 (74%)	119±2 (96±1%)	6±2 (4±1%)	31	80
2	B	13/21 (62%)	13±0 (98±3%)	0±0 (2±3%)	54	92

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	4416/6080 (73%)	4229 (96%)	187 (4%)	33 82

5 of 33 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	185	ARG	24
1	A	123	HIS	23
1	A	256	VAL	20
1	A	180	LEU	15
1	A	208	THR	14

6.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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](#)

The completeness of assignment taking into account all chemical shift lists is 85% for the well-defined parts and 85% 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 [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	2544
Number of shifts mapped to atoms	2544
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	13

7.1.2 Chemical shift referencing [i](#)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction \pm precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	211	-0.52 ± 0.11	Should be applied
$^{13}\text{C}_\beta$	199	0.28 ± 0.09	None needed (< 0.5 ppm)
$^{13}\text{C}'$	201	-0.15 ± 0.05	None needed (< 0.5 ppm)
^{15}N	197	0.20 ± 0.22	None needed (< 0.5 ppm)

7.1.3 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 85%, i.e. 1669 atoms were assigned a chemical shift out of a possible 1964. 32 out of 33 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	732/751 (97%)	291/299 (97%)	297/306 (97%)	144/146 (99%)
Sidechain	832/1079 (77%)	511/634 (81%)	319/388 (82%)	2/57 (4%)

Continued on next page...

Continued from previous page...

	Total	¹ H	¹³ C	¹⁵ N
Aromatic	105/134 (78%)	55/71 (77%)	47/57 (82%)	3/6 (50%)
Overall	1669/1964 (85%)	857/1004 (85%)	663/751 (88%)	149/209 (71%)

7.1.4 Statistically unusual chemical shifts [i](#)

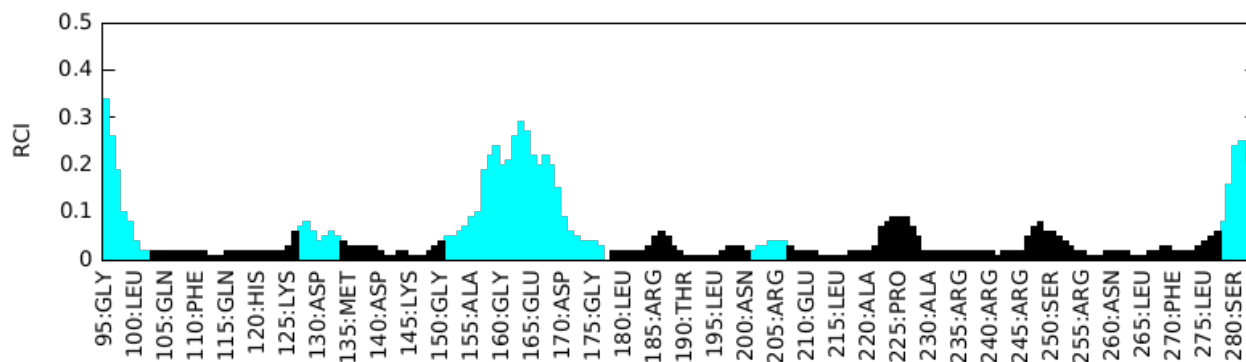
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	136	PRO	HD3	0.34	5.52 – 1.72	-8.6
1	A	136	PRO	HG3	-0.10	3.56 – 0.26	-6.1
1	A	197	TYR	CD1	125.19	139.11 – 126.41	-6.0
1	A	106	GLN	HG3	0.61	3.75 – 0.85	-5.8
1	A	194	VAL	HG21	-0.77	2.20 – -0.60	-5.6
1	A	194	VAL	HG22	-0.77	2.20 – -0.60	-5.6
1	A	194	VAL	HG23	-0.77	2.20 – -0.60	-5.6
1	A	245	ARG	HD3	1.71	4.36 – 1.86	-5.6
1	A	149	LEU	HD11	-0.68	2.16 – -0.64	-5.1
1	A	149	LEU	HD12	-0.68	2.16 – -0.64	-5.1
1	A	149	LEU	HD13	-0.68	2.16 – -0.64	-5.1
1	A	197	TYR	CD2	125.19	140.11 – 125.31	-5.1
1	A	245	ARG	CD	38.77	47.57 – 38.77	-5.0

7.1.5 Random Coil Index (RCI) plots [i](#)

The images below report *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:



Random coil index (RCI) for chain B:

