Full wwPDB NMR Structure Validation Report (i)

Apr 27, 2016 – 03:23 AM BST

PDB ID	:	2MUJ
Title	:	Plasmodium falciparum SERA protein peptide analogues having short helical regions induce protection against malaria
Authors	:	Cubillos, M.; Alba, M.; Bermudez, A.; Trujillo, M.; Patarroyo, M.E.
Deposited on	:	2014-09-10

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
http://wwpdb.org/validation/2016/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (i)) were used in the production of this report:

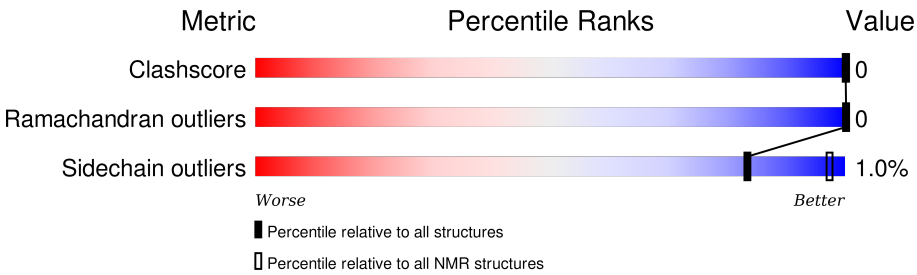
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

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 50%.

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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	21	33% 5% 62%

2 Ensemble composition and analysis (i)

This entry contains 13 models. Model 9 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:7-A:14 (8)	0.06	9

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 3 clusters and 1 single-model cluster was found.

Cluster number	Models
1	1, 4, 6, 11, 12
2	5, 8, 10, 13
3	2, 3, 9
Single-model clusters	7

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 352 atoms, of which 176 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Serine-repeat antigen protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	21	352	111	176	28	36	1	0

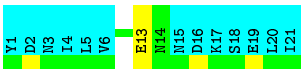
4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA 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-repeat antigen protein

Chain A:



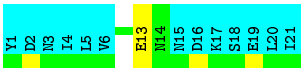
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: Serine-repeat antigen protein

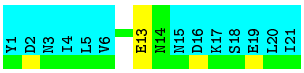
Chain A:



4.2.2 Score per residue for model 2

- Molecule 1: Serine-repeat antigen protein

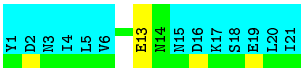
Chain A:



4.2.3 Score per residue for model 3

- Molecule 1: Serine-repeat antigen protein

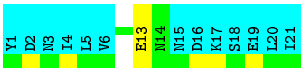
Chain A:



4.2.4 Score per residue for model 4

- Molecule 1: Serine-repeat antigen protein

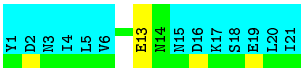
Chain A:



4.2.5 Score per residue for model 5

- Molecule 1: Serine-repeat antigen protein

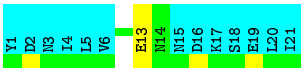
Chain A:



4.2.6 Score per residue for model 6

- Molecule 1: Serine-repeat antigen protein

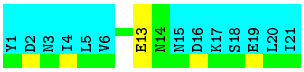
Chain A:



4.2.7 Score per residue for model 7

- Molecule 1: Serine-repeat antigen protein

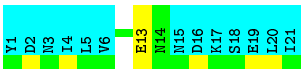
Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: Serine-repeat antigen protein

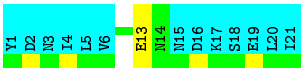
Chain A: 33% 5% 62%



4.2.9 Score per residue for model 9 (medoid)

- Molecule 1: Serine-repeat antigen protein

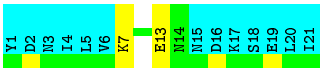
Chain A: 33% 5% 62%



4.2.10 Score per residue for model 10

- Molecule 1: Serine-repeat antigen protein

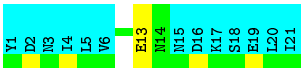
Chain A: 29% 10% 62%



4.2.11 Score per residue for model 11

- Molecule 1: Serine-repeat antigen protein

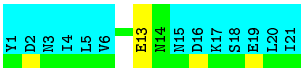
Chain A: 33% 5% 62%



4.2.12 Score per residue for model 12

- Molecule 1: Serine-repeat antigen protein

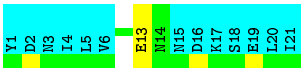
Chain A: 33% 5% 62%



4.2.13 Score per residue for model 13

- Molecule 1: Serine-repeat antigen protein

Chain A:



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DGSA-DISTANCE GEOMETRY SIMULATED ANNEALING.*

Of the 50 calculated structures, 13 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
InsightII	geometry optimization	
InsightII	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)	2muj_cs.str
Number of chemical shift lists	1
Total number of shifts	153
Number of shifts mapped to atoms	153
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	50%

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

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	1.68±0.01	1±0/70 (1.4±0.0%)	1.03±0.02	0±0/92 (0.0±0.0%)
All	All	1.68	13/910 (1.4%)	1.03	0/1196 (0.0%)

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	13	GLU	CD-OE2	10.93	1.37	1.25	4	13

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
All	All	897	884	897	-

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 -.

There are no clashes.

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	8/21 (38%)	8±0 (96±6%)	0±0 (4±6%)	0±0 (0±0%)	100	100
All	All	104/273 (38%)	100 (96%)	4 (4%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	8/21 (38%)	8±0 (99±3%)	0±0 (1±3%)	83	97
All	All	104/273 (38%)	103 (99%)	1 (1%)	83	97

All 1 unique residues with a non-rotameric sidechain are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	7	LYS	1

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 carbohydrates 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 50% for the well-defined parts and 47% for the entire structure.

7.1 Chemical shift list 1

File name: 2muj_cs.str

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	153
Number of shifts mapped to atoms	153
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

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 50%, i.e. 55 atoms were assigned a chemical shift out of a possible 111. 0 out of 0 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	16/40 (40%)	16/16 (100%)	0/16 (0%)	0/8 (0%)
Sidechain	35/62 (56%)	35/37 (95%)	0/21 (0%)	0/4 (0%)
Aromatic	4/9 (44%)	4/5 (80%)	0/4 (0%)	0/0 (—%)
Overall	55/111 (50%)	55/58 (95%)	0/41 (0%)	0/12 (0%)

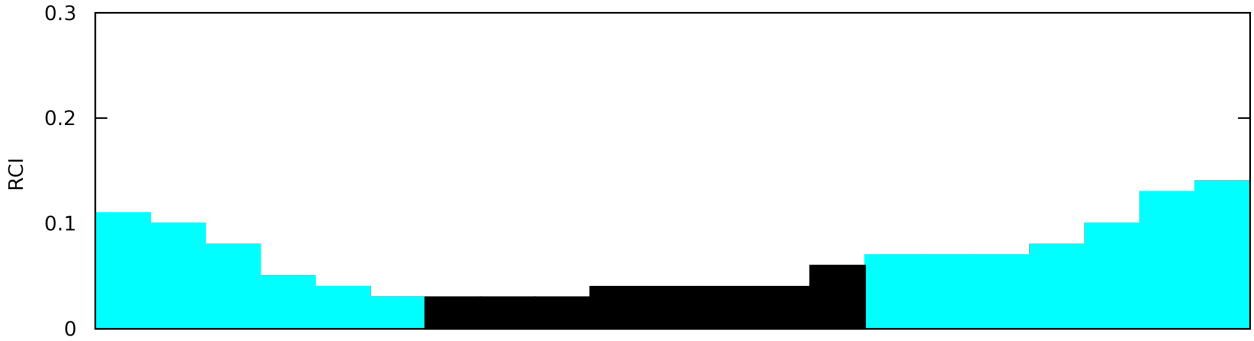
The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 47%, i.e. 129 atoms were assigned a chemical shift out of a possible 274. 0 out of 3 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	41/105 (39%)	41/42 (98%)	0/42 (0%)	0/21 (0%)
Sidechain	84/152 (55%)	84/88 (95%)	0/57 (0%)	0/7 (0%)
Aromatic	4/17 (24%)	4/9 (44%)	0/8 (0%)	0/0 (—%)
Overall	129/274 (47%)	129/139 (93%)	0/107 (0%)	0/28 (0%)

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.

Random coil index (RCI) for chain A: