Screening and characterization of hypothetical proteins of *Plasmodium falciparum* as novel vaccine candidates in the fight against malaria using reverse vaccinology

Claire Aguttu[1*] ![ID], Brenda Apio Okech[2], Ambrose Mukisa[1] and George William Lubega[3]

Abstract

Background: *Plasmodium falciparum* is the most deadly and leading cause of morbidity and mortality in Africa. About 90% of all malaria deaths in the world today occur in Sub-Saharan Africa especially in children aged < 5 years. In 2018, it was reported that there were 228 million malaria cases that resulted in 405,000 deaths from 91 countries. Currently, a fully effective and long-lasting preventive malaria vaccine is still elusive therefore more effort is needed to identify better effective vaccine candidates. The aim of this study was to identify and characterize hypothetical proteins as vaccine candidates derived from *Plasmodium falciparum* 3D7 genome by reverse vaccinology.

Results: Of the 23 selected hypothetical proteins, 5 were predicted on the extracellular localization by WoLFPSORTv.2.0 program and all the 5 had less than 2 transmembrane regions that were predicted by TMHMMv2.0 and HMMTOP programs at default settings. Two out of the five proteins lacked secretory signal peptides as predicted by SignalP program. Among the 5 extracellular proteins, 3 were predicted to be antigenic by VaxiJen (score ≥ 0.5) and had negative GRAVY values ranging from − 1.156 to − 0.440. B cell epitope prediction by ABCpred and BCpred programs revealed a total of 15 antigenic epitopes. A total of 13 cytotoxic T cells were predicted from the 3 proteins using CTLPred online server. Only 2 out of the 13 CTL were antigenic, immunogenic, non-allergenic, and non-toxic using VaxiJen, IEDB, AllergenFp, and Toxinpred servers respectively in that order. Five HTL peptides from XP_001351030.1 protein are predicted inducers of all the three cytokines. STRING protein–protein network analysis of HPs revealed XP_001350955.1 closely interacts with nucleoside diphosphate kinase (PF13-0349) at 0.704, XP_001351030.1 interacts with male development protein1 (Mdv-1) at 0.645, and XP_001351047.1 with an uncharacterized protein (MAL8P1.53) at 0.400.

* Correspondence: aguttu2003@yahoo.com
[1]Department of Biochemistry and Sports Science, College of Natural Sciences, Makerere University, P.O. Box 7062, Kampala, Uganda
Full list of author information is available at the end of the article

Conclusion: Reverse vaccinology is a promising strategy for the screening and identification of antigenic antigens with potential capacity to elicit cellular and humoral immune responses against *P. falciparum* infection. In this study, potential vaccine candidates of *Plasmodium falciparum* were identified and screened using standard bioinformatics tools. The vaccine candidates contained antigenic and immunogenic epitopes which could be considered for novel and effective vaccine targets. However, we strongly recommend in vivo and in vitro experiments to validate their immunogenicity and protective efficacy to completely decipher the vaccine targets against malaria.

Keywords: Malaria, Epitopes, *Plasmodium falciparum*, Vaccine, Reverse vaccinology, Hypothetical protein

Background

Malaria is caused by protozoan parasites of the genus Plasmodium: *Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale, Plasmodium malariae*, and *Plasmodium knowlesi* transmitted to people through a bite of an infected female Anopheles mosquito vector. However, *Plasmodium falciparum* is the most deadly and leading cause of morbidity and mortality predominantly in Africa. About 90% of all *malaria* deaths in the world today occur in *Sub*-Saharan Africa especially in children aged < 5 years [1]. In 2018, it was reported that there were 228 million malaria cases that resulted in 405,000 deaths from 91 countries [2]. Some of the malaria symptoms may include body weakness, headache, fever, and shivers [3]. In case of misdiagnosis coupled with delayed treatment, the patient may develop anemia, kidney failure, cerebral malaria, retinopathy, and convulsions. Malaria is commonly managed through the use of antimalarial drugs mainly artemisinin-based combination therapy and indoor residual spraying. Unfortunately, the drug and vector control intervention are being threatened by the ever emerging antimalarial drug and insecticidal resistance which has resulted in an increase of malaria transmission worldwide [4].To-date, there is no efficacious vaccine available globally so far against malaria. Currently, a number of vaccines for malaria are in both pre-clinical and clinical development, targeting both children and pregnant women [5]. These are categorized as pre-erythrocytic vaccines, blood-stage vaccines, transmission-blocking vaccines, and combination vaccines targeting the different stages of the malaria parasite's life-cycle. Some of the prime candidates include the merozoite surface antigens like merozoite surface protein-1 and apical membrane antigen-1 which have shown moderate effects against the malaria parasite [6, 7]. Malarial vaccine development is hampered by factors such as multiple stages of the life-cycle, multiple antigens per stage, multiple epitopes per antigen, multiple arms of the immune system, multiple immune responses in different hosts, and multiple strains of the parasite [8]. RTS,S is the most advanced malaria vaccine candidate and is based on a virus-like particle containing central repeat and C-terminal epitopes of the major sporozoite surface antigen, circumsporozoite protein.

However, it has limitation of waning vaccine efficacy over time with a significant reduction by 3 years post-immunization [9]. Another noticeable limitation of the RTS,S vaccine is the incapability to induce CD8+ T cell responses, which represent an efficient anti-parasite mechanism that eliminates malaria liver stages (reviewed in [10]. It is therefore acceptable that identifying new targets which may be more efficacious is paramount. Initially, the *P. falciparum* 3D7 nuclear genome contained 5300–5400 protein-coding genes and 60% (3208) had unknown functions [11]. However, the number of Plasmodium-predicted genes has since risen to 5438 [12] and approximately 50% have no ascribed function [13, 14] and are also known as hypothetical proteins (HP). Hypothetical proteins are sequences with little to no experimental evidence for their function's existence being characterized by a low identity to proteins with known function [15]. Two groups of HPs exist: uncharacterized protein families and domains of unknown function. Many studies have identified and characterized hypothetical proteins from different microorganisms which appear to be of great importance [16–21]. Reverse vaccinology (RV) is a new approach to identify drug target and vaccine candidates without the need for culturing the parasite [22]. Through the use of online bioinformatics algorithms, potential peptide-based vaccine antigens notably the serogroup B Neisseria meningitides vaccine and later staphylococcus vaccine were identified and developed successful [23, 24]. Reverse vaccinology analyzes the entire parasites' protein repertoire using bioinformatics tools to prioritize potential targets for experimental validation both either in in vitro or in vivo. Thus, identifying new target antigens is the another way of boosting up new malaria vaccine development [25]. Moreover, for the probable antigens to be potentially good vaccine candidates, they must be surface exposed and able to be recognized by the host's immune system [26]. This study was designed to employ RV and immunoinformatics approaches to identify potential vaccine targets with their epitopes that can produce the B and T cell-mediated immunity. These predicted epitopes could be considered as promising candidates for effective peptide-based vaccine against malaria.

Methods

Protein selection and retrieval

Hypothetical proteins were searched in National Center for Biotechnology information (NCBI) database by typing the keywords "conserved hypothetical proteins *Plasmodium falciparum*" and a total of 23 protein sequences (accession nos. XP_002808602.1, XP_001350996.2, XP_002808605.1, XP_001350997.1, XP_001351014.2, XP_001350955.1, XP_001351004.1, XP_001351002.1, XP_001351011.2, XP_001351013.1, XP_024328987.1, XP_001351030.1, XP_001351040.1, XP_002808611.1, XP_001351044.1, XP_001351047.1, XP_001351049.1, XP_001351045.1, XP_001350986.1, XP_001350982.1, XP_002808604.1, XP_001350978.2, XP_002808603.2) of conserved hypothetical proteins of *Plasmodium falciparum* 3D7 were selected for this study. The protein sequences were selected based on the criteria that they had to be conserved and their function was unknown hence hypothetical. The proteins were then characterized by several bioinformatics tools including WoLFPSORT, SignalP, TMHMM, BLASTP, VaxiJen, and ProtParam. For immunoinformatics, BCpred and ABCpred were used for B cell epitope prediction while CTLpred was employed for T cell epitope prediction.

Subcellular localization of the hypothetical proteins

Predicting the subcellular location is one of the major criteria for designing a vaccine as immune cells do readily recognize surface exposed proteins on a pathogen. Therefore, subcellular locations of the 23 proteins were predicted using WoLFPSORTv2.0 [27] which is a free online server localized at www.wolfpsort.org. WoLFPSORT is an extension of the PSORT II program 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 method groups proteins in more than 10 locations with an estimated sensitivity and specificity of around 70% for nucleus, mitochondria, cytosol, plasma membrane, extracellular, and chloroplast. Only those proteins that were localized on the extracellular site of the pathogen were selected for further analysis.

Antigenicity analysis

Antigenicity of the 5 extracellular proteins chosen from the previous step was checked using the VaxiJen2.0 online server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html). VaxiJen is an alignment-free approach for antigen prediction with an accuracy of 70 to 89% hence a crucial tool in reverse vaccinology. The method is based on auto cross covariance transformation of protein sequences into uniform vectors of principal amino acid properties. The method threshold value was set to 0.5%. Hence, any protein that had an antigenic score above 0.5% was selected for further analysis [28].Proteins with VaxiJen score less than 0.5% were considered non antigenic and were therefore discarded.

Prediction of transmembrane helices (TM) and signal peptide

The three antigenic (VaxiJen ≥ 0.5%) hypothetical proteins selected from the previous step were characterized for transmembrane domains using TMHMM based on hidden Markov model (http://www.cbs.dtu.dk/services/TMHMM/) [29] and HMMTOP (http://www.enzim.hu/hmmtop/) [30] at a default setting of the parameters. Proteins having ≤ 1 transmembrane helices by both methods were selected as they are considered to be good targets because of their easy to clone and express during experimental validation studies.

SignalP ver.5.0 server (http://www.cbs.dtu.dk/services/SignalP/) [31] was used to identify the location of signal peptide within the selected proteins. Proteins with predicted signal peptide were analyzed further.

Identification of non-human homologous proteins

It is important that potential vaccine targets are not human homologs to avoid autoimmune reactions as the immune system targets cells and proteins it considers "non-self" under normal conditions. In this regard, the three proteins chosen from the previous steps were subjected to a blast analysis using NCBI-BLASTp (https://blast.ncbi.nlm.nih.gov/Blast) against the human proteome as described by Altshul and co-workers [32]. The expectation value (E value) which assesses the statistical significance of BLAST was kept at 0.005 and identity at < 35%. Proteins with E value above 0.005 and < 35% identity were considered non-human homologs and are expected not to interfere with normal host immune mechanism when used as vaccine candidates [33, 34]

Identification of conserved identity with other Plasmodium strains

Proteins screened from the previous steps were assessed for conservation in the different related Plasmodium strains (*Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae*, and *Plasmodium yoeli*) using BLASTp analysis on the NCBI server. This analysis serves to identify functionally conserved proteins which are shared by two or more species. The identity percentage and minimum query coverage were set to 80% and 50% respectively. Hence, all proteins with a sharing percentage ≥ 80% were considered as orthologous conserved.

Allergenicity and antibody production predictions

Allergenicity was checked by two different methods including AllerTOP.v2.0 (http://www.pharmfac.net/allertop) and AllergenFP.v1.0 (http://ddg-pharmfac.net/

Allergen FP). AllergenFP.v1.0 uses amino acid E-descriptors and auto- and cross-covariance transformation of protein sequences into uniform equal-length vectors to predict allergens [35]. Proteins not having allergic properties by all the two prediction methods were considered for further analysis. IgPred [36] does predict the potential antibody (Ab) isotype which can be elicited by a particular protein with an accuracy of around 80%. We employed IgPred online server (http://crdd.osdd.net/raghava/igpred/) to predict the different antibody subtypes that might be elicited by the selected hypothetical proteins.

Physico-chemical parameters analysis

The physicochemical properties, amino acid composition, molecular weight (Mw), theoretical isoelectric point (pI), instability index (II), extinction coefficient (EC), half-life, and grand average of hydropathy (GRAVY) of the non-allergenic proteins were analyzed using ProtParam server (https://web.expasy.org/protparam/) [37]. Instability index predicts protein's stability in the test tube whereby an II value (< 40) is said to be stable and vice versa. Aliphatic index value explains vaccines thermostability and is defined as the relative volume occupied by the aliphatic side chain amino acids. GRAVY values explain the hydrophilic or hydrophobic nature of the protein and are calculated as the sum of all hydropathy values of all the amino acids divided by the number of residues in the sequence [38].

Prediction of B and CTL antigenic epitopes

Accurate identification of antigenic epitopes on a protein is important for the development of immunodiagnostic kits, synthetic peptide vaccines, and antibody production [39]. B cell epitopes were predicted on the three selected hypothetical proteins using prediction methods namely ABCpred (http://crdd.osdd.net/raghava/abcpred/) [40] and BCpred software (https://omictools.com/bcpreds-tool).The length of the B cell epitopes was fixed at 16 and the cutoff at 0.51 in ABCpred. For BCpreds predictions, 20 mers peptides were identified at a specificity of 70%. ABCpred uses artificial neural network (ANN) which is a machine learning system inspired by biological neural network to find patterns in a given dataset. BCpreds contains two methods based on different algorithms namely amino acid pair (AAP) antigenicity method and BCpreds method using subsequence kernel [41]. The B cell epitopes resulting from the three algorithms were assembled and the overlapping regions were selected as predicted B cell epitopes. Subsequently, the selected B cell epitopes were screened for their antigenicity, allergenicity, and toxicity using VaxiJen v2.0, AllergenFP v1.0, and ToxinPred server (http://crdd.osdd.net/raghava/toxinpred/) respectively. CTLPred server

(http://crdd.osdd.net/raghava/ctlpred/) [42] a direct method for predicting CTL epitopes from an antigenic sequence was used to predict cytotoxic T cell epitopes by a combined approach of artificial neural network (ANN) and support vector machine (SVM) learning technique at a cutoff score of 0.51 and 0.36, respectively, above which peptides are considered to be antigenic. The selected T cell epitopes were analyzed for their antigenicity, immunogenicity, allergenicity, and toxicity using VaxiJen2.0, IEDB (http://tools.iedb.org/immunogenicity/) programs, AllergenFPv 1.0, and Toxinpred servers, respectively.

Prediction of helper T-lymphocyte (HTL) epitopes

Helper T-lymphocyte (HTL) induces both humoral and cellular immune responses. Hence, HTL epitopes are most likely to play a significant role in preventive and immunotherapeutic vaccines. We applied the IEDB MHC-II binding tool (http://tools.iedb.org/mhcii/) to predict 15 amino acid long HTL epitopes using NN-align method [43]. NN-align method generated a percentile rank by comparing peptide's binding affinity with a comprehensive set of randomly selected peptides from the Swiss-Prot database. For this study, peptides with a percentile rank ≤ 5 were considered for further analysis [44]. The selected HTL peptides were assessed for antigenicity and cytokine induction particularly interferon-gamma (IFN-γ), interleukin-4 (IL-4), and interleukin-10 (IL-10). For predicting antigenicity, interleukin-4 (IL-4) and interleukin-10 (IL-10), VaxiJen, IL4pred (http://crdd.osdd.net/raghava/il4pred/), and IL10pred (http://crdd.osdd.net/raghava/IL-10pred/) servers, respectively, were used [45, 46]. In order to predict IFN-γ inducing HTL epitopes, we employed IFNepitope server (http://crdd.osdd.net/raghava/ifnepitope/) using a hybrid method (Motif and SVM) along with IFN-gamma versus non-IFN-gamma model [47].

Protein–protein interaction analysis

This was aimed at understanding the functional pathway and interaction of the hypothetical proteins with closely related proteins. STRINGv10.5 web server (https://string-db.org/) was used to predict this interaction by choosing the query sequences and protein–protein interaction networks were generated [48].

Results
Protein sequence retrieval and subcellular localization analysis

Twenty three hypothetical proteins of *Plasmodium falciparum* with amino acid length ranging from 81 to 2221 were retrieved from NCBI. These were then submitted to WoLFPSORT web server for subcellular localization. The prediction revealed 9(39%), 4(18%), 5(22%), 1(4%), 1(4%), and 3(13%) are localized in the cytoplasm, nucleus, extracellular, plasma membrane,

Fig. 1 Subcellular localization prediction using WOLFPsort web server of the 23 hypothetical proteins of *P. falciparum*
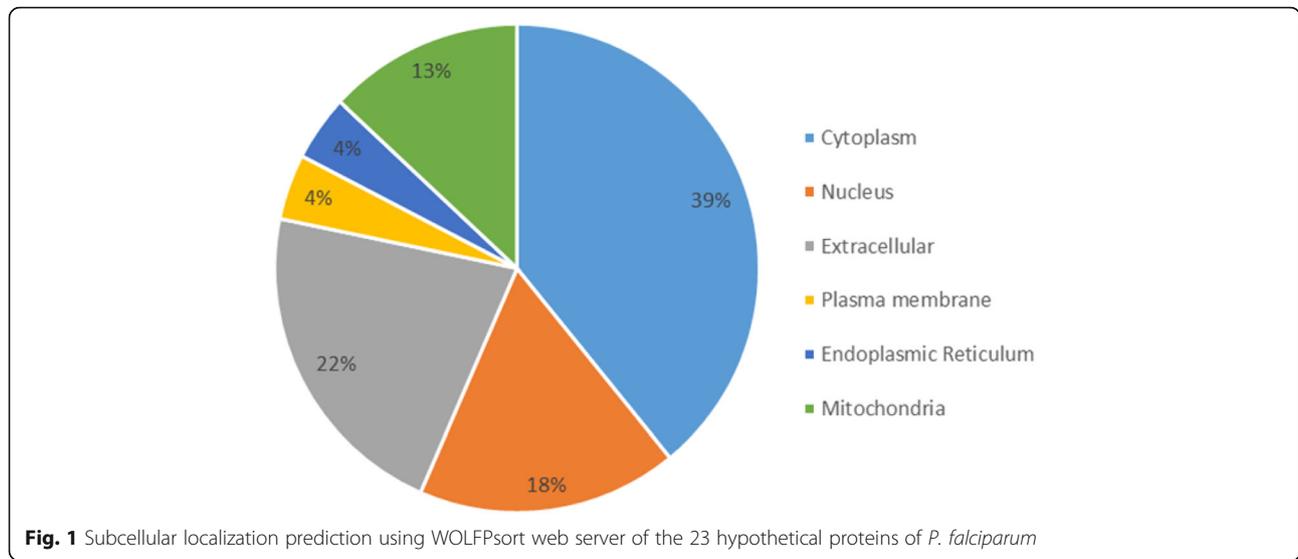
endoplasmic reticulum, and mitochondria, respectively. The results of subcellular localization analysis are given in Fig. 1.

Antigenicity analysis

The antigenicity of the 5 extracellular hypothetical proteins was calculated using VaxiJen ver. 2.0. Of these, 3 extracellular proteins were found to have antigenicity score above the threshold value of 0.5 (antigenic). Hypothetical proteins with VaxiJen score above 0.5 are shown in Table 1. Two extracellular hypothetical proteins XP_001351049.1 and XP_001350982.1 were eliminated at this step for having an antigenicity score lower than 0.5 which were considered as non-antigens.

Prediction of transmembrane domains and Signal peptide

Characteristic of transmembrane helices in proteins was predicted using TMHMM based on hidden Markov model and HMMTOP programs at a default setting of the parameters. As per the predictions, all the 3 antigenic extracellular hypothetical proteins were observed to contain none or 1 transmembrane domain (Table 1). SignalPv5.0 predicted a signal peptide on two proteins (NCBI: XP_001350955.1 and XP_001351030.1) and no signal peptide was found on XP_001351047.1 protein (Table 1). By using AllerTop and AllergenFP webservers to predict allergenic proteins and IgPred to predict the

immunoglobulin subtype induced by the proteins, all the three hypothetical proteins were non-allergens. Hypothetical proteins XP_001350955.1 and XP_001351030.1 were predicted to induce IgG while for XP_001351047.1 no immunoglobulin subtype (Table 1).

Screening for non-human homologs

In order to avoid interference against host immune mechanism, it is critical that potential vaccine candidates are non-human homologous. Consequently, the 3 hypothetical proteins selected from the previous steps were subjected to BLASTp search against human proteome. All the 3 extracellular proteins, namely XP_001350955.1, XP_001351030.1, and XP_001351047.1 had no significant similarity with human proteome (Table 2).

Analysis of conserved identity with other Plasmodium strains

This step was carried out in order to identify antigens which can provide cross-protection among Plasmodium species. Here, a BLASTp analysis was performed to assess the individual sharing of the selected hypothetical proteins among *Plasmodium vivax*, *Plasmodium ovale*, and *Plasmodium yoeli*. The alignment showed that protein, XP_001351047.1 from *Plasmodium falciparum* shared significant sequence identity, i.e., 80%, 77.78%, and 72.15% with *Plasmodium vivax*, *Plasmodium ovale*,

Table 1 Characteristics of 3 selected extracellular hypothetical proteins of *Plasmodium falciparum*

No.	NCBI accession number	No. of amino acids	Subcellular localization	TM helices	SignalP	VaxiJen score	Allergenicity	Ig subtype
1	XP_001350955.1	117	Extracellular	0	Yes	0.7376	No	IgG
2	XP_001351030.1	234	Extracellular	1	Yes	0.5191	No	IgG
3	XP_001351047.1	81	Extracellular	0	No	0.8264	No	-

Tm Transmembrane, *Ig* immunoglobulin, *IgG* immunoglobulin G

Table 2 Homology and conservation characteristics of 3 extracellular hypothetical proteins of *Plasmodium falciparum*

No. NCBI accession No.	Amino acid sequence homology percentage in Plasmodium strains (%)			Human proteome homology
	P. vivax	*P. ovale*	*P. yoeli*	
1. XP_001350955.1	-	-	-	No significant similarity found
2.XP_001351030.1	36.2	-	-	No significant similarity found
3.XP_001351047.1	80.25	77.78	72.15	No significant similarity found

No amino acid sequence homology found in other Plasmodium strains

and *Plasmodium yoeli*, respectively. XP_001351030.1 protein had 36.2% homology to *Plasmodium vivax* while *Plasmodium falciparum* protein XP_001350955.1 did not show any sequence similarity with other Plasmodium species (Table 2).

Protein physico-chemical parameter analysis
ProtParam analysis show that the molecular weight, pI, II, EC, and GRAVY of the 3 hypothetical proteins ranged between 9.58 and 27.85 kDa, 5.01 and 6.71, 24.47 and 57.78, 62.59 and 98.21, 5960 and 23380, and − 1.156 to − 0.440, respectively. All the three hypothetical proteins had a half-life of 10 h in bacterial host (*Escherichia coli*). The grand average of hydropathy (GRAVY) value for a peptide or protein is the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence. A GRAVY value is a good indicator of the hydrophobicity of the protein. The lower GRAVY values of our HPs indicated that they have better interaction with water [49].

B cell epitope prediction
For this analysis, three algorithms namely BCpreds server (BCpred and amino acid pair prediction methods) and ABCpred were utilized. BCpred algorithms generated 20-mer sequences of B cell epitopes with specificity of 70% whereas ABCpred generated 16mer B cell epitopes at a score of 0.51. The combination of BCpred, ABCpred, and VaxiJen servers allowed the prediction of 21 overlapping antigenic B cell epitopes from three hypothetical proteins. Out of the 21 antigenic B cell epitope, 15 were neither allergenic nor toxic. Antigenic B cell epitopes from the selected hypothetical proteins of *P. falciparum* are presented in Table 4.

Prediction of T cell epitopes
CTLPred server predicted a total of 19 cytotoxic T cell epitopes from the three HPs studied. A total of 13 out of 19 cytotoxic T cell epitope regions were predicted as antigens by VaxiJen server. Of these 13 antigenic epitopes, 7 were found in XP_001350955.1, 4 and 2 epitopes were in XP_001351030.1 and XP_001351047.1, respectively. Out of the 13 antigenic CTL epitopes, 8 epitopes were immunogenic (Table 5). And of the 8 antigenic and immunogenic CTLs, only 2 epitopes (bolded in Table 5) were neither allergenic nor toxic.

### Prediction of helper T-lymphocyte (HTL) epitopes
The IEDB MHC-II binding tool using NN-align method predicted a total of 61 HTL epitopes from the three hypothetical proteins. Thirty out of 61 were antigenic. Twelve out of 30 antigenic HTL epitopes were predicted to induce at least 2 cytokines (interleukin 4, interleukin 10, and interferon gamma). Five HTL peptides (bolded) from XP_001351030.1 protein are predicted inducers of all the three cytokines while no epitope from XP_001350955.1 and XP_001351047.1 proteins was able to induce at least two cytokines (Table 6).

Protein–protein function prediction
Protein–protein interaction networks were analyzed by STRING 10.5 server and revealed 10, 3, and 1 potential interacting protein associates (Fig. 2A–C) for XP_001350955.1, XP_001351030.1, and XP_001351047.1 based on network parameters including text mining, gene fusion, co-occurrence, co-expression, neighborhood, and databases. Because proteins function by interacting with other proteins where they form protein complexes and networks, understanding the complex protein interactions give important clues as to the function of novel proteins. Exploring this type of STRING generated predicted interaction networks can guide future experimental research, e.g., prediction of the possible cellular pathway of the protein of interest. Similar protein–protein interaction study has been previously worked out [50, 51]. XP_001350955.1 closely interacts with nucleoside diphosphate kinase (PF13-0349) at 0.704, XP_001351030.1 interacts with male development protein1 (Mdv-1) at 0.645, and XP_001351047.1 with an uncharacterized protein (MAL8P1.53) at 0.400.

Discussion
Malaria due to *Plasmodium falciparum* is still a major cause of mortality particularly in the developing countries of Africa and Asia. Until now, research efforts to develop an efficacious malaria vaccine have not yielded. Over the years, there has been rapid development of low-cost sequencing techniques which has led to generation of huge amounts of genomic and proteomic data; however, research on hypothetical proteins (HP) is yet to keep pace with. Currently, over 50% of the *Plasmodium*

Fig. 2 Protein–protein interaction network of three hypothetical proteins of *Plasmodium falciparum* generated by STRING web server. **A** XP_001350955. **B** XP_001351030. **C** XP_001351047.1
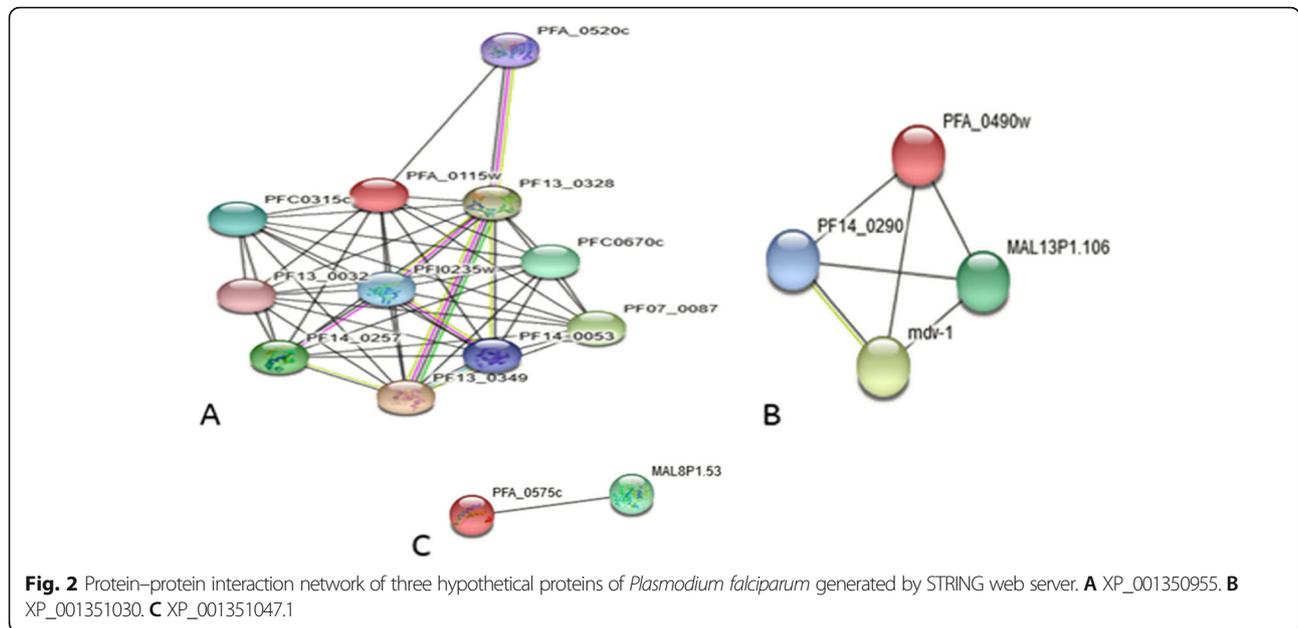
falciparum proteins have no ascribed function. Characterization of HP may be useful in better understanding the organism's metabolic pathways, disease progression, drug development, and disease control strategies [52]. With a complete *Plasmodium falciparum* genome sequence [11] and advancement in bioinformatics, it is now possible to identify potential vaccine candidates using reverse vaccinology which reduces the time and cost of designing and identifying vaccine candidates [53]. This study utilized several bioinformatics and immunoinformatic tools for identification and characterization of hypothetical proteins of *P. falciparum* for vaccine development. For each protein, different properties and their epitopes were analyzed for possible immune response. For this study, the properties of a good vaccine candidate considered were (1) they should be extracellular surface or cell surface localized to increase their accessibility to immune system surveillance, (2) they must be antigenic, (3) they must not show homology with the human proteins to avoid generation of autoimmune response, (4) they lack or possess one transmembrane (TM) regions to facilitate expression, and (5) they must be non-allergenic. Furthermore, secreted or cell surface antigens are considered good targets for developing vaccine as they are

usually antigenic and are responsible for the initial host-pathogen interaction [54]. Secondly, cell surface antigens are easily recognized and do elicit an immune response when used as the target antigens for a vaccine [55] with respect to those pathogens against which a strong B cell response is critical. A signal peptide motif serves to direct the intracellular protein to the extracellular surface of either the plasma membrane and or apical surfaces [56]. Proteomic and immunoinformatic tools revealed hypothetical proteins that could be valuable targets for vaccine development. Based on subcellular localization, antigenicity (VaxiJen score > 0.5), non-relatedness to human proteome (E value = 0.005 and identity at < 35%) and number of transmembrane helices predictions (less than 2), 3 out of 23 hypothetical proteins were identified as potential vaccine candidates against *P. falciparum* malaria. These three HPs include NCBI accession no. XP_001350955.1, XP_001351030.1, and XP_001351047.1. Subcellular localization of a hypothetical protein is useful to provide insights into their function as different cellular locations represent different functions. The HPs were predicted to be extracellular by WoLF PSORT server, which has a high accuracy in predicting subcellular localization of proteins in eukaryotic organisms [27]. These extracellular

Table 3 Physicochemical properties of the hypothetical proteins of *Plasmodium falciparum*

NCBI accession no.	Amino acid	Molecular weight	pI	Instability index (II)	Aliphatic index (AI)	Extinction coefficient (EC)	GRAVY	Half life (h)
XP-001350955.1	117	13581.48	5.01	38.84	98.21	5960	−0.440	10
XP-001351030.1	234	27846.86	6.55	57.78	70.38	23380	−0.824	10
XP-001351047.1	81	9581.76	6.71	24.47	62.59	12950	−1.156	10

pI isoelectric point, *GRAVY* grand average of hydropathy

Table 4 Final list of potential B-cell epitopes of the three selected hypothetical proteins of *Plasmodium falciparum*

Protein ID	B-cell epitope	Start position	ABCpred	VaxiJen	AllergenFP	ToxinPred
XP_001350955.1	TNIIDESEQENSIDLS	85	0.89	0.5752	Non-allergen	Non-toxin
	TQPISKKTSVEEKLKE	56	0.80	0.7490	Non-allergen	Non-toxin
	SIDLSDYTNIGSSLNP	96	0.77	1.0217	Non-allergen	Non-toxin
	LAQCLNTQPISKKTS	49	0.65	0.5782	Non-allergen	Non-toxin
	FSSIKIFLFFIVLEIL	4	0.52	0.7748	Non-allergen	Non-toxin
XP_001351030.1	YDEEDIDTEKASPLNP	95	0.90	0.5239	Non-allergen	Non-toxin
	RQKNKKKVLHDASLFD	192	0.82	0.6868	Non-allergen	Non-toxin
	PLNPYLAPSFVEMQSK	107	0.72	0.5705	Non-allergen	Non-toxin
	DQHIKKMNNKRICKNG	207	0.69	0.8737	Non-allergen	Non-toxin
	LLIISCALFFLFKRSE	13	0.69	0.6245	Non-allergen	Non-toxin
	DNMNNNDDNNNVKKYK	79	0.64	0.8050	Non-allergen	Non-toxin
XP_001351047.1	QNTIKAHVEANDECKE	10	0.91	1.3329	Non-allergen	Non-toxin
	KKEKYLKCFNNWYKNN	26	0.89	0.5078	Non-allergen	Non-toxin
	TQACDDYYEDYQICVL	48	0.81	0.8927	Non-allergen	Non-toxin
	CFNNWYKNNFLKGDLT	33	0.63	1.2766	Non-allergen	Non-toxin

proteins can be considered as vaccine targets. However, there is need to update and confirm their exact localization using immunoflourescent assays of electron microscope. VaxiJen server also showed that the selected HPs were immunogenic. The transmembrane localization of the protein positions itself to interact directly with the host's immune system; therefore, the number of transmembrane domains (TM) is seen as one of the selection criteria for a potential vaccine candidate. However, vaccine targets should possess ≤ 1 TM as it is usually difficult to clone, express, and purify proteins with more than one TM spanning regions. We predicted TM regions using TMHMM and HMMTOP programs and all the three HPs had less than 2 TM regions (Table 1). Since vaccine candidates with similar sequence to the hosts (e.g., human and mouse)

may cause autoimmunity [57]. It is therefore imperative that the probable vaccine candidates have no human homologs and hence exclusively present in pathogens and absent in humans. The three selected HPs were submitted to NCBI-BLASTp and all did not show significant similarity with human host (Table 2), suggesting that they could be used for vaccine development without causing autoimmunity. The appropriate physico-chemical properties and stable structure of the potential vaccine candidates are needed to evoke an immune response [58]. The GRAVY value for a peptide or protein is calculated as the sum of hydropathy values of all the amino acids divided by the number of residues in the sequence [38, 59]. All the three hypothetical proteins analyzed had negative GRAVY values (Table 3) clearly indicating their hydrophilic

Table 5 List of cytotoxic T cell epitopes predicted from the three hypothetical proteins of *Plasmodium falciparum*

Protein ID	Peptide	Start	Score (ANN/SVM) (0. 51/ 0.36)	Antigenicity score (>0.5)	Immunogenicity score (IEDB>0)	Allergenicity	Toxicity
XP_001350955.1	**RQNIPIRSL**	**42**	**0.88/0.96567173**	**0.7996**	**0.19332**	**No**	**No**
	FIVLEILLL	13	0.87/0.70629735	1.1767	0.20918	Yes	No
	DRNEKQTNI	79	0.91/0.58847506	1.6421	-0.19141	No	No
	FLFFIVLEI	10	0.49/0.92826689	1.5459	0.3734	Yes	No
	CLNTQPISK	53	0.91/0.38273127	0.7369	-0.07062	Yes	No
	AQCLNTQPI	51	0.89/0.35780961	0.9096	-0.10266	Yes	No
	KKTSVEEKL	61	0.80/0.42325026	0.7333	-0.06092	No	No
XP_001351030.1	VESIFYENV	57	0.66/0.96673418	0.8440	0.27146	Yes	No
	ASLFDQHIK	**203**	**0.97/0.54674556**	**1.3009**	**0.13182**	**No**	**No**
	ASPLNPYLA	105	0.75/0.69145562	0.9599	-0.0411	No	No
	ISCALFFLF	16	0.82/0.57119502	1.2939	0.21359	Yes	No
XP_001351047.1	YYEDYQICV	54	0.75/0.41137609	1.2069	0.023	Yes	No
	FNNWYKNNF	34	0.58/0.3902168	1.0489	0.00495	Yes	No

nature and good water solubility property. This information might be useful for localizing these proteins. The molecular weight, isoelectric point, and extinction coefficient of proteins are important in setting-up purification and crystallization experiments [60]. Furthermore, molecular weight is also important in characterizing protein function. Our HPs; XP_001350955.1, XP_001351030.1, and XP_001351047.1 had Mw 13581.48 Da, 27846.86Da, and 9581.76Da respectively. The extinction coefficient of our hypothetical proteins at 280 nm ranges from 5960 to 12,950 M cm with respect to the concentration of cysteine (Cys), Tryptophane (Trp), and Tyrosine (Tyr). The high extinction coefficient of hypothetical proteins is an indicator of presence of high concentration of Cys, Trp, and Tyr. It is defined as a measurement of how strongly a protein absorbs light at a given wavelength. The computed extinction coefficients aid in the quantitative study of protein–protein and protein–ligand interactions in solution [16]. Instability index is a measure of the in vivo stability of a protein and therefore an instability index smaller than 40 is believed to be stable [60, 61]. Two, XP_001350955.1 and XP_001351047.1 of our HPs had an instability index of 38.84 and 24.47 respectively hence are thus likely to be stable, while XP_001351030.1 which had instability index of 57.78 is considered unstable. The aliphatic index is estimated based on the number of aliphatic residues (alanine (Ala), valine (Val), isoleucine (Ile), and leucine (Leu)) in the protein and higher values indicate higher thermo stability over a wide temperature range [62]. Aliphatic index for the hypothetical protein sequences ranged from 62.59 to 98.21. The very high aliphatic index of the protein sequences indicates that these proteins may be stable for a wide temperature range.

Thus, all the calculated physicochemical properties could be important for further experimental studies of these HPs. Several reports [63–66] indicate that most of the malaria vaccines work mainly by inducing protective serum antibodies and to some extent CD4+ T cells which is often a sufficient component of vaccine efficacy. Unlike antibodies, however, CD8 T cells alone are also capable of conferring complete sterilizing protection, demonstrating their critical role in pre-erythrocytic immunity [67, 68]. Therefore, both the antigenic B and T cell epitopes are essential for obtaining the maximum immune response through humoral and cell-mediated immunity. The B cell epitopes were identified through ABCpred and Bcpred servers while CTL and HTL cell epitopes were predicted using CTLPred and IEDB-MHC11 web servers respectively and were further validated against antigenic property through VaxiJen server. This is based on the idea that the development of a peptide vaccine largely relays on identifying immunodominant epitopes that can induce specific immune responses without the need of involving whole microorganism. From the three HPs, a number of antigenic B, cytotoxic and helper T cell epitopes were identified which could potentially be used for designing an epitope based vaccine against *P. falciparum* malaria (Tables 4, 5, and 6).

The characterization of protein–protein interactions provides insights into their biological and cellular functions in the cell. Generally, the function and activity of a protein are often modulated by other proteins with which it interacts. A typical example are the molecular processes of DNA replication, transcription, translation, cell signaling, and cell cycle control among others which

Table 6 Helper T-lymphocyte epitopes predicted from the three hypothetical proteins of *Plasmodium falciparum*

Protein	HTL peptide	Position	NN-align P. rank	Antigenicity	IL10pred	IL4pred	IFNepitope
XP_001350955.1	LSDYTNIGSSLNPDD	29–43	3.5	1.3058	Non-inducer	Non-inducer	Negative
	SDYTNIGSSLNPDDI	30–44	3.7	1.0481	Non-inducer	Inducer	Negative
	DLSDYTNIGSSLNPD	28–42	4.1	1.2586	Non-inducer	Non-inducer	Negative
XP_001351030.1	**NEAKMDSFIYQIYMM**	**91–105**	**2.8**	**1.3318**	**Inducer**	**Inducer**	**Positive**
	EAKMDSFIYQIYMMK	**92–106**	**1.9**	**1.4314**	**Inducer**	**Inducer**	**Positive**
	AKMDSFIYQIYMMKS	**93–107**	**1.8**	**1.0428**	**Inducer**	**Inducer**	**Positive**
	KMDSFIYQIYMMKSE	**94–108**	**1.7**	**1.0074**	**Inducer**	**Inducer**	**Positive**
	MDSFIYQIYMMKSEF	**95–109**	**1.5**	**1.0892**	**Inducer**	**Inducer**	**Positive**
	EYNEAKMDSFIYQIY	89–103	3.4	1.1536	Inducer	Inducer	Negative
	YNEAKMDSFIYQIYM	90–104	3.3	1.3256	Inducer	Inducer	Negative
	DSFIYQIYMMKSEFD	96–110	0.46	0.7763	Non-inducer	Inducer	Positive
	SFIYQIYMMKSEFDK	97–111	0.18	0.6950	Non-inducer	Inducer	Positive
	FIYQIYMMKSEFDKN	98–112	0.18	1.0801	Non-inducer	Inducer	Positive
	IYQIYMMKSEFDKNL	99–113	0.33	0.7683	Non-inducer	Inducer	Positive
	YQIYMMKSEFDKNLN	100–114	0.63	0.9757	Inducer	Inducer	Negative
XP_001351047.1	KYLKCFNNWYKNNFL	29–43	4.9	0.6516	Non-inducer	Inducer	Negative

IL4 interleukin 4, *IL10* interleukin 10, *IFN* interferon

are performed by large number of proteins organized by their protein–protein interactions [69]. Currently, protein–protein interaction databases are increasingly becoming important resource for investigating biological networks and pathways in cells. For functional protein–protein networks, STRINGv10.0 was used for the prediction of the interaction between our hypothetical proteins with other partners (Fig. 2). The protein frameworks are derived from various experimental data, analysis of gene, the gene fusion neighborhood, co-occurrence, and co-expression that is curated from various pathway databases [70]. The top partner proteins with an interaction score > 0.4 were applied to construct the PPI networks to query hypothetical proteins. Protein XP_001350955.1 interacts with 10 proteins: nucleoside diphosphate kinase (NDK), proliferating cell nuclear antigens (PCNA),

uncharacterized protein (PF07_0087), acidic leucine-rich nuclear phosphoprotein 32-related protein, uncharacterized protein (PFC0670c), uncharacterized protein (PFC0315c), replication factor A-related protein putative, ribonucleotide reductase small subunit, chromatin assembly factor 1 protein WD40 domain putative, and uncharacterized protein; hydrolase putative. Nucleoside diphosphate kinases are enzymes required for the synthesis of nucleoside triphosphates. Proliferating cell nuclear antigen (PCNA) plays an essential role in DNA replication and repair machinery as the processivity factor for DNA polymerase δ and ε [71]. Protein XP_001351030.1 partners with male development protein1, which is important in female gametocyte activation [72]. It also interacts with putative uncharacterized protein (MAL13P1.106) and uncharacterized protein (PF14_

Fig. 3 Screenshots from STRING server of predicted functional partners with query sequences. Top interacting partners have been displayed for **A** XP_001350955. **B** XP_001351030. **C** XP_001351047
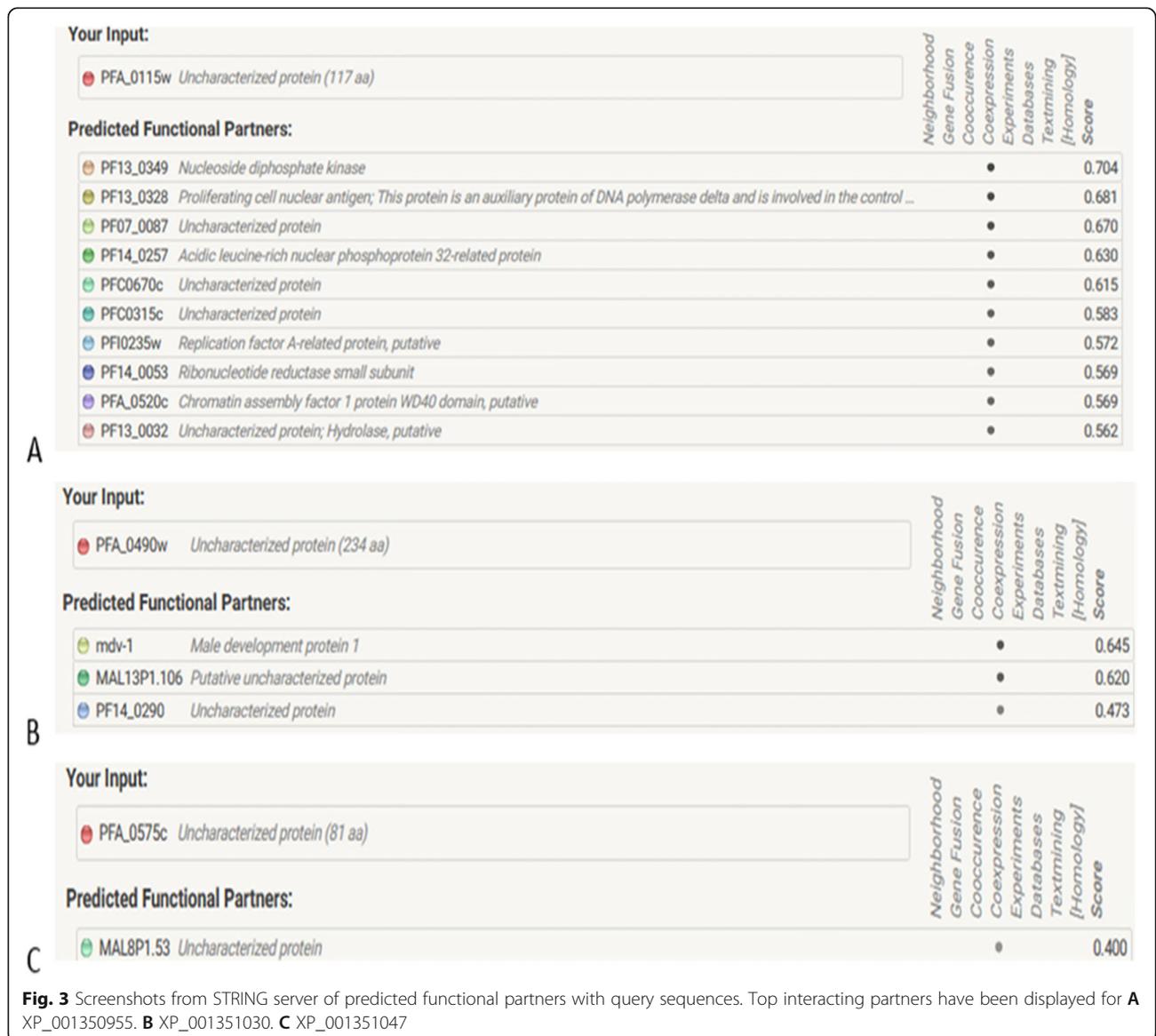
0290). Protein XP_001351047.1 is found to interact with only one protein: uncharacterized protein (MAL8P1.53). The predicted functional partner proteins, alongside their confidence scores for each hypothetical protein involved in this study, are summarized in Fig. 3. The protein–protein interactions are critical for almost every process in a living cell; therefore, information generated herein about the interactions of our HPs with other proteins could shed insight into understanding the parasite pathogenesis and can provide the basis for novel vaccine approaches.

However it is essential that the selected vaccine candidates along with their epitopes be further validated for their immunogenicity and protective efficacy experimentally if they are to be used for future vaccine development against *P. falciparum* malaria.

Conclusion

Reverse vaccinology is a promising strategy for the screening and identification of antigenic antigens with potential capacity to elicit cellular and humoral immune responses against *P. falciparum* infection. In this study, three hypothetical proteins were selected through computational methods and verified as potential vaccine candidates against *P. falciparum* malaria. We therefore recommend further in-depth immunoinformatics and structural biology approaches together with in vitro and in vivo experiments to validate their immunogenicity and protective efficacy to completely decipher the vaccine targets against malaria.

Abbreviations
RV: Reverse vaccinology; NCBI: National center for information and biotechnology; TM: Transmembrane helices; Ab: Antibody; GRAVY: Grand average of hydropathy; HTL: Helper T cells; AAP : Amino acid pair; IEDB: Immune epitope database; CTL : Cytotoxic T-lymphocytes; MHC: Major histocompatibility complex; BLAST: Basic local alignment search tool; IL: Interleukin; IFN-ɣ: Interferon gamma; ANN: Artificial neural network; SVM: Support vector machine; TMHMM: Trans membrane hidden Markov model; Mw: Molecular weight; PI: Isoelectric point; II: Instability index; EC: Extinction coefficient; HP: Hypothetical proteins; Cys: Cysteine; Try: Tryptophan; Tyr : Tyrosine; Ala: Alanine; Val: Valine; Ile : Isoleucine; NDK: Nucleoside diphosphate kinase; PCNA: Proliferating cell nuclear antigen

Authors' contributions
CA selected the scope of the study, collected, analyzed and interpreted data, wrote draft manuscript, and revised the final draft. BO designed the study, analyzed and interpreted data, and revised the final draft. AM collected, analyzed, and interpreted the data and revised final draft. GWL interpreted the data, revised the final draft, and supervised the study. All authors read and approved the final manuscript.

Availability of data and materials
All the data and material generated and analyzed in this study have been included in this manuscript.

Declarations

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable
Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Biochemistry and Sports Science, College of Natural Sciences, Makerere University, P.O. Box 7062, Kampala, Uganda. [2]Uganda Virus Reseach Institute, P.O. Box 49, Entebbe, Uganda. [3]Department of Bio-molecular Resources and Bio-lab Sciences, School of Biosecurity, Biotechnology and Laboratory Sciences (SBLS), College of Veterinary Medicine, Animal Resources and Biosecurity, Makerere University, P.O Box 7062, Kampala, Uganda.

References
1. WHO, U (2016) World malaria report. World Health Organization, pp 1–186
2. Organization, WH (2018) High burden to high impact: a targeted malaria response. World Health Organization
3. Beare NA et al (2006) Malarial retinopathy: a newly established diagnostic sign in severe malaria. Am J Trop Med Hyg 75(5):790–797. https://doi.org/10.4269/ajtmh.2006.75.790
4. Jones TR, Hoffman SL (1994) Malaria vaccine development. Clin Microbiol Rev 7(3):303–310. https://doi.org/10.1128/CMR.7.3.303
5. Coelho CH et al (2017) Advances in malaria vaccine development: report from the 2017 malaria vaccine symposium. NPJ Vaccines 2(1):34
6. Draper SJ, Angov E, Horii T, Miller LH, Srinivasan P, Theisen M, Biswas S (2015) Recent advances in recombinant protein-based malaria vaccines. Vaccine 33(52):7433–7443. https://doi.org/10.1016/j.vaccine.2015.09.093
7. Ouattara A, Barry AE, Dutta S, Remarque EJ, Beeson JG, Plowe CV (2015) Designing malaria vaccines to circumvent antigen variability. Vaccine 33(52):7506–7512. https://doi.org/10.1016/j.vaccine.2015.09.110
8. Beeson JG et al (2019) Challenges and strategies for developing efficacious and long-lasting malaria vaccines. Sci Transl Med 11(474)
9. Rts S (2015) 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. Lancet 386(9988):31–45
10. Radtke AJ, Tse SW, Zavala F (2015) From the draining lymph node to the liver: the induction and effector mechanisms of malaria-specific CD8+ T cells. Semin Immunopathol 37(3):211–220. https://doi.org/10.1007/s00281-015-0479-3
11. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shallom SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DMA, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrell B (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 419(6906):498–511. https://doi.org/10.1038/nature01097
12. Böhme U, Otto TD, Sanders M, Newbold CI, Berriman M (2019) Progression of the canonical reference malaria parasite genome from 2002–2019. Wellcome Open Res 4. https://doi.org/10.12688/wellcomeopenres.15194.1
13. Briquet S, Ourimi A, Pionneau C, Bernardes J, Carbone A, Chardonnet S, Vaquero C (2018) Identification of Plasmodium falciparum nuclear proteins by mass spectrometry and proposed protein annotation. PLoS One 13(10):e0205596. https://doi.org/10.1371/journal.pone.0205596
14. Tang Y, Meister TR, Walczak M, Pulkoski-Gross MJ, Hari SB, Sauer RT, Amberg-Johnson K, Yeh E (2019) A mutagenesis screen for essential plastid biogenesis genes in human malaria parasites. PLoS Biol 17(2):e3000136. https://doi.org/10.1371/journal.pbio.3000136
15. Bharat Siva Varma P, Adimulam YB, Kodukula S (2015) In silico functional annotation of a hypothetical protein from Staphylococcus aureus. J Infect Public Health 8(6):526–532. https://doi.org/10.1016/j.jiph.2015.03.007

16. Mohan R, Venugopal S (2012) Computational structural and functional analysis of hypothetical proteins of Staphylococcus aureus. Bioinformation 8(15):722–728. https://doi.org/10.6026/97320630008722

17. Verma A, Singh VK, Gaur S (2016) Computational based functional analysis of Bacillus phytases. Comput Biol Chem 60:53–58. https://doi.org/10.1016/j.compbiolchem.2015.11.001

18. Shahbaaz M et al (2015) In silico approaches for the identification of virulence candidates amongst hypothetical proteins of Mycoplasma pneumoniae 309. Comput Biol Chem 59(Pt A):67–80

19. Islam MS, Shahik SM, Sohel M, Patwary NIA, Hasan MA (2015) In silico structural and functional annotation of hypothetical proteins of Vibrio cholerae O139. Genomics Inform 13(2):53–59. https://doi.org/10.5808/GI.2015.13.2.53

20. Pritam M, Singh G, Swaroop S, Singh AK, Singh SP (2019) Exploitation of reverse vaccinology and immunoinformatics as promising platform for genome-wide screening of new effective vaccine candidates against Plasmodium falciparum. BMC Bioinform 19(13):468. https://doi.org/10.1186/s12859-018-2482-x

21. Singh SP, Verma V, Mishra BN (2015) Characterization of Plasmodium falciparum proteome at asexual blood stages for screening of effective vaccine candidates: an immunoinformatics approach. Immunol Immunogenet Insights 7:III.S24755

22. Sette A, Rappuoli R (2010) Reverse vaccinology: developing vaccines in the era of genomics. Immunity 33(4):530–541. https://doi.org/10.1016/j.immuni.2010.09.017

23. Serruto D, Bottomley MJ, Ram S, Giuliani MM, Rappuoli R (2012) The new multicomponent vaccine against meningococcal serogroup B, 4CMenB: immunological, functional and structural characterization of the antigens. Vaccine 30(Suppl 2):B87–B97. https://doi.org/10.1016/j.vaccine.2012.01.033

24. O'Ryan M, Stoddard J, Toneatto D, Wassil J, Dull PM (2014) A multi-component meningococcal serogroup B vaccine (4CMenB): the clinical development program. Drugs 74(1):15–30. https://doi.org/10.1007/s40265-013-0155-7

25. Dellagostin OA et al (2017) Reverse vaccinology: an approach for identifying leptospiral vaccine candidates. Int J Mol Sci 18(1):158. https://doi.org/10.3390/ijms18010158

26. Lin CS, Uboldi AD, Marapana D, Czabotar PE, Epp C, Bujard H, Taylor NL, Perugini MA, Hodder AN, Cowman AF (2014) The merozoite surface protein 1 complex is a platform for binding to human erythrocytes by Plasmodium falciparum. J Biol Chem 289(37):25655–25669. https://doi.org/10.1074/jbc.M114.586495

27. Horton P et al (2007) WoLF PSORT: protein localization predictor. Nucleic Acids Res 35(suppl_2):W585–W587

28. Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinform 8(1):4. https://doi.org/10.1186/1471-2105-8-4

29. Krogh A, Larsson B, Heijne GV, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580

30. Tusnady GE, Simon I (2001) The HMMTOP transmembrane topology prediction server. Bioinformatics 17(9):849–850. https://doi.org/10.1093/bioinformatics/17.9.849

31. Nielsen H (2017) Predicting secretory proteins with SignalP. In: Protein function prediction. Springer, pp 59–73

32. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

33. Naz A, Awan FM, Obaid A, Muhammad SA, Paracha RZ, Ahmad J, Ali A (2015) Identification of putative vaccine candidates against Helicobacter pylori exploiting exoproteome and secretome: a reverse vaccinology based approach. Infect Genet Evol 32:280–291. https://doi.org/10.1016/j.meegid.2015.03.027

34. 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(1):83–93. https://doi.org/10.1016/j.ygeno.2013.12.002

35. Dimitrov I, Naneva L, Doytchinova I, Bangov I (2014) AllergenFP: allergenicity prediction by descriptor fingerprints. Bioinformatics 30(6):846–851. https://doi.org/10.1093/bioinformatics/btt619

36. Gupta S et al (2013) Identification of B-cell epitopes in an antigen for inducing specific class of antibodies. Biol Direct 8(1):1–15

37. Gasteiger E et al (2005) Protein identification and analysis tools on the ExPASy server. In: The proteomics protocols handbook. Springer, pp 571–607

38. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157(1):105–132. https://doi.org/10.1016/0022-2836(82)90515-0

39. Shirai H et al (2014) Antibody informatics for drug discovery. Biochimica et Biophysica Acta (BBA)-Proteins and. Proteomics 1844(11):2002–2015

40. Saha S, Raghava GPS (2006) Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 65(1):40–48. https://doi.org/10.1002/prot.21078

41. EL-Manzalawy Y, Dobbs D, Honavar V (2008) Predicting linear B-cell epitopes using string kernels. J Mol Recognit 21(4):243–255. https://doi.org/10.1002/jmr.893

42. Bhasin M, Raghava GP (2004) Prediction of CTL epitopes using QM, SVM and ANN techniques. Vaccine 22(23-24):3195–3204. https://doi.org/10.1016/j.vaccine.2004.02.005

43. Nielsen M, Lund O (2009) NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. BMC Bioinform 10(1):1–10

44. Paul S et al (2016) TepiTool: a pipeline for computational prediction of T cell epitope candidates. Curr Protoc Immunol 114(1):18.19. 1–18.19. 24

45. Dhanda SK, Gupta S, Vir P, Raghava GPS (2013) Prediction of IL4 inducing peptides. Clin Dev Immunol 2013:1–9. https://doi.org/10.1155/2013/263952

46. Nagpal G et al (2017) Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. Sci Rep 7(1):1–10

47. Dhanda SK, Vir P, Raghava GP (2013) Designing of interferon-gamma inducing MHC class-II binders. Biol Direct 8(1):1–15

48. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C (2015) STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43(D1):D447–D452. https://doi.org/10.1093/nar/gku1003

49. da Costa WLO, Araújo CLA, Dias LM, Pereira LCS, Alves JTC, Araújo FA, Folador EL, Henriques I, Silva A, Folador ARC (2018) Functional annotation of hypothetical proteins from the Exiguobacterium antarcticum strain B7 reveals proteins involved in adaptation to extreme environments, including high arsenic resistance. PLoS ONE 13(6):e0198965. https://doi.org/10.1371/journal.pone.0198965

50. Goñi J, Esteban FJ, de Mendizábal N, Sepulcre J, Ardanza-Trevijano S, Agirrezabal I, Villoslada P (2008) A computational analysis of protein-protein interaction networks in neurodegenerative diseases. BMC Syst Biol 2(1):52. https://doi.org/10.1186/1752-0509-2-52

51. Gao P, Wang QP, Chen L, Huang T (2012) Prediction of human genes' regulatory functions based on proteinprotein interaction network. Protein Pept Lett 19(9):910–916. https://doi.org/10.2174/092986612802084528

52. Sen T, Verma NK (2020) Functional annotation and curation of hypothetical proteins present in a newly emerged serotype 1c of Shigella flexneri: emphasis on selecting targets for virulence and vaccine design studies. Genes 11(3):340. https://doi.org/10.3390/genes11030340

53. Mora M, Veggi D, Santini L, Pizza M, Rappuoli R (2003) Reverse vaccinology. Drug Discov Today 8(10):459–464. https://doi.org/10.1016/S1359-6446(03)02689-8

54. Oprea M, Antohe F (2013) Reverse-vaccinology strategy for designing T-cell epitope candidates for Staphylococcus aureus endocarditis vaccine. Biologicals 41(3):148–153. https://doi.org/10.1016/j.biologicals.2013.03.001

55. Kindt TJ et al (2007) Kuby immunology. Macmillan

56. Duffaud GD et al (1985) Chapter 2 Structure and Function of the Signal Peptide. In: Bronner F (ed) Current Topics in Membranes and Transport. Academic Press, pp 65–104

57. Salam MA (2009) Prospects of vaccine in leishmaniasis. Bangladesh J Med Microbiol 3(2):40–46

58. Singh SP, Mishra BN (2009) Identification and characterization of merozoite surface protein 1 epitope. Bioinformation 4(1):1–5. https://doi.org/10.6026/97320630004001

59. Gasteiger E et al (2005) Protein identification and analysis tools on the ExPASy server. The proteomics protocols handbook, pp 571–607

60. Gill SC, von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem 182(2):319–326. https://doi.org/10.1016/0003-2697(89)90602-7

61. Guruprasad K, Reddy BVB, Pandit MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng Des Sel 4(2):155–161. https://doi.org/10.1093/protein/4.2.155

62. Ikai A (1980) Thermostability and aliphatic index of globular proteins. J Biochem 88(6):1895–1898

63. Kester KE, Cummings JF, Ofori-Anyinam O, Ockenhouse CF, Krzych U, Moris P, Schwenk R, Nielsen RA, Debebe Z, Pinelis E, Juompan L, Williams J, Dowler M, Stewart VA, Wirtz RA, Dubois MC, Lievens M, Cohen J, Ballou WR, Heppner, Jr DG, RTS,S Vaccine Evaluation Group (2009) Randomized, double-blind, phase 2a trial of falciparum malaria vaccines RTS,S/AS01B and RTS,S/AS02A in malaria-naive adults: safety, efficacy, and immunologic associates of protection. J Infect Dis 200(3):337–346

64. Miura K (2016) Progress and prospects for blood-stage malaria vaccines. Expert Rev Vaccines 15(6):765–781. https://doi.org/10.1586/14760584.2016.1141680

65. Douglas AD, Baldeviano GC, Lucas CM, Lugo-Roman LA, Crosnier C, Bartholdson SJ, Diouf A, Miura K, Lambert LE, Ventocilla JA, Leiva KP, Milne KH, Illingworth JJ, Spencer AJ, Hjerrild KA, Alanine DGW, Turner AV, Moorhead JT, Edgel KA, Wu Y, Long CA, Wright GJ, Lescano AG, Draper SJ (2015) A PfRH5-based vaccine is efficacious against heterologous strain blood-stage Plasmodium falciparum infection in aotus monkeys. Cell Host Microbe 17(1):130–139. https://doi.org/10.1016/j.chom.2014.11.017

66. Payne RO, Milne KH, Elias SC, Edwards NJ, Douglas AD, Brown RE, Silk SE, Biswas S, Miura K, Roberts R, Rampling TW, Venkatraman N, Hodgson SH, Labbé GM, Halstead FD, Poulton ID, Nugent FL, de Graaf H, Sukhtankar P, Williams NC, Ockenhouse CF, Kathcart AK, Qabar AN, Waters NC, Soisson LA, Birkett AJ, Cooke GS, Faust SN, Woods C, Ivinson K, McCarthy JS, Diggs CL, Vekemans J, Long CA, Hill AVS, Lawrie AM, Dutta S, Draper SJ (2016) Demonstration of the Blood-Stage Plasmodium falciparum Controlled Human Malaria Infection Model to Assess Efficacy of the P. falciparum Apical Membrane Antigen 1 Vaccine, FMP2.1/AS01. J Infect Dis 213(11): 1743–1751. https://doi.org/10.1093/infdis/jiw039

67. Cockburn IA, Tse SW, Zavala F (2014) CD8+ T cells eliminate liver-stage Plasmodium berghei parasites without detectable bystander effect. Infect Immun 82(4):1460–1464. https://doi.org/10.1128/IAI.01500-13

68. Van Braeckel-Budimir N, Harty JT (2014) CD8 T-cell-mediated protection against liver-stage malaria: lessons from a mouse model. Front Microbiol 5:272

69. Phizicky EM, Fields S (1995) Protein-protein interactions: methods for detection and analysis. Microbiol Rev 59(1):94–123. https://doi.org/10.1128/mr.59.1.94-123.1995

70. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C (2009) STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res 37(Database issue):D412–D416. https://doi.org/10.1093/nar/gkn760

71. Kelman Z (1997) PCNA: structure, functions and interactions. Oncogene 14(6):629–640. https://doi.org/10.1038/sj.onc.1200886

72. Lal K, Delves MJ, Bromley E, Wastling JM, Tomley FM, Sinden RE (2009) Plasmodium male development gene-1 (mdv-1) is important for female sexual development and identifies a polarised plasma membrane during zygote development. Int J Parasitol 39(7):755–761. https://doi.org/10.1016/j.ijpara.2008.11.008

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.