



# Full wwPDB X-ray Structure Validation Report ⓘ

Mar 9, 2018 – 12:50 pm GMT

PDB ID : 5NQF  
Title : Crystal structure of Plasmodium falciparum AMA1 in complex with a 39 aa PvRON2 peptide  
Authors : Vulliez-le Normand, B.; Saul, F.A.; Faber, B.W.; Bentley, G.A.  
Deposited on : 2017-04-20  
Resolution : 1.90 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : trunk30967  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Refmac : 5.8.0158  
CCP4 : 7.0 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk30967

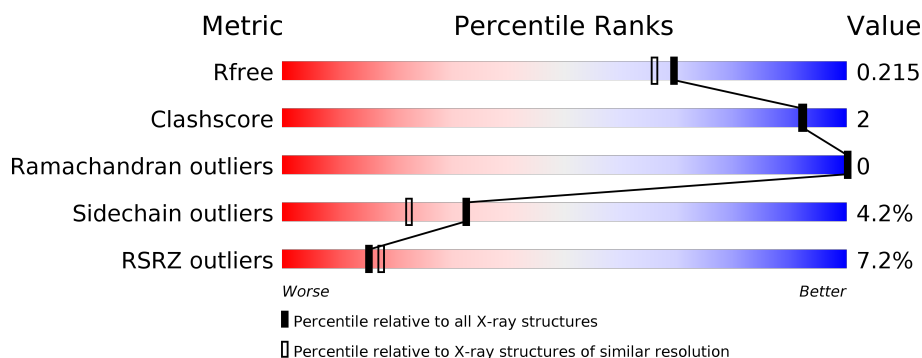
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.90 Å.

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)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	111664	5502 (1.90-1.90)
Clashscore	122126	6115 (1.90-1.90)
Ramachandran outliers	120053	6048 (1.90-1.90)
Sidechain outliers	120020	6048 (1.90-1.90)
RSRZ outliers	108989	5379 (1.90-1.90)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	371	<div> <div>4%</div> <div> <div></div> <div>74%</div> <div>6%</div> <div>20%</div> </div> </div>
2	B	39	<div> <div>23%</div> <div> <div></div> <div>82%</div> <div>8%</div> <div>10%</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 2970 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Apical membrane antigen 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	298	2410	1520	410	464	16	0	3	0

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	95	GLU	-	expression tag	UNP A0A024UZE1
A	96	PHE	-	expression tag	UNP A0A024UZE1
A	162	LYS	ASN	engineered mutation	UNP A0A024UZE1
A	288	VAL	THR	engineered mutation	UNP A0A024UZE1
A	373	ASP	SER	engineered mutation	UNP A0A024UZE1
A	422	ASP	ASN	engineered mutation	UNP A0A024UZE1
A	423	LYS	SER	engineered mutation	UNP A0A024UZE1
A	443	GLY	-	expression tag	UNP A0A024UZE1
A	444	LEU	-	expression tag	UNP A0A024UZE1
A	445	GLU	-	expression tag	UNP A0A024UZE1
A	446	GLN	-	expression tag	UNP A0A024UZE1
A	447	LYS	-	expression tag	UNP A0A024UZE1
A	448	LEU	-	expression tag	UNP A0A024UZE1
A	449	ILE	-	expression tag	UNP A0A024UZE1
A	450	SER	-	expression tag	UNP A0A024UZE1
A	451	GLU	-	expression tag	UNP A0A024UZE1
A	452	GLU	-	expression tag	UNP A0A024UZE1
A	453	ASP	-	expression tag	UNP A0A024UZE1
A	454	LEU	-	expression tag	UNP A0A024UZE1
A	455	ASN	-	expression tag	UNP A0A024UZE1
A	456	SER	-	expression tag	UNP A0A024UZE1
A	457	ALA	-	expression tag	UNP A0A024UZE1
A	458	VAL	-	expression tag	UNP A0A024UZE1
A	459	ASP	-	expression tag	UNP A0A024UZE1
A	460	HIS	-	expression tag	UNP A0A024UZE1
A	461	HIS	-	expression tag	UNP A0A024UZE1
A	462	HIS	-	expression tag	UNP A0A024UZE1

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	463	HIS	-	expression tag	UNP A0A024UZE1
A	464	HIS	-	expression tag	UNP A0A024UZE1
A	465	HIS	-	expression tag	UNP A0A024UZE1

- Molecule 2 is a protein called Rhoptry neck protein 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	35	Total	C	N	O	S	0	1	0
			253	159	43	48	3			

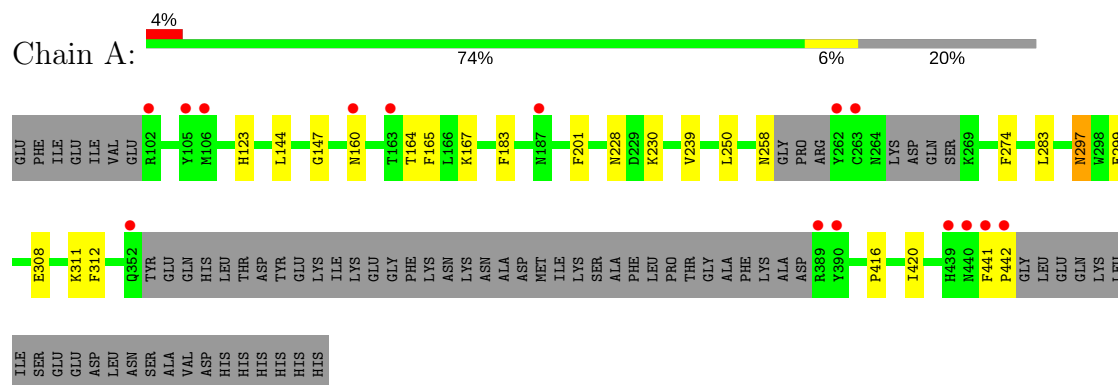
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	290	Total	O	0	0
			290	290		
3	B	17	Total	O	0	0
			17	17		

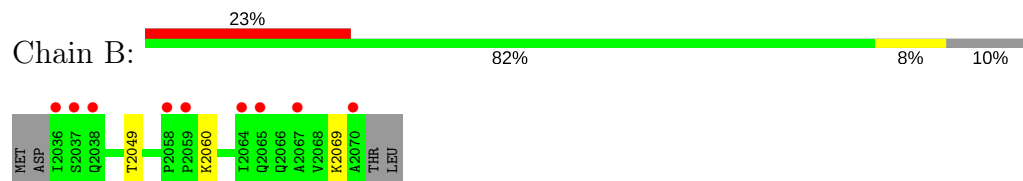
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-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. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

#### • Molecule 1: Apical membrane antigen 1



#### • Molecule 2: Rhoptry neck protein 2



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	71.48Å 38.25Å 72.36Å 90.00° 98.55° 90.00°	Depositor
Resolution (Å)	35.78 – 1.90 35.78 – 1.90	Depositor EDS
% Data completeness (in resolution range)	95.9 (35.78-1.90) 95.9 (35.78-1.90)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.62 (at 1.89Å)	Xtriage
Refinement program	BUSTER 2.10.2	Depositor
R, $R_{free}$	0.169 , 0.205 0.178 , 0.215	Depositor DCC
$R_{free}$ test set	1511 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	20.4	Xtriage
Anisotropy	0.573	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 56.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.024 for l,-k,h	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	2970	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.65% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality

### 5.1 Standard geometry

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 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	0.51	0/2473	0.63	0/3348
2	B	0.45	0/258	0.66	0/352
All	All	0.51	0/2731	0.63	0/3700

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 the 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 within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2410	0	2255	9	0
2	B	253	0	253	2	0
3	A	290	0	0	0	0
3	B	17	0	0	0	0
All	All	2970	0	2508	9	0

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 2.

All (9) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:201:PHE:HZ	2:B:2060[B]:LYS:HE2	1.61	0.65

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:297:ASN:HD22	1:A:297:ASN:H	1.47	0.62
1:A:311:LYS:HG2	1:A:420:ILE:HB	1.97	0.47
1:A:164:THR:HG23	1:A:167:LYS:HE2	1.98	0.46
1:A:228:ASN:HD22	2:B:2049:THR:HG22	1.81	0.45
1:A:123:HIS:CE1	1:A:147:GLY:HA3	2.51	0.45
1:A:239:VAL:HG23	1:A:250:LEU:HD11	1.98	0.45
1:A:441:PHE:HA	1:A:442:PRO:C	2.40	0.42
1:A:312:PHE:HB3	1:A:416:PRO:HB3	2.03	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 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 X-ray entries followed by that with respect to entries of similar resolution.

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	293/371 (79%)	281 (96%)	12 (4%)	0	100	100
2	B	34/39 (87%)	34 (100%)	0	0	100	100
All	All	327/410 (80%)	315 (96%)	12 (4%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 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 X-ray entries followed by that with respect to entries of similar resolution.

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	263/329 (80%)	251 (95%)	12 (5%)	29	19

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	B	28/33 (85%)	27 (96%)	1 (4%)	38	28
All	All	291/362 (80%)	278 (96%)	13 (4%)	32	20

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	144	LEU
1	A	160	ASN
1	A	165	PHE
1	A	183	PHE
1	A	230	LYS
1	A	258	ASN
1	A	274	PHE
1	A	283	LEU
1	A	297	ASN
1	A	299[A]	GLU
1	A	299[B]	GLU
1	A	308	GLU
2	B	2069	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	187	ASN
1	A	228	ASN
1	A	297	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

## 6 Fit of model and data ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

In the following table, the column labelled ‘#RSRZ > 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q < 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	298/371 (80%)	-0.13	15 (5%) 29 32	13, 20, 53, 88	0
2	B	35/39 (89%)	1.34	9 (25%) 0 0	26, 45, 68, 79	0
All	All	333/410 (81%)	0.03	24 (7%) 15 17	13, 22, 57, 88	0

All (24) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	2036	ILE	9.0
1	A	262	TYR	5.4
1	A	352	GLN	4.8
2	B	2065	GLN	4.7
2	B	2059	PRO	4.7
1	A	160	ASN	4.6
1	A	263	CYS	4.5
1	A	442	PRO	4.5
1	A	102	ARG	4.3
1	A	389	ARG	4.0
1	A	187	ASN	3.4
2	B	2037	SER	3.4
1	A	163	THR	3.4
1	A	439	HIS	3.4
2	B	2067	ALA	3.1
2	B	2070	ALA	3.1
1	A	440	ASN	2.8
1	A	105	TYR	2.5
1	A	106	MET	2.4
1	A	441	PHE	2.3
2	B	2038	GLN	2.3
2	B	2058	PRO	2.2
1	A	390	TYR	2.2
2	B	2064	ILE	2.0

## 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

## 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.