Full wwPDB X-ray Structure Validation Report (i)

May 13, 2020 – 06:34 am BST

PDB ID	:	5NQG
Title	:	Crystal structure of Plasmodium vivax 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	:	2.15 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

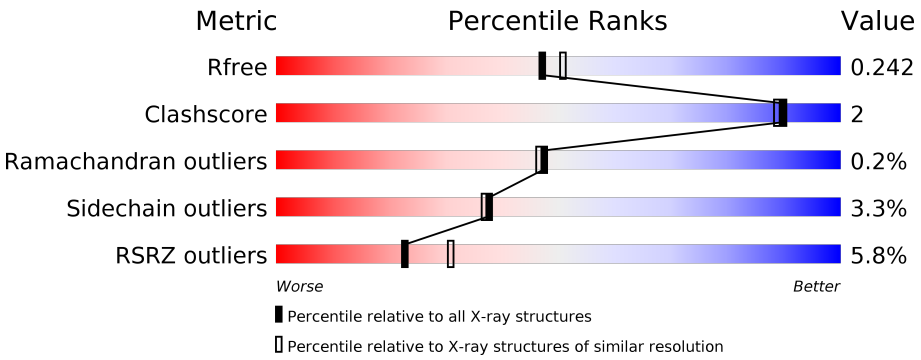
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.15 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}	130704	1479 (2.16-2.16)
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)
RSRZ outliers	127900	1456 (2.16-2.16)

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 >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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% 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	471	4% 75% 6% 19%
2	B	39	8% 82% • 15%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 3558 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 merozoite antigen 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	382	3088	1938	538	586	26	0	3	0

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	39	GLU	-	expression tag	UNP O61130
A	40	ALA	-	expression tag	UNP O61130
A	41	SER	-	expression tag	UNP O61130
A	42	ILE	-	expression tag	UNP O61130
A	119	GLU	GLN	engineered mutation	UNP O61130
A	178	ASN	SER	engineered mutation	UNP O61130
A	226	ASP	ASN	engineered mutation	UNP O61130
A	441	GLN	ASN	engineered mutation	UNP O61130
A	488	LEU	-	expression tag	UNP O61130
A	489	GLU	-	expression tag	UNP O61130
A	490	GLN	-	expression tag	UNP O61130
A	491	LYS	-	expression tag	UNP O61130
A	492	LEU	-	expression tag	UNP O61130
A	493	ILE	-	expression tag	UNP O61130
A	494	SER	-	expression tag	UNP O61130
A	495	GLU	-	expression tag	UNP O61130
A	496	GLU	-	expression tag	UNP O61130
A	497	ASP	-	expression tag	UNP O61130
A	498	LEU	-	expression tag	UNP O61130
A	499	ASN	-	expression tag	UNP O61130
A	500	SER	-	expression tag	UNP O61130
A	501	ALA	-	expression tag	UNP O61130
A	502	VAL	-	expression tag	UNP O61130
A	503	ASP	-	expression tag	UNP O61130
A	504	HIS	-	expression tag	UNP O61130
A	505	HIS	-	expression tag	UNP O61130
A	506	HIS	-	expression tag	UNP O61130

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	507	HIS	-	expression tag	UNP O61130
A	508	HIS	-	expression tag	UNP O61130
A	509	HIS	-	expression tag	UNP O61130

- Molecule 2 is a protein called RON2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	33	236	145	40	48	3	0	1	0

- Molecule 3 is water.

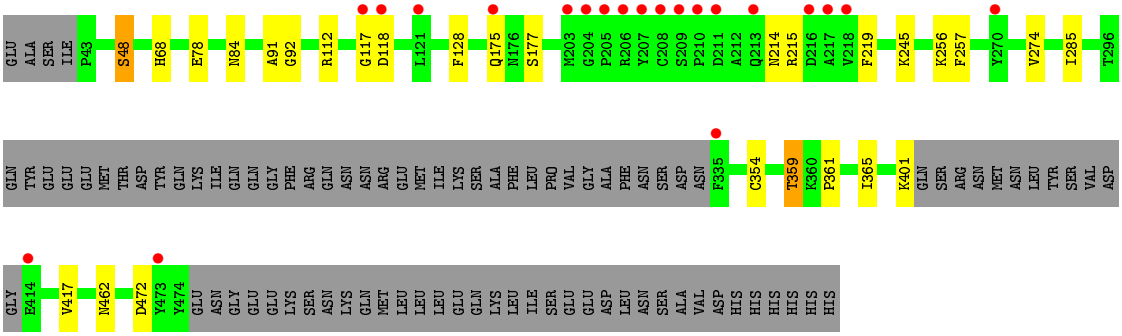
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	223	223	223	0	0
			Total	O		
3	B	11	11	11	0	0

3 Residue-property plots (i)

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.

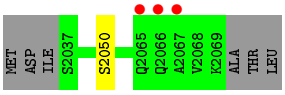
- Molecule 1: Apical merozoite antigen 1

Chain A:



- Molecule 2: RON2

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	164.46Å 54.00Å 62.13Å 90.00° 105.71° 90.00°	Depositor
Resolution (Å)	39.58 – 2.15 39.58 – 2.15	Depositor EDS
% Data completeness (in resolution range)	97.3 (39.58-2.15) 97.6 (39.58-2.15)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.37 (at 2.16Å)	Xtriage
Refinement program	BUSTER 2.10.2	Depositor
R, R_{free}	0.174 , 0.226 0.184 , 0.242	Depositor DCC
R_{free} test set	1641 reflections (5.85%)	wwPDB-VP
Wilson B-factor (Å^2)	35.9	Xtriage
Anisotropy	0.603	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 51.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	3558	wwPDB-VP
Average B, all atoms (Å^2)	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.54% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 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 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	0/3163	0.66	0/4260
2	B	0.46	0/241	0.65	0/332
All	All	0.51	0/3404	0.66	0/4592

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 (i)

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	3088	0	2986	12	0
2	B	236	0	216	0	0
3	A	223	0	0	0	0
3	B	11	0	0	0	0
All	All	3558	0	3202	12	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 (12) 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:274:VAL:HG21	1:A:285:ILE:HG21	1.72	0.71

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:274:VAL:HG23	1:A:354:CYS:HB3	1.78	0.64
1:A:78:GLU:HB2	1:A:91:ALA:HB2	1.84	0.59
1:A:359:THR:HG21	1:A:462:ASN:HD22	1.69	0.57
1:A:68:HIS:CE1	1:A:92:GLY:HA3	2.42	0.55
1:A:175:GLN:C	1:A:177:SER:H	2.11	0.54
1:A:256:LYS:HG2	1:A:365:ILE:HB	1.92	0.51
1:A:48:SER:HB2	1:A:417:VAL:HG12	1.96	0.48
1:A:359:THR:HG21	1:A:462:ASN:ND2	2.30	0.47
1:A:401:LYS:HE3	1:A:417:VAL:H	1.82	0.45
1:A:359:THR:CG2	1:A:462:ASN:HD22	2.30	0.43
1:A:257:PHE:HB3	1:A:361:PRO:HB3	2.04	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	379/471 (80%)	367 (97%)	11 (3%)	1 (0%)	41	37
2	B	32/39 (82%)	31 (97%)	1 (3%)	0	100	100
All	All	411/510 (81%)	398 (97%)	12 (3%)	1 (0%)	47	46

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	117	GLY

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	342/423 (81%)	331 (97%)	11 (3%)	39	38
2	B	26/33 (79%)	25 (96%)	1 (4%)	33	31
All	All	368/456 (81%)	356 (97%)	12 (3%)	38	37

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	48	SER
1	A	84	ASN
1	A	112	ARG
1	A	118	ASP
1	A	128	PHE
1	A	214	ASN
1	A	215	ARG
1	A	219	PHE
1	A	245	LYS
1	A	359	THR
1	A	472	ASP
2	B	2050	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	142	ASN
1	A	445	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates (i)

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 (i)

6.1 Protein, DNA and RNA chains (i)

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, 95th 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^2)	Q<0.9
1	A	382/471 (81%)	0.12	21 (5%)	25	34	29, 43, 81, 114	0
2	B	33/39 (84%)	0.57	3 (9%)	9	13	43, 60, 84, 93	0
All	All	415/510 (81%)	0.15	24 (5%)	23	31	29, 45, 83, 114	0

All (24) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	216	ASP	5.8
1	A	207	TYR	5.1
1	A	206	ARG	4.5
1	A	205	PRO	4.4
1	A	203[A]	MET	4.3
1	A	118	ASP	4.2
1	A	210	PRO	4.1
1	A	217	ALA	3.9
1	A	204	GLY	3.8
1	A	208	CYS	3.7
1	A	117	GLY	3.4
2	B	2065	GLN	3.2
1	A	335	PHE	3.1
1	A	211	ASP	2.9
1	A	121	LEU	2.8
1	A	213	GLN	2.6
1	A	209	SER	2.6
1	A	218	VAL	2.4
1	A	270	TYR	2.4
2	B	2066	GLN	2.4
1	A	414	GLU	2.3
1	A	175	GLN	2.3
1	A	473	TYR	2.2
2	B	2067	ALA	2.1

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.