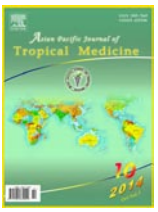
Contents lists available at ScienceDirect

Asian Pacific Journal of Tropical Medicine

journal homepage:www.elsevier.com/locate/apjtm



Document heading doi: 10.1016/S1995-7645(14)60146-2

Conserved regions of *Plasmodium vivax* potential vaccine candidate antigens in Sri Lanka: Conscious *in silico* analysis of prospective conformational epitope regions

Shanika Amarasinghe[1], Hashendra Kathriarachchi[1], Preethi Udagama[2*]

[1]*Department of Plant Sciences, Faculty of Science, University of Colombo, CumarathungaMunidasaMawatha, Colombo 03, Sri Lanka*
[2]*Department of Zoology, Faculty of Science, University of Colombo, CumarathungaMunidasaMawatha, Colombo 03, Sri Lanka*

ARTICLE INFO

Article history:
Received 1 May 2014
Received in revised form 20 June 2014
Accepted 22 August
Available online 20 October 2014

Keywords:
Plasmodium vivax
*Pv*MSP-1_{19}
*Pv*AMA1-D Ⅱ
*Pv*DBP Ⅱ
Conserved regions
Conformational epitopes
Immunogenicity

ABSTRACT

Objective: To do mapping and modeling of conformational B cell epitope regions of highly conserved and protective regions of three merozoitecandidate vaccine proteins of *Plasmodium vivax* (*P. vivax*) , *ie*. merozoite purface protein-1 (*Pv*MSP-1), apical membrane antigen -1 domain Ⅱ (*Pv*AMA1-D Ⅱ) and region Ⅱ of the Duffy binding protein (*Pv*DBP Ⅱ), and to analyze the immunogenic properties of these predicted epitopes. **Methods:** 3-D structures of amino acid haplotypes from Sri Lanka (available in GeneBank) of *Pv*MSP-1_{19} (n=27), *Pv*AMA1-D Ⅱ (n=21) and *Pv*DBP Ⅱ (n=33) were modeled. SEPPA, selected as the best online server was used for conformational epitope predictions, while prediction and modeling of protein structure and properties related to immunogenicity was carried out with Geno3D server, SCRATCH Protein Server, NetSurfP Server and standalonesoftware, Genious 5.4.4. **Results:** SEPPA revealed that regions of predicted conformational epitopes formed 4 clusters in *Pv*MSP-I_{19}, and 3 clusters each in *Pv*AMA1-D Ⅱ and *Pv*DBP Ⅱ, all of which displayed a high degree of hydrophilicity, contained solvent exposed residues, displayed high probability of antigenicity and showed positive antigenic propensity values, that indicated high degree of immunogenicity. **Conclusions:** Findings of this study revealed and confirmed that different parts of the sequences of each of the conserved regions of the three selected potential vaccine candidate antigens of *P. vivax* are important with regard to conformational epitope prediction that warrants further laboratory experimental investigations in in vivo animal models.

1. Introduction

The three main obstacles for eradication of malaria are the steady rise in drug resistant parasite strains, insecticide resistance of the vector mosquito and lack of successful vaccine(s)[1], which emphasize the importance of vaccination strategies for malaria. The multifaceted life cycle of the parasite, and widespread genetic diversity of the different parasite stages are adverse obstacles to successful malaria control[2].

Plasmodium vivax (*P. vivax*) is more important than conventionally thought for reasons that this parasite has a wider geographical range (Asia and America), potentially exposing more people to risk of infection, is less amendable to control, and most importantly, infections with *P. vivax* can cause severe clinical syndromes[3].Additionally, vivax malaria could be as fatal in a similar way to severe *Plasmodium falciparum (P. falciparum)* malaria and is malignant and common[4]. Unlike *P. falciparum*, difficulty of maintaining *P. vivax* in continuous in vitro culture with the exception of short term ex vivo cultures, to conduct laboratory research has lead *P. vivax* to be successfully tested in Non Human Primate models[5]. The occurrence of *P. vivax* relapses is detected in 20%-80% of patients which guarantee that up to 20% of an endemic population may undergo symptomatic infection each year, even in low-transmission areas, which in turn will collectively

*Corresponding author: Prof. Preethi V. Udagama, Department of Zoology, Faculty of Science, University of Colombo, No. 94 CumarathungaMunidasaMawatha, Colombo 03, Sri Lanka, 00300.
Tel.: +94 71 4416050
E-mail: dappvr@yahoo.commailto:dappvr@yahoo.com

result in 10-30 episodes of malaria during the lifetime of an individual[6]. The failure of Primaquine treatment, the only therapeutic option against lethal relapses aggravates this situation[1].

Thus effective, stable malaria control may depend on developing reasonable, general defensive anti-malarial vaccines[7]. Anti-parasite vaccine strategies developed so far are in line with the parasite life cycle, and can be broadly categorized as pre-erythrocytic, blood stage and sexual stage vaccines[8]. An effective vaccine against the erythrocytic stages of the malaria parasite would aim to limit parasite multiplication, thereby reducing morbidity and mortality of drug resistant parasites[9].

Three merozoite proteins have been considered as prime targets for blood stage malaria vaccine(s) since they are exposed to effective immune mechanisms leading to an interruption of the parasite's erythrocytic cycle, namely the merozoite surface protein-1 (MSP-1), the apical membrane Antigen (AMA1) and the Duffy binding protein (DBP)[10].

While lack of availability of an effective vaccine improves the success of continuing the parasite burden, an important step in the fundamental design of a novel vaccine is the meticulous determination of conformational epitopes of neutralizing antibodies. The polymorphism associated with conformational B cell epitopes is important to understand the mechanism of antibody recognition and the reaction of the host immune system, which in turn will obstruct the parasite's ability to elude the host immune response[11].

An accurate immune response could be induced by a correctly chosen mixture of pertinent epitopes focused on conserved and highly immunogenic regions of an antigen or several antigens. This selection can be based on aspects such as immunogenicityand conserved nature of the antigens. Furthermore, epitope vaccines have attracted considerable importance as these present with multiple advantages together with their validity in personalized medicine[12-14].

A research programme conducted by the Department of Zoology, University of Colombo, Sri Lanka, led to the identification of locally and globally conserved regions of three potential *P. vivax* vaccine candidate antigens, *ie*. MSP-1 (*Pv*MSP-1)[15], AMA1-Ⅰ (*Pv*AMA1-DⅠ)[16], and DBPⅡ (*Pv*DBPⅡ)[17]. The respective T cell and linear B cell epitope predictions for these molecules were reported therein.

Though a vast amount of data exist on linear epitopes, as discovery of conformational B-cell epitopes is a challenge, the information on the latter are scarce[11]. X ray crystallography structures and methods applying NMR cross-saturation with TROSY detection have revealed several epitopes on malaria vaccine candidates with respect to *P. falciparum*[18-21]. Conversely, such information for *P. vivax* is lacking.

Epitopes can be discriminated from non-epitope regions of a protein with characteristics related to structural and functional aspects[22], and certain physicochemical properties of amino acids have shown relationship to the locations of linear epitopes within protein sequences[12]. Based on these observations, some of the epitope related characteristics proposed to be studied were hydrophilicity, hydrophobicity, surface accessibility, and antigenicity[23].

The main focus of the current study was the in silico prediction and mapping of conformational B cell epitopes of the highly conserved and protective regions of MSP-1(PvMSP-1_{19}), AMA-1 (PvAMA1-DⅠ) and DBP (PvDBPⅡ)[10], both in local and global *P. vivax* parasite isolates, using bioinformatics tools. Furthermore, the molecular immunogenicity of the predicted epitopes was evaluated in order to augment the abovefindings. The outcome of this study would assist in the rational design of avivax malaria vaccine construct based on these 3 potential asexual vaccine candidate antigens to be tested in immunization trials.

2. Materials and methods

2.1. Haplotype retrieval and extracting conserved regions for the 3 proteins

The different Sri Lankan haplotype sequences uploaded to the GeneBank (http://www.ncbi.nlm.nih.gov/genbank/) were retrieved for PvMSP-1_{19} (n=27), PvAMA1-DⅠ (n=21) and PvDBPⅡ (n=33) via accession numbers reported previously and the conserved regions were identified according to previous studies[15-17].

2.1.1. Modeling 3D structures of the P. vivax haplotypes

Due to lack of 3D structures of given sequences, except for PvMSP-1_{19} which is conserved globally[15]and of which the crystal structure is available on the Protein Data Bank (PDB) under PDB ID 2NPR[24], the homology modeling technique was used to generate the 3D structures of the local isolates of PvAMA1-DⅠ and PvDBPⅡ using the PDB structures 1W81 and 2C6J respectively[25,26], which were required as the inputs to the conformational epitope predicting server Geno3D[27]. Geno3D is an automatic web server (http://geno3d-pbil.ibcp.fr/cgi-bin/geno3d_automat.pl?page=/GENO3D/geno3d_home.html) for protein molecular modeling. Starting with a query of protein sequence, the server identifies a homologous protein through a sequence similarity search using PSI-BLAST for protein sharing less that 95% pair wise identity. The template can be chosen from a special interface provided by the server, which leads to the point where modeling could be initiated. The built 3D protein model using distance geometry, simulated annealing and energy minimization algorithms, was made available with atomic coordinates of each model with best satisfying

spatial restraints[27].

2.1.2. Selection of best server for predicting conformational B cell epitopes

Several freely available online servers were utilized since different servers are based on different properties of an epitope. Thus, CBTOPE[28], EPSVR[29], Epitopia[30], Discotope[31] and SEPPA[32] were used in order to retrieve the most reliable prediction of the conformational B Cell epitopes of these haplotype sequences.

Using already confirmed and published conformational epitope data available in the Conformational Epitope Database (CED)[33], the best server was detected by identifying the Matthews correlation coefficient (MCC), sensitivity (S) and precision (P) calculations of the five servers by the reproduced results with reference to the method followed previously[34]. The previously reported conformational epitope data of *P. falciparum* MSP1 (Wellcome isolate) stored in the CED (CE0120, CE0121, CE0122) were utilized for this purpose[18, 19, 21].

It was apparent that CBTOPE predicted all of the epitope residues of the three conformational epitopes, while SEPPA came second with one missing residue of the CE0121 epitope. Furthermore, epitopia predicted 3, 6, and 8 residues correctly for epitopes CE0120, CE0121 and CE0122 respectively, and thus was placed third among the five servers used.

The calculation of MCC, sensitivity and precision of the servers using the true positives (which are the real epitope residues), true negatives (which are real non epitope residues), false positives (epitopes predicted by the server but which are not actual epitope residues), and false negatives (actual epitope residues which are not predicted by the server), clearly evidenced that of the five tested, SEPPA was the superior server. Though CBTOPE showed the maximum sensitivity of 1, the precision of this server was as low as 0.2644. Conversely, SEPPA showed the second highest sensitivity of 0.928 57, but with high precision of 0.722 22 and the highest MCC as well (0.733 7). Therefore, this undoubtedly was the best server of the five used for conformational epitope prediction. Thus the output of the SEPPA server for predicted conformational epitope clusters were further analyzed for their immunogenic properties.

2.2. Conformational B cell epitope prediction for *PvMSP-1*$_{19}$, *PvAMA1-D* Ⅱ and *PvDBP* Ⅱ

Conserved regions of the three selected proteins were inserted into the SEPPA server with the optimal threshold value for determining the epitope residues of 1.8 for which a sensitivity of 0.568 was received while the specificity was 0.74 on the SEPPA training dataset[32]. However, as the above server outputs did not generate a consistent 3D view of the predicted epitope regions of conformational epitopes, the predictions were then modeled on to the respective molecule (*Pv*MSP-1$_{19}$, *Pv*AMA1-D Ⅱ and *Pv*DBP Ⅱ), using Jmol viewer (http://www.jmol.org/).

2.3. Analysis of immunogenic properties of the predicted epitopes

Two different servers, ANTIGENPro software provided by SCRATCH Protein Predictor and NetSurfP Server[35, 36], and the stand alone software Genious 5.4.4[37], were used to assess the properties related to immunogenicity of predicted epitope regions, *ie*. surface/solvent accessibility, level of hydrophilicity, disulfide bonding, antigenicity and antigenic propensity (a value given for the inclination to be antigenic). Furthermore, Hopp and Woods scale was utilized to check for the property of hydrophilicity[38].

3. Results

3.1. PVMSP-1$_{19}$

3.1.1. Conformational epitope server output for PvMSP-1$_{19}$

SEPPA server predicted four clusters of conformational epitopes (Figure 1a). The 4 clusters were localized in the top left (Cluster *Pv*MSP-1$_{19}$A), top right (Cluster *Pv*MSP-1$_{19}$B), bottom left (Cluster *Pv*MSP-1$_{19}$B), and bottom right (Cluster *Pv*MSP-1$_{19}$D) corners of the molecule. If these clusters were numbered, cluster *Pv*MSP-1$_{19}$A would contain residues 1-4, 7, 9-10, 24-25, *Pv*MSP-1$_{19}$B the 39th residue, *Pv*MSP-1$_{19}$C the 68th residue and the *Pv*MSP-1$_{19}$D with residues from 76th up to the 80th.

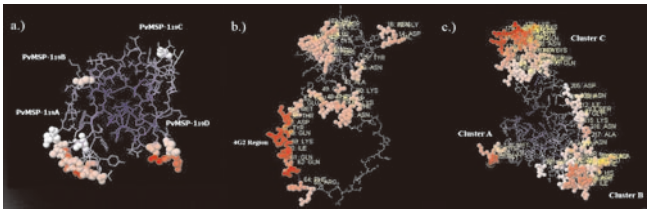


Figure 1. 3D view of molecules.
(a) *Pv*MSP-1$_{19}$ (b) *Pv*AMA1-D Ⅱ (where the 4G2 region is illustrated on the image) and (c) *Pv*DBP Ⅱ with the predicted epitope regions of SEPPA server modeled by Jmol viewer.
The spheres in red: the epitope residues with highest antigenicity (>90%), salmon pink: with 60%-90% antigenicity, light pink: 50% -60% antigenicity and white: antigenicity less than 50%.

3.1.2. Analysis of predicted epitopes for PvMSP-1$_{19}$

The residues of each cluster associated with their antigenicity, hydrophilicity, antigenic propensity and solvent accessibility are summarized in Figure 2, which clearly

depicts that at least one residue of a given cluster is marked as antigenic by Geneious Pro 5.4.4. The hydrophilicity of the regions as predicted by Hopp & Woods hydrophobicity scaleand Geneious Pro 5.4.4, suggest that the hydrophilic residues are greater in number in comparison to the hydrophobic residues (8 hydrophilic residues, 3 neutral residues, and 5 hydrophobic residues). Also, it is evident that the largest epitope region, *Pv*MSP-1_{19}A, contains four residues which are antigenic as predicted by the Geneious 5.4.4 software while the second largest epitope cluster, *Pv*MSP-1_{19}D, is limited to only one residue predicted as antigenic. The solvent accessibility of the residuesas suggested by NetSurfP serverimplies that with the exception of the first Methionine (M) residue of *Pv*MSP-1_{19}. The rest of the residues lie on the surface of the molecule.

Cluster	Residue number	SEPPA prediction	Surface accessibility	Hydrophilicity	Predicted antigenic regions
*Pv*MSP-1_{19}A	1 615	T	E	-0.4	
	1 617	M	B	-1.3	
	1 618	S	E	0.3	
	1 619	S	E	0.3	
	1 622	T	E	-0.4	
	1 624	I	E	-1.8	
	1 625	D	E	3	
	1 639	D	E	3	
	1 640	G	E	0	
*Pv*MSP-1_{19}B	1 654	G	E	0	
*Pv*MSP-1_{19}C	1 686	S	E	0.3	
*Pv*MSP-1_{19}D	1 691	T	E	-0.4	
	1 692	K	E	3	
	1 693	E	E	3	
	1 694	G	E	0	
	1 695	S	E	0.3	

Figure 2. Summary of the protein structure properties of the conformational epitope clusters selected by the SEPPA server for *Pv*MSP-1_{19}.
Column 1: Name of the cluster, Column 2: The number of the residue[24], Column 3: The predicted epitope residues by SEPPA, Column 4: The solvent accessibility of each residue (E: Exposed, B: Buried), Column 5: The hydrophilicity value according to Hopp and Woods calculations; Blue-highest hydrophilic residue ; Purple-neutral ; Red-hydrophobic, Column 6: Antigenic residues according to the Geneious 5.4.4 software colored maroon.

The predicted probability of antigenicity of 0.749 932 by the ANTIGENPro server provided by SCRATCH protein predictor (http://scratch.proteomics.ics.uci.edu) indicates high probability of *Pv*MSP-1_{19} being antigenic. The scores associated with the predicted epitope residues resulting from the SEPPA server showed that the *Pv*MSP-1_{19}A and D clusters had the highest and the second highest antigenic propensity values, respectively. This fact reiterated the antigenic propensity values calculated by the SCRATCH Protein Predictor using the ANTIGENPro software.

An attempt was made to align the double epidermal growth factor (EGF) like Domains of *P. vivax* according to the *P. falciparum* EGF factor like Domain sequence downloaded from Protein Data Bank (http://www.pdb.org/pdb/home/home.do) (PDB ID: 1CEJ). Figure 3 shows the results of the alignment, which confirms a previous report by having the conserved positions for the cysteins of the double EGF like domains in *P. vivax* aligned with those of *P. falciparum* EGF like domain sequences[39].

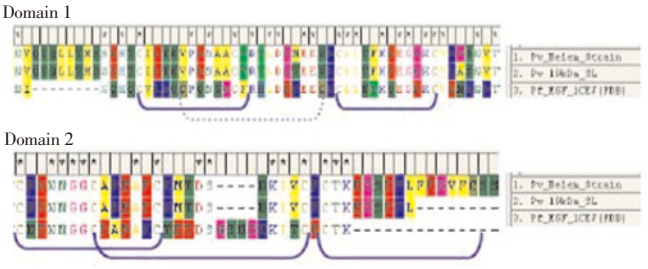


Figure 3. Alignment of the amino acid sequences of the double EGF like Domains of *P. vivax* Salvador/Belem strain MSP-1 19 kDa region and *P. falciparum* MSP-1 (Canonical cysteine residues and prominent inter domain contact residues are colored).

The results show that the EGF like Domain 1 (MSSEHTCID TNVPDNAACYRYLDGMEEWRCLLTFKEEGGKCVPGSNVT) is detected on the sequence of *Pv*MSP-1_{19} with the cystein positions at 7, 18, 30, 41 (with 1 615 residue in the reference salvador strain as 1st residue of *Pv*MSP-1_{19}) which aligns with the exact locationswhere it is on *P. falciparum* MAD20 EGF domain. EGF like domain 2 (CKDNNGGCAPEAECKMTD SNKIVCKCTKEGSEPL) is found on the sequence of *Pv*MSP-1_{19} where the conserved cystein residues lie in the 49, 56, 62, 72, 74 positions.

3.2. PvAMA1-D Ⅱ

3.2.1. Conformational epitope server output for PvAMA1-D Ⅱ

In order to predict the conformational epitopes of *Pv*AMA-1 domain Ⅱ, both domains Ⅰ and Ⅱ were entered to the structure preparation of the Sri Lankan haplotypes. SEPPA server predicted 3 clusters of conformational epitopes of *Pv*AMA1-Ⅱ (Figure 1b). The numbering of the amino acids is according to work of Coley and co workers where the 48th residue of reference Salvador Strain AMA protein was numbered as 1st in our study[37]. The summary of the solvent accessibility, antigenicity and hydrophilicity of the predicted epitopes are displayed in Figure 4.

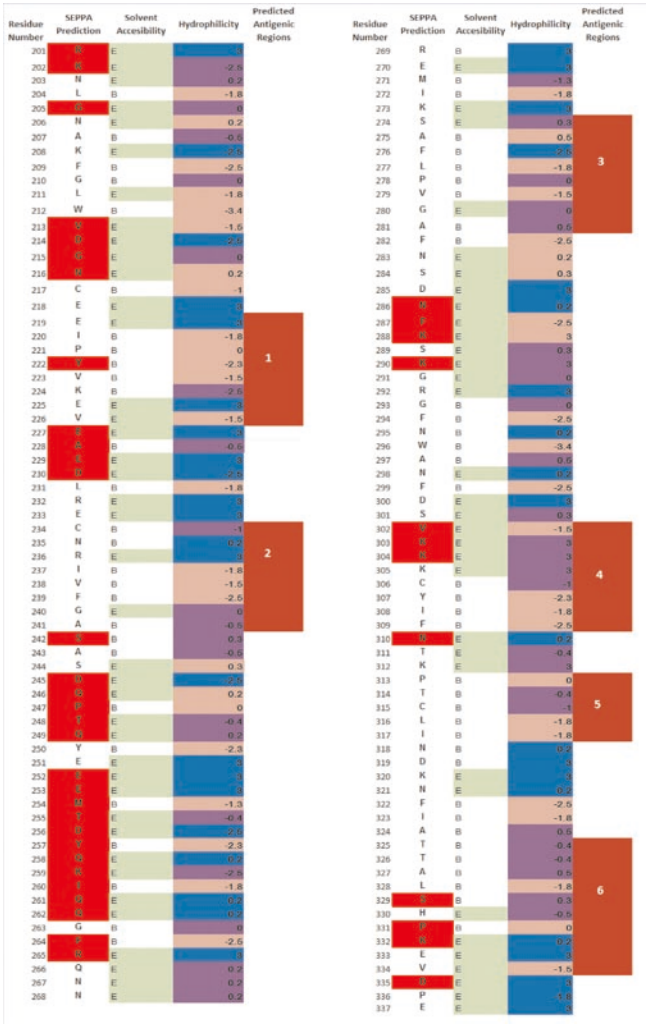


Figure 4. Summary on the protein analysis of the conformational epitope clusters selected by the SEPPA server for *Pv*AMA1-D Ⅱ. Column 1: The number of the residue (according to Ramarque *et al*, 2007), Column 2: The predicted epitope residues by SEPPA server (colored red), Column 3: The solvent accessibility of each residue (E: Exposed, B: Buried), Column 4: The hydrophilicity value according to Hopp and Woods calculations; Blue-highest hydrophilic residue; Purple-neutral ; Red-hydrophobic, Column 5: The antigenic residues according to Geneious 5.4.4 software colored maroon and labeled 1-6.

3.2.2. Analysis of predicted epitopes for PvAMA1-D Ⅱ

SEPPA server predicted 43 residues as epitope residues out of 137 residues of *Pv*AMA1-Ⅱ. Geneious Pro 5.4.4 and the Hopp and Woods hydrophobicity scale showed that more than 50% of the residues of this molecule were hydrophilic which suggests that the molecule is solvent exposed in many areas (Figure 4). NetSurfP server predictions on solvent accessibility reiterated these results by showing exposed residues on all of the other predicted epitope residues except 7 residues. The predicted probability of antigenicity by the SCRATCH protein predictor was 0.863 770, indicative of a high probability of the molecule containing antigenic residues. Geneious Pro 5.4.4 predicted 6 clusters of antigenic residues (Antigenicity column of Figure 4). Unlike the upshot for *Pv*MSP-1_{19} where antigenic regions were predicted in all the 4 clusters, for *Pv*AMA1-D Ⅱ the predicted antigenic regions were not satisfactorily aligned with the predicted epitope regions. In fact, the 2nd and 3rd regions predicted as antigenic by the Geneious Pro 5.4.4 software liedadjacent to the conformational epitope region of 4G2 region predicted by SEPPA (Figure 4). The NetSurfP server also predictedthe pH value at isoelectric point and the net charge of the molecule to be 5.33 and -2.1, respectively. The secondary structures predicted by PsiPred suggested that the 4G2 region mainly contained helices and coils.

SCRATCH Protein predictor also revealed that two disulfide bonds could be formed between the 218th and the 235th residues, and 307th and 316th residues, the latter having a higher probability than the first to form a disulfide bond. These areas in fact lie in the zone predicted by Geneious Pro 5.4.4 software to contain the 1st, 2nd, 4th 5th and 6th antigenic regions.

3.3. PvDBP Ⅱ

3.3.1. Conformational epitope server output for PvDBP Ⅱ

As predicted by SEPPA server, 3 epitope clusters were found on the *Pv*DBP Ⅱ (Figure 1c) which were named Clusters A, B and C. The net charge at pH 7 was 11, and the isoelectric point for this molecule was 8.9.

3.3.2. Analysis of predicted epitopes for PvDBP Ⅱ

In cluster A, all residues were hydrophilic with the exception of four residues (Figure 5). Cluster B did not surprisingly show a clear bias towards hydrophilic residues. Nevertheless it is random and the hydrophilicity values were also not strong. Comparatively, cluster C showedhighly hydrophilic residues in the predicted epitope regions (Figure 5). The overall hydrophilicity as predicted by Hopp Woods calculator was only 0.3 and a mere 49% of all the residues of the sequence were hydrophilic.

Residues 250, 253-254, 257, 258 of cluster A were predicted antigenic by the Geneious 5.4.4 software (Figure 5), but the short region from residue 169 to 171 of the same cluster was not predicted as antigenic. However, the regions from 161-167 and 173-190 which were not included in the epitope region of cluster A, were found to be antigenic. A significant part of Cluster B (from residue 51, 52, 54, 56-65, 136, 138-155) was not scored as antigenic. Except for the center residues (191, 192) of cluster C, the residues from 206th to 240th were predicted as strongly antigenic. This may have a relationship with the hydrophobicity of each cluster as it is evident that cluster B showed a weak hydrophilicity and an antigenicity compared with the other 2 clusters.

There are residues that have been predicted as exposed to solvents as well as buried within the molecule. For each cluster, it is clearly visible that exposed residues are more in number compared with the buried ones (Figure 5). Unlike the other two protein molecules, further investigation of

the *Pv*DBP Ⅱ showed that there could be glycosylation sites (residues) in the sequence at positions 54, 151 and 220 (Figure 6).

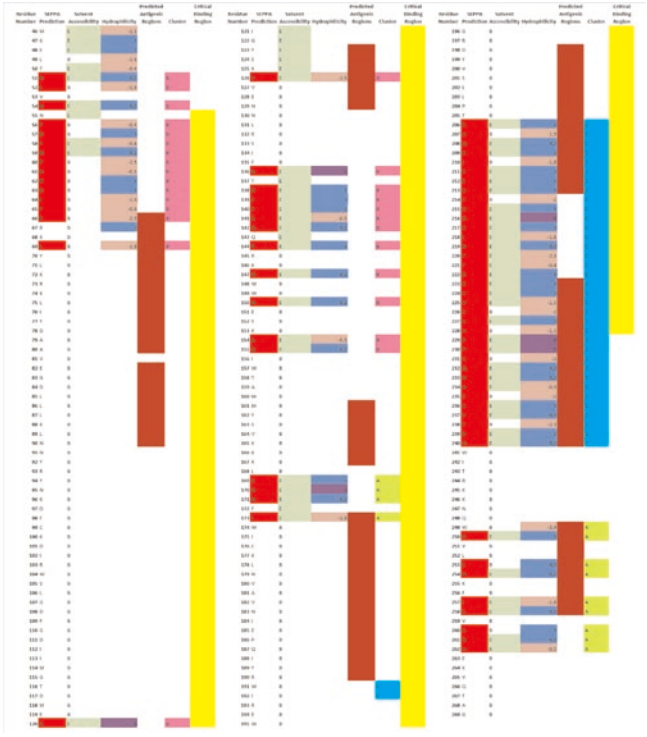


Figure 5. Summary of the protein analysis of the conformational epitope clusters selected by the SEPPA server for *Pv*DBP Ⅱ.
Column 1: The number of the residue (according to Singh *et al*, 2006), Column 2: The predicted epitope residues by SEPPA server (colored red), Column 3: The solvent accessibility of each residue (E: Exposed, B: Buried), Column 4: The hydrophilicity value according to Hopp and Woods calculations, Blue-highest hydrophilic residue; Purple-neutral; Red-hydrophobic, Column 5: The antigenic residues according to Geneious 5.4.4 software colored maroon, Column 6: The Cluster (A, B or C) to which the epitope residues belong, Column 7: The residues in the critical binding motif.

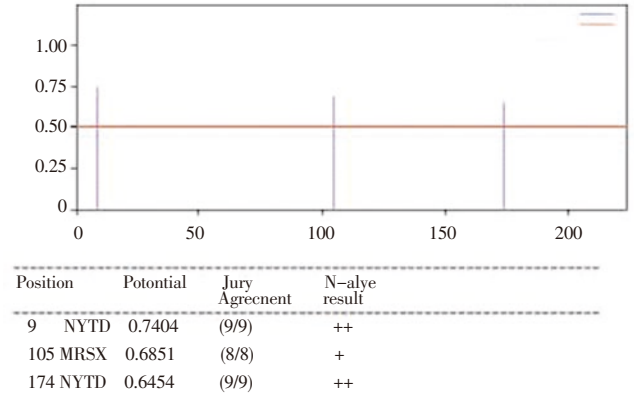


Figure 6. Predicted *N*-Glycosylation sites of *Pv*DBP Ⅱ.
The residue numbers are given as 1st position for the 46th of the sequence of Singh *et al*[26].

The critical binding motif (CBM) of *Pv*DBP Ⅱ consists of 173 residues, encompassing the 55th residue up to the 228th residue. There are 3 disulfide bonds projected on the sequence according to work by Singh and coworkers[38], the first from the 100th cysteine residue that extends to the 227th cysteine residue, the second from the 178th – 232th cysteine residue, and the third from the 215th to the 236th cysteine residue. This is predicted by the Disulfide Bonding State and Cysteine Connectivity Prediction Server), offered by the PredictProtein Server (http://www.predictprotein.org/). The CBM continues to the sections except for the 51-54 region in Cluster B, 229-240 region in cluster C and 249-262 region of the Cluster A. This may perhaps provide more information on the importance of the CBM for the binding of the parasite to the erythrocyte, where the epitopes must play a crucial role in identifying the proteins on the surface of the erythrocyte.

4. Discussion

Antigen-specific epitope-based vaccines supported by new strategies to design novel advances using immunoinformatics and other computational techniques, combined with the expansive adaptability in the devise and synthesis of genetic (DNA) vaccines seem to be the answer to many diseases caused by pathogens refractive to conventional vaccines[44]. The in silico prediction of epitopes can significantly narrow down the list of candidate epitopes based upon the increasing number of bioinformatics tools and antigen sequences available in freely accessible public databases for detecting pathogen peptides[44]. Such a selected set of candidate vaccines could initiate molecular laboratory experiments to authenticate the vaccine targets based on the biological function of the selected antigen sequences[44].

There is evolving confirmation on the importance of conformational epitopes in the immune response. Therefore, characterization of surface proteins in their indigenous conformations is critical in the development of new inventions for therapeutic and vaccine targets[45]. Most biologically important epitopes are habitually uncovered to interact with the external environment, and studies show that classification of intact protein complexes is a decisive step in the exploration for prospective vaccine candidates[46]. While most hydrophilic residues that interact with solvent lie on the surface of the molecule, the most hydrophobic ones lie in the interior[38]. Thus antigenic determinants that lie on the surface of the molecules that interact with the other proteins as well as a solvent (the biological medium is mostly aqueous) consist of charged hydrophilic residues[38]. This opens up an opportunity for investigating the possibility that at least some antigenic determinants may be associated with stretches of amino acid sequences that contain a large number of charged and polar residues and are lacking in

large hydrophobic residues[38].

Compilation of a conformational epitope data set from 76 X-ray structures of complexes between antibodies and protein antigens and analyzing the data confirmed that most epitopes are composed by diminutive fractions of the antigen sequence which facilitate the binding of conformational antibodies[46].

Finding the conformational epitopes are important as previous studies have shown that hyper immunization of animals with intact MSP-1 or the *C*-terminal 42-kDa fragments primarily induced antibodies to conserved B cell epitopes at the *C*-terminal 19-kDa region, MSP-119[47]. In the current study, conformational epitope prediction from SEPPA server gave rise to 4 epitope clusters on the *Pv*MSP-1_{19} molecule that were localized in the top left, top right, bottom left and bottom right areas. These areas were strictly compatible with the regions discovered by NMR cross-saturation with transverse relaxation optimized spectroscopy detection and X Ray Diffraction[18,19] from MSP-1 of *P. falciparum*. The surface localization of these epitopes may be indicative of the role of these epitopes, *ie*. that of binding the merozoite molecule to the receptors on the surface of the host erythrocyte.

It may be plausible to infer that the epitope cluster on the bottom surface of the *Pv*MSP-1_{19} molecule showing high antigenicity and encompassing a large surface area may indeed be an important epitope. It was revealed that these predicted conformational epitopes did not fall on the previously predicted linear epitope regions[15], but that the two residues of the first cluster did overlap with a sequence of the previously predicted T cell epitope region, only common for the DRB_1010 allele. The second largest conformational epitope cluster located at the bottom of the molecule overlap with a linear B cell epitope region[15] but was outlying the previously predicted most common T cell epitope region (WRCLLTFKE sequence fragment with a promiscuous binder of W[4] beginning from 1 644, with reference to the Sal-1 strain).

The double EGF domains of PfMSP-1 are considered a vital target of protective antibodies both in individuals naturally exposed, and in vaccinated and experimentally infected animals[47]. Consequently, the MSP-1_{19} EGF domains incorporated recombinant proteins are leading candidates for inclusion in a vaccine for the control of the most important cause of human malaria, *P. falciparum*[39]. Furthermore, it was previously recorded that EGF-like domains typically function in binding, and same has been assumed as the role for those at the C terminus of MSP-1. It has been demonstrated that the EGF like domains is conserved across divergent *Plasmodium* species[39]. Hence, if the MSP-1 double EGF like domains may bind efficiently to receptor orthologs from broad host origins, disregard of the *Plasmodium* species with the mediation of an essential interaction with the erythrocyte surface[39]. It was evident that the EGF like domains were conserved on the *Pv*MSP-1_{19} segment and therefore important from a vaccine design point of view for a universal vaccine candidate construct since this may indicate that the nature of the function(s) of the EGF like modules in *P. falciparum* and *P. vivax*are indeed similar. Further, if assessment of proportion and probability of the predicted epitopes being located in the EGF like domain regions of *Pv*MSP-1_{19} could be assessed experimentally, it may reveal the conservation of the epitopes globally since the conservation of the positions of the cysteine residues suggests that the conformation of the MSP-1_{19} is indeed similar, even though the sequences of the molecule differ in different *Plasmodium* species[18, 19]. The conservation of the 3D structure may imply the functional similarity of the MSP-119 sequence across the Genus *Plasmodium*[39].

It is noteworthy that with *Pv*AMA1-D Ⅱ, examining only the domain Ⅱ sequence was inadequate to obtain an accurate judgment of the functionality of this domain[40]. The input of both domains Ⅰ and Ⅱ of *Pv*AMA-1 was required to obtain an accurate output, which was substantiated by examining the previously reported conformational epitope region of 4G2 of *P. falciparum*, in the given outputs[20] *ie*. positioned within Loop Ⅱ of domain Ⅱ. The SEPPA server showed high antigenicity of the region analogous to the 4G2 conformational epitope. The residue positions for the 4G2 region were 298,299, 304-315, 332, and 335-336[40]. There were two more clusters also predicted by this server describing the potential of having conformational epitopes nevertheless were antigenically low in strength than the 4G2 region.

The two disulfide bonds present in the *Pv*AMA D-Ⅱ most probably would facilitate bonding between and strengthen the bond by forming the hydrophobic core between the two proteins which may indicate that without forming two more epitope clusters as mentioned[41,42], the 4G2 epitope is being exposed (which lies between the 2nd and 3rd antigenic regions predicted by Geneious Pro 5.4.4 software) and stabilized by the rest of the amino acid chain by these disulfide bonds.

Nevertheless, previously identified residues on the rest of the sequence that may be potential epitope regions, warrant investigations in an experimental model. The conformational epitope regions however do not fall within the previously predicted T cell epitope regions but on a linear B cell epitope region spanning 40-59 and 77-97 (4G2 region spans region 43-83)[16].

The predicted conformational B cell epitope regions of the *Pv*DBP Ⅱ molecule could be divided mainly into 3 clusters. The clusters were formed by the random arrangement of the primary sequence on the 3D arrangement of the molecule. Based upon previous predictions, the two T cell epitope regions lie side by side with the two Linear B cell epitope regions[17]. However, though the conformational B cell epitope regions did not superimpose on the exact

residues of the T cell epitope regions, it was clear that both these epitope types were on the same region of the molecule. Several of the seven linear B cell epitope regions previously predicted, showed partial overlapping with the conformational epitope regions.

Glycosylation is an important post transcriptional modification related to physical properties of a given protein such as folding, trafficking, packing, stabilization, protease protection, quaternary structure and organization of water structure which in turn aid the recognition and biological triggering[43]. These residues are distributed, two in cluster B and one in cluster Cof *Pv*DBP Ⅱ. Nevertheless, there are no signal peptides located and therefore the motif for glycosylation may not function because protein without signal peptides are unlikely to be exposed to the *N*-glycosylation machinery in vivo[43].

A technical constraint of the current study was the inability to use all of the online available servers for epitope prediction and for protein structure evaluation, for the reason that all were not freely available. The other limitation was that only a selected number of servers were used within the time limitation of this study. The main drawback at the beginning of the study, which we successfully overcame, was the unavailability of a 3D model on Sri Lankan haplotypes for the proteins of *Pv*AMA1-D Ⅱ and *Pv*DBP Ⅱ. These had to be generated using Geno3D server to model the molecules and selecting the most probable structures using the Ramachandran Plot analysis[48].

Furthermore, if a similar analysis was conducted of global *P. vivax* sequences of the three selected parasite proteins, the association between local and global isolates in order to fathom global conservation of the predicted epitope regions would have been established.

The analysis and findings of the current study revealed that the different parts of the sequences of each of the three selected potential vaccine candidates are indeed important in conformational epitope prediction. It is plausible that the use of only a part of a sequence may not produce the exact response which the whole molecule would otherwise produce. This needs to be experimentally investigated by using the entire molecule and the predicted regions of the molecule in a non-human primate model to carry out comparative immunity studies. The inclusion of the linear epitope regions in concert with the conformational epitopes on to a virtual model and testing of the 3D arrangement can be further analyzed for the selected epitopes. This study further insinuates that a cocktail of the three potential vaccine candidates would be much more promising to raise a better response than the use of a single vaccine candidate. Future studies warrant the rational design of a malaria vaccine construct based on the homologous sequences of the 3 potential vaccine candidate antigens of *Plasmodium cynomolgi*, the simian counterpart of *P. vivax*, to be tested in the *Plasmodium cynomolgi*-Macacasinica simian model.

Conflict of interest statement

The authors declare to have no competing interests.

References

[1] Restrepo-Montoya D, Becerra D, Carvajal-Patiño JG, Mongui A, Niño LF, et al. Identification of *Plasmodium vivax* proteins with potential role in invasion using sequence redundancy reduction and profile hidden markov models. *PLoS ONE* 2011; **6**(10): e25189. doi:10.1371/journal.pone.0025189.

[2] World Health Organization. *World malaria report*. Geneva, Switzerland: World Health Organization; 2010.

[3] Guerra CA, Howes RE, Patil AP, Gething PW, Van Boeckel TP, Temperley WH, et al. The international limits and population at risk of *Plasmodium vivax* transmission in 2009. *PLoS Negl Trop Dis* 2010; **4**(8): e774. doi:10.1371/journal.pntd.0000774.

[4] Baird JK. Neglect of *Plasmodium vivax* malaria. *Trends Parasitol* 2007; **23**: 533-539.

[5] Galinski MR, Meyer EV, Barnwell JW. *Plasmodium vivax*: modern strategies to study a persistent parasite's life cycle. *Adv Parasitol* 2013; **81**: 1-26.

[6] Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, Anstey NM. Vivax malaria: neglected and not benign. *Am J Trop Med Hyg* 2007; **77**: 79-87.

[7] Gardiner DL, Dixon MW, Spielmann T, Skinner-Adams TS, Hawthorne PL. Implication of a *Plasmodium falciparum* gene in the switch between asexual reproduction and gemetocytogenesis. *Mol Biochem Parasitol* 2005; **140**: 153-160.

[8] Nussenzweig RS, Long CA. Malaria vaccines: Multiple targets. *Sci New Series* 1994; **265**(5177): 1381-1383.

[9] Alan F. Cowman, Deborah L. Functional analysis of *Plasmodium falciparum* merozoite antigens: Implications for erythrocyte invasion and vaccine development. *Philosophical Transactions: Biol Sci* 2002; **357**(1417): 25-33.

[10] Galinski MR, Barnwell JW. *Plasmodium vivax*: who cares? *Malar J* 2008; **11**: 7.

[11] Ansari HR, Raghava GPS. Identification of conformational B-cell epitopes in an antigen from its primary sequence. *Immunome Res* 2012; **6**: 6.

[12] EL-Manzalawy Y, Honavar V. Recent advances in B-cell epitope prediction methods. *Immunome Res* 2010; **6**(Suppl2): S2.

[13] Geourjon C, Combet C. Identification of related proteins with weak sequence identity using secondary structure information. *Protein Sci* 2001; **10**: 788-797.

[14] Toussaint NC, Kohlbacher O. OptiTope-a web server for the selection of an optimal set of peptides for epitope-based vaccines. *Nucleic Acids Res* 2009; **37**(Web Server issue): W617-W622.

[15] Dias S, Longacre S. Genetic diversity and recombination at

the C-terminal fragment of the merozoite surface protein-1 of *Plasmodium vivax* (*Pv*MSP-1) in Sri Lanka. *Infect Genet Evol* 2011a; **11**(1): 145-156.

[16]Dias S, Somarathna M. Evaluation of the genetic diversity of domain Ⅱ of *Plasmodium vivax* apical membrane antigen 1 (*Pv*AMA-1) and the ensuing strain-specific immune responses in patients from Sri Lanka. *Vaccine* 2011b; **29**(43): 7491-7504.

[17]Premaratne PH, Aravinda BR. Genetic diversity of *Plasmodium vivax* duffy binding protein Ⅱ (*Pv*DBP Ⅱ) under unstable transmission and low intensity malaria in Sri Lanka. *Infect Genet Evol* 2011; **11**(6): 1327-1339.

[18]Pizarro JC, Chitarra V. Crystal structure of a Fab complex formed with *Pf*MSP1-19, the *C*-terminal fragment of merozoite surface protein 1 from *Plasmodium falciparum*: a malaria vaccine candidate. *J Mol Biol* 2003; **328**(5): 1091-1093.

[19]Morgan WD, Lock MJ. Malaria parasite-inhibitory antibody epitopes on *Plasmodium falciparum* merozoite surface protein-1(19) mapped by TROSY NMR. *Mol Biochem Parasitol* 2004; **138**(1): 29-36.

[20]Pizarro JC, Le Normandn V. Crystal structure of the malaria vaccine candidate apical membrane Antigen 1. *Science* 2005; **308**: 411.

[21]Morgan WD, Frenkiel TA, Lock MJ, Grainger M, Holder AA. Precise epitope mapping of malaria parasite inhibitory antibodies by TROSY NMR cross-saturation. *Biochemistry* 2005; **18**(2): 518-523.

[22]Ambroise J, Giard J, Gala JL, Macq B. Identification of relevant properties for epitopes detection using a regression model. *IEEE/ACM Trans Comput Biol Bioinform* 2011; **8**(6):1700-1707. doi: 10.1109/TCBB.2011.77.

[23]Ponomarenko JV, van Regenmortel MHV. Gu J, Philip E (eds) . Bourne B-Cell epitope prediction, structural bioinformatics. 2nd ed. Hoboken: John Wiley & Sons, Inc.; 2009

[24]Babon JJ. Structural studies on *Plasmodium vivax* merozoite surface protein-1. *Mol Biochem Parasitol* 2007; **153**(1): 31-40.

[25]Coley AM. Structure of the malaria antigen AMA1 in complex with a growth-inhibitory antibody. *PLoS Pathogens* 2007; **3**(9): e138

[26]Singh SK. Structural basis for Duffy recognition by the malaria parasite Duffy-binding-like domain. *Nature* 2005; **439**(7077): 741-744.

[27]Combet C, Jambon M, Delέage G, Geourjon C. Geno3D: automatic comparative molecular modeling of protein. *Bioinformatics* 2002; **18**(1): 213-214.

[28]Sweredoski MJ, Pierre B. PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. *Bioinform Appl Note* 2008; **24**(12): 1459-1460.

[29]Liang S, Zheng D. EPMeta: prediction of antigenic epitopes using support vector regression and multiple server results. *BMC Bioinformatics* 2010; **11**: 381.

[30]Rubinstein ND, Mayrose I. Epitopia: a web-server for predicting B-cell epitopes. *BMC Bioinformatics* 2009; **10**: 287.

[31]Haste P, Nielsen M. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci* 2006; **15**: 2558-2567.

[32]Sun J, Wu D. SEPPA: a computational server for spatial epitope prediction of protein antigens. *Nucleic Acids Res* 2009; **37**: W612-W616.

[33]Huang J, Honda W. CED: a conformational epitope database. *BMC Immunol* 2006; **7**: 7.

[34]Huang YX. Pep-3D-Search: a method for B-cell epitope prediction based on mimotope analysis. *BMC Bioinformatics* 2008; **9**(1): 538.

[35]Magnan CN, Zeller M, Kayala MA, Vigil A, Randall A. High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics* 2010; **26**: 2936-2943.

[36]Petersen B. A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Structural Biol* 2009; **9**(1): 51.

[37]Drummond AJ, Ashton B, Buxton S. Geneious v5.4.[Online] Available from: http://www.geneious.com.

[38]Hopp TP, Woods KR. Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci USA* 1981; **78**(6): 3824-3828.

[39]Drew DR, O'Donnell, Rebecca A. A common cross-species function for the double epidermal growth factor-like modules of the highly divergent *Plasmodium* surface proteins MSP-1 and MSP-8. *J Biol Chem* 2004; **279**(19): 20147-20153.

[40]Remarque EJ, Bart W. Faber BW. Apical membrane antigen 1: a malaria vaccine candidate in review. *Trends Parasitol* 2007; **9**(24): 1471-4922.

[41]Sevier CS, Chris AK. Formation and transfer of disulphide bonds in living cells. *Nat Rev Mol Cell Biol* 2002; **3**(11): 836-847.

[42]Grzegorz B. Formation of disulfide bonds in proteins and peptides. *Biotechnol Adv* 2005; **23**(1): 87-92.

[43]Van den SP. Concepts and principles of O-linked glycosylation. *Crit Rev Biochem Mol Biol* 1998; **33**(3): 151-208.

[44]Khan AM, Miottob O, Heiny AT, Salmon J, Srinivasan KN, Nascimento E, et al. A systematic bioinformatics approach for selection of epitopebased vaccine targets. *Cell Immunol* 2006; **244**(2): 141-147.

[45]Rappuoli R. *Vaccine design: Innovative approaches and novel strategies.* Norfolk Caister Academic Press; 2011.

[46]Andersen PH, Nielsen M. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci* 2006; **15**: 2558-2567.

[47]Hui George SN, Nikaido C. Dominance of conserved B-cell epitopes of the *Plasmodium falciparum* merozoite surface protein, MSP1, in blood-stage infections of naive aotus monkeys. *Infect & Immu* 1996; **5**: 1502-1509.

[48]Kleywegt GJ, Jones TA. Phi/Psi-chology: Ramachandran revisited. *Structure* 1996; **4**: 1395-1400.