Tropical Biomedicine 39(3): 394-401 (2022)
https://doi.org/10.47665/tb.39.3.009

RESEARCH ARTICLE

In silico structural modeling and quality assessment of *Plasmodium knowlesi* apical membrane antigen 1 using comparative protein models

Haron, F.N.[1], Azazi, A.[1], Chua, K.H.[2], Lim, Y.A.L.[3], Lee, P.C.[4,5], Chew, C.H.[1*]

[1]Faculty of Health Sciences, Universiti Sultan Zainal Abidin, 21300 Kuala Nerus, Terengganu, Malaysia
[2]Department of Biomedical Science, Faculty of Medicine, Universiti Malaya, 50603 Kuala Lumpur, Malaysia
[3]Department of Parasitology, Faculty of Medicine, Universiti Malaya, 50603 Kuala Lumpur, Malaysia
[4]Faculty of Science and Natural Resources, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia
[5]Biotechnology Research Institute, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia
*Corresponding author: chewch@unisza.edu.my; chinghoong80@yahoo.com

ARTICLE HISTORY

Received: 15 May 2022
Revised: 25 July 2022
Accepted: 26 July 2022
Published: 30 September 2022

ABSTRACT

Plasmodium knowlesi is the most common zoonotic parasite associated with human malaria infection in Malaysia. Apical membrane antigen 1 (AMA1) protein in the parasite plays a critical role in parasite invasion into host cells. To date, there is no complete three-dimensional ectodomain structure of *P. knowlesi* AMA1 (PkAMA1) protein. The knowledge of a protein structure is important to understand the protein molecular functions. Three *in silico* servers with respective structure prediction methods were used in this study, i.e., SWISS-MODEL for homology modeling and Phyre2 for protein threading, which are template-based modeling, while I-TASSER for template-free *ab initio* modeling. Two query sequences were used in the study, i.e., native ectodomain of PkAMA1 strain H protein designated as PkAMA1-H and a modified PkAMA1 (mPkAMA1) protein sequence in adaptation for *Pichia pastoris* expression. The quality of each model was assessed by ProSA-web, QMEAN and SAVES v6.0 (ERRAT, Verify3D and Ramachandran plot) servers. Generated models were then superimposed with two models of *Plasmodium* AMA1 deposited in Protein Data Bank (PDB), i.e., PkAMA1 (4UV6.B) and *Plasmodium vivax* AMA1 (PvAMA1, 1W81) protein structures for similarity assessment, quantified by root-mean-square deviation (RMSD) value. SWISS-MODEL, Phyre2 and I-TASSER server generated two, one and five models, respectively. All models are of good quality according to ProSA-web assessment. Based on the average values of model quality assessment and superimposition, the models that recorded highest values for most parameters were selected as best predicted models, i.e., model 2 for both PkAMA1-H and mPkAMA1 from SWISS-MODEL as well as model 1 of PkAMA1-H and model 3 of mPkAMA1 from I-TASSER. Template-based method is useful if known template is available, but template-free method is more suitable if there is no known available template. Generated models can be used as guidance in further protein study that requires protein structural data, i.e., protein-protein interaction study.

Keywords: *Plasmodium knowlesi*; apical membrane antigen 1 (AMA1); *in silico*; three-dimensional (3D) modeling; model quality assessment.

INTRODUCTION

Malaria is a parasitic disease caused by *Plasmodium* parasite, which is transmitted by *Anopheles* mosquitoes. Around half of the global populations are still affected by malaria (WHO, 2021). *Plasmodium knowlesi* is the most common zoonotic parasite for human malaria infection in Malaysia due to its high incidence where it accounted for most of the national malaria cases (Chew *et al*., 2012; Goh *et al*., 2013; Lee *et al*., 2015; Zaw & Lin, 2019; Cooper *et al*., 2020). Its unique erythrocytic life cycle of 24 hours instead of 48 hours or more in other human *Plasmodium* species would cause high parasitaemia in a short time in a malaria patient and potentially resulting in fatal complications (Rajahram *et al*., 2019; Chin *et al*., 2020). Thus, prompt diagnosis, treatment, and novel antimalarial and/or vaccine are extremely crucial.

The complex *Plasmodium* life cycle possesses various surface proteins in which some of them were widely studied for vaccine and antimalarial drug development (Devine *et al*., 2016; Sirima *et al*., 2017; Blank *et al*., 2020; Molina-Franky *et al*., 2020). One of the well-studied surface proteins is apical membrane antigen 1 (AMA1), a transmembrane protein located in the merozoite stage of the *Plasmodium* parasite and conserved in *Apicomplexan* parasites. The ectodomain of AMA1 is divided into three domains, i.e., domain I (DI), domain II (DII) and domain III (DIII), which is defined by the eight pairs of cysteine-cysteine disulfide bonds (Macraild *et al*., 2011). Studies have shown that merozoite without AMA1 or blocked AMA1 interaction would result in failed merozoite invasion into host erythrocyte (Srinivasan *et al*., 2013; Wang *et al*., 2014; Yang *et al*., 2017), making it an important surface protein for malaria vaccine and treatment studies.

The three-dimensional (3D) structure of AMA1 is important in malaria studies to further understand its molecular functions. Computational or *in silico* modeling for protein structure prediction served as an alternative tool for protein-protein interaction studies instead of the more sophisticated, expensive, and time-consuming 3D structure determination by X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy (Qiu *et al*., 2020). Nowadays, various *in silico* protein structure prediction tools have been developed, either in software or web server platforms due to the high demand for protein-protein and/or protein-peptide interaction studies, such as protein docking (Chew *et al*., 2017).

Currently, there are several *in silico* 3D protein structure prediction tools available either as webserver or downloadable software. Generally, there were three principal methods used for protein structure prediction, i.e., homology modeling, protein threading or fold-recognition modeling, and *ab initio* protein modeling methods. Homology modeling and protein threading were template-based structure prediction methods, while *ab initio* protein modeling depends on algorithmic process based on physical principles (sequence information) rather than on previously solved structures as the template. The present study aims to predict the protein structures of native PkAMA1 for *P. knowlesi* strain H (PkAMA1-H) and a modified PkAMA1 (mPkAMA1) that was designed in adaptation for *Pichia pastoris* expression (Haron *et al*., 2020). The level of accuracy was further compared to the currently available *Plasmodium* AMA1 protein structures deposited in Protein Data Bank (PDB), which is experimentally determined by X-ray crystallography or NMR.

In this study, SWISS-MODEL (Waterhouse *et al*., 2018), Protein Homology/analogY Recognition Engine version 2.0 (Phyre2) (Kelley *et al*., 2015), and Iterative Threading ASSEmbly Refinement (I-TASSER) (Yang & Zhang, 2015) servers were used as the homology modeling, protein threading, and *ab initio* structure prediction methods, respectively. Homology modeling applies the rule that similar sequences from the same evolutionary family would adopt similar tertiary structures with >25% identity. In the protein threading method, however, the target protein was compared with known protein structures instead of the protein sequences, relying on the fact that protein structure is highly conservative in evolution and has a limited number of unique structural folds. In contrast, the *ab initio* method commonly did not rely on known protein structures but rather using an algorithmic approach in building predicted models (Deng *et al*., 2018).

MATERIALS AND METHODS

Protein sequence

The native ectodomain of AMA1 protein sequence of *P. knowlesi* strain H (GenBank accession number: XP_002259339.1) designated as PkAMA1-H was used for protein structure prediction. Additionally, an experimental modified PkAMA1 designated as mPkAMA1 with substitution of five N-glycosylation sites in adaptation for *Pichia pastoris* expression system (Haron *et al*., 2020) was also used for structural prediction. The two PkAMA1 proteins comprise of the entire ectodomain (42-487 amino acid residues), i.e., domain I (DI), domain II (DII) and domain III (DIII) (Figure 1).

```
Apical merozoite antigen 1 [Plasmodium knowlesi strain H]
Sequence ID: XP_002259339.1
Amino acid length: 563

Query  1    GPIIERSIRMSNPWKAFMEKYDLERAHNSGIRIDLGEDAEVGNSKYRIPAGKCPVFGKGI  60
Sbjct  42   GPIIERSIRMSNPWKAFMEKYDLERAHNSGIRIDLGEDAEVGNSKYRIPAGKCPVFGKGI  101

Query  61   VIENSKVSFLTPVATGAQRLKEGGFAFPNADDHISPITIANLKERYKENADLMKLNDIAL  120
Sbjct  102  VIENSNVSFLTPVATGAQRLKEGGFAFPNADDHISPITIANLKERYKENADLMKLNDIAL  161
                N107K
Query  121  CKTHAASFVIAEDQNTNYRHPAVYDEKEKTCYMLYLSAQENMGPRYCSPDSQNKDAMFCF  180
Sbjct  162  CKTHAASFVIAEDQNTSYRHPAVYDEKNKTCYMLYLSAQENMGPRYCSPDSQNKDAMFCF  221
                          S178N     N189E
Query  181  KPDKNEKFDNLVYLSKNVRNDWENKCPRKNLGNAKFGLWVDGNCEEIPYVNEVEARSLRE  240
Sbjct  222  KPDKNEKFDNLVYLSKNVSNDWENKCPRKNLGNAKFGLWVDGNCEEIPYVNEVEARSLRE  281
                          S240R
Query  241  CNRIVFEASASDQPRQYEEELTDYEKIQEGFRQNNRDMIKSAFLPVGAFNSDNFKSKGRG  300
Sbjct  282  CNRIVFEASASDQPRQYEEELTDYEKIQEGFRQNNRDMIKSAFLPVGAFNSDNFKSKGRG  341

Query  301  YNWANFDSVNNKCYIFNTKPTCLINDKNFFATTALSHPQEVDNEFPCSIYKDEIEREIKK  360
Sbjct  342  YNWANFDSVNNKCYIFNTKPTCLINDKNFFATTALSHPQEVDNEFPCSIYKDEIEREIKK  401

Query  361  QSRNMNLYSVDKERIVLPRIFISTDKESIKCPCEPEHISQSTCNFYVCNCVEKRAEIKEN  420
Sbjct  402  QSRNMNLYSVDKERIVLPRIFISTDKESIKCPCEPEHISNSTCNFYVCNCVEKRAEIKEN  461
                                              N441Q
Query  421  NEVIIKEEFKEDYENPDGKHKKKMLL  446
Sbjct  462  NEVIIKEEFKEDYENPDGKHKKKMLL  487

Bold = Amino acid substitution to avoid N-glycosylation (Haron et al., 2020)
Query = Modified PkAMA1 (mPkAMA1)
Sbjct = XP_002259339.1 (PkAMA1-H)
```

Figure 1. NCBI blast alignment of ectodomain of *Plasmodium knowlesi* AMA1 for strain H (PkAMA1-H) and modified PkAMA1 (mPkAMA1) protein sequences. Domain I (DI, amino acid positions of 42-248), domain II (DII, amino acid positions of 249-385) and domain III (DIII, amino acid positions of 386-487) in yellow, cyan, and magenta, respectively.

Protein structure prediction

The 3D protein structure of the entire ectodomain of two PkAMA1 proteins were predicted with three structure prediction servers, i.e., SWISS-MODEL (https://swissmodel.expasy.org/), Phyre2 (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) and I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) servers. By default, the number of model(s) generated by each server depends on each server algorithm. For SWISS-MODEL server, the number of generated model(s) depends on the number of comparative structures in PDB. The SWISS-MODEL server finds PDB templates with BLAST from the user query sequence. Upon BLAST screening, two *Plasmodium* AMA1 protein templates from PDB were chosen by the server, i.e., 1W81 (*Plasmodium vivax* AMA1 encompassing entire ectodomain, DI-DII-DIII) and 4UV6 (PkAMA1 encompassing DI-DII) to generate two homology models. The Phyre2 server generated only one top model from a 100 PDB structures hit, while I-TASSER server generated five models based on five highest C-score, which is highly dependent on the number of simulation decoys, computed by the I-TASSER algorithm to construct each I-TASSER model.

Model quality and similarity assessment

The quality for each protein model was analyzed with Protein Structure Analysis (ProSA-web) (https://prosa.services.came.sbg.ac.at/prosa.php) (Wiederstein & Sippl, 2007), Qualitative Model Energy ANalysis (QMEAN) version 4.3.0 (https://swissmodel.expasy.org/qmean/) (Benkert *et al*., 2011) and SAVES v6.0 (https://saves.mbi.ucla.edu/) (Dym *et al*., 2012) servers. The SAVES v6.0 server includes ERRAT, Verify3D and Ramachandran plot analysis.

For model similarity assessment, the experimentally determined 3D structure of PkAMA1 protein was retrieved from PDB (https://www.rcsb.org/) for superimposition purpose. The PkAMA1 structure from PDB (ID: 4UV6) encompassed only domain I to domain II (DI-II) (Vulliez-Le Normand *et al*., 2015) consisted of two chains with very slight differences, i.e., chain A and chain B, in which chain B (4UV6.B) was used in the present study for structural comparison. Other *Plasmodium* AMA1, i.e., AMA1 of *P. vivax* (PvAMA1, PDB ID: 1W81) (Pizarro *et al*., 2005), which is the only 3D protein structure that encompassed entire ectodomain (DI-DII-DIII) was also retrieved from PDB for model similarity assessment. Superimposed complexes were quantified with root-mean-square deviation (RMSD) value. Model visualization and superimposition were performed with PyMOL 2.4.0 Molecular Graphics System software (Schrodinger, Inc.) (https://pymol.org/2/).

RESULTS

Both PkAMA1-H and mPkAMA1 shared generated models of the same amino acid position by all three *in silico* server (Table 1). The number of amino acids successfully predicted into a protein structure by SWISS-MODEL and Phyre2 depends on the size of the PDB protein template used. For instance, model 1 generated by SWISS-MODEL and the only model generated by Phyre2 were based on PDB template of PvAMA1 (1W81) that encompass DI-II-III, while model 2 from SWISS-MODEL was based on PkAMA1 (4UV6) template that encompass DI-II. Meanwhile, as I-TASSER generates model using *ab initio* method, the whole ectodomain size was predicted into a whole ectodomain protein structure.

The quality of generated models by the three structure prediction servers was evaluated by ProSA-web, QMEAN and SAVES v6.0 servers (Table 2). The ProSA z-score was calculated by ProSA-web while the QMEAN4 score was calculated by QMEAN server. The ERRAT, Verify3D, overall G-factor and percentage of residues in allowed region were calculated by the SAVES v6.0 server, in which the latter two parameters were obtained from Ramachandran plot.

In ProSA-web server, the z-score value was displayed in a plot that contains the z-scores of all experimentally determined proteins (X-ray crystallography and NMR spectroscopy) in PDB. The model z-score obtained within the range of the scores typically found for native proteins of similar size indicates good model quality while a score outside the range indicates erroneous structure (Wiederstein & Sippl, 2007). According to the ProSA-web plot graph, the z-scores for PDB protein structures with 300-500 residues were in the range of -1 to -13. All generated models of PkAMA1-H and mPkAMA1 were in the z-score range of experimentally-determined protein structures, indicating good model quality (Figure 2). In QMEAN server, the score is an estimate of the 'degree of nativeness' and indicates the generated model comparable quality with experimentally determined structures. The resulting score is a QMEAN4 value in which higher score indicates better quality model (Benkert *et al*., 2011). In the current study, models generated by SWISS-MODEL server recorded higher QMEAN4 scores for both PkAMA1-H and mPkAMA1, followed by Phyre2 model and I-TASSER models. SWISS-MODEL model 2 for both PkAMA1-H and mPkAMA1 recorded the highest QMEAN4 score among other generated models.

Model 1 of PkAMA1-H and model 3 of mPkAMA1 predicted by I-TASSER server were the only acceptable models according to ERRAT parameter as ERRAT value of 95% or higher indicates good high-resolution model (Tran *et al*., 2015). Verify3D value of ≥80% indicates good model quality (Singh *et al*., 2019), signifying all generated models were of good model quality according to Verify3D parameter except for model 5 of PkAMA1-H and models 4 and 5 of mPkAMA1 predicted by I-TASSER with values of < 80%. For Ramachandran plot assessment, the overall G-factor of > -0.5 indicates a good model quality (Tran *et al*., 2015). All models were shown as good quality models except for I-TASSER model 4 and 5 of PkAMA1-H and I-TASSER model 5 of mPkAMA1. All models were acceptable based on the Ramachandran map as a value of over 90% in the allowed region indicates good model quality (Singh *et al*., 2019). Based on the average quality assessments, all models were of good quality models except for model 4 and 5 of I-TASSER server for both PkAMA1-H and mPkAMA1. When taken ERRAT parameter into consideration, only I-TASSER model 1 of PkAMA1-H and I-TASSER model 3 of mPkAMA1 were the only models of good quality.

Table 1. Amino acid positions of PkAMA1 ectodomain for each predicted model

PkAMA1-H				mPkAMA1			
In silico server	Model	Amino acid position	Domain	*In silico* server	Model	Amino acid position	Domain
SWISS-MODEL	1	43-474	DI-DII-DIII	SWISS-MODEL	1	43-474	DI-DII-DIII
	2	51-387	DI-DII		2	51-387	DI-DII
Phyre2	1	43-475	DI-DII-DIII	Phyre2	1	43-475	DI-DII-DIII
I-TASSER	1	42-487	DI-DII-DIII	I-TASSER	1	42-487	DI-DII-DIII
	2				2		
	3				3		
	4				4		
	5				5		

Table 2. PkAMA1 models quality assessment by ProSA-web, QMEAN and SAVES v6.0 server

PkAMA1-H							
In silico server	Model	ProSA z-score	QMEAN4 score	SAVES v6.0			
				ERRAT (%)	Verify3D (%)	Overall G-factor	Residues in allowed region (%)
SWISS-MODEL	1	-6.74	-7.34	88.89	80.32	-0.12	100
	2	-7.02	-0.79	89.26	94.05	0.28	100
Phyre2	1	-7.09	-2.43	77.40	82.53	-0.34	98.9
I-TASSER	1	-7.96	-7.34	95.43	86.55	-0.45	99.3
	2	-8.31	-6.02	94.27	86.32	-0.45	97.5
	3	-7.42	-6.16	94.70	80.27	-0.49	97.8
	4	-6.68	-8.52	89.95	80.27	-0.61	97.5
	5	-6.42	-8.04	92.45	72.87	-0.59	97.5
mPkAMA1							
In silico server	Model	ProSA z-score	QMEAN4 score	SAVES v6.0			
				ERRAT (%)	Verify3D (%)	Overall G-factor	Residues in allowed region (%)
SWISS-MODEL	1	-6.71	-2.12	88.46	84.95	-0.11	99.7
	2	-6.88	-0.79	88.99	94.05	-0.15	100
Phyre2	1	-7.00	-2.45	77.40	82.53	-0.34	98.9
I-TASSER	1	-7.98	-5.68	94.75	96.19	-0.46	97.5
	2	-7.35	-7.18	94.04	86.10	-0.49	96.5
	3	-7.81	-6.17	95.42	85.65	-0.46	98.0
	4	-7.38	-6.86	93.61	76.68	-0.50	97.5
	5	-5.69	-9.07	89.96	78.03	-0.69	96.0

Good quality models threshold: ProSA z-score = [-5< *x* < -13], QMEAN4 score = higher score indicates better quality, ERRAT (%) = *x* ≥ 95%, Verify3D (%) = *x* ≥ 80%, overall G-factor = *x* > -0.5 and residues in allowed region (%) = *x* > 90%.

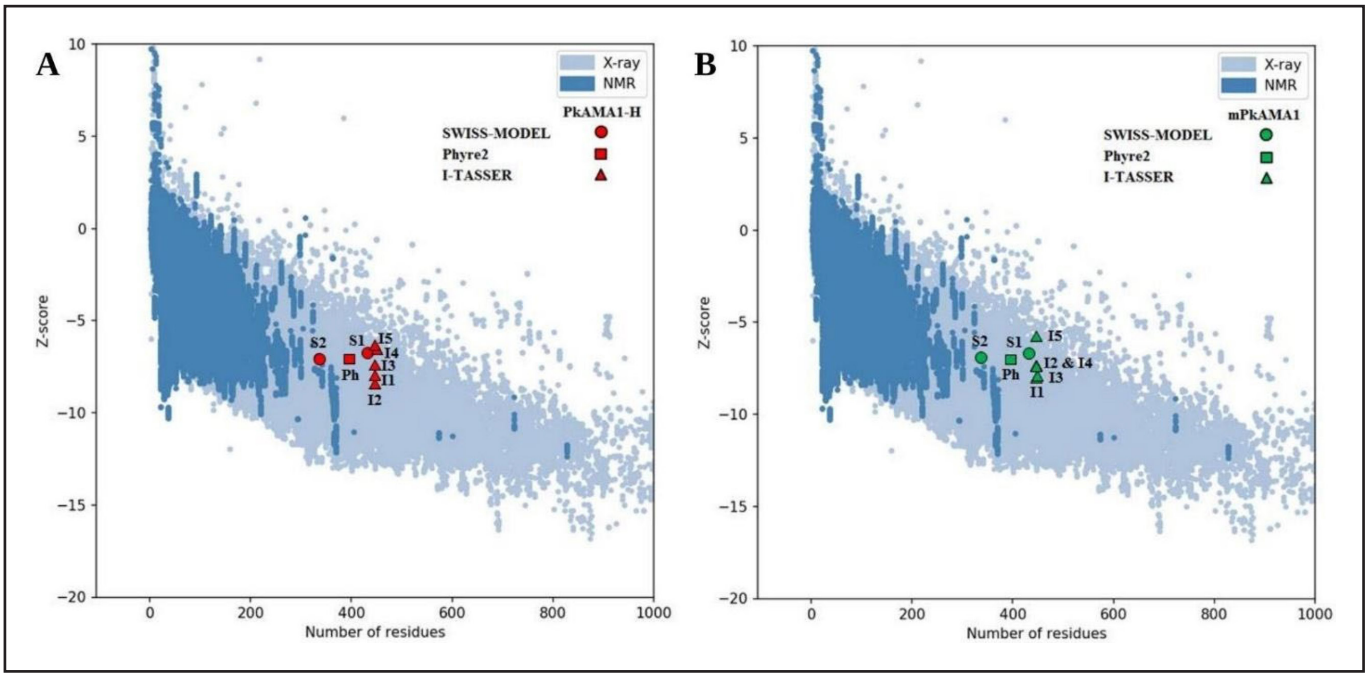

Figure 2. Graph of Protein Data Bank (PDB) protein structures (blue), PkAMA1-H (red) and mPkAMA1 (green) protein structures with z-scores computed by ProSA-web server. PDB structures determined by X-ray crystallography and nuclear magnetic resonance (NMR) were shown in light blue and dark blue, respectively. Models generated by SWISS-MODEL, Phyre2 and I-TASSER server were shown in circle, square and triangle, respectively, with respective model number. Abbreviations, S= SWISS-MODEL, Ph= Phyre2, and I= I-TASSER.

The degree of similarity between the generated model with published PDB AMA1 protein structure was quantified by superimposition in the PyMOL software (Table 3). Low RMSD value indicates the closest similarity between the two superimposed protein structures. In comparison with PDB PkAMA1 (4UV6.B) structure, both model 2 from SWISS-MODEL server which utilized template-based method recorded the lowest RMSD values for both PkAMA1-H and mPkAMA1 as model 2 consisted of partial ectodomain (DI-II) similar to PDB PkAMA1 (4UV6.B) model (Table 1). When compared with PDB PvAMA1 (1W81) structure, both models from Phyre2 server as well as model 1 from SWISS-MODEL server recorded the lowest RMSD values for both PkAMA1-H and mPkAMA1 as all three PkAMA1 predicted models were based on PvAMA1 (1W81), which is the only complete PDB structure of AMA1 ectodomain (DI-II-III) (Table 1 and Table 3). As I-TASSER uses *ab initio* method in generating protein structure, it is expected to record higher RMSD value than both template-based SWISS-MODEL and Phyre2 servers. The generated I-TASSER models also encompass full-length PkAMA1 ectodomain (Table 1). Therefore, the RMSD values by I-TASSER were also taken into account despite the high RMSD values as compared to the other two servers. For I-TASSER server, model 2 of PkAMA1-H and model 3 of mPkAMA1 shared lowest RMSD among other I-TASSER models when superimposed with PDB PkAMA1 (4UV6.B). Meanwhile, model 2 of PkAMA1-H and model 1 of mPkAMA1 shared the lowest RMSD value among other I-TASSER models when compared with PDB PvAMA1 (1W81).

From the overall quality assessment values by ProSA-web, QMEAN, SAVES v6.0 server and RMSD value by superimposition, both SWISS-MODEL model 2 of PkAMA1-H and mPkAMA1 were selected as the best models which are generated by homology modeling as the representative for template-based method. As the I-TASSER models have the most complete amino acids sequence in their ectodomain structures, the best model was also selected from I-TASSER as the representative for template-free method, in which model 1 for PkAMA1-H and model 3 for mPkAMA1 (Figure 3).

DISCUSSION

In the current study, one webserver was chosen for each prediction method, i.e., SWISS-MODEL for homology modeling, Phyre2 for protein threading and I-TASSER for *ab initio* methods. Homology modeling or comparative modeling of protein used experimentally determined protein structures deposited in PDB by X-ray crystallography or NMR spectroscopy to construct the target protein based on the amino acid sequences similarity. SWISS-MODEL was the first fully automated homology modeling server and has been used for 25 years. The server ranked among the top servers in Continuous Automated Model Evaluation (CAMEO), a blind prediction assessment based on sequences from PDB and is one of the most used structure prediction servers globally (Kuhlman & Bradley, 2019). Protein threading is a protein fold-based method that used both experimentally protein sequences and structures as the protein template. Protein threading also makes a prediction based on the structure information. Critical Assessment of Protein Structure Prediction (CASP) is under the management of Protein Structure Prediction Center (https://predictioncenter.org/index.cgi) that conducts experiments to measure the protein structure modeling methods which takes place every two years since 1994. The Phyre2 server is a successor of Phyre server and ranked tenth out of 45 groups in the CASP and the major priority of Phyre2 is the ease of use, especially to researchers without computational knowledge background (Kelley *et al*., 2015). *Ab initio* method refers to the uses of an algorithmic process to predict tertiary structure from the amino acid primary sequence. I-TASSER was built based on iterative fragment assembly simulations and has ranked number one in Protein Structure Prediction Center from CASP8 (2008) to CASP14 (2020) with exception of CASP9 (2010) that ranked in number two. All three *in silico* servers are user-friendly, where the user only need to submit the target protein sequence in FASTA format.

In terms of model quality assessment, the ProSA-web used the potential of mean force (PMF) to locate the region in the generated protein structure of any improper or unsuitable geometries. In ProSA-web, the energy of deposited protein structures was shown in the z-score. Comparing the z-score of generated models with PDB deposited protein structures provides a method to determine the viability of the generated model (Wiederstein & Sippl, 2007). Meanwhile, the QMEAN server estimates the generated model quality by comparing its structural features to experimental structures of similar size deposited in PDB (Benkert *et al*., 2011). The QMEAN server is the top ranking in the CASP13 evaluation for the estimation of model accuracy (EMA) methods performances (Cheng *et al*., 2019; Chen & Siu, 2020). ERRAT determines six nonbonded distance-related interactions, i.e., carbon-carbon (CC), nitrogen-nitrogen (NN), oxygen-oxygen (OO) carbon-oxygen (CO), carbon-nitrogen (CN), and nitrogen-oxygen (NO) atoms that occur in all protein structures. It classifies the distance of the atoms in a proposed structure with statistical analysis. The average and standard deviation of each atom's distance based on known protein structures of various fold classification are used to determine the generated model quality. The Verify3D analysis scores generated protein 3D structure by constructing probability tables. The table assesses the probability of each amino acid residue that would be located in the 3D protein structure. A higher score indicates higher viability of the generated protein model. The geometry factors (G-factors) determine the "normality" of each residue's stereochemical properties based on the analysis of 163 non-homologous protein structures that were determined by X-ray crystallography (Esposito *et al*., 2006).

Table 3. Model superimposition with *Plasmodium* AMA1 from Protein Data Bank (PDB)

In silico server	Model	Superimposition of PkAMA1-H with:		Superimposition of mPkAMA1 with:	
		PkAMA1 (4UV6.B), (Å)	PvAMA1 (1W81), (Å)	PkAMA1 (4UV6.B), (Å)	PvAMA1 (1W81), (Å)
SWISS-MODEL	1	0.538	0.235	0.505	0.237
	2	0.141	0.513	0.142	0.512
Phyre2	1	0.514	0.230	0.519	0.230
I-TASSER	1	0.561	0.461	0.589	0.419
	2	0.526	0.457	0.635	0.536
	3	0.650	0.555	0.417	0.435
	4	0.606	0.477	0.616	0.530
	5	0.627	0.676	0.618	0.424

Å = root-mean-square deviation (RMSD) value.

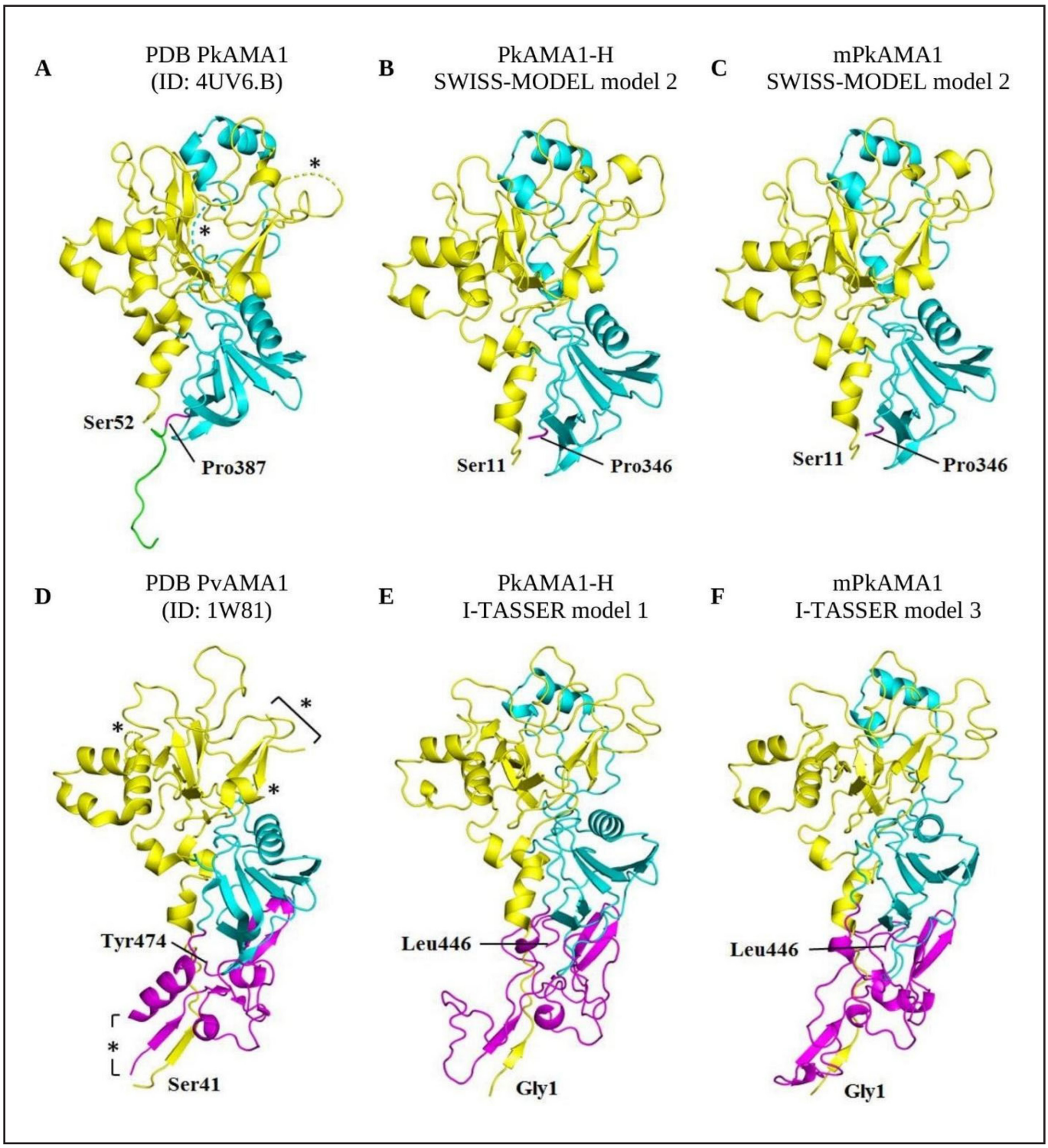

Figure 3. *Plasmodium* AMA1 3D protein structures from Protein Data Bank and the best predicted models of PkAMA1-H and mPkAMA1 protein structures. (A) PDB PkAMA1 (ID: 4UV6.B), (B) SWISS-MODEL model 2 of PkAMA1-H, (C) SWISS-MODEL model 2 of mPkAMA1, (D) PDB PvAMA1 (ID: 1W81), (E) I-TASSER model 1 of PkAMA1-H, and (F) I-TASSER model 3 of mPkAMA1. Domain I (DI), domain II (DII) and domain III (DIII) were shown in yellow, cyan, and magenta, respectively, while green for c-myc tail in PDB PkAMA1 (4UV6.B) structure. The first and last protein residues were labeled accordingly. Asterisk symbol signifies unmodeled loops i.e., Ser212-Ala217 and Gly328-Ser332 in PDB PkAMA1 structure (A) while Met171-Asp174, Pro205-Val218, Pro295-Asn334, and Lys401-Asp412 for PDB PvAMA1 structure (D).

The RMSD is the most basic quantification for deviation measurement between two superimposed protein structures (Kufareva & Abagyan, 2012). As a rule of thumb, using RMSD for the quantification of superimposition complexes is a good measure for predicted model that is close to native model (x <3Å) (Chothia & Lesk, 1986). Generally, a RMSD value of 1.5Å to 2Å indicates detailed atomic accuracy meanwhile RMSD of around 4Å to 6Å indicates models with only correct backbone orientation. RMSD value of more than 6Å is regarded as very poor accuracy models (Shin *et al*., 2017). In current study, predicted models were close to the native AMA1 structures from PDB, which is appropriate to use the basic RMSD measurement and does not need a more complex quantification, i.e., S-score and Global Distance Score (GDT), which is more accurate but only needed when the superimposed complexes have significant difference in their structures (x >3Å) (Kufareva & Abagyan, 2012).

A single amino acid variation could change protein structure significantly (Schaefer & Rost, 2012). Despite only five amino acids differences between the native PkAMA1-H and the modified mPkAMA1, the values obtained from model quality assessments and superimposition were vary, except for models by Phyre2 tool which were consistent for both PkAMA1-H and mPkAMA1. The Phyre2 tool algorithm seems to be less sensitive to any amino acids substitution in the protein sequence, as opposed to SWISS-MODEL and I-TASSER algorithms, which are more sensitive to amino acids substitution as seen in the varying values obtained between PkAMA1-H and mPkAMA1. Based on the findings, amino acids substitution has greatly impacted the model construction by SWISS-MODEL and I-TASSER algorithm, which subsequently impacted the qualities of models.

In terms of predicted model similarity with experimentally determined structures, homology modeling is the most accurate method to construct a protein model when there is known modeling template, however *ab initio* is preferable when known template is not available. Both template-based SWISS-MODEL and Phyre2 tools showed impressive values in superimposition, but did not offer additional information on PkAMA1 structure, in contrast to the template-free I-TASSER which provides full-length PkAMA1 structure. The DIII structure of PkAMA1 has yet to be determined experimentally, thus relying on the DIII of the closely-related PvAMA1 structure as model construction template. The accuracy of the predicted DIII of PkAMA1 will remain vague unless determined experimentally. The PvAMA1 (1W81) is the only available experimentally determined 3D *Plasmodium* AMA1 model that encompassed the entire ectodomain (DI-II-III) to date. Apart from that, the significance of PvAMA1 (1W81) as template model for superimposition study was due to close phylogenetic relationship between *P. vivax* and *P. knowlesi* amongst the *Plasmodium* species. The phylogenetic tree of the *P. knowlesi* strain H showed a close relationship with the PvAMA1 gene (bootstrap value: 100%) in which these two *Plasmodium* species contain high levels of identical amino acid alignment (Herman *et al*., 2018). *In silico* PkAMA1 epitope prediction also indicated that PkAMA1 and PvAMA1 shared high sequence similarity and it is expected to obtain epitopes that shared significant consensus sequence (Azazi *et al*., 2021).

The protein structure prediction can be useful in the study of protein-protein interaction, i.e., ligand binding site determination, which otherwise is laborious when studied experimentally. More *in silico* prediction tools with diversified algorithms representing each prediction method, i.e., homology modeling, protein threading and *ab initio* methods can be utilized to obtain extensive data on the effect of each prediction method on the constructed model quality. The inclusive of more *in silico* tools in a protein structure prediction study will also be beneficial to elucidate the effect of amino acids substitution on the resulting constructed model.

CONCLUSION

Based on the overall values of model quality and similarity assessment, the best predicted models were chosen from SWISS-MODEL model 2 for both PkAMA1-H and mPkAMA1 as template-based method representatives while I-TASSER model 1 of PkAMA1-H as well as model 3 of mPkAMA1 as best template-free method representatives. These generated models can be used as guidance in further protein studies that require protein structural data, i.e., protein-protein interaction study.

Conflict of interests

The author declares that they have no conflict of interests.

ACKNOWLEDGEMENT

This study was supported by the Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme (FRGS) (Project no. FRGS/1/2015/SKK12/UNISZA/02/1).

REFERENCES

Azazi, A., Haron, F.N., Chua, K.H., Lim, Y.A.L., Lee, P.C. & Chew, C.H. (2021). Bioinformatics characterization of *Plasmodium knowlesi* apical membrane antigen 1 (PkAMA1) for multi-epitope vaccine design. *Tropical Biomedicine* **38**: 265-275. https://doi.org/10.47665/tb.38.3.067

Benkert, P., Biasini, M. & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27**: 343-350. https://doi.org/10.1093/bioinformatics/btq662

Blank, A., Fürle, K., Jäschke, A., Mikus, G., Lehmann, M., Hüsing, J., Heiss, K., Giese, T., Carter, D., Böhnlein, E. *et al*. (2020). Immunization with full-length *Plasmodium falciparum* merozoite surface protein 1 is safe and elicits functional cytophilic antibodies in a randomized first-in-human trial. *npj Vaccines* **5**: 10. https://doi.org/10.1038/s41541-020-0160-2

Chen, J. & Siu, S. (2020). Machine Learning Approaches for Quality Assessment of Protein Structures. *Biomolecules* **10**: 626. https://doi.org/10.3390/biom10040626

Cheng, J., Choe, M.H., Elofsson, A., Han, K.S., Hou, J., Maghrabi, A., McGuffin, L.J., Menɪndez-Hurtado, D., Olechnovič, K., Schwede, T. *et al*. (2019). Estimation of model accuracy in CASP13. *Proteins* **87**: 1361-1377. https://doi.org/10.1002/prot.25767

Chew, C.H., Lim, Y.A.L. & Chua, K.H. (2017). Heterologous expression of *Plasmodium vivax* apical membrane antigen 1 (PvAMA1) for binding peptide selection. *PeerJ* **5**: e3794. https://doi.org/10.7717/peerj.3794

Chew, C.H., Lim, Y.A., Lee, P.C., Mahmud, R. & Chua, K.H. (2012). Hexaplex PCR detection system for identification of five human *Plasmodium* species with an internal control. *Journal of Clinical Microbiology* **50**: 4012-4019. https://doi.org/10.1128/JCM.06454-11

Chin, A.Z., Maluda, M.C.M., Jelip, J., Jeffree, M.S., Culleton, R. & Ahmed, K. (2020). Malaria elimination in Malaysia and the rising threat of *Plasmodium knowlesi*. *Journal of Physiological Anthropology* **39**: 36. https://doi.org/10.1186/s40101-020-00247-5

Chothia, C. & Lesk, A. (1986). The relation between the divergence of sequence and structure in proteins. *The EMBO Journal* **5**: 823-826. https://doi.org/10.1002/j.1460-2075.1986.tb04288.x

Cooper, D.J., Rajahram, G.S., William, T., Jelip, J., Mohammad, R., Benedict, J., Alaza, D.D., Malacova, E., Yeo, T.W., Grigg, M.J. *et al*. (2020). *Plasmodium knowlesi* malaria in Sabah, Malaysia, 2015-2017: ongoing increase in incidence despite near-elimination of the human-only *Plasmodium* species. *Clinical Infectious Diseases* **70**: 361-367. https://doi.org/10.1093/cid/ciz237

Deng, H., Jia, Y. & Zhang, Y. (2018). Protein structure prediction, *International Journal of Modern Physics B* **32**: 1840009. https://doi.org/10.1142/S021797921840009X

Devine, S.M., MacRaild, C.A., Norton, R.S. & Scammells, P.J. (2016). Antimalarial drug discovery targeting apical membrane antigen 1. *Medicinal Chemistry Communications* **8**: 13-20. https://doi.org/10.1039/c6md00495d

Dym, O., Eisenberg, D. & Yeates, T.O. (2012). Detection of errors in protein models. In: International Tables for Crystallography Volume F: Crystallography of biological macromolecules, Arnold, E., Himmel, D.M. & Rossmann, M.G. (editors) 2nd edition. Indianapolis: John Wiley & Sons, pp. 677-679.

Esposito, E.X., Tobi, D. & Madura, J.D. (2006). Chapter 2: Comparative Protein Modeling. Reviews in Computational Chemistry. Lipkowitz, K.B., Cundari, T.R., Gillet, V.J. & Boyd, D.B. (editors). John Wiley & Sons. *Reviews in Computational Chemistry* **22**: 57-168. https://doi.org/10.1002/0471780367.ch2

Goh, X.T., Lim, Y.A.L, Vythilingam, I., Chew, C.H., Lee, P.C., Ngui, R., Tan, T.C., Yap, N.J., Nissapatorn, V. & Chua, K.H. (2013). Increased detection of *Plasmodium knowlesi* in Sandakan division, Sabah as revealed by PlasmoNex. *Malaria Journal* **12**: 264. https://doi.org/10.1186/1475-2875-12-264

Haron, F.N., Azazi, A., Chua, K.H., Lim, Y.A.L, Lee, P.C. & Chew, C.H. (2020). Small-scale optimization of recombinant *Plasmodium knowlesi* apical membrane antigen 1 (rPkAMA1) expression in *Pichia pastoris*. *Asian Journal of Medicine and Biomedicine* **4**: 79-90.

Herman, L.S., Fornace, K., Phelan, J., Grigg, M.J., Anstey, N.M., William, T., Moon, R.W., Blackman, M.J., Drakeley, C.J. & Tetteh, K.K. (2018). Identification and validation of a novel panel of *Plasmodium knowlesi* biomarkers of serological exposure. *PLoS Neglected Tropical Diseases*, **12**(6): e0006457. https://doi.org/10.1371/journal.pntd.0006457

Kelley, L., Mezulis, S., Yates, C., Wass, M. & Sternberg, M. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**: 845-858. https://doi.org/10.1038/nprot.2015.053

Kufareva, I. & Abagyan, R. (2012). Methods of protein structure comparison. *Methods in Molecular Biology* **857**: 231-257. https://doi.org/10.1007/978-1-61779-588-6_10

Kuhlman, B. & Bradley, P. (2019). Advances in protein structure prediction and design. *Nature Reviews Molecular Cell Biology* **20**: 681-697. https://doi.org./10.1038/s41580-019-0163-x

Lee, P.C., Chong, E.T.J., Anderios, F., Lim, Y.A., Chew, C.H. & Chua, K.H. (2015). Molecular detection of human *Plasmodium* species in Sabah using PlasmoNex™ multiplex PCR and hydrolysis probes real-time PCR. *Malaria Journal* **14**: 1-7. https://doi.org/10.1186/s12936-015-0542-5

Macraild, C.A., Anders, R.F., Foley, M. & Norton, R.S. (2011). Apical membrane antigen 1 as an anti-malarial drug target. *Current Topics in Medicinal Chemistry* **11**: 2039-2047. https://doi.org/10.2174/156802611796575885

Molina-Franky, J., Cuy-Chaparro, L., Camargo, A., Reyes, C., Gómez, M., Salamanca, D.R., Patarroyo, M.A. & Patarroyo, M.E. (2020). *Plasmodium falciparum* pre-erythrocytic stage vaccine development. *Malaria Journal* **19**: 56. https://doi.org/10.1186/s12936-020-3141-z

Pizarro, J.C., Vulliez-Le Normand, B., Chesne-Seck, M.L., Collins, C.R., Withers-Martinez, C., Hackett, F., Blackman, M.J., Faber, B.W., Remarque, E.J., Kocken, C.H. *et al*. (2005). Crystal structure of the malaria vaccine candidate apical membrane antigen 1. *Science,* **308**: 408-411. https://doi.org/10.1126/science.1107449

Qiu, Y., Li, X., He, X., Pu, J., Zhang, J. & Lu, S. (2020). Computational methods-guided design of modulators targeting protein-protein interactions (PPIs). *European Journal of Medicinal Chemistry* **207**: 112764. https://doi.org/10.1016/j.ejmech.2020.112764

Rajahram, G.S., Cooper, D.J., William, T., Grigg, M.J., Anstey, N.M. & Barber, B.E. (2019). Deaths from *Plasmodium knowlesi* malaria: case series and systematic review. *Clinical Infectious Diseases* **69**: 1703-1711. https://doi.org/10.1093/cid/ciz011

Schaefer, C. & Rost, B. (2012). Predict impact of single amino acid change upon protein structure. *BMC Genomics* **13**: S4. https://doi.org/10.1186/1471-2164-13-S4-S4

Shin, W.H., Kang, X., Zhang, J. & Kihara, D. (2017). Prediction of local quality of protein structure models considering spatial neighbors in graphical models. *Scientific Reports* **7**: 40629. https://doi.org/10.1038/srep40629

Singh, R., Gurao, A., Rajesh, C., Mishra, S.K., Rani, S., Behl, A., Kumar, V. & Kataria, R.S. (2019). Comparative modeling and mutual docking of structurally uncharacterized heat shock protein 70 and heat shock factor-1 proteins in water buffalo. *Veterinary World* **12**: 2036-2045. https://doi.org/10.14202/vetworld.2019.2036-2045

Sirima, S.B., Durier, C., Kara, L., Houard, S., Gansane, A., Loulergue, P., Bahuaud, M., Benhamouda, N., Nebié, I., Faber, B. *et al*. (2017). Safety and immunogenicity of a recombinant *Plasmodium falciparum* AMA1-DiCo malaria vaccine adjuvanted with GLA-SE or Alhydrogel(R) in European and African adults: A phase 1a/1b, randomized, double-blind multi-centre trial. *Vaccine* **35**: 6218-6227. https://doi.org/10.1016/j.vaccine.2017.09.027

Srinivasan, P., Yasgar, A., Luci, D.K., Beatty, W.L., Hu, X., Andersen, J., Narum, D.L., Moch, J. K., Sun, H., Haynes, J.D. *et al*. (2013). Disrupting malaria parasite AMA1-RON2 interaction with a small molecule prevents erythrocyte invasion. *Nature Communications* **4**: 2261. https://doi.org/10.1038/ncomms3261

Tran, N.T., Jakovlic, I. & Wang, W. (2015). *In silico* characterisation, homology modelling and structure-based functional annotation of blunt snout bream (*Megalobrama amblycephala*) Hsp70 and Hsc70 proteins. *Journal of Animal Science and Technology* **57**: 44. https://doi.org/10.1186/s40781-015-0077-x

Vulliez-Le Normand, B., Faber, B.W., Saul, F.A., van der Eijk, M., Thomas, A.W., Singh, B., Kocken, C.H. & Bentley, G.A. (2015). Crystal structure of *Plasmodium knowlesi* apical membrane antigen 1 and its complex with an invasion-inhibitory monoclonal antibody. *PLoS ONE* **10**: e0123567. https://doi.org/10.1371/journal.pone.0123567

Wang, G., MacRaild, C.A., Mohanty, B., Mobli, M., Cowieson, N.P., Anders, R.F., Simpson, J.S., McGowan, S., Norton, R.S. & Scanlon, M.J. (2014). Molecular insights into the interaction between *Plasmodium falciparum* apical membrane antigen 1 and an invasion-inhibitory peptide. *PLoS ONE* **9**: e109674. https://doi.org/10.1371/journal.pone.0109674

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T., Rempfer, C., Bordoli, L. *et al*. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **46**: W296-W303. https://doi.org/10.1093/nar/gky427

Wiederstein, M. & Sippl, M.J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research* **35**: W407-W410. https://doi.org/10.1093/nar/gkm290

WHO (World Health Organization). (2021). World Malaria Report 2021. Issue December. Geneva: World Health Organization, pp. 1-322.

Yang, A., Lopaticki, S., O'Neill, M.T., Erickson, S.M., Douglas, D.N., Kneteman, N.M. & Boddey, J.A. (2017). AMA1 and MAEBL are important for *Plasmodium falciparum* sporozoite infection of the liver. *Cellular Microbiology* **19**: 10. https://doi.org/10.1111/cmi.12745

Yang, J. & Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Research* **43**: W174-W181. https://doi.org/10.1093/nar/gkv342

Zaw, M.T. & Lin, Z. (2019). Human *Plasmodium knowlesi* infections in South-East Asian countries. *Journal of Microbiology, Immunology and Infection* **52**: 679-684. https://doi.org/10.1016/j.jmii.2019.05.012