wwPDB NMR Structure Validation Summary Report (i)

Jun 12, 2024 – 07:33 PM EDT

PDB ID	:	2LR1
BMRB ID	:	18340
Title	:	Structural Mechanism for Bax Inhibition by Cytomegalovirus Protein vMIA
Authors	:	Ma, J.
Deposited on	:	2012-03-20

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:

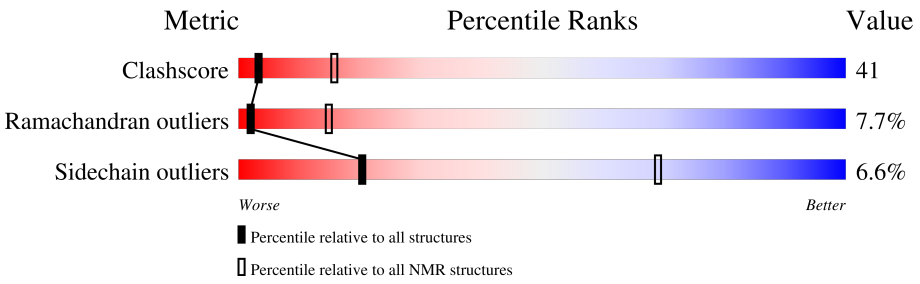
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson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

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	192	45% 33% 8% • 13%
2	B	21	100%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 19 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:15-A:36, A:43-A:47, A:53-A:192 (167)	0.09	19

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 3 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 3376 atoms, of which 1696 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Apoptosis regulator BAX.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	192	2977	949	1489	252	277	10	0

- Molecule 2 is a protein called Immediate early glycoprotein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	21	399	119	207	46	26	1	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: Apoptosis regulator BAX

Chain A: 45% 33% 8% • 13%

M1 D2 G3 S4 G5 E6 Q7 P8 R9 G10 G11 G12 P13 T14 S15 I19 L25 L26 L27 Q28 G29 F30 I31 Q32 R37 M38 G39 G40 E41 A42 P43 E44 L45 A46 L47 D48 P49 V50 P51 Q52 D53 L59 L63 K64 R65 I66 G67 D68 E69 L70 D71 M79 I80

V83 D84 S87 P88 V91 F92 V95 A96 A97 D98 M99 F100 S101 D102 G103 N104 F105 N106 W107 G108 R109 V110 V111 A112 L113 F114 K119 L120 V121 L122 K123 A124 L125 C126 T127 K128 V129 I133 W139 T140 L141 R147 L148 L149 I152 L161 T167 P168 T169 W170

Q171 T172 V173 T174 I175 A178 G179 V180 L181 T182 I187 W188 G192

• Molecule 2: Immediate early glycoprotein

Chain B: 100%

C130 E131 A132 L133 K134 K135 A136 L137 R138 R139 H140 R141 F142 L143 W144 Q145 R146 R147 Q148 R149 A150

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

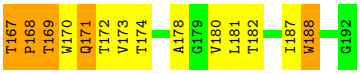
The representative model is number 19. Colouring as in section 4.1 above.

• Molecule 1: Apoptosis regulator BAX

Chain A: 46% 32% 8% • 13%

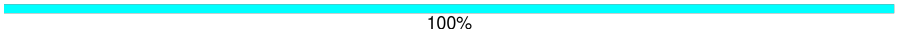
M1 D2 G3 S4 G5 E6 Q7 P8 R9 G10 G11 G12 P13 T14 S15 I19 L25 L26 L27 Q28 G29 F30 I31 Q32 R37 M38 G39 G40 E41 A42 P43 E44 L45 A46 L47 D48 P49 V50 P51 Q52 L59 L63 K64 R65 I66 G67 D68 E69 L70 D71 M74 Q77 R78

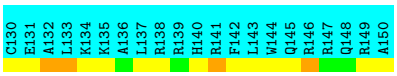
M79 I80 V83 D84 S87 P88 V91 F92 V95 A96 A97 D98 M99 F100 S101 D102 G103 N104 F105 N106 W107 G108 R109 V110 V111 A112 L113 F114 K119 L120 V121 L122 K123 A124 T127 K128 V129 I133 W139 T140 L141 R147 L148 L149 I152 W158 L161



• Molecule 2: Immediate early glycoprotein

Chain B:





5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DGSA-distance geometry simulated annealing.*

Of the 500 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**
xplor	structure solution	
xplor	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	173
Number of shifts mapped to atoms	173
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	0%

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	0.60±0.00	0±0/1344 (0.0± 0.0%)	0.82±0.01	0±1/1815 (0.0± 0.0%)
All	All	0.60	0/26880 (0.0%)	0.82	4/36300 (0.0%)

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	0.1±0.2
All	All	0	1

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(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	47	LEU	CA-C-O	-6.49	106.48	120.10	14	1
1	A	168	PRO	N-CD-CG	6.23	112.54	103.20	16	1
1	A	47	LEU	O-C-N	5.44	131.41	122.70	14	1
1	A	47	LEU	CB-CA-C	5.34	120.35	110.20	14	1

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	47	LEU	Mainchain	1

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	1317	1332	1332	109±8
2	B	0	0	0	0±0
All	All	26340	26640	26639	2184

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 41.

5 of 299 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:147:ARG:O	1:A:149:LEU:N	1.16	1.79	16	20
1:A:167:THR:OG1	1:A:168:PRO:HD2	1.11	1.43	15	19
1:A:15:SER:O	1:A:19:ILE:CD1	1.00	2.10	12	15
1:A:45:LEU:HD22	1:A:45:LEU:N	0.94	1.75	12	5
1:A:46:ALA:O	1:A:47:LEU:HD23	0.94	1.63	2	1

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	166/192 (86%)	128±1 (77±1%)	25±1 (15±1%)	13±1 (8±1%)	2 15
2	B	0	-	-	-	-
All	All	3320/4260 (78%)	2569 (77%)	497 (15%)	254 (8%)	2 15

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)
1	A	30	PHE	20

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	84	ASP	20
1	A	103	GLY	20
1	A	104	ASN	20
1	A	106	ASN	20

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	140/157 (89%)	131±1 (93±1%)	9±1 (7±1%)	20 69
2	B	0	-	-	-
All	All	2800/3500 (80%)	2615 (93%)	185 (7%)	20 69

5 of 24 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	167	THR	20
1	A	109	ARG	19
1	A	148	LEU	19
1	A	69	GLU	17
1	A	59	LEU	16

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 0% for the well-defined parts and 6% 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	173
Number of shifts mapped to atoms	173
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	11

7.1.2 Chemical shift referencing (i)

No chemical shift referencing corrections were calculated (not enough data).

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 0%, i.e. 0 atoms were assigned a chemical shift out of a possible 2314. 0 out of 32 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

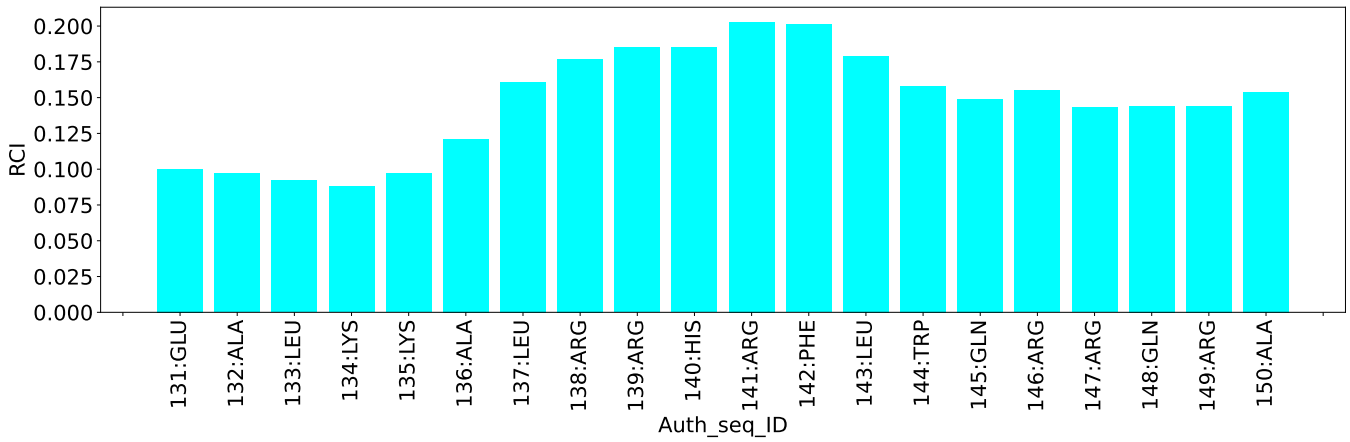
	Total	**^{1}H**	**^{13}C**	**^{15}N**
Backbone	0/841 (0%)	0/344 (0%)	0/334 (0%)	0/163 (0%)
Sidechain	0/1283 (0%)	0/841 (0%)	0/396 (0%)	0/46 (0%)
Aromatic	0/190 (0%)	0/94 (0%)	0/90 (0%)	0/6 (0%)
Overall	0/2314 (0%)	0/1279 (0%)	0/820 (0%)	0/215 (0%)

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	B	143	LEU	CG	57.21	21.37 – 32.19	28.1
1	B	137	LEU	CG	57.08	21.37 – 32.19	28.0
1	B	135	LYS	CD	59.43	23.50 – 34.42	27.9
1	B	134	LYS	CG	55.25	19.35 – 30.45	27.3
1	B	135	LYS	CG	55.24	19.35 – 30.45	27.3
1	B	137	LEU	CD1	55.14	16.71 – 32.55	19.3
1	B	143	LEU	CD1	55.02	16.71 – 32.55	19.2
1	B	137	LEU	CD2	55.14	15.73 – 32.47	18.5
1	B	143	LEU	CD2	55.02	15.73 – 32.47	18.5
1	B	145	GLN	CG	24.79	28.36 – 39.21	-8.3
1	B	148	GLN	CG	24.79	28.36 – 39.21	-8.3

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 B: