American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)
ISSN (Print) 2313-4410, ISSN (Online) 2313-4402

http://asrjetsjournal.org/

A Potential Malaria Vaccine Candidate Identified Using an Insilico Approach

Claire Aguttu[a*], Brenda Apio Okech[b], Ambrose Mukisa[c], George W. Lubega[d]

[a,c]*Makerere University, College of Natural Sciences, Department of Biochemistry and Sport Science. P.O Box 7062 Kampala, Uganda*

[b]*Uganda Virus Research Institute. P.O Box 45 Entebbe, Uganda*

[d]*Makerere University, College of Veterinary Medicine, animal resource and biosecurity, Department of Biomolecular sciences and biolabs. P.O Box 7062 Kampala, Uganda*

[a]*Email: caguttu@covab.mak.ac.ug,* [c]*Email: amukisa@conas.mak.ac.ug*

[b]*Email: bokech@gmail.com,* [d]*Email: glubega@imul.com*

Abstract

The search for an effective malaria vaccine has yielded no success yet. Unfortunately, resistance to post-infection treatments is on the increase hence the need to develop an effective vaccine. The aim of many reverse vaccinology studies is to identify novel proteins found exposed on the surface. Many malaria vaccine candidates can be effective tools against malaria but gross allelic polymorphism is a major hindrance which could be overcome by using highly conserved proteins. Also, peptide-based vaccines can be of great importance in fighting malaria, however, this is limited by HLA restriction which can be maneuvered by using promiscuous peptides. In the current work, our objective was to computationally identify conserved hypothetical, antigenic, surface proteins in pathogenic *Plasmodium falciparum parasite*. So in this study, we employed an *in silico* approach to screen the proteins on the basis of surface localization, non-homology with host proteome, and MHC class I and II binding promiscuity. The analyses reported XP_001351004.1 an uncharacterized protein as a novel vaccine candidate. Generation of the 3D model of the protein was done using the RaptorX server. Furthermore, the B-cell and T-cell epitopes were also predicted. B-cell epitopes were predicted using ABCpred and Kolanskar and Tongaonkar antigenicity method while the T-cell epitopes were predicted using CTLpred. Five peptides were selected based on their hydrophobicity. Results from this study could be extended to *in-vivo* and *in-vitro* experiments for future vaccine development.

Key words: Malaria; Plasmodium falciparum parasite; vaccine candidate; hypothetical proteins.

---

* Corresponding author.

1. Introduction

Human malaria is caused by parasites of the genus *Plasmodium* which include: *P.falciparum, P.vivax, P.ovale, P.malariae and, P.knowlesi* transmitted to humans through a bite of an infected female *Anopheles* mosquito. *P.falciparum* malaria still remains the leading cause of morbidity and mortality in sub-Saharan Africa [1]. Globally approximately 855,000 deaths were attributed to malaria in 2013 most of whom were children from African continent [2] while in 2014, the WHO reported about 198 Million new cases of malaria [3], a high morbidity rate which negatively impacts on the economies of affected countries. Malaria symptoms may include; high fever, chills, headache, and shivers among others [4]. Moreover, improper diagnosis and delayed treatment may cause anemia, kidney failure, cerebral malaria, retinopathy [4] and convulsions.

In Uganda, malaria is majorly controlled and managed by use of antimalarial drugs mainly artemisinin combination therapy (ACT) and quinine as first and second line drugs respectively. Unfortunately the drug intervention is being threatened by the ever-emerging drug resistance. Other control measures like the use of insecticide-treated mosquito nets, prophylactic drug treatment and indoor residual spraying have and are still being applied but with debatable success. It should, however, be noted that use of indoor pesticides has the potential of causing adverse side effects for instance; headache and nausea[5], immune toxicity[6] and childhood leukemia[7] among others to both humans and animals. Hence the availability of an efficacious malaria vaccine would be of paramount importance in solving the problems associated with the use of antimalarial drugs as well as insecticides.

To-date there is not a commercially available malaria vaccine although, in the past 30 years, a good number of vaccine candidates have been investigated for vaccine development with little success, as many fail at Phase II of clinical trial. Some of the prime candidates include; the merozoite surface antigens like merozoite surface protein1 (MSP1) and apical membrane antigen1 (AMA1) which have shown moderate effects against the malaria parasite ([8,9,10,11] even in combination (RESA, MSP1, and MSP2), that formed the so-called Combination B vaccine, anti-parasitic effect was also limited. RTS/S, the only vaccine candidate that has shown promise of being the first ever malaria vaccine confers moderate protection in both clinical and severe malaria vaccines. Also for it to be of substantial importance to malaria control, it has been recommended that it is used in combination with other control measures [12]. While this may provide a glimpse of hope for people in malaria-endemic areas, a more efficacious vaccine is still greatly anticipated. It is believed that the limits of the anti-infection vaccine, for example, the RTS/S[13] and some blood-stage vaccines e.g AMA1 [14] may be attributed to gross allelic polymorphism and this could be overcome through the use of highly conserved vaccine targets.

The complete *P. falciparum* 3D7 nuclear genome is 23 Megabase distributed among 14 chromosomes. The genome contains about 5300 protein-coding genes with an A+T rich content of more than 97% [15] many of which encode uncharacterized proteins with no recognizable signal peptides and have no similarities with proteins in existing databases.

Of the 5300 predicted proteins, approximately 60% (3208) are hypothetical proteins (proteins with no similarity

to proteins in other organisms). 31% (1631) have one or more transmembrane domains while 17.3% (911) possess putative signal peptide or signal anchor [15].

Conserved hypothetical proteins have been defined as large fractions of genes in sequenced genomes that encode proteins found in organisms from several phylogenetic lineages but have not been functionally characterised and described at the protein chemical level [17]. These hypothetical proteins have obscure functions which pose a challenge in functional genomics as well as general biology [16, 17] although presumably, they may have great importance to complete genomic and proteomic information [18]. Moreover, HPs possibly play important roles in the survival of a pathogen [19,20] and may also serve as markers and pharmacological targets [21].

Currently, bioinformatics has endowed research with a number of algorithms that can be used to mine and predict and provide targets of potential vaccine antigens from the available comprehensive genomic, proteomic and transcriptomic datasets of human and pathogens [22]. This idea was first proven with the successful development of serogroup B *Neisseria meningitides* vaccine and later *staphylococcus* vaccine [23].

In the identification of novel proteins as revealed by genome analysis, several criteria have been implemented for the selection of additional antigens for vaccine inclusion. These include; essentiality to the parasite, most especially survival as evidenced by resistance to genetic disruption; surface exposure for antibody binding. Also the antigen should be able to induce protective immunity in rodents or primates as well as having the ability to induce antibodies that affect parasite replication *in vitro*; and sero-epidemiological correlations of protection in endemic populations. More importantly though is the antigens' capacity to induce broadly reactive antibody and T-cell responses in humans [24].

To further establish more insight into the role of bioinformatics in studying vaccine candidate antigens, Immunoinformatics approach has been used to identify a potential vaccine candidate. The screening strategies used in this study were; subcellular localization, sequence non-homology to humans and MHC binding potential.

2. Methods

2.1. Protein sequence retrieval

Sequences of hypothetical conserved proteins (HCP) were downloaded from the National Center for Biotechnology Information (NCBI) database in FASTA format. NCBI is a resource center that provides access to biomedical and genomic information. We retrieved 20 sequences characterized as "hypothetical conserved protein" (www.ncbi.nlm.nih.gov/protein). These were first and foremost analysed for the presence of signal peptides and transmembrane domains. The signal peptide was predicted using SignalP4.1 an internet software predicts presence and location of signal peptide cleavage sites in residues of the proteins with server-recommended cutoff valves as described by Petersen [25] (www.cbs.dtu.dk/services/signalP). A signal peptide is a short sequence of 5-30 amino acids, present in proteins destined for the endoplasmic reticulum or secretion. While transmembrane helixes were identified using a prediction method of TMHMM (Transmembrane protein

topology with hidden markov model which has an accuracy of 97-98% [26].

2.2. Subcellular localization of the proteins

Because the immune system readily recognizes surface-exposed proteins on the pathogen, predicting the subcellular localization of the proteins serves as one of the major criteria for designing a vaccine candidate. Subcellular localization of the proteins was predicted using WoLFPSORTv2 [27] WOLFPSORT is an extension of the PSORT II program for protein subcellular location prediction which converts protein amino acid sequences into numerical localization features; based on sorting signals, amino acid composition and functional motifs such as DNA-binding motifs. The results from WOLFPSORT were validated using CELLO v.2.5 server [28]. CELLO uses a two- level support vector machine (SVM) system for supervised learning model on a set of training levels. Firstly, based on the amino acid composition, partitioned amino acid composition, dipeptide composition and the sequence composition based on the physicochemical properties of amino acids. Secondly, the data is then processed to produce the probability distribution for protein localization as confidence score.

2.3. Non –homologous proteins to the human proteome

The protein sequence that was predicted to be on the extracellular surface by CELLO was taken for further analysis. This step was performed using NCBI- BLASTp as described by Altshul and co-workers [29]. The expectation value (*E* value) which assesses the statistical significance of BLAST was kept at 0.005 above which the protein was considered non-homologous [30]. BLASTp is a local alignment method that detects the best regions of similarity between the query and the target. This step was imperative because it aids the selection of proteins which don't share homology with human proteome since the immune system targets cells and proteins it deems "non-self'' under healthy conditions.

2.4. Prediction of the binding capacity to MHC class1 and 11 proteins

To increase the power of our work to find a good vaccine candidate, the selected protein sequence was subjected to different computational methods to try and predict its binding capacity to MHC proteins since presentation of antigens by major histocompatibility complex (MHC) proteins is essential for adaptive immunity.

The two major classes of MHC ie MHC class I and II play a pivotal role in activating cytotoxic and helper T-cells respectively against foreign antigens. MHC class I binding regions in the protein were predicted using ProPred1 which uses BIMAS server and literature to predict MHC binders for 47 class I alleles [31]. ProPred was used to predict MHC class II binding regions in the protein sequences. The server facilitates in locating promiscuous binding of nine residue-long peptide regions on a protein that can bind to 51 different HLA-DR alleles and are therefore useful in selecting vaccine candidates. Hence the peptides with the highest binding score for any MHC alleles were identified.

2.5. Prediction of antigenic epitopes

Antigenic epitopes are regions on protein surfaces that are preferentially recognized by B-cell antibodies. Accurate prediction of B-cell antigenic epitopes is important for immunologic research for instance, development of immunodiagnostic kits and synthetic peptide vaccines [32]. B-cell and cytotoxic T-cell epitopes were predicted on the identified protein using ABCpred and CTLPred respectively. For this predictions, the length of the B-cell epitopes was fixed at 16 and the cutoff at 0.51.CTLPred was used to predict cytotoxic cell epitopes by a combined approach of artificial neural network and support vector machine technique at a cutoff score of 0.51 and 0.36 respectively.

The antigenic B-cell epitopes were validated using the Kolanskar and Tongaonkar antigenicity prediction method, a tool found at the IEDB analysis resource. This is a semi-empirical method that predicts antigenic determinants on protein sequences based on the physicochemical properties of amino acid residues and frequencies of occurrences [33]. Therefore, the epitopes that were overlapping in both the ABCpred and Kolanskar and Tongaonkar method were selected for further analysis.

2.6. 3D structure prediction

Predicting the structure of a hypothetical protein from its amino acid sequence allows to predict its function. The structure of our protein was predicted using RaptorX [34]. This prediction server excels at predicting the 3D structures of protein sequences with less than 30% sequence identity with solved structures in the protein data bank. RaptorX predicts secondary and tertiary structures, contacts, solvent accessibility, disordered regions, and binding sites.

3. Results

Several subunit vaccines against human prokaryotic [35] and eukaryotic pathogens have been identified using reverse vaccinology [23]. However, no efficacious vaccine for malaria has been found even with a large number of uncharacterized (hypothetical) proteins in the *plasmodium falciparum* genome. Therefore, prioritizing *in silico* and experimental analyses of the hypothetical proteins could reveal new and important vaccine targets.

3.1. Protein sequence retrieval

In this study, we focused on using *in silico* tools to identify and characterize hypothetical conserved proteins of *plasmodium falciparum* of vaccine importance. A search for these proteins in the NCBI database using a query "hypothetical conserved protein *plasmodium falciparum*" revealed five (5) hits which were then retrieved in FASTA format and a local database of all the proteins created (Table1).

Table 1: Hypothetical conserved proteins of *Plasmodium falciparum* parasite

S/No	NCBI gene id Number	Plasmodb ID	No. of amino acids in protein
1	gi|124505783|ref|XP_001351005.1		155
2	gi|124505781|ref|XP_001351004.1	PF3D7_0107300	399
3	gi|296005128|ref|XP_002808899.1		72
4	gi|296005126|ref|XP_002808898.1		66
5	gi|124511812|ref|XP_001349039.1	PF3D7_0713200	223

3.2. Signal peptide prediction

Signal peptide characterization was employed to classify the five hcps. Upon examining them, hypothetical conserved proteins (hcps) for the presence of targeting signals, only two sequences (Table 2) had positive signal peptide prediction with signalP4.1 [36]. This suggests that these proteins are secretory and could play a functional role. Secondly, presence of a target signal in a predicted amino acid sequence indicates that the protein is certainly expressed [19].

3.3 Transmembrane prediction

The goal of this study was to identify a suitable malaria vaccine candidate, bearing in mind that a good vaccine must have the ability to elicit a strong immune response. And that such must preferably be a membrane protein (either integral or peripheral) because they are exposed to the immune system surveillance.

Therefore, using TMHMM a transmembrane prediction tool, we also analysed our hcps for presence of transmembrane helices which was meant to inform whether the protein is predicted to be on the cell membrane or soluble. The TMHMM method has the potential to discriminate between soluble and membrane proteins with specificity and sensitivity better than 99%.

Additionally, it also demonstrates a single sequence prediction accuracy of 97-98% correct topology [26]. Accordingly, four out of the 5 hcps sequences analysed, had at least one transmembrane helices (TMHs) (Table 2).The sequences XP_001351005.1 and XP_001351004.1 had both signal peptide and TMHs predicted.

3.4. Subcellular Localisation prediction

To further characterize the five hcps, we also analysed the protein sequences for subcellular localization being one of the major prerequisites for a potential vaccine antigen and of major importance to this study are proteins located on the extracellular (Extr) and plasma membrane (plas). This is because they are targets for the host's immune system. WoLF PSORT [27] (www.genscript.com/tools/wolf-psort), an extension of the PSORT11

program was used to achieve this objective and accordingly, results show that 2 of the 5 hcps proteins were predicted on the extracellular surface while one sequence on plasma membrane the of the parasite (Table 3). A validation done by CELLOv2.5 also predicted sequence XP_001351004.1 to be on the extracellular surface.

Table 2: Structural characteristics of hypothetical conserved proteins.

S/No	Ncbi gene id	Signal Peptide	No.TMHs	No. AA[b] residues of TMH
1	gi|124505783|ref|XP_001351005.1	present	2(7-29,116-138)*	21,22
2	gi|124505781|ref|XP_001351004.1	present	1(7-29)*	21
3	gi|296005128|ref|XP_002808899.1	absent	1(15-34)	19
4	gi|296005126|ref|XP_002808898.1	absent	1(24-41)	17
5	gi|124511812|ref|XP_001349039.1	absent	0	

* For the full amino acid residues refer to table 1 [b] Amino acids

Table 3: Subcellular localization of hypothetical conserved proteins

Ncbi gene id	**Subcellular locations by WOLFPSORT**	**Subcellular Locations by CELLOv2.5**
gi|124505783|ref|XP_001351005.1*	extr: 21, E.R.: 4, plas: 3, lyso: 2	Nuclear 1.956
gi|124505781|ref|XP_001351004.1*	extr: 19, E.R.: 5.5, E.R._golg: 3.5, lyso: 3, plas: 2	Nuclear 3.616 Extracellular 0.433
gi|296005128|ref|XP_002808899.1	mito: 18, nucl: 5, extr: 4, cyto: 4	Mitochondrion 2.099
gi|296005126|ref|XP_002808898.1	plas: 12, extr: 9, mito: 7, lyso: 4	Mitochondrion 2.044
gi|124511812|ref|XP_001349039.1	cyto: 16, extr: 6, nucl: 5, cysk: 4	Chloroplast 1.823

3.5. Non- Homology search

It is also paramount that a protein vaccine candidate should not show any homology with the host proteins and in this case human proteins, a blast search was performed on the protein sequence that was predicted on the extracellular surface by both WOLFPSORT and CELLOv2.5. Subsequently, sequence XP_001351004.1 was submitted to NCBI- BLASTp [29] and showed no significant similarity with the human proteins.

3.6. Affinity of the protein sequence to MHC class 1 and 11 proteins

Propred1 and Propred were used to predict for MHC binding prediction for 47 MHC class1 and 51 MHC class11 alleles respectively. The protein sequence; XP_001351004.1 that was predicted on the extracellular surface by both WOLFPSORT and CELLOv2.5 was subjected to Propred1 and Propred servers respectively. In Propred1, XP_001351004.1 had two peptides in the 80-100% range (ie MHC alleles HLA-A24, MHC-KA) while in Propred, it had eight peptides at a range of 80-90% with DRBI-1101,DRBI-1311,DRBI-1321,DRBI-1104,DRBI-1106,DRBI-1128,DRBI-1305 and no peptides at 90-100% binding score. These results aided in emphasizing the importance of the selected sequence as an ideal vaccine candidate.

3.7. Selection of antigenic peptides

Here B and Cytotoxic T-cell epitopes were predicted on protein XP_001351004.1 using ABCpred and Kolanskar and Tongaonkar antigenicity prediction methods for B-cell epitopes while CTLPred for T-cell epitopes. Five B-cell epitopes were predicted with the highest score being 0.94 for sequence LKIIDTDDLKRNKNNN (Table3). Mapping of antigenic determinant epitopes on protein vaccine candidates is crucial in predicting their possible immunogenic regions and potential for both humoral and cellular immunity. We also obtained a hydrophilicity plot for XP_001351004.1 protein using Kolanskar and Tongaonkar antigenicity prediction method.

Table 4: B-cell epitopes of XP_001351004.1 predicted by ABCpred

Rank	Sequence	Start position	Score
1	LKIIDTDDLKRNKNNN	160	0.94
2	THIEDLYKEDDKKRKN	256	0.93
3	VGSNTYHILEICKCSK	360	0.90
4	NVEYNDSGVFDRLLLN	314	0.90
5	PEYISFYCSQRVICNT	241	0.88

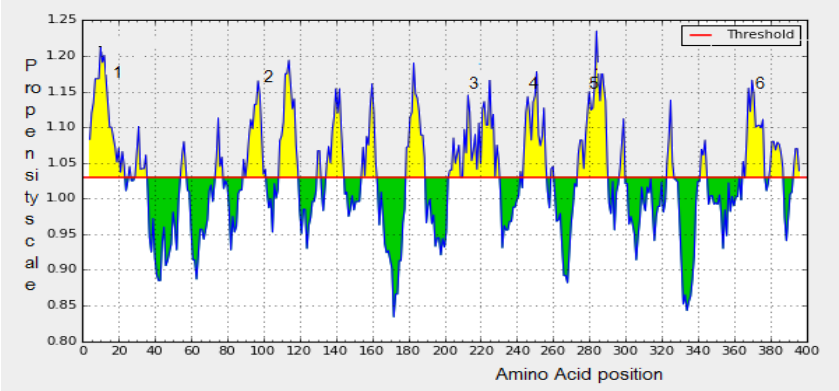

Figure1: Hydrophilicity plot, Y-axis (antigenic propensity scale), X-axis (Amino acid positions).The yellow peaks above the threshold indicate the antigenic (hydrophilic) portions of the protein while the green peaks below the threshold indicate the non-antigenic (hydrophobic) portions of the protein.Cytotoxic T lymphocyte (CTL) epitopes are potential candidates for subunit vaccine design for various diseases. CTLPred was used to predict cytotoxic T-cell epitopes. This server combines artificial neural network and support vector machine techniques with the scores of 0.51 and 0.36 respectively. Four epitopes were predicted with KYVFNYEQL as the highest ranking epitope at 203th residue of the protein (Table4).

3.8. Structure prediction of XP_001351004.1 protein

We further analysed the structure of the protein which gives insight into the function of the protein. Using RaptorX, the 3D structure model of XP_001351004.1 was predicted as having two major and one minor alpha helixes and nine beta sheets.

Table 5: T-cell epitopes of XP_001351004.1 predicted by CTLpred

Peptide Rank	Start Position	Sequence	Score (ANN/SVM)	
1	203	KYVFNYEQL	0.95/1.6985349	Epitope
2	300	KLYNKEKEI	0.99/0.94263252	Epitope
3	84	NEHFTLHNI	0.50/1.3613789	Epitope
4	206	FNYEQLAFL	0.89/0.83586063	Epitope

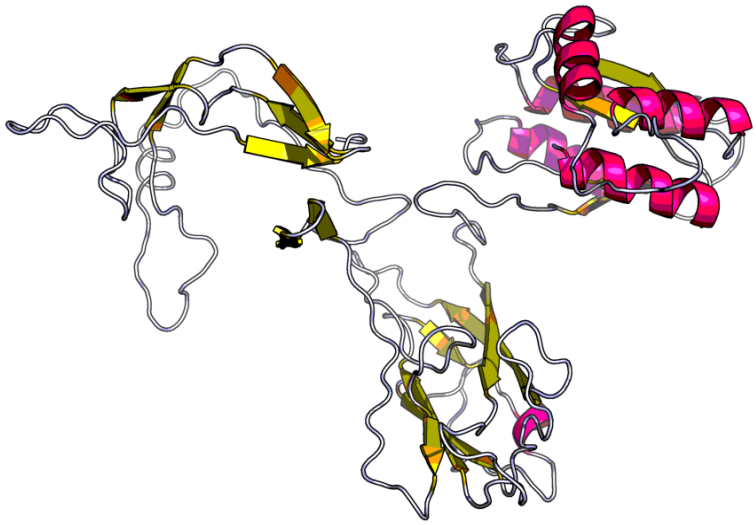

Figure2: 3D structure model of XP_001351004.1. The predictions showed that the protein has alpha helices and beta sheets. The input predicted as **2** domain(s), Best template: **3prxB**, p-value **1.73e-03, 399**(**100%**) residues are modeled, **23**(**5%**) positions predicted as disordered, Secondary struct: **21%** H, **25%** E, **and 52% C** Solvent access: **28%** E, **42%** M, **and 28% B**

4. Discussion

Malaria caused predominately by the *P. falciparum* parasite continues to be one of the major causes of mortality particularly in the developing countries of Africa and Asia. For decades now, research has not yet yielded a vaccine against malaria. Therefore, newer antigens that elicit better immune response need to be researched from a bulky of proteins that remain uncharacterized in the parasite's proteome. This study utilized immunoinformatic tools to predict a potential vaccine candidate in *P. falciparum* from available databases. The proteins were screened for parameters that included among others; signal peptide, transmembrane helices and subcellular location which was meant to inform whether the protein is on the surface of a pathogen or not. This is because surface proteins are easily recognized and liable to elicit an immune response when used as the target

antigens for a vaccine [37,38]. Proteins which are secretory are known to be immunogenic which portrays our protein as a potential vaccine candidate following SignalP prediction that showed that our protein had a signal peptide. The protein was also predicted by TMHMM to have a transmembrane helix suggesting its presence on the surface membrane.

The *P. falciparum* protein was then presented to WOLFPSORTv3 and validated by CELLO v2.5 to predict its subcellular localization. Accordingly, this protein was found to be located on the plasma membrane and extracellular by WOLFPSORT and CELLO respectively. The same protein was also forwarded to NCBI-BLASTp to screen for sequence non-homology against the human database this is because, in order to raise an immune response inside the body, the antigen must be recognized as non-self by the body. The ability to tolerate self-antigens arises during lymphocyte development where immature lymphocytes are exposed to self-antigens and those showing responses to self-antigens are eliminated [37]. Thus, the immunogenicity of the antigen depends on the degree of its foreignness or greater phylogenetic distance. The protein showed no significant homology to human proteins. In the next step of the procedure, we further analysed our protein for MHC class I and class II promiscuity in Propred1 and Propred, respectively. MHC proteins are called human leukocyte antigen (HLA) proteins in humans whose products play essential roles in intercellular immune recognition and discrimination between self and non-self.

Class I MHC proteins associates with endogenous antigens e.g., normal cellular proteins, viral proteins, bacterial proteins inside infected cells and present them to CD8+ cytotoxic T-cells hence terminating the pathogen or the infected cell [39]. On the other hand, Class II MHC proteins do associate with exogenous antigens processed by antigen-presenting cells (APCs) and present them to CD4+ helper T-cells which activate other immune cells capable of cellular and humoral responses [40]. In this study, it was observed that XP_001351004.1 protein has high MHC promiscuity for both MHC class I and II alleles. The selected protein was submitted for B-cell and T-cell epitope prediction. This is because the development of a peptide vaccine is based on identifying immunodominant epitopes that can induce specific immune responses without the need to deal with the whole microorganism. An epitope is a part of the antigenic molecule which can be distinguished by antibodies produced by T-cells, and B-cells [37]. We used ABCpred, which is an algorithm that was created by Saha and Raghava [41] in 2006 and is able to predict epitopes with 66% accuracy, 67% sensitivity, and 65% specificity. The ABCpred algorithm also has a better predictive performance compared to various physicochemical properties, including hydrophilicity [42], flexibility [43] and accessibility [44].The scores of B and T-cell epitopes in the predictions suggested that they have sequence similarity with known epitopes in the databases and would make an excellent target for creating an epitope vaccine. Structural genomics is a tool used to study microorganisms at the molecular level. Structural genomics aims to determine unique structures of proteins in organisms by X-ray crystallography [45]. The protein XP_001351004.1 being an uncharacterized protein, lacked in structural as well as functional annotation. And according to results obtained from a BLASTp search against Protein Data Bank there was no suitable template for 3D structure predictions through homology modeling Study. Here RaptorX was used to predict the structure of the selected protein.

5. Conclusion and recommendation

In conclusion, a plethero of malaria vaccines have been researched with many showing very low efficacy. This study used computational methods to screen NCBI database and from which a suitable vaccine candidate for plasmodium parasites was identified. XP_001351004.1 (UniProtKB - Q8I269 (Q8I269_PLAF7), uncharacterized protein was reported as the novel vaccine candidate. The structure of this uncharacterized protein was created, and the B cell and T cell epitopes were predicted. Therefore, it is suggested that XP_001351004.1 protein can be advanced for in vitro and in vivo experiments to formulate a new subunit malaria vaccine.

6. Limitation of the study

Extensive validation of the results was not done due to limited access to most bioinformatics software programs.

Acknowledgements

We acknowledge the African development bank (AfDB) for funding this piece of work and Dr. Julius Mulindwa for his bioinformatics expertise received.

Ethical considerations

This study did not require the use of human and or human specimens. This study employed the use of immunoinformatic and bioinformatic softwares which did not require ethical clearance.

Reference

[1] Lopez, A.D., et al., Measuring the Global Burden of Disease and Risk Factors, 1990-2001, in Global Burden of Disease and Risk Factors, A.D. Lopez, et al., Editors. 2006: Washington (DC).

[2] Murray, C.J., et al., Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet, 2014. **384**(9947): p. 1005-70.

[3] Organization, W.H., World malaria report 2014. 2014. Avarilable: http://www. who. int/malaria/publications/world_malaria_report_2014/en/. Accessed, 2015. **24**.

[4] Beare, N.A., et al., Malarial retinopathy: a newly established diagnostic sign in severe malaria. Am J Trop Med Hyg, 2006. **75**(5): p. 790-7.

[5] Titlic, M., Z. Josipovic-Jelic, and A. Punda, [Headache caused by pesticides--a review of the literature]. Acta Med Croatica, 2008. **62**(2): p. 233-6.

[6] Banerjee, B.D., The influence of various factors on immune toxicity assessment of pesticide chemicals. Toxicol Lett, 1999. **107**(1-3): p. 21-31.

[7] Van Maele-Fabry, G., et al., Residential exposure to pesticides and childhood leukaemia: a systematic review and meta-analysis. Environ Int, 2011. **37**(1): p. 280-91.

[8] Hill, A.V., Vaccines against malaria. Philos Trans R Soc Lond B Biol Sci, 2011. **366**(1579): p. 2806-14

[9] Draper, S.J., et al., Recent advances in recombinant protein-based malaria vaccines. Vaccine, 2015. **33**(52): p. 7433-43.

[10] Sheehy, S.H., et al., ChAd63-MVA-vectored blood-stage malaria vaccines targeting MSP1 and AMA1: assessment of efficacy against mosquito bite challenge in humans. Mol Ther, 2012. **20**(12): p. 2355-68.

[11] Ouattara, A., et al., Lack of allele-specific efficacy of a bivalent AMA1 malaria vaccine. Malar J, 2010. **9**: p. 175.

[12]Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: final results of a phase 3, individually randomised, controlled trial. The Lancet. **386**(9988): p. 31-45.

[13] Olotu, A., et al., Efficacy of RTS,S/AS01E malaria vaccine and exploratory analysis on anti-circumsporozoite antibody titres and protection in children aged 5–17 months in Kenya and Tanzania: a randomised controlled trial. The Lancet Infectious Diseases, 2011. **11**(2): p. 102-109.

[14] Marshall, V.M., et al., Diversity of the vaccine candidate AMA-1 of Plasmodium falciparum. Molecular and Biochemical Parasitology, 1996. **77**(1): p. 109-113.

[15] Gardner, M.J., et al., Genome sequence of the human malaria parasite Plasmodium falciparum. Nature, 2002. **419**(6906): p. 498-511

[16] Y, G.M., Conserved ‘Hypothetical’ Proteins: New Hints and New Puzzles. International Journal of Genomics, 2001. **2**(1): p. 14-18.

[17] Galperin, M.Y. and E.V. Koonin, 'Conserved hypothetical' proteins: prioritization of targets for experimental study. Nucleic Acids Research, 2004. **32**(18): p. 5452-5463.

[18] Loewenstein Y, Raimondo D, Redfern OC, Watson J, et al. (2009). Protein function annotation by homology-based inference. Genome Biol. 10: 207

[19] Desler C, Suravajhala P, Sanderhoff M, Rasmussen M, et al.(2009). In Silico screening for functional candidates amongst hypothetical proteins. BMC Bioinformatics10: 289

[20] Kumar K, Prakash A, Tasleem M, Islam A, Ahmad F, Hassan MI. Functional annotation of putative hypothetical proteins from Candida dubliniensis. Gene. 2014;543:93–100. doi: 10.1016/j.gene.2014.03.060.

[21] G. Lubec, L. Afjehi-Sadat, J.W. Yang, J.P. John.Searching for hypothetical proteins: theory and practice based upon original data and literature Prog. Neurobiol., 77 (2005), pp. 90-127

[22] Tomar N, De RK (2010) Immunoinformatics: an integrated scenario.Immunology 131:153–168

[23] Rappuoli, R., Reverse vaccinology. Current Opinion in Microbiology, 2000. **3**(5): p. 445-450
[24] Plebanski, M., et al., Malaria vaccines: into a mirror, darkly? Trends Parasitol, 2008. **24**(12): p. 532-6.

[25] Petersen, T.N., et al., SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods, 2011. **8**(10): p. 785-6.

[26] Krogh, A., et al., Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol, 2001. **305**(3): p. 567-80.

[27] Horton, P., et al., WoLF PSORT: protein localization predictor. Nucleic Acids Res, 2007. **35**(Web Server issue): p. W585-7.

[28] Yu CS, Chen YC, Lu CH, Hwang JK (2006) Prediction of protein subcellular localization. Proteins: Structure, Function and

Bioinformatics 64:643–651

[29] Altschul, S.F., et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research, 1997. **25**(17): p. 3389-3402.

[30] Chawley P, Samal HB, Prava J, Suar M, Mahapatra RK (2014) Comparative genomics study for identification of drug and vaccine targets in Vibrio cholerae: MurA ligase as a case study. Genomics 103:83–93

[31] Singh H, Raghava GPS (2003) ProPred1: prediction of promiscuous MHC class-I binding sites. Bioinformatics 19:1009–1014

[32] Chen J, Liu H, Yang J, Chou KC (2007) Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. Amino Acids 33(3):423–428

[33] Kolaskar, A.S. and P.C. Tongaonkar, A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Letters, 1990. **276**(1–2): p. 172-174.

[34] Källberg M, Margaryan G,Wang S, Ma J, Xu J (2014) RaptorX server: aresource for template-based protein structure modeling. Methods Mol Biol 1137:17–27.

[35] Ariel, N., et al., Search for potential vaccine candidate open reading frames in the Bacillus anthracis virulence plasmid pXO1: in silico and in vitro screening. Infection and immunity, 2002. **70**(12): p. 6817-6827.

[36] Nielsen, H., et al., Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Engineering, Design and Selection, 1997. **10**(1): p. 1-6.

[37] Kindt TJ, Goldsby RA, Osborne BA (2007) B-cell generation, activation, and differentiation. Kuby immunology. WH Freeman and Company, New York, pp 271–301

[38]Vetrivel, U., Subramanian, G., Fau-Dorairaj, S., Dorairaj, S., 2005. A novel in silico approach to identify potential therapeutic targets in human bacterial pathogens (1877-6566 (Electronic)) doi: D - NLM: PMC3238024 OTO - NOTNLM

[39] Klein J, Horejsi V. Immunology. Oxford, Blackwell Science; 1997.

[40] Dengjel J, Schoor O, Fischer R, Reich M, Kraus M, et al. Autophagy promotes MHC class II presentation of peptides from intracellular source proteins. Proc Nat Acad Sci USA. 2005;102:7922–7927. doi: 10.1073/pnas.0501190102.

[41] Saha S and Raghava GP: Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 65: 40-48, 2006.

[42] .Parker JM, Guo D and Hodges RS: New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: Correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry 25: 5425-5432, 1986.

[43] Karplus PA and Schulz GE: Prediction of chain flexibility in proteins. Naturwissenschaften 72: 212-213, 1985.

[44] Emini EA, Hughes JV, Perlow DS and Boger J: Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J Virol 55: 836-839, 1985

[45] Lesley, S. A., P. Kuhn, A. Godzik, A. M. Deacon, I. Mathews, A. Kreusch, G. Spraggon, H. E. Klock, D. McMullan, T. Shin, J. Vincent, A. Robb, L. S. Brinen, M. D. Miller, T. M. McPhillips, M. A. Miller, D. Scheibe, J. M. Canaves, C. Guda, L. Jaroszewski, T. L. Selby, M.-A. Elslinger, S. S. Taylor, K. O. Hodgson, I. A. Wilson, P. G. Schultz, and R. A. Stevens. 2002. Structural genomics of the Thermotoga maritime proteome implemented in a high-throughput structure determination pipeline. Proc. Natl. Acad. Sci. USA 99:11664-11669.