Full wwPDB NMR Structure Validation Report (i)

Jun 15, 2024 – 03:32 PM EDT

PDB ID	:	2MIZ
BMRB ID	:	19699
Title	:	Structure of the m04/gp34 mouse Cytomegalovirus Immunoevasin core domain
Authors	:	Sgourakis, N.G.; Natarajan, K.; Margulies, D.H.; Bax, A.
Deposited on	:	2013-12-21

This is a Full wwPDB NMR Structure Validation 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:

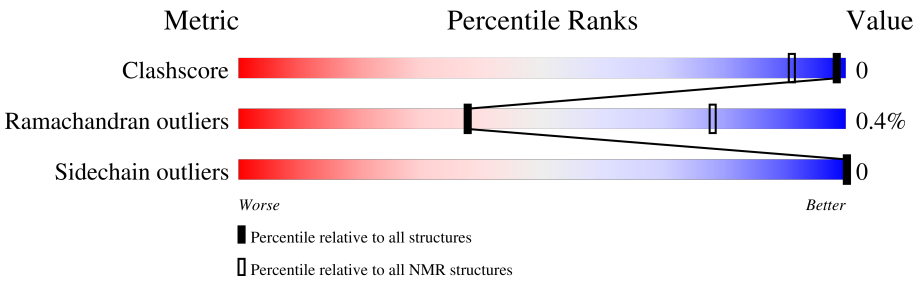
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 34%.

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	200	69% 9% 22%

2 Ensemble composition and analysis (i)

This entry contains 10 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 10 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:22-A:35, A:42-A:93, A:105-A:176 (138)	0.48	10

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, 8, 9
2	4, 7, 10
Single-model clusters	3; 5; 6

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2428 atoms, of which 1204 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called m04 immunoevasin.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	156	2428	766	1204	216	233	9	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	MET	-	EXPRESSION TAG	UNP A2Q6L0
A	-1	ALA	-	EXPRESSION TAG	UNP A2Q6L0
A	0	SER	-	EXPRESSION TAG	UNP A2Q6L0
A	7	SER	CYS	ENGINEERED MUTATION	UNP A2Q6L0
A	44	VAL	ILE	ENGINEERED MUTATION	UNP A2Q6L0

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: m04 immunoevasin

Chain A: 69% 9% 22%

MET ALA SER SER ARG ASN ASP ASN GLU SER GLU LYS MET GLN LYS GLU TYR LYS GLU LYS MET LYS TYR R21 V36 P37 S38 S39 D40 P41 F94 T95 P96 M97 H98 G99 L100 R101 T102 K103 N104 G176 VAL PRO ASP LEU THR PRO PRO ARG ARG ASP GLU PHE TYR THR LYS

ASN GLU SER PRO ASN THR

4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: m04 immunoevasin

Chain A: 68% 9% 22%

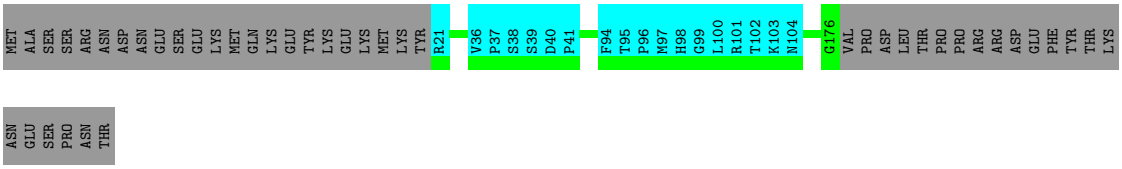
MET ALA SER SER ARG ASN ASP ASN GLU SER GLU LYS MET GLN LYS GLU TYR LYS GLU LYS MET LYS TYR R21 V36 P37 S38 S39 D40 P41 V73 F94 T95 P96 M97 H98 G99 L100 R101 T102 K103 N104 G176 VAL PRO ASP LEU THR PRO PRO ARG ARG ASP GLU PHE TYR

THR LYS ASN GLU SER PRO ASN THR

4.2.2 Score per residue for model 2

• Molecule 1: m04 immunoevasin

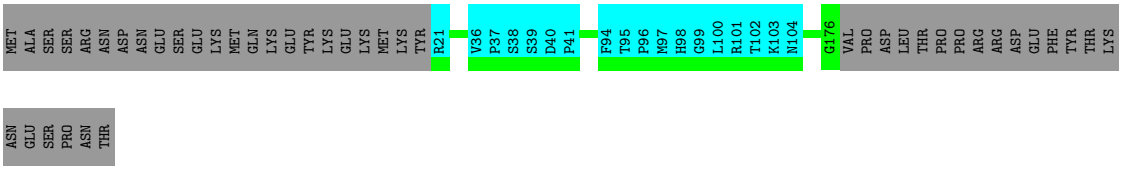
Chain A: 69% 9% 22%



4.2.3 Score per residue for model 3

- Molecule 1: m04 immunoevasin

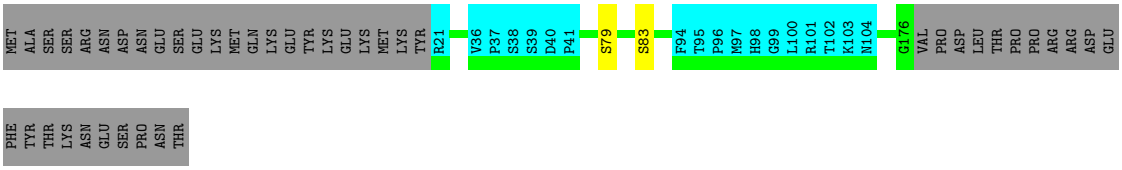
Chain A:



4.2.4 Score per residue for model 4

- Molecule 1: m04 immunoevasin

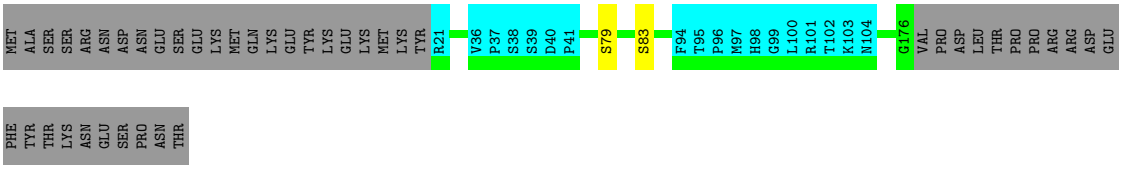
Chain A:



4.2.5 Score per residue for model 5

- Molecule 1: m04 immunoevasin

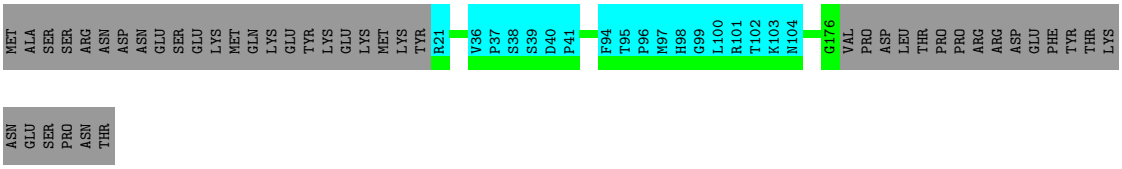
Chain A:



4.2.6 Score per residue for model 6

- Molecule 1: m04 immunoevasin

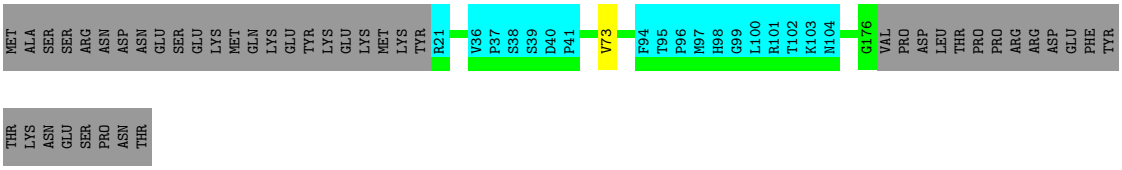
Chain A:



4.2.7 Score per residue for model 7

- Molecule 1: m04 immunoevasin

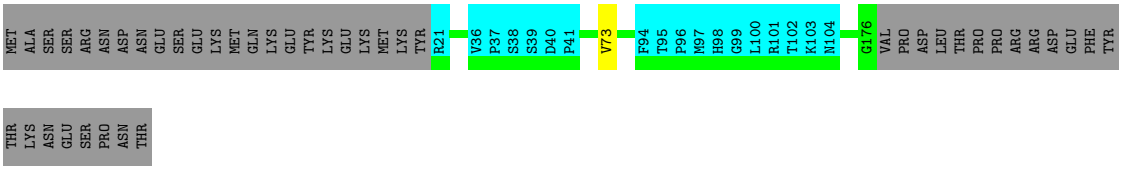
Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: m04 immunoevasin

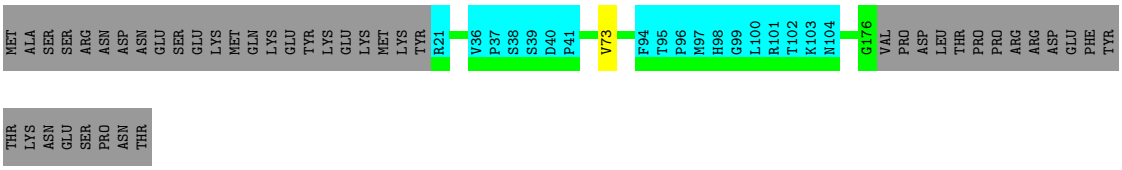
Chain A:



4.2.9 Score per residue for model 9

- Molecule 1: m04 immunoevasin

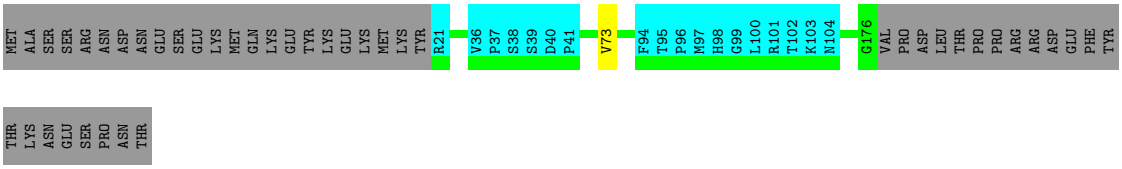
Chain A:



4.2.10 Score per residue for model 10 (medoid)

- Molecule 1: m04 immunoevasin

Chain A:


MET
ALA
SER
SER
ARG
ASN
ASP
ASN
GLU
SER
GLU
LYS
MET
GLN
LYS
GLU
TYR
LYS
GLU
LYS
MET
LYS
TYR
R21
V36
P37
S38
S39
D40
P41
V73
F94
T95
P96
M97
H98
G99
L100
R101
T102
K103
N104
G176
VAL
PRO
ASP
LEU
THR
PRO
PRO
ARG
ARG
ASP
GLU
PHE
TYR

THR
LYS
ASN
GLU
SER
PRO
ASN
THR

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *RASREC*.

Of the 10000 calculated structures, 10 were deposited, based on the following criterion: *target function*.

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

Software name	Classification	Version
CS-Rosetta	structure solution	3.0
CS-Rosetta	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	697
Number of shifts mapped to atoms	696
Number of unparsed shifts	0
Number of shifts with mapping errors	1
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	34%

6 Model quality (i)

6.1 Standard geometry (i)

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

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	1082	1062	1062	0±0
All	All	10820	10620	10620	2

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 0.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:79:SER:OG	1:A:83:SER:O	0.46	2.34	5	2

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	137/200 (68%)	133±1 (97±1%)	3±1 (2±1%)	0±0 (0±0%)	38 78

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1370/2000 (68%)	1332 (97%)	33 (2%)	5 (0%)	38 78

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	73	VAL	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	A	125/185 (68%)	125±0 (100±0%)	0±0 (0±0%)	100 100
All	All	1250/1850 (68%)	1250 (100%)	0 (0%)	100 100

There are no protein residues with a non-rotameric sidechain to report.

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 34% for the well-defined parts and 33% 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	697
Number of shifts mapped to atoms	696
Number of unparsed shifts	0
Number of shifts with mapping errors	1
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. All 1 occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	21	ARG	H1	7.333	0.020	1

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	138	0.33 ± 0.24	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	128	1.02 ± 0.24	Should be checked
$^{13}C'$	145	0.19 ± 0.16	None needed (< 0.5 ppm)
^{15}N	143	-1.44 ± 0.26	Should be applied

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 34%, i.e. 636 atoms were assigned a chemical shift out of a possible 1847. 0 out of 27 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	518/682 (76%)	130/276 (47%)	258/276 (93%)	130/130 (100%)
Sidechain	118/1030 (11%)	0/669 (0%)	118/318 (37%)	0/43 (0%)
Aromatic	0/135 (0%)	0/65 (0%)	0/64 (0%)	0/6 (0%)
Overall	636/1847 (34%)	130/1010 (13%)	376/658 (57%)	130/179 (73%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 33%, i.e. 696 atoms were assigned a chemical shift out of a possible 2094. 0 out of 29 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	568/767 (74%)	142/310 (46%)	283/312 (91%)	143/145 (99%)
Sidechain	128/1174 (11%)	0/762 (0%)	128/361 (35%)	0/51 (0%)
Aromatic	0/153 (0%)	0/74 (0%)	0/71 (0%)	0/8 (0%)
Overall	696/2094 (33%)	142/1146 (12%)	411/744 (55%)	143/204 (70%)

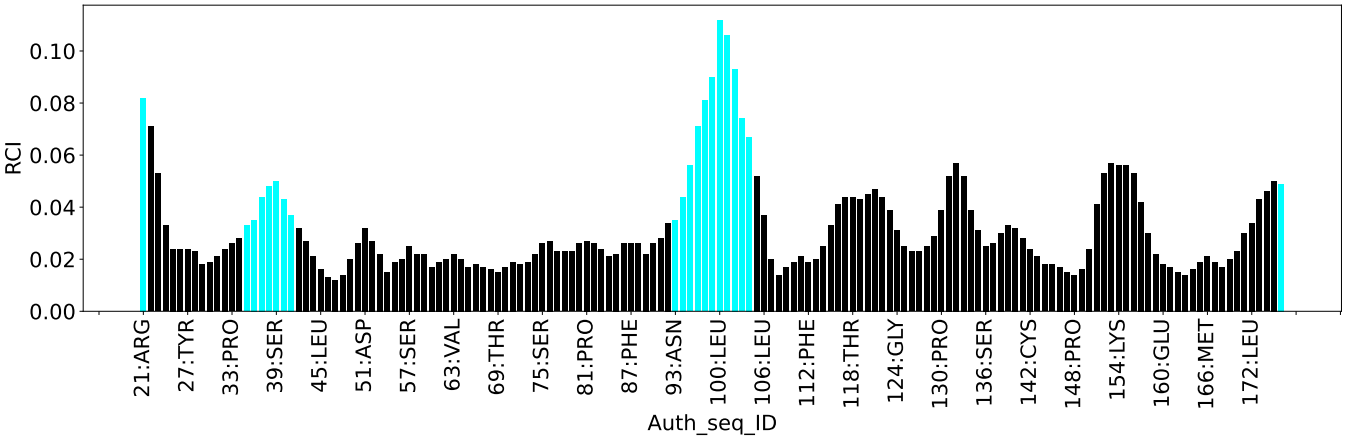
7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

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:


RCI
0.00
0.02
0.04
0.06
0.08
0.10
21:ARG
27:TYR
33:PRO
39:SER
45:LEU
51:ASP
57:SER
63:VAL
69:THR
75:SER
81:PRO
87:PHE
93:ASN
100:LEU
106:LEU
112:PHE
118:THR
124:GLY
130:PRO
136:SER
142:CYS
148:PRO
154:LYS
160:GLU
166:MET
172:LEU
Auth_seq_ID