wwPDB NMR Structure Validation Summary Report

Apr 20, 2024 – 08:39 AM EDT

PDB ID	:	6MI5
BMRB ID	:	30515
Title	:	NMR solution structure of lanmodulin (LanM) complexed with yttrium(III) ions
Authors	:	Cook, E.C.; Featherson, E.R.; Showalter, S.A.; Cotruvo Jr., J.A.
Deposited on	:	2018-09-19

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:

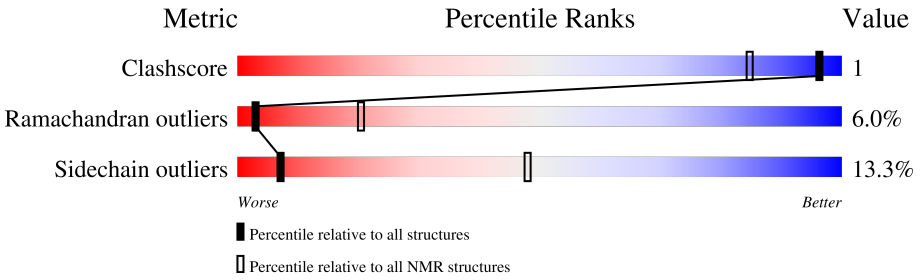
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson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

1 Overall quality at a glance

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

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	X	117	67% 23% • 9%

2 Ensemble composition and analysis

This entry contains 12 models. Model 2 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *minimized average structure*.

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	X:28-X:134 (107)	1.02	2

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 4 single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 6, 9
2	4, 5, 11
Single-model clusters	7; 8; 10; 12

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1739 atoms, of which 854 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Lanmodulin.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	X	117	1736	543	854	155	184	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
X	134	HIS	-	expression tag	UNP C5B164
X	135	HIS	-	expression tag	UNP C5B164
X	136	HIS	-	expression tag	UNP C5B164
X	137	HIS	-	expression tag	UNP C5B164
X	138	HIS	-	expression tag	UNP C5B164
X	139	HIS	-	expression tag	UNP C5B164

- Molecule 2 is YTTRIUM (III) ION (three-letter code: YT3) (formula: Y).

Mol	Chain	Residues	Atoms	
			Total	Y
2	X	3	3	3

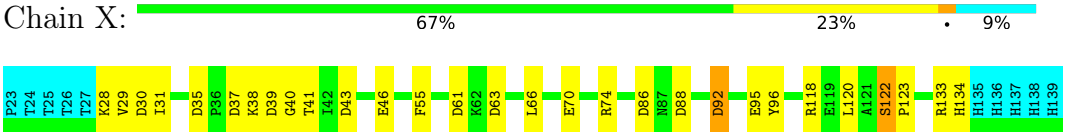
4 Residue-property plots

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: Lanmodulin

Chain X:

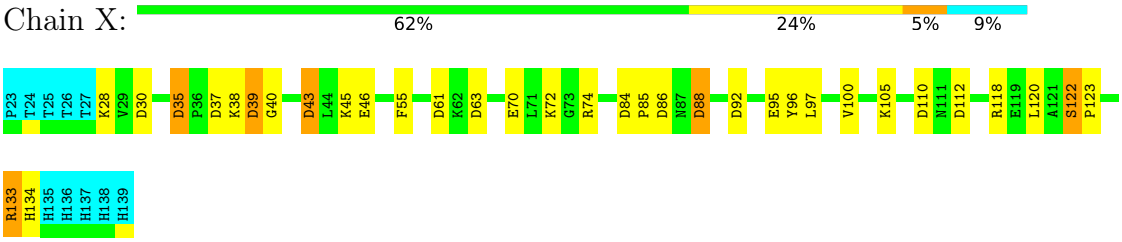


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 2. Colouring as in section 4.1 above.

- Molecule 1: Lanmodulin

Chain X:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *molecular dynamics*.

Of the 12 calculated structures, 12 were deposited, based on the following criterion: *structures with acceptable covalent geometry*.

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

Software name	Classification	Version
Amber	refinement	
Amber	structure calculation	14

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	998
Number of shifts mapped to atoms	998
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 (i)

6.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: YT3

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	X	0.66±0.01	0±0/806 (0.0± 0.0%)	1.58±0.06	20±2/1090 (1.8± 0.2%)
All	All	0.66	0/9672 (0.0%)	1.58	237/13080 (1.8%)

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	X	0.0±0.0	6.3±1.9
All	All	0	76

There are no bond-length outliers.

5 of 42 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
1	X	46	GLU	OE1-CD-OE2	-13.40	107.22	123.30	7	12
1	X	92	ASP	CB-CG-OD2	12.47	129.52	118.30	7	7
1	X	61	ASP	CB-CG-OD1	12.36	129.42	118.30	9	9
1	X	84	ASP	CB-CG-OD2	12.34	129.41	118.30	10	3
1	X	37	ASP	CB-CG-OD2	12.30	129.37	118.30	2	11

There are no chirality outliers.

5 of 28 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	X	122	SER	Peptide	9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
1	X	41	THR	Peptide	8
1	X	66	LEU	Peptide	6
1	X	120	LEU	Peptide	6
1	X	39	ASP	Peptide	4

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	X	796	784	784	1±1
All	All	9588	9408	9408	16

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 1.

5 of 13 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:X:122:SER:H	1:X:123:PRO:CD	0.60	2.08	4	2
1:X:29:VAL:HG22	1:X:30:ASP:H	0.53	1.62	8	3
1:X:44:LEU:HD23	1:X:104:PHE:CD1	0.50	2.41	6	1
1:X:96:TYR:CZ	1:X:131:LEU:HA	0.46	2.45	8	1
1:X:122:SER:H	1:X:123:PRO:HD3	0.45	1.71	4	1

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	X	107/117 (91%)	86±2 (80±2%)	15±3 (14±3%)	6±2 (6±2%)	3	20
All	All	1284/1404 (91%)	1027 (80%)	180 (14%)	77 (6%)	3	20

5 of 23 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	X	122	SER	9
1	X	123	PRO	9
1	X	38	LYS	8
1	X	134	HIS	6
1	X	40	GLY	5

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	X	82/92 (89%)	71±1 (87±2%)	11±1 (13±2%)	7 48
All	All	984/1104 (89%)	853 (87%)	131 (13%)	7 48

5 of 43 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	X	31	ILE	9
1	X	43	ASP	9
1	X	55	PHE	9
1	X	92	ASP	8
1	X	35	ASP	7

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)

Of 3 ligands modelled in this entry, 3 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 67% 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	998
Number of shifts mapped to atoms	998
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	5

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	107	2.65 ± 0.07	Should be checked
$^{13}C_{\beta}$	100	3.00 ± 0.09	Should be checked
$^{13}C'$	0	—	None (insufficient data)
^{15}N	100	0.98 ± 0.51	None needed (imprecise)

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. 974 atoms were assigned a chemical shift out of a possible 1356. 0 out of 16 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	401/532 (75%)	203/216 (94%)	102/214 (48%)	96/102 (94%)
Sidechain	566/778 (73%)	380/501 (76%)	186/251 (74%)	0/26 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	7/46 (15%)	6/23 (26%)	1/22 (5%)	0/1 (0%)
Overall	974/1356 (72%)	589/740 (80%)	289/487 (59%)	96/129 (74%)

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	X	41	THR	HG1	4.78	0.08 – 2.19	17.2
1	X	42	ILE	CG1	14.28	19.24 – 36.26	-7.9
1	X	123	PRO	CD	42.65	45.11 – 55.58	-7.3
1	X	123	PRO	HG3	3.98	0.33 – 3.48	6.6
1	X	93	LYS	HD2	0.35	0.58 – 2.64	-6.1

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 X: