Full wwPDB NMR Structure Validation Report (i)

Apr 27, 2016 – 03:34 AM BST

PDB ID	:	2MUD
Title	:	Structural modifications to a high-activity binding peptide located whitin the PfEMP1 NTS domain induce protection against P. falciparum malaria in Aotus monkeys
Authors	:	Curtidor, H.; Torres, M.; Alba, M.; Patarroyo, M.
Deposited on	:	2014-09-08

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:

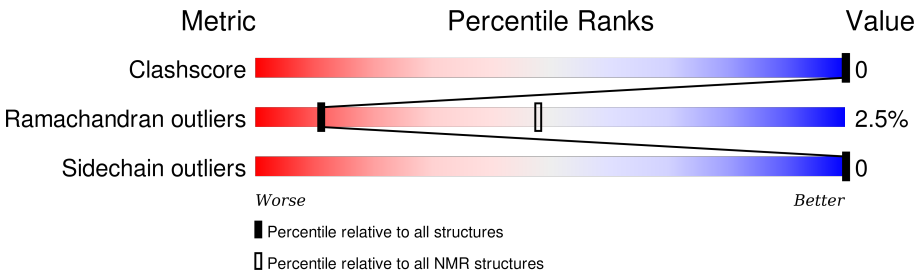
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 35%.

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	22	64% 27% 9%

2 Ensemble composition and analysis (i)

This entry contains 29 models. Model 28 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:1-A:19 (19)	0.06	28

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 4 clusters and 4 single-model clusters were found.

Cluster number	Models
1	1, 4, 5, 6, 7, 11, 13, 14, 16, 17, 23, 27, 28, 29
2	10, 18, 19, 21
3	22, 24, 25, 26
4	2, 8, 15
Single-model clusters	3; 9; 12; 20

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 341 atoms, of which 170 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Variant-specific surface protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	22	341	107	170	30	33	1	1

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1B	ACE	-	ACETYLATION	UNP Q26031
A	21	NH2	-	AMIDATION	UNP Q26031

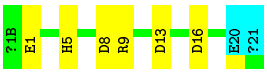
4 Residue-property plots (i)

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: Variant-specific surface protein

Chain A:



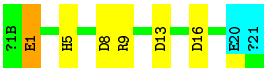
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: Variant-specific surface protein

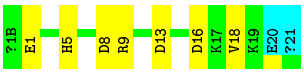
Chain A:



4.2.2 Score per residue for model 2

- Molecule 1: Variant-specific surface protein

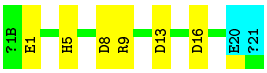
Chain A:



4.2.3 Score per residue for model 3

- Molecule 1: Variant-specific surface protein

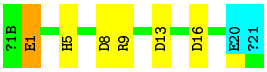
Chain A:



4.2.4 Score per residue for model 4

- Molecule 1: Variant-specific surface protein

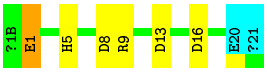
Chain A:



4.2.5 Score per residue for model 5

- Molecule 1: Variant-specific surface protein

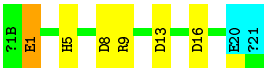
Chain A:



4.2.6 Score per residue for model 6

- Molecule 1: Variant-specific surface protein

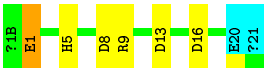
Chain A:



4.2.7 Score per residue for model 7

- Molecule 1: Variant-specific surface protein

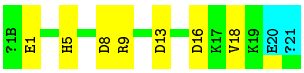
Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: Variant-specific surface protein

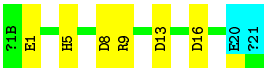
Chain A:



4.2.9 Score per residue for model 9

- Molecule 1: Variant-specific surface protein

Chain A:



4.2.10 Score per residue for model 10

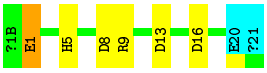
- Molecule 1: Variant-specific surface protein

Chain A:

4.2.11 Score per residue for model 11

- Molecule 1: Variant-specific surface protein

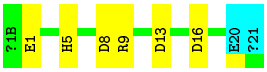
Chain A:



4.2.12 Score per residue for model 12

- Molecule 1: Variant-specific surface protein

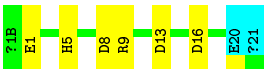
Chain A:



4.2.13 Score per residue for model 13

- Molecule 1: Variant-specific surface protein

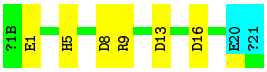
Chain A:



4.2.14 Score per residue for model 14

- Molecule 1: Variant-specific surface protein

Chain A:



4.2.15 Score per residue for model 15

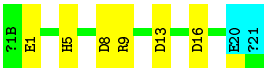
- Molecule 1: Variant-specific surface protein

Chain A:

4.2.16 Score per residue for model 16

- Molecule 1: Variant-specific surface protein

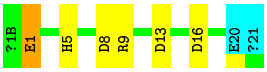
Chain A:



4.2.17 Score per residue for model 17

- Molecule 1: Variant-specific surface protein

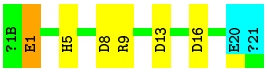
Chain A:



4.2.18 Score per residue for model 18

- Molecule 1: Variant-specific surface protein

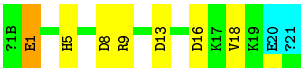
Chain A:



4.2.19 Score per residue for model 19

- Molecule 1: Variant-specific surface protein

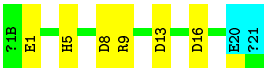
Chain A:



4.2.20 Score per residue for model 20

- Molecule 1: Variant-specific surface protein

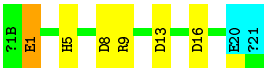
Chain A:



4.2.21 Score per residue for model 21

- Molecule 1: Variant-specific surface protein

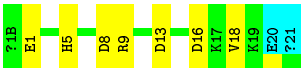
Chain A:



4.2.22 Score per residue for model 22

- Molecule 1: Variant-specific surface protein

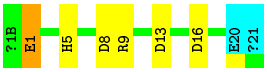
Chain A:



4.2.23 Score per residue for model 23

- Molecule 1: Variant-specific surface protein

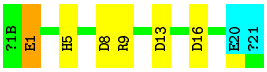
Chain A:



4.2.24 Score per residue for model 24

- Molecule 1: Variant-specific surface protein

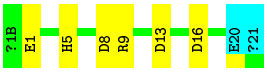
Chain A:



4.2.25 Score per residue for model 25

- Molecule 1: Variant-specific surface protein

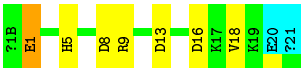
Chain A:



4.2.26 Score per residue for model 26

- Molecule 1: Variant-specific surface protein

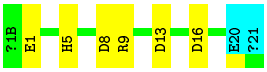
Chain A:



4.2.27 Score per residue for model 27

- Molecule 1: Variant-specific surface protein

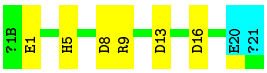
Chain A:



4.2.28 Score per residue for model 28 (medoid)

- Molecule 1: Variant-specific surface protein

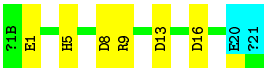
Chain A:



4.2.29 Score per residue for model 29

- Molecule 1: Variant-specific surface 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, 29 were deposited, based on the following criterion: *structures with the least restraint violations.*

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)	2mud_cs.str
Number of chemical shift lists	1
Total number of shifts	109
Number of shifts mapped to atoms	109
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	35%

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)

Bond lengths and bond angles in the following residue types are not validated in this section: ACE, NH2

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	1.52±0.01	4±0/162 (2.5±0.0%)	1.88±0.02	8±0/215 (3.8±0.2%)
All	All	1.52	116/4698 (2.5%)	1.88	239/6235 (3.8%)

All unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	1	GLU	CD-OE2	10.95	1.37	1.25	26	29
1	A	13	ASP	CG-OD2	5.37	1.37	1.25	5	29
1	A	16	ASP	CG-OD2	5.37	1.37	1.25	9	29
1	A	8	ASP	CG-OD2	5.37	1.37	1.25	13	29

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	9	ARG	NE-CZ-NH1	7.85	124.23	120.30	24	29
1	A	16	ASP	CB-CG-OD2	-6.83	112.15	118.30	2	29
1	A	13	ASP	CB-CG-OD2	-6.82	112.16	118.30	6	29
1	A	8	ASP	CB-CG-OD2	-6.79	112.19	118.30	14	29
1	A	8	ASP	CB-CG-OD1	5.71	123.44	118.30	9	29
1	A	16	ASP	CB-CG-OD1	5.71	123.44	118.30	12	29
1	A	13	ASP	CB-CG-OD1	5.68	123.41	118.30	6	29
1	A	5	HIS	CG-ND1-CE1	-5.43	98.64	105.70	4	29
1	A	18	VAL	CA-CB-CG2	5.20	118.70	110.90	26	7

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	4669	4669	4698	-

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	19/22 (86%)	12±1 (64±5%)	6±1 (33±6%)	0±0 (3±3%)	11 48
All	All	551/638 (86%)	354 (64%)	183 (33%)	14 (3%)	11 48

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	1	GLU	14

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	17/18 (94%)	17±0 (100±0%)	0±0 (0±0%)	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	493/522 (94%)	493 (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 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 35% for the well-defined parts and 36% for the entire structure.

7.1 Chemical shift list 1

File name: 2mud_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	109
Number of shifts mapped to atoms	109
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 35%, i.e. 89 atoms were assigned a chemical shift out of a possible 251. 0 out of 2 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	38/95 (40%)	38/38 (100%)	0/38 (0%)	0/19 (0%)
Sidechain	41/132 (31%)	41/78 (53%)	0/47 (0%)	0/7 (0%)
Aromatic	10/24 (42%)	10/13 (77%)	0/10 (0%)	0/1 (0%)
Overall	89/251 (35%)	89/129 (69%)	0/95 (0%)	0/27 (0%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 36%, i.e. 95 atoms were assigned a chemical shift out of a possible 263. 0 out of 2 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	40/100 (40%)	40/40 (100%)	0/40 (0%)	0/20 (0%)
Sidechain	45/139 (32%)	45/82 (55%)	0/50 (0%)	0/7 (0%)
Aromatic	10/24 (42%)	10/13 (77%)	0/10 (0%)	0/1 (0%)
Overall	95/263 (36%)	95/135 (70%)	0/100 (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: