Article

An In Silico Analysis of Malaria Pre-Erythrocytic-Stage Antigens Interpreting Worldwide Genetic Data to Suggest Vaccine Candidate Variants and Epitopes

Amed Ouattara [1,2,*], Ankit Dwivedi [3], Matthew Adams [1], Amadou Niangaly [2], Matthew B. Laurens [1], Myaing M. Nyunt [1], Christopher V. Plowe [1], Abdoulaye Djimde [2], Shannon Takala-Harrison [1] and Joana C. Silva [3]

1   Malaria Research Program, Center for Vaccine Development and Global Heath, School of Medicine, University of Maryland School of Medicine, 685 West Baltimore St., Baltimore, MD 21201, USA; madams@som.umaryland.edu (M.A.); mlaurens@som.umaryland.edu (M.B.L.); myaingnyunt@gmail.com (M.M.N.); plowe.chris@gmail.com (C.V.P.); stakala@som.umaryland.edu (S.T.-H.)
2   Malaria Research and Training Center, University of Sciences, Techniques and Technologies of Bamako, Bamako P.O. Box 1805, Mali; niangaly@icermali.org (A.N.); adjimde@icermali.org (A.D.)
3   Institute for Genome Sciences, University of Maryland School of Medicine, 670 West Baltimore St., Baltimore, MD 21201, USA; adwivedi@som.umaryland.edu (A.D.); jcsilva@som.umaryland.edu (J.C.S.)
*   Correspondence: amouattara@som.umaryland.edu

Abstract: Failure to account for genetic diversity of antigens during vaccine design may lead to vaccine escape. To evaluate the vaccine escape potential of antigens used in vaccines currently in development or clinical testing, we surveyed the genetic diversity, measured population differentiation, and performed in silico prediction and analysis of T-cell epitopes of ten such *Plasmodium falciparum* pre-erythrocytic-stage antigens using whole-genome sequence data from 1010 field isolates. Of these, 699 were collected in Africa (Burkina Faso, Cameroon, Guinea, Kenya, Malawi, Mali, and Tanzania), 69 in South America (Brazil, Colombia, French Guiana, and Peru), 59 in Oceania (Papua New Guinea), and 183 in Asia (Cambodia, Myanmar, and Thailand). Antigens surveyed include cell-traversal protein for ookinetes and sporozoites, circumsporozoite protein, liver-stage antigens 1 and 3, sporozoite surface proteins P36 and P52, sporozoite asparagine-rich protein-1, sporozoite microneme protein essential for cell traversal-2, and upregulated-in-infectious-sporozoite 3 and 4 proteins. The analyses showed that a limited number of these protein variants, when combined, would be representative of worldwide parasite populations. Moreover, predicted T-cell epitopes were identified that could be further explored for immunogenicity and protective efficacy. Findings can inform the rational design of a multivalent malaria vaccine.

Keywords: plasmodium; liver-stage; antigen; polymorphism; variants; global diversity; multivalent vaccine

1. Introduction

Malaria eradication will require powerful new tools that interrupt transmission [1–3]. Vaccination is the control measure with the greatest impact on past disease eradication efforts [4,5]. Selecting vaccine candidate variants based on laboratory strains has not been a successful strategy for designing malaria vaccines with broad efficacy [6]. Phase-2 trials of a *Plasmodium falciparum* apical membrane antigen 1 (AMA1)-based malaria vaccine [7,8], and a phase-3 trial of the RTS,S vaccine [9] have shown that vaccine efficacy is highest against homologous strains and lower against heterologous strains. This limits the overall efficacy of these subunit vaccines in endemic areas. These findings highlight the fact that the variability of malaria antigens, particularly in epitopes, represents an obstacle for development of a broadly efficacious vaccine, as has been observed with vaccines against pneumococcus [10–12] and human immunodeficiency virus (HIV) [13,14].

Over the past decade, genetic and genomic studies have been conducted to assess the diversity of *Plasmodium falciparum* antigens [15]. Most of these studies were limited to specific countries [16–18], regions [19,20], or continents [21]. Moreover, these molecular and genomic epidemiology studies were limited to a few antigens mostly present in the blood stage and involved relatively small sample sizes. A further limitation of previous studies was the lack of resolution to epitope-level diversity. We hypothesized that looking at genetic diversity of predicted epitopes may be a more informative option than overall genetic diversity within a given antigen, which may be confounded by immunologically neutral or by highly conserved protein regions that may mask the genetic diversity signal of epitopes. To address all these limitations, we aimed to assess the distribution of global diversity in ten major liver- and sporozoite-stage antigens included in the formulation of vaccines in the development pipeline. Antibodies to the cell-traversal protein for ookinetes and sporozoites (CelTOS, PF3D7_1216600) [22], the circumsporozoite protein (CSP, PF3D7_0304600) [23], and the liver-stage antigen-1 and 3 (LSA1, PF3D7_1036400; LSA3, PF3D7_0220000) [24,25] have all been associated with protection from clinical malaria in humans and other vertebrates. In addition, deletion of the genes encoding P36 (PF3D7_040440) [26], P52 (PF3D7_0404500), [27] upregulated-in-infectious-sporozoites 3 (UIS3, PF3D7_1302200), [28] and upregulated-in-infectious-sporozoites 4 (UIS4, PF3D7_1016900) [29] proteins have been shown to prevent parasite egress from the liver, inhibiting development of blood-stage infection. Finally, liver-stage infections have been prevented by deletion of the sporozoite asparagine-rich protein-1 (SAP1, PF3D7_1147000) [30], while the sporozoite microneme protein SPECT2 (PF3D7_0408700), as the name suggests, has been shown to be important for cell traversal before hepatocyte invasion [31] (Supplemental Table S1).

Many *P. falciparum* antigens, evolving under selective pressure exerted by the host's immune system, are known to be very diverse [32,33]. Understanding and characterizing selection pressures acting on protein-encoding genes could help identify variants to include in candidate vaccines with broad protection. On the other hand, geographic structuring as observed at the whole-genome level for *P. falciparum* populations, may prevent the design of a single, global malaria vaccine [34,35], limiting vaccine efficacy to specific regions defined by the geographic distribution of strains similar to the vaccine strain. The public availability of *P. falciparum* whole-genome sequences from around the world opens the door to comprehensive analyses of selection within, and population differentiation between, geographic populations. Inferences from these analyses could facilitate the design of broadly protective, multivariant malaria vaccines. By using whole-genome sequence data and associated metadata on geographic location to screen for candidate antigen variants, this study aimed to better understand each antigen's natural variation in the context of geographic distribution. By characterizing the extent of genetic diversity in genes that encode liver-stage antigens, we hoped to identify the most frequent alleles and/or conserved epitopes of each protein that may be integrated in the design of broadly protective malaria vaccines.

2. Materials and Methods

2.1. Antigen Selection

Vaccine candidate antigens selected in this study are in preclinical development or in early-stage clinical trials, excepting CSP, which forms the basis for the RTS,S vaccine, which underwent a large-scale, pilot implementation trial in Africa [36] and has been endorsed by World Health Organization (WHO).

2.2. Allelic Sequence Generation

Sequence reads from whole-genome sequencing datasets were aligned to the reference 3D7 genome using Bowtie2 [37]. Base recalibration was conducted with the MalariaGEN polymorphism panel, and single-nucleotide polymorphism (SNP) calls were made using GATK's UnifiedGenotyper v4.1.8.1 [38,39] with the major allele called for each polymorphic

site. The high-confidence SNPs were selected using the filter "DP < 12 || QUAL < 50 || FS > 14.5 || (MQ0 $\geq$ 2 && (MQ0/(1.0 $\times$ DP)) > 0.1)". For each sample, the nucleotide sequences for the ten antigen-encoding genes were reconstructed by using the 3D7 *P. falciparum* alleles as the backbone on which the SNP calls (synonymous and nonsynonymous) for each sample were introduced, using GATK's v3.7 FastaAlternateReferenceMaker [38]. The missing genotypes were considered as reference alleles. The sequences were recovered in correct orientation corresponding to the localization of antigens on the specific strand. Sequences from Brazil (*n* = 22), Burkina Faso (*n* = 56), Cambodia (*n* = 129), Colombia (*n* = 16), Cameroon (*n* = 130), French Guiana (*n* = 36), Guinea (*n* = 123), Kenya (*n* = 57), Malawi (*n* = 150), Mali (*n* = 99), Myanmar (*n* = 18), Papua New Guinea (*n* = 59), Peru (*n* = 11), Tanzania (*n* = 68), and Thailand (*n* = 36), as reported in [35] or obtained from MalariaGEN (Parasite genomic surveillance: www.malariagen.net/data/pf3k-5 (accessed on 9 March 2019)) (Supplemental Table S2), were used in the analysis. The analyses did not consider structural and copy number variations and focused only on SNP variants.

2.3. Data Analyses

2.3.1. Nucleotide Polymorphism

To assess if the candidate vaccine antigens would evolve under a neutral model, genetic diversity statistics were assessed separately for each sample set, and at the worldwide level, when warranted, using DnaSP [40]. The parameters calculated included the number of polymorphic sites (S), nucleotide diversity per synonymous (π_S) and nonsynonymous (π_N) site, overall nucleotide diversity (π), and the number of unique haplotypes for each given antigen.

2.3.2. Haplotype Diversity and Distribution

Haplotypes were based on the whole coding sequence for each gene without their untranslated regions. The frequencies of the three major haplotypes of each antigen were compared between countries and regions using a chi-square test in SAS 9.4®. The mean haplotype diversity was compared between countries and regions using the non-parametric, 2-sided Mann–Whitney test. Antigens with three haplotypes representing more than 80% of a country's parasite population were considered "priority" vaccine candidates. Immunogenic amino acid fragments that are conserved across countries could be ideal cross-protective vaccine candidates. Thus, for each antigen, conserved fragments, defined as a stretch of at least 20 amino acids without polymorphic residues in the 3D7 reference strain, were identified from the worldwide diversity dataset, and variant proportions were compared using a chi-square test in SAS 9.4®. This metric could identify short amino acid sequences that could be used in a multifragment malaria vaccine candidate.

Haplotype networks are commonly used to visualize the relationship between haplotypes from different sampling sites. Haplotype networks were generated with sequences both within sampling locations (indicating the unit of geographic sampling, e.g., country, region, continent) and for the global dataset by median-joining network analysis using PopART (Population Analysis with Reticulate Trees) software [41]. The dataset used was based only on nonsynonymous nucleotides (compared to the 3D7 reference), and the between-country, -region, and -continent frequencies of subpopulations were compared using chi-squared tests (SAS 9.4®).

2.3.3. Population Structure

To assess the null hypothesis that allele frequencies do not vary between countries, regions, or continents, we computed F_{ST} (gene fasta files), as implemented in DnaSP v6 [42], with significance assessed by permutation of sequences among geographic locations, with 10,000 replicates.

2.3.4. Cytotoxic T-Lymphocyte (CTL) Epitope Identification and Polymorphisms

To assess if polymorphisms in candidate genes were located in putative epitopes, we used the T-cell epitope prediction algorithm (NetMHCpan-3.0, http://www.cbs.dtu.dk/services/NetMHCIIpan-3.0 (accessed on 28 October 2021)) and common human leukocyte antigen (HLA) frequencies in Mali, a country in West Africa with high malaria incidence (HLA-A*23:01, HLA-A*30:01, HLA-B*35:01, DRB1_1304, DRB1_0701, DRB1_1101, DRB1_0120) [43,44] to screen all candidate gene sequences of the reference strain 3D7 for potential CD4+ and CD8+ T-cell epitopes. These HLA types are part of the three supertypes (HLA-A2, -A3, and -B7) [45,46]. A putative epitope cutoff was set with binding strengths of up to 50 nM indicating strong-binding epitopes and strengths from 51 to 500 nM indicating intermediate binders [47]. All epitopes identified through these analyses were used to screen for potential SNPs in the homologous genomic regions in 1010 samples from 15 countries included in this study.

2.3.5. B-Cell Epitopes and Polymorphisms

Structural epitopes are groups of amino acid residues that are antibody targets. Most B-cell epitopes are discontinuous, comprising from 15 to 25 nonsequential residues [48], and these conformational B-cell epitopes represent more than 90% of functional B-cell epitopes [49]. We used CBTOPE to predict conformational B-cell epitopes using the 3D7 strain [50]. The approach uses support vector machines (SVM) to predict discontinuous epitopes with an accuracy greater than 85% and a sensitivity assessed as an area under curve (AUC) of 0.9. Following B-cell epitope residue prediction, we concatenated discontinuous residues into single fragments (epitope fragment) that define individual samples and assessed their polymorphism in the complete dataset. Thus, each sample was defined by its polymorphic residues within regions orthologous to predicted B-cell epitopes in the reference 3D7. The frequencies of these epitope fragments were compared between countries and regions using a chi-square test in SAS 9.4$^{®}$.

2.3.6. Polymorphisms in Epitopes and Protein Function

We used the PROVEAN (Protein Variation Effect Analyzer) software [51] to identify SNPs predicted to be functionally important. Briefly, sequences are initially clustered based on sequence alignment scores. The scores are averaged within and across clusters to generate the PROVEAN score. If the PROVEAN score is equal to or below a predefined threshold of −2.5, the SNP(s) within the gene's coding sequence is(are) predicted to have a "deleterious" effect on protein function, while a score above this threshold is equivalent to a "neutral" effect. The −2.5 cutoff has been shown to have 81% sensitivity, 75% specificity, and 75.17% balanced accuracy using UniProt nonhuman protein variant datasets [52].

2.3.7. Secondary Predicted Structure and Solvent Accessibility

The NetSurfP 2.0 server (https://services.healthtech.dtu.dk/service.php?NetSurfP-2.0 (accessed on 30 October 2021)) was used to predict each candidate protein's structural features, including their secondary predicted structure and surface accessibility, using nucleotide sequences in fasta format. The sequence of each gene in the reference 3D7 strain was used in COACH (Consensus Approach) and COACH-D (Consensus Approach 3D) to predict proteins' structures [53,54]. This approach uses S-SITE [53] and TM-SITE [53] comparative method data combined with results from COFACTOR [55], FINDSITE [56], and ConCavity approaches [57] to predict proteins' structures and ligand binding sites.

3. Results

To identify malaria vaccine candidate variants, or, more specifically, epitopes and protein fragments, that will not be the target of allele-specific efficacy, we surveyed the genetic diversity in the genes encoding ten *P. falciparum* pre-erythrocytic-stage antigens currently in development or in clinical testing. We obtained their DNA and/or protein sequences from 1010 *P. falciparum* whole-genome sequence datasets. These datasets (Supplemental

Table S2) (Parasite genomic surveillance: www.malariagen.net/data/pf3k-5 (accessed on 9 March 2019)) were obtained from the public domain (*n* = 721) or generated in-house (*n* = 289). Each of these vaccine candidates was analyzed separately to select potential infection-blocking vaccine candidates that could enter a vaccine development pipeline.

3.1. Protein Diversity and Mode of Evolution

Malaria vaccine antigen diversity can negatively impact efficacy in endemic areas [9,58–62]. To identify proteins that are less likely to be susceptible to allele-specific efficacy, we estimated, for each locus, haplotype diversity (*Hd*) and nucleotide diversity (π), following the rationale that the proteins with the fewest variants are the least likely to lead to vaccine evasion. While *Hd* estimates the probability that two randomly sampled alleles are different, π estimates the average number of nucleotide differences per site between two randomly sampled alleles. The analyses showed that *Hd* estimates ranged from 0.21 for UIS3, the least diverse protein, to 0.99 for the most diverse protein (SAP1), and the π statistic for the ten candidate proteins screened ranged from 0.038% for UIS3 to 1.530% for CelTOS (Table 1).

Table 1. Genetic diversity parameters of 10 infection-blocking vaccine candidates.

Antigens	*Hd*	π Per 100 bp	Breaks in Sequence Conservation across Haplotypes
SAP1	0.9990	0.181	58/2940
LSA3	0.9989	0.358	16/1558
CSP	0.9952	0.940	2/397
CelTOS	0.9884	1.530	3/182
SPECT2	0.9856	0.194	7/842
LSA1	0.9845	0.338	10/1162
LSA3	0.9989	0.358	5/478
UIS4	0.7190	0.451	1/108
P36	0.4100	0.071	2/379
UIS3	0.2100	0.038	2/229

To determine if the genetic diversity observed was similarly represented across geographic regions, we estimated F_{ST} between all pairs of geographic regions (Supplemental Table S3). F_{ST} statistics estimated using samples from the four regions showed low-to-great differentiation (F_{ST} values ranging from 0005 to 0.68) between regions. Moreover, our analyses showed that for each antigen, F_{ST} values ranged from a lack of clustering to a complete subdivision at the country level (Supplemental Table S4), with UIS3 protein showing a lack of subdivision between countries (F_{ST} values ranging from 0 to 0.14).

To determine the diversity of each protein, we started by estimating haplotype diversity. Most proteins except P36 and UIS3 showed considerable polymorphism suggestive of an excess of rare alleles (Table 1).

To assess whether conserved regions exist in protein-encoding genes, we aligned all the sequences from the 15 countries and identified fragments that were invariable in the whole dataset of 1010 sequences for each antigen. Using a chi-square test for differences of proportions, there was a significant difference in the proportion of conserved fragments (number of conserved fragments divided by the reference protein length) obtained from each antigen (*p* = 0.02) with only 1 fragment identified in UIS4 and 58 fragments in SAP1 (Table 1).

3.2. Vaccine Antigen Geographic Distribution and Haplotype Diversity

Determining if protein variants in each of the candidate vaccine loci are specific to a given country or if, instead, they are distributed worldwide, together with their relative frequency, can provide invaluable information for the design of a malaria vaccine that is cross-protective. To visualize the worldwide distribution of the sequences of each of the most conserved antigens, a haplotype distribution pie chart and a plot (Figures 1 and 2)

were built. While moderate substructuring (F_{ST} values ranging from 0.05 to 0.15) and higher substructuring (F_{ST} values ranging from 0.16 to 0.25) of variants were observed for most antigens, UIS3, UIS4, and P36 had the same predominant haplotype in all continents (Supplemental Table S5). These analyses showed that the greatest polymorphism in UIS3, UIS4, and P36 was observed in Africa, with the predominant variant of each of these antigens representing, respectively, 95.02%, 33.67%, and 91.4% of all variants sampled. Overall, the most frequent antigen variants of UIS3 and UIS4 were the same in all continents, while the major variant of the P36 antigen in Africa and South America was the second most frequent in Asia and Oceania and vice versa (Figure 1).
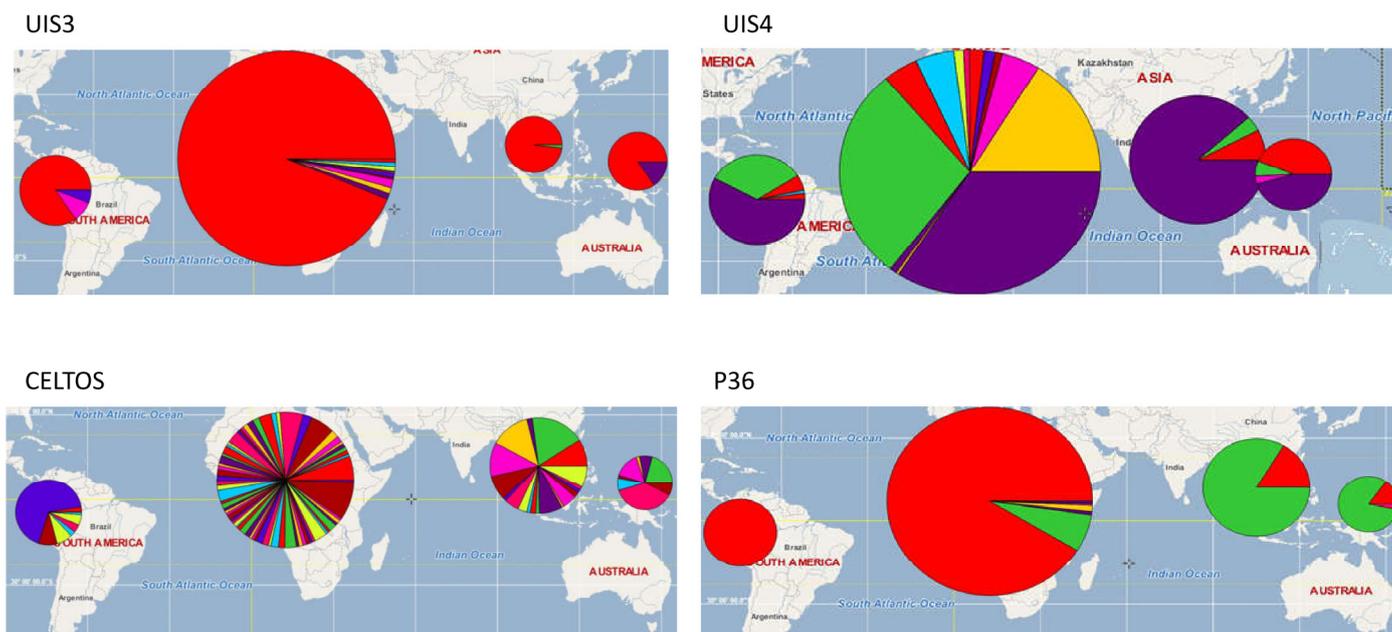

Figure 1. Worldwide distribution of amino acid sequences of four malaria vaccine candidates. Panel A represents haplotypes for upregulated-in-infectious-sporozoites 3 (UIS3), while panel B shows the worldwide distribution of UIS4 sequences. Panel C represents P36 sequences, and panel D is a world map view of the distribution of sequences of the cell-traversal protein for ookinetes and sporozoites (CelTOS). The size of the pie is proportional to the sample size used, and each slice of the pie represents a haplotype with color consistency across panels. CelTOS is the most diverse protein, while P36 is the most conserved. Whereas the major haplotype of UIS3 is consistent across continents, the major haplotype of P36 in Africa is the minor haplotype in other continents.

To identify protein variants that, when combined, could represent a large portion of worldwide diversity within these vaccine candidate proteins, we measured the prevalence of each protein variant at the country, continent, and worldwide levels. While the main variants of most proteins were less than 4% of the global parasite population studied, the worldwide frequencies of the most frequent variant for UIS3, UIS4, and P36 were 96.1%, 46.5%, and 74.1%, respectively (Figure 2), suggesting that a combination of two or three variants of these candidates could cover the global antigenic diversity. Overall, seven out of the ten proteins were highly diverse (Table 1), potentially limiting their effectiveness as malaria vaccine candidate antigens due to the high risk of vaccine-induced selection when deployed in the field, as happened with the pneumococcus vaccine PCV13 [63].
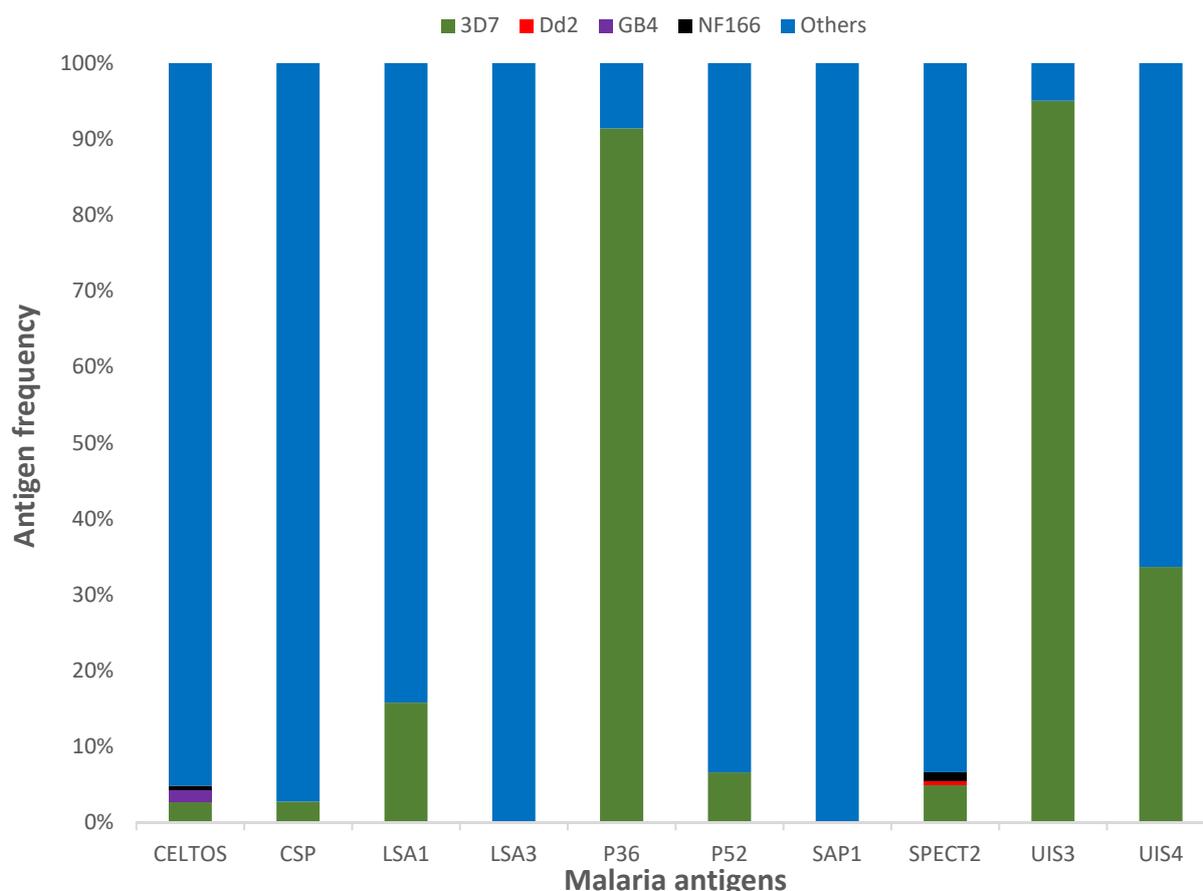
Figure 2. Frequencies of ten liver-stage malaria vaccine candidates. Antigens' names are listed on the *x* axis, while frequencies are on the *y* axis. The frequencies are stacked to 100%, and each color represents a variant or a group of variants. The laboratory strain 3D7 is represented in green.

3.3. Identification of T-Cell Epitope and Polymorphisms

To identify potential constituents of a multiepitope-based malaria vaccine containing the most conserved sequences, protein regions encoding putative CD4+ and CD8+ T-cell epitopes in the reference 3D7 alleles were characterized for polymorphism using global isolates. These putative epitope sequences were identified using the NetMHCpan-4.1 prediction algorithm and common human leukocyte antigen (HLA) frequencies in Mali [44,64], a centrally located country in West Africa, one of the regions in the world with the most intense malaria transmission. Predicted epitopes were distributed across all proteins (Supplemental Table S6), and a multiple sequence alignment of these regions was used to identify SNPs located in each epitope. Results show a great variability in the average number (sum of CD4+ and CD8+ epitopes/ number of amino acids) of medium- and strong-binding T-cell epitopes in each protein ($p < 0.001$). While P36 had the highest average number of CD4+ and CD8+ linear T-cell epitopes (9.3), few CD4+ and CD8+ (1.4) epitopes were predicted in LSA1 (Supplemental Table S6). CelTOS and P36 had completely conserved CD4+ and CD8+ epitopes across the sequences used in this study. CSP, LS1, P36, and P52 had no polymorphisms in their CD4+ T-cell epitopes, but some variation was detected in the CD8+ epitopes of these proteins. CD8+ epitopes in SAP1, SPECT2, and UIS3 were invariant in the whole dataset. Finally, UIS4 had only one CD4+ epitope with mutations. Moreover, UIS3 and UIS4 had no detectable CD8+ epitopes at the medium and strong binder cutoff values used and the HLA alleles investigated. SAP1 and LS3 had no CD4+ epitopes that were strong or medium binders, again with the same potential caveats regarding cutoff values and HLA genotypes.

Overall, for each antigen, we detected significant differences in the number of epitopes that could be used as components of a multiepitope malaria vaccine, with CelTOS and P36 having both CD4+ and CD8+ conserved T-cell epitopes.

3.4. B-Cell Conformational Epitopes and Polymorphisms

Using the conformational B-cell epitope (CBTOPE) server for conformational epitope prediction and epitope sequence analysis tools, we identified conformational epitopes of all ten candidates and generated a concatenated fragment of individual nucleotides/epitopes. The resulting major epitope-containing fragment of CelTOS represented only 7.03% of the whole dataset sequences, while two and three major epitope-containing fragments of SPECT2 and CSP represented, respectively, 87% and 61.4% of the dataset's epitope-containing fragments. LSA1 (90%), LSA3 (97%), P36 (99.8%), P52 (99.9%), and SAP1 (90.4%) each had one major epitope-containing fragment, representing at least 85% of the dataset variability. Furthermore, an assessment between countries and continents showed no differences in epitope-containing fragment distribution at the country and continent level (all *p* values greater than 0.05). Although eight out of ten of these antigens had relatively conserved B-cell epitope-containing fragments, reported polymorphic residues were scattered across antigen-encoding genes with polymorphisms located in binding sites of LSA1, LSA3, P52, CelTOS, SAP1, and CSP (Supplemental Figure S1). Interestingly, P36, UIS4, and UIS3 had no polymorphic residues in antibody binding sites.

3.5. T-Cell Epitope Biological Relevance and Solvent Accessibility

The impact of amino acid polymorphisms on protein/peptide function was evaluated with PROVEAN (Protein Variation Effect Analyzer) [51,52]. None of the 36 amino acid polymorphisms identified in putative LSA1 epitopes were predicted to have a deleterious effect on protein function. The same findings were obtained for polymorphisms identified in putative epitopes of CSP, LSA3, UIS3, SPECT2, SAP1, P52, and P36. In contrast, PROVEAN predicted that nonsynonymous changes in putative epitopes of UIS4, UIS3, SAP1, and P36 likely have deleterious effects on protein function, suggesting that these epitope-encoding protein regions can tolerate some variation despite potentially negative fitness effects on the parasite, or the proteins may be functionally constrained. Finally, all putative epitopes encoded in the *celtos* gene were conserved, potentially allowing for an immune response to target an epitope whose sequence may be under strong functional constraints.

Disordered regions in proteins/peptides (regions that do not fold into a fixed three-dimensional structure), as well as their surface accessibility to the solvent and secondary structure, were predicted using NetSurfP. Except for P36 and P52, disordered regions were distributed across all proteins, while, excepting a few protein fragments, all protein surfaces were accessible to the solvent (Supplemental Figure S2). Our results show that, with a few exceptions, putative epitopes generated from P36 and UIS4 proteins had no deleterious amino acids (Figure 3), were accessible to the solvent, and were mostly not in a helix-forming region (Table 2). Conserved CelTOS, P52, SPECT2, CSP, UIS4, UIS3, LSA1, and LSA3 epitopes were exposed to solvent but were in disordered and/or helix-forming regions (Supplemental Figure S2). Finally, all except for one peptide of the 33 SAP1 conserved peptides were in disordered regions (Supplemental Figure S2). Predicted 3D structure of malaria vaccine candidates were generated using COACH and COACH-D meta servers (Figure 4 and Supplemental Figure S3), followed by an identification of hypothetical protein-ligand binding sites. Probable binding residues [54] were identified based on their C-score, which is a measure of the confidence score of the prediction. All antigens have low-reliable putative binding site(s).
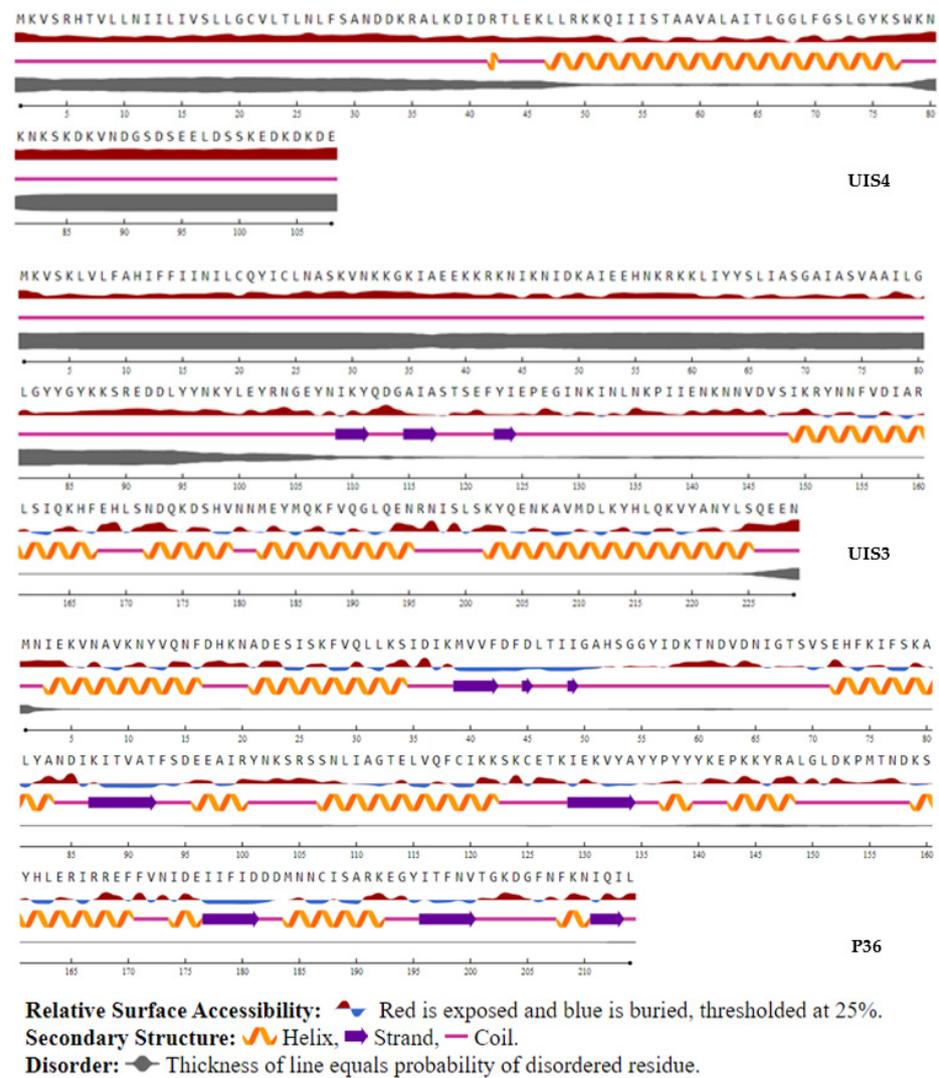
Figure 3. Predicted secondary structure and surface accessibility of P36, UIS3, and UIS4 malaria vaccine candidates.

Table 2. Relative surface accessibility of putative CD4+ and CD8+ T-cell epitopes.

Antigen	Strong-Binding CD4+ Epitope	Relative Surface Accessibility CD4+	Strong-Binding CD8+ Epitope	Relative Surface Accessibility CD8+
CELTOS	MNALRRLPVICS	Exposed	LPVICSFLVF	Exposed
	KLAILSVSSFLF	Exposed	-	-
CSP	LAILSVSSFLFV	Exposed	SSFLFVEALF *	Exposed
	ENWYSLKKNSRS	Exposed	-	-
	TNFKSLLRNLGV	Buried	-	-
	NFKSLLRNLGVS	Buried	-	-
	QTNFKSLLRNLG	Buried	-	-
LSA1	FKSLLRNLGVSE	Buried	KFIKSLFHIF *	Buried
	NFKSLLRNLGVS	Buried	-	-
	TNFKSLLRNLGV	Buried	-	-
	ISFYFILVNLLI	Buried	-	-
	SFYFILVNLLIF	Buried	-	-
LSA3	None	-	ASYVVGFFTF *	Buried
-	-	-	SYVVGFFTFS *	Buried
-	-	-	PFYSFVFDIF *	Buried
-	-	-	KVKNFVKKYK	Exposed

Table 2. *Cont.*

Antigen	Strong-Binding CD4+ Epitope	Relative Surface Accessibility CD4+	Strong-Binding CD8+ Epitope	Relative Surface Accessibility CD8+
LSA3	-	-	KVDKNNKVPK *	Exposed
-	-	-	KTRKKAQRPK *	Buried
-	-	-	KVFAAPFISA *	Buried
-	-	-	KINKYFFLIK	Exposed
-	IRYNKSRSSNLI	Buried	-	-
-	AIRYNKSRSSNL	Buried	-	-
-	KFVQLLKSIDIK	Buried	-	-
-	RYNKSRSSNLIA	Buried	-	-
P36	FVQLLKSIDIKM	Buried	-	-
-	AIRYNKSRSSNL	Buried	KSKCETKIEK	Buried
-	EAIRYNKSRSSN	Buried	-	-
-	EEAIRYNKSRSS	Buried	-	-
-	IRYNKSRSSNLI	Buried	-	-
-	SKFVQLLKSIDI	Buried	-	-
-	MCYHFTMKRKKL	Exposed	-	-
-	HMCYHFTMKRKK	Exposed	-	-
-	NLFGLSSSKYIL	Buried	-	-
-	QNLFGLSSSKYI	Exposed	-	-
-	NININFVCSNVI	Buried	KYILFNNFLI	Buried
-	ININFVCSNVIQ	Buried	ILFNNFLILF *	Buried
P52	CYHFTMKRKKLF	Exposed	VYFIFLSFII *	Exposed
-	YHFTMKRKKLFV	Exposed	KVKHIMRINI	Buried
-	LFGLSSSKYILF	Buried	RTRTFWQNLF	Exposed
-	GTMIIYTKNINS	Buried	KLSRNHSFSS	Buried
-	MIIYTKNINSLM	Buried	NPSNCFHDVY	Buried
-	TMIIYTKNINSL	Buried	-	-
-	VGTMIIYTKNIN	Buried	-	-
-	FGLSSSKYILFN	Buried	-	-
-	-	-	VKYFNKPIQF	Exposed
-	-	-	YKYIQNIILF	Buried
-	-	-	YFMPKNDLNF	Buried
-	-	-	KYIQNIILFL	Buried
-	-	-	NYMPQNYYHI	Buried
SAP1	None	-	RIFFSFFSYF	Buried
-	-	-	RFKLTCNFKF	Buried
-	-	-	KLKNFFLNYK	Buried
-	-	-	KMTKNYNINA	Exposed
-	-	-	YTRAVWLLKK	Buried
-	-	-	MPKNDLNFIF	Buried
-	-	-	MPQNYYHINY	Buried
-	KLRILKKHYYVV *	Exposed	LYFIGIGYNL	Buried
-	LRILKKHYYVF *	Exposed	IYVLCVDTTI	Buried
SPECT2	MKLRILKKHYYV *	Exposed	KRSKKTFLVK	Buried
-	MKLRILKKHYYV *	Exposed	KVVMFGFSLK	Buried
-	KLRILKKHYYVV *	Exposed	RSKKTFLVKS	Buried
-	LRILKKHYYVVF *	Exposed	KKIKHSFNLA	Exposed
-	-	-	YIPWDKTTAY	Buried
-	-	-	-	-
-	-	-	-	-
-	KYHLQKVYANYL *	Buried	-	-
-	YHLQKVYANYLS *	Buried	-	-
-	MEYMQKFVQGLQ *	Buried	-	-
-	NMEYMQKFVQGL *	Buried	-	-

Table 2. *Cont.*

Antigen	Strong-Binding CD4+ Epitope	Relative Surface Accessibility CD4+	Strong-Binding CD8+ Epitope	Relative Surface Accessibility CD8+
UIS3	NNMEYMQKFVQG *	Buried	None	-
-	VNNMEYMQKFVQ *	Buried	-	-
-	LIYYSLIASGAI *	Exposed	-	-
-	IYYSLIASGAIA *	Exposed	-	-
	KQIIISTAAVAL *	Exposed	-	-
	QIIISTAAVALA *	Exposed	-	-
UIS4	RTLEKLLRKKQI *	Exposed	None	-
	DRTLEKLLRKKQ	Exposed	-	-
	LEKLLRKKQII *	Exposed	-	-

* Has one or more mutations relative to 3D7 epitope sequence. <u>Underlined</u>: mutations are predicted to influence peptide function.
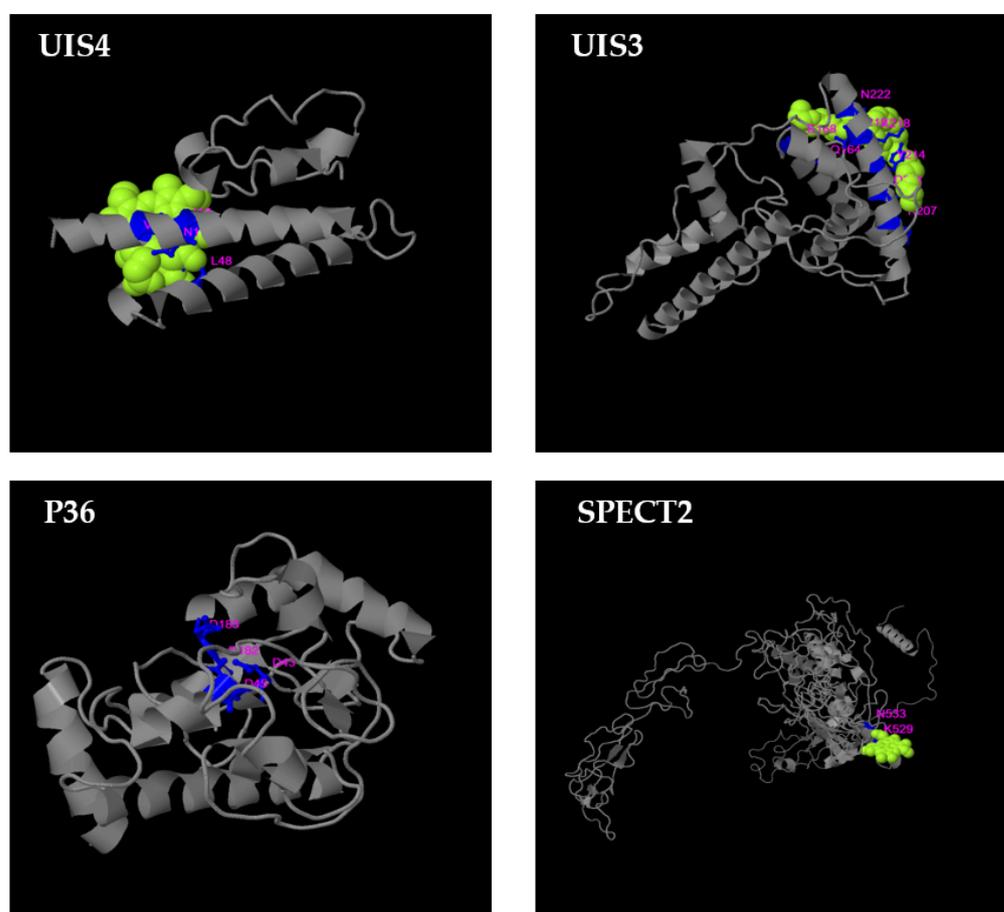
Figure 4. Predicted 3D structure of four malaria vaccine candidates with hypothetical protein–ligand binding sites shown in light purple.

4. Discussion

This study characterized the genetic diversity in ten pre-erythrocytic *P. falciparum* antigens that are currently included in the formulation of malaria vaccines in clinical development. Thanks to the public availability of whole-genome data for thousands of *P. falciparum* isolates from around the world [65], it is now possible to conduct genome-wide analyses in silico without the need to generate additional sequence data, especially from geographically distributed regions that represent diverse malaria transmission patterns. Malaria vaccinologists thus have at their disposal tools and information that can be used to inform vaccine design, long before field trials are planned. To generate a baseline

understanding of the background genetic variation in promising vaccine antigens, we investigated the geographic distribution of genetic variants of 10 putatively sporozoite-/liver-stage antigens. As some antigens are expressed during both liver and blood stages, we aimed to evaluate their potential effectiveness in preventing infection as well as any potential synergistic role they could play in preventing blood-stage disease [66].

The *P. falciparum* liver stage is an ideal parasite developmental stage for the investigation of potential vaccine targets as efficacy directed against this stage can provide protection against both infection and disease [67]. While the most advanced malaria vaccine candidate [68] is based on CSP, the major protein expressed on the sporozoite surface, CSP diversity [9] may have led, in part, to low efficacy in the field [69,70]. While new immunization approaches aiming to boost cellular and humoral responses are being used to improve RTS,S-induced protection [71–73], multiple additional liver-stage antigens are still in the vaccine development pipeline [74–76].

We have identified the specific variants of UIS3, UIS4, and P36 that cover worldwide diversity and, if appropriately immunogenic, could be used to formulate a cross-protective malaria vaccine. Furthermore, the lack of polymorphisms in binding and structurally relevant sites within these antigens suggests that potential mutations are highly detrimental to parasite fitness, which may prevent emergence of vaccine-resistant strains [77,78].

Molecular epidemiology and population genetic analyses from a total of 1010 allelic sequences indicated that UIS3 with one variant, P36 with one variant, and UIS4 with three variants represented more than 90% of the worldwide diversity and were the most conserved of the ten proteins analyzed. Importantly, the allele in the 3D7 reference strain, on which in-development vaccines incorporating these antigens are based [27,75,79], was the predominate variant in all countries. Consistent with published reports [20,80,81], our analyses of worldwide samples have shown high nucleotide diversity in African samples compared to Asian, South American, and Oceanic samples. In addition, haplotype diversity and the distribution of the frequencies of major alleles were higher in Africa compared to the rest of the world. Many factors could explain these high frequencies. Although malaria incidence has declined worldwide since 2010, Africa still has the highest burden of the disease [82]. The high transmission intensity [82], the large effective population size of *P. falciparum* [83], and selective pressure due to acquired immunity may explain the high diversity observed in Africa. In addition, as multiplicity of infection is more frequent in Africa compared to other continents [84,85], the probability of mosquitoes carrying multiple alleles is increased, potentiating more intragenic recombination leading to more diversity in the parasite population. However, among the 10 liver-stage antigens evaluated, only UIS3, UIS4, and P36 were conserved across countries and continents, making them ideal for vaccine design.

Additional analyses based on F_{ST}, haplotype geographic distribution, supported a low-to-moderate subdivision of these three antigens by country and continent. Parallel to our findings, previous studies [80,81] that focused on the erythrocyte stage of parasite development also found subpopulations and substructuring at the country and continent levels. These population subdivisions could be explained by the same immune mechanism happening during the liver and erythrocyte stages.

Although P36 showed significant differentiation between continents, this substructuring at the continent level was clearly explained by the high prevalence of the 3D7 variant of P36 in Africa compared to Asia and the Americas. However, in Asia and the Americas, the 3D7 variant of P36 was the second most frequent variant, which may allow the design of a bivalent P36-based vaccine that would cover at least 80% of the worldwide parasite population. As studies have shown that polymorphic regions of proteins may be involved in immune escape [60,86,87], the high haplotype diversity and the relatively high abundance of minor variants in seven out of the ten antigen candidates may make it difficult to design a global vaccine based on those candidates.

As the expression of full-length malaria antigens has been challenging [88], the consideration of epitope-based vaccines has become more prevalent [89]. Contributing to

the popularity of epitope-based vaccines is the fact that protein expression systems have various limitations when expressing *P. falciparum* proteins. The lack of post-translational modifications in addition to its inefficiency in expressing higher-molecular-weight proteins has limited the use of *Escherichia coli* in malaria vaccinology [90]. Although insect cell and yeast expression systems are attractive for their post-translational modifications and large yield, and mammalian cell expression systems are appealing because of their ability to recognize all protein synthesis signals [90,91], these systems are very expensive, limiting their use in initial antigen screening. Thus, fragment- or epitope-based vaccines comprising UIS3, UIS4, and P36 fragments may be good alternatives to whole proteins, if they can elicit the necessary inhibitory capacity.

CD8+ T cells play a critical role in an effective immune response against pre-erythrocytic-stage malaria parasites [92]. This protection is dependent on CD4+ T cells [93,94] and may also require neutralizing antibodies [95]. Furthermore, studies have shown that polymorphisms in relevant epitopes could alter their binding to receptors and affect protein recognition by the immune system [96]. To identify epitopes that could trigger effective B- and T-cell responses and thus replace whole proteins in a multivalent epitope-based vaccine, we screened each vaccine candidate to select linear T-cell epitopes and linear and conformational B-cell epitopes. The P36 antigen had strong-binding, highly conserved CD4+ and CD8+ epitopes; while the binding strength of UIS3 and UIS4 epitopes was lower than that of P36, they contained few polymorphisms. The selection of high-score B-cell epitopes using CBTOPE showed that P36, UIS3, and UIS4 were the three candidates without polymorphisms in any of the predicted epitopes. Although antibodies may not recognize these targets (lack of antigenicity), the design of a vaccine with immunogenic B- and T-cell epitopes could be based on these three antigens by covalently attaching the candidate protein(s) to an immunogenic carrier. In vitro microarray studies are being conducted to characterize and assess the immunogenicity of these candidates in humans.

Although ex vivo assays are ideal to evaluate the phenotype and function of T cells in peripheral blood along with antibody response, in silico analyses could be used as screening techniques to rule out peptides with a low binding affinity for their targeted ligand and those with polymorphic residues that are detrimental to the receptor and ligand interaction. Hence, binding residues were predicted and polymorphisms that may have an impact on protein binding to its receptors were identified. These analyses indicate P36 was the single antigen with the strongest binding peptides, and the residues of these peptides were conserved across samples from all continents surveyed.

Although a large dataset was used to assess the worldwide diversity of these ten liver-stage candidates, the recent release of Pf7k (www.malariagen.net/pf7k (accessed on 9 March 2019)), which includes more sequences from these countries, may help improve haplotype frequency estimates. Hence, additional analyses are being conducted to validate P36 findings. Major histocompatibility complex (MHC) and HLA [97] polymorphisms are the major hurdles in the prediction of T-cell epitopes [98]. Thus, a limitation of this study was the limited selection of HLA genotypes used, which cannot be considered representative of the worldwide HLA distribution. However, three supertypes, including the ones used in this study, have been shown to cover at least 88% of the worldwide diversity. As the reliability and accuracy of the different prediction algorithms used could only be confirmed by in vitro and/or in vivo testing, we are conducting in vitro invasion inhibition assays and ex vivo microarray assays using serum from infected patients to assess their functional relevance. Finally, it is noteworthy to mention that B-cell epitope prediction is usually less reliable than T-cell epitope prediction [99].

Overall, our findings suggest that among all ten antigens tested, P36, UIS3, and UIS4 were the top candidates for a successful strain-transcending vaccine based on protein residue conservation and the existence of many B- and/or T-cell epitopes. However, P36 stands out as the best candidate for the conservation of its epitopes and the lack of polymorphisms at functionally relevant residue positions. The essential role the protein plays during hepatocyte invasion [27] makes it a good target for a liver-stage vaccine, as a

knockout of the gene encoding P36 has led to parasite developmental arrest at the early stage of hepatocyte infection [27]. Since preclinical testing in the mid-2000s [26], new data have not been available on its use in early clinical trials. We have generated full proteins and peptides of P36 to assess its relevance in ex vivo assays and immunogenicity in in vivo analyses.

5. Conclusions

We have presented herein findings from in silico analyses for the screening and selection of malaria vaccine candidates. These results indicate that instead of CelTOS, CSP, LSA1, LSA3, SPECT2, or SAP1, a few variants of P36, UIS3, or UIS4 could be used to cover the worldwide diversity of these proteins and, were they to induce a protective response, prevent vaccine escape when the vaccine is deployed in the field. Moreover, epitope prediction and analysis, in addition to ongoing functional assays, will validate the use of these approaches in epitope-based vaccine design. These approaches applied systematically may accelerate malaria vaccine discovery.

References

1. Tran, T.M.; Portugal, S.; Draper, S.J.; Crompton, P.D. Malaria Vaccines: Moving Forward After Encouraging First Steps. *Curr. Trop. Med. Rep.* **2015**, *2*, 1–3. [CrossRef] [PubMed]
2. Tanner, M.; de Savigny, D. Malaria eradication back on the table. *Bull. World Health Organ.* **2008**, *86*, 82. [CrossRef] [PubMed]
3. mal ERACGoV. A research agenda for malaria eradication: Vaccines. *PLoS Med.* **2011**, *8*, e1000398. [CrossRef]
4. Henderson, D.A. The miracle of vaccination. *Notes Rec. R. Soc. Lond.* **1997**, *51*, 235–245. [CrossRef] [PubMed]
5. Delany, I.; Rappuoli, R.; De, G.E. Vaccines for the 21st century. *EMBO Mol. Med.* **2014**, *6*, 708–720. [CrossRef] [PubMed]
6. Genton, B.; Al-Yaman, F.; Betuela, I.; Anders, R.F.; Saul, A.; Baea, K.; Mellombo, M.; Taraika, J.; Brown, G.V.; Pye, D.; et al. Safety and immunogenicity of a three-component blood-stage malaria vaccine (MSP1, MSP2, RESA) against Plasmodium falciparum in Papua New Guinean children. *Vaccine* **2003**, *22*, 30–41. [CrossRef]

15 of 18

7. Thera, M.A.; Doumbo, O.K.; Coulibaly, D.; Laurens, M.B.; Ouattara, A.; Kone, A.K.; Guindo, A.B.; Traore, K.; Traore, I.; Kouriba, B.; et al. A field trial to assess a blood-stage malaria vaccine. *N. Engl. J. Med.* **2011**, *365*, 1004–1013. [CrossRef]

8. Ouattara, A.; Takala-Harrison, S.; Thera, M.A.; Coulibaly, D.; Niangaly, A.; Saye, R.; Tolo, Y.; Dutta, S.; Heppner, D.G.; Soisson, L.; et al. Molecular basis of allele-specific efficacy of a blood-stage malaria vaccine: Vaccine development implications. *J. Infect. Dis.* **2013**, *207*, 511–519. [CrossRef]

9. Neafsey, D.E.; Juraska, M.; Bedford, T.; Benkeser, D.; Valim, C.; Griggs, A.; Lievens, M.; Abdulla, S.; Adjei, S.; Agbenyega, T.; et al. Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. *N. Engl. J. Med.* **2015**, *373*, 2025–2037. [CrossRef]

10. Obaro, S.K. Confronting the pneumococcus: A target shift or bullet change? *Vaccine* **2000**, *19*, 1211–1217. [CrossRef]

11. Obaro, S.K. The new pneumococcal vaccine. *Clin. Microbiol. Infect.* **2002**, *8*, 623–633. [CrossRef] [PubMed]

12. Diggle, M.A.; Clarke, S.C. Increased genetic diversity of Neisseria meningitidis isolates after the introduction of meningococcal serogroup C polysaccharide conjugate vaccines. *J. Clin. Microbiol.* **2005**, *43*, 4649–4653. [CrossRef] [PubMed]

13. Rolland, M.; Tovanabutra, S.; de Camp, A.C.; Frahm, N.; Gilbert, P.B.; Sanders-Buell, E.; Heath, L.; Magaret, C.A.; Bose, M.; Bradfield, A.; et al. Genetic impact of vaccination on breakthrough HIV-1 sequences from the STEP trial. *Nat. Med.* **2011**, *17*, 366–371. [CrossRef] [PubMed]

14. Edlefsen, P.T.; Rolland, M.; Hertz, T.; Tovanabutra, S.; Gartland, A.J.; de Camp, A.C.; Magaret, C.A.; Ahmed, H.; Gottardo, R.; Juraska, M.; et al. Comprehensive sieve analysis of breakthrough HIV-1 sequences in the RV144 vaccine efficacy trial. *PLoS Comput. Biol.* **2015**, *11*, e1003973. [CrossRef] [PubMed]

15. Barry, A.E.; Schultz, L.; Buckee, C.O.; Reeder, J.C. Contrasting population structures of the genes encoding ten leading vaccine-candidate antigens of the human malaria parasite, Plasmodium falciparum. *PLoS ONE* **2009**, *4*, e8497. [CrossRef] [PubMed]

16. Gandhi, K.; Thera, M.A.; Coulibaly, D.; Traore, K.; Guindo, A.B.; Ouattara, A.; Takala-Harrison, S.; Berry, A.A.; Doumbo, O.K.; Plowe, C.V. Variation in the circumsporozoite protein of Plasmodium falciparum: Vaccine development implications. *PLoS ONE* **2014**, *9*, e101783. [CrossRef] [PubMed]

17. Barry, A.E.; Schultz, L.; Senn, N.; Nale, J.; Kiniboro, B.; Siba, P.M.; Mueller, I.; Reeder, J.C. High levels of genetic diversity of Plasmodium falciparum populations in Papua New Guinea despite variable infection prevalence. *Am. J. Trop. Med. Hyg.* **2013**, *88*, 718–725. [CrossRef]

18. Amambua-Ngwa, A.; Tetteh, K.K.; Manske, M.; Gomez-Escobar, N.; Stewart, L.B.; Deerhake, M.E.; Cheeseman, I.H.; Newbold, C.I.; Holder, A.A.; Knuepfer, E.; et al. Population genomic scan for candidate signatures of balancing selection to guide antigen characterization in malaria parasites. *PLoS Genet.* **2012**, *8*, e1002992. [CrossRef]

19. Mobegi, V.A.; Duffy, C.W.; Amambua-Ngwa, A.; Loua, K.M.; Laman, E.; Nwakanma, D.C.; MacInnis, B.; Aspeling-Jones, H.; Murray, L.; Clark, T.G.; et al. Genome-wide analysis of selection on the malaria parasite Plasmodium falciparum in West African populations of differing infection endemicity. *Mol. Biol. Evol.* **2014**, *31*, 1490–1499. [CrossRef]

20. Duffy, C.W.; Amambua-Ngwa, A.; Ahouidi, A.D.; Diakite, M.; Awandare, G.A.; Ba, H.; Tarr, S.J.; Murray, L.; Stewart, L.B.; D'Alessandro, U.; et al. Multi-population genomic analysis of malaria parasites indicates local selection and differentiation at the gdv1 locus regulating sexual development. *Sci. Rep.* **2018**, *8*, 15763. [CrossRef]

21. Shelton, J.M.; Corran, P.; Risley, P.; Silva, N.; Hubbart, C.; Jeffreys, A.; Rowlands, K.; Craik, R.; Cornelius, V.; Hensmann, M.; et al. Genetic determinants of anti-malarial acquired immunity in a large multi-centre study. *Malar. J.* **2015**, *14*, 333. [CrossRef] [PubMed]

22. Bergmann-Leitner, E.S.; Mease, R.M.; De La Vega, P.; Savranskaya, T.; Polhemus, M.; Ockenhouse, C.; Angov, E. Immunization with pre-erythrocytic antigen CelTOS from Plasmodium falciparum elicits cross-species protection against heterologous challenge with Plasmodium berghei. *PLoS ONE* **2010**, *5*, e12294. [CrossRef]

23. John, C.C.; Moormann, A.M.; Pregibon, D.C.; Sumba, P.O.; McHugh, M.M.; Narum, D.L.; Lanar, D.E.; Schluchter, M.D.; Kazura, J.W. Correlation of high levels of antibodies to multiple pre-erythrocytic Plasmodium falciparum antigens and protection from infection. *Am. J. Trop. Med. Hyg.* **2005**, *73*, 222–228. [CrossRef] [PubMed]

24. John, C.C.; Tande, A.J.; Moormann, A.M.; Sumba, P.O.; Lanar, D.E.; Min, X.M.; Kazura, J.W. Antibodies to pre-erythrocytic Plasmodium falciparum antigens and risk of clinical malaria in Kenyan children. *J. Infect. Dis.* **2008**, *197*, 519–526. [CrossRef] [PubMed]

25. Bang, G.; Prieur, E.; Roussilhon, C.; Druilhe, P. Pre-Clinical Assessment of Novel Multivalent MSP3 Malaria Vaccine Constructs. *PLoS ONE* **2011**, *6*, e28165. [CrossRef]

26. van Dijk, M.R.; Douradinha, B.; Franke-Fayard, B.; Heussler, V.; van Dooren, M.W.; van, S.B.; van Gemert, G.J.; Sauerwein, R.W.; Mota, M.M.; Waters, A.P.; et al. Genetically attenuated, P36p-deficient malarial sporozoites induce protective immunity and apoptosis of infected liver cells. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12194–12199. [CrossRef]

27. Van Buskirk, K.M.; O'Neill, M.T.; De La, V.; Maier, A.G.; Krzych, U.; Williams, J.; Dowler, M.G.; Sacci, J.B., Jr.; Kangwanrangsan, N.; Tsuboi, T.; et al. Preerythrocytic, live-attenuated Plasmodium falciparum vaccine candidates by design. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 13004–13009. [CrossRef]

28. Mueller, A.K.; Labaied, M.; Kappe, S.H.; Matuschewski, K. Genetically modified Plasmodium parasites as a protective experimental malaria vaccine. *Nature* **2005**, *433*, 164–167. [CrossRef]

29. Mueller, A.K.; Camargo, N.; Kaiser, K.; Andorfer, C.; Frevert, U.; Matuschewski, K.; Kappe, S.H. Plasmodium liver stage developmental arrest by depletion of a protein at the parasite-host interface. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 3022–3027. [CrossRef]

30. Aly, A.S.; Mikolajczak, S.A.; Rivera, H.S.; Camargo, N.; Jacobs-Lorena, V.; Labaied, M.; Coppens, I.; Kappe, S.H. Targeted deletion of SAP1 abolishes the expression of infectivity factors necessary for successful malaria parasite liver infection. *Mol. Microbiol.* **2008**, *69*, 152–163. [CrossRef]

31. Kappe, S.H.; Buscaglia, C.A.; Bergman, L.W.; Coppens, I.; Nussenzweig, V. Apicomplexan gliding motility and host cell invasion: Overhauling the motor model. *Trends Parasitol.* **2004**, *20*, 13–16. [CrossRef] [PubMed]

32. Takala, S.L.; Escalante, A.A.; Branch, O.H.; Kariuki, S.; Biswas, S.; Chaiyaroj, S.C.; Lal, A.A. Genetic diversity in the Block 2 region of the merozoite surface protein 1 (MSP-1) of Plasmodium falciparum: Additional complexity and selection and convergence in fragment size polymorphism. *Infect. Genet. Evol.* **2006**, *6*, 417–424. [CrossRef] [PubMed]

33. Takala, S.L.; Coulibaly, D.; Thera, M.A.; Batchelor, A.H.; Cummings, M.P.; Escalante, A.A.; Ouattara, A.; Traore, K.; Niangaly, A.; Djimde, A.A.; et al. Extreme polymorphism in a vaccine antigen and risk of clinical malaria: Implications for vaccine development. *Sci. Transl. Med.* **2009**, *1*, 2ra5. [CrossRef] [PubMed]

34. Neafsey, D.E.; Schaffner, S.F.; Volkman, S.K.; Park, D.; Montgomery, P.; Milner, D.A., Jr.; Lukens, A.; Rosen, D.; Daniels, R.; Houde, N.; et al. Genome-wide SNP genotyping highlights the role of natural selection in Plasmodium falciparum population divergence. *Genome Biol.* **2008**, *9*, R171. [CrossRef] [PubMed]

35. Moser, K.A.; Drabek, E.F.; Dwivedi, A.; Stucke, E.M.; Crabtree, J.; Dara, A.; Shah, Z.; Adams, M.; Li, T.; Rodrigues, P.T.; et al. Strains used in whole organism Plasmodium falciparum vaccine trials differ in genome structure, sequence, and immunogenic potential. *Genome Med.* **2020**, *12*, 6. [CrossRef]

36. Adepoju, P. RTS,S malaria vaccine pilots in three African countries. *Lancet* **2019**, *393*, 1685. [CrossRef]

37. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef]

38. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef]

39. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; Del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **2013**, *43*, 1–33. [CrossRef]

40. Rozas, J.; Ferrer-Mata, A.; Sanchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sanchez-Gracia, A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [CrossRef]

41. Bandelt, H.J.; Forster, P.; Rohl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **1999**, *16*, 37–48. [CrossRef] [PubMed]

42. Rozas, J. DNA sequence polymorphism analysis using DnaSP. *Methods Mol. Biol.* **2009**, *537*, 337–350. [CrossRef] [PubMed]

43. Lyke, K.E.; Fernandez-Vina, M.A.; Cao, K.; Hollenbach, J.; Coulibaly, D.; Kone, A.K.; Guindo, A.; Burdett, L.A.; Hartzman, R.J.; Wahl, A.R.; et al. Association of HLA alleles with Plasmodium falciparum severity in Malian children. *Tissue Antigens* **2011**, *77*, 562–571. [CrossRef] [PubMed]

44. Cao, K.; Moormann, A.M.; Lyke, K.E.; Masaberg, C.; Sumba, O.P.; Doumbo, O.K.; Koech, D.; Lancaster, A.; Nelson, M.; Meyer, D.; et al. Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci. *Tissue Antigens* **2004**, *63*, 293–325. [CrossRef]

45. Sette, A.; Sidney, J. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics* **1999**, *50*, 201–212. [CrossRef]

46. Sette, A.; Sidney, J. HLA supertypes and supermotifs: A functional perspective on HLA polymorphism. *Curr. Opin. Immunol.* **1998**, *10*, 478–482. [CrossRef]

47. Sette, A.; Vitiello, A.; Reherman, B.; Fowler, P.; Nayersina, R.; Kast, W.M.; Melief, C.J.; Oseroff, C.; Yuan, L.; Ruppert, J.; et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J. Immunol.* **1994**, *153*, 5586–5592.

48. Rubinstein, N.D.; Mayrose, I.; Halperin, D.; Yekutieli, D.; Gershoni, J.M.; Pupko, T. Computational characterization of B-cell epitopes. *Mol. Immunol.* **2008**, *45*, 3477–3489. [CrossRef]

49. Sivalingam, G.N.; Shepherd, A.J. An analysis of B-cell epitope discontinuity. *Mol. Immunol.* **2012**, *51*, 304–309. [CrossRef]

50. Ansari, H.R.; Raghava, G.P. Identification of conformational B-cell Epitopes in an antigen from its primary sequence. *Immunome Res.* **2010**, *6*, 6. [CrossRef]

51. Choi, Y.; Chan, A.P. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **2015**, *31*, 2745–2747. [CrossRef] [PubMed]

52. Choi, Y.; Sims, G.E.; Murphy, S.; Miller, J.R.; Chan, A.P. Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* **2012**, *7*, e46688. [CrossRef] [PubMed]

53. Yang, J.; Roy, A.; Zhang, Y. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics* **2013**, *29*, 2588–2595. [CrossRef] [PubMed]

54. Wu, Q.; Peng, Z.; Zhang, Y.; Yang, J. COACH-D: Improved protein-ligand binding sites prediction with refined ligand-binding poses through molecular docking. *Nucleic Acids Res.* **2018**, *46*, W438–W442. [CrossRef]

55. Roy, A.; Yang, J.; Zhang, Y. COFACTOR: An accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res.* **2012**, *40*, W471–W477. [CrossRef]

56. Brylinski, M.; Skolnick, J. A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 129–134. [CrossRef]

57. Capra, J.A.; Laskowski, R.A.; Thornton, J.M.; Singh, M.; Funkhouser, T.A. Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. *PLoS Comput. Biol.* **2009**, *5*, e1000585. [CrossRef]

58. Ouattara, A.; Barry, A.E.; Dutta, S.; Remarque, E.J.; Beeson, J.G.; Plowe, C.V. Designing malaria vaccines to circumvent antigen variability. *Vaccine* **2015**, *33*, 7506–7512. [CrossRef]

59. Ouattara, A.; Mu, J.; Takala-Harrison, S.; Saye, R.; Sagara, I.; Dicko, A.; Niangaly, A.; Duan, J.; Ellis, R.D.; Miller, L.H.; et al. Lack of allele-specific efficacy of a bivalent AMA1 malaria vaccine. *Malar. J.* **2010**, *9*, 175. [CrossRef]

60. Ouattara, A.; Niangaly, A.; Adams, M.; Coulibaly, D.; Kone, A.K.; Traore, K.; Laurens, M.B.; Tolo, Y.; Kouriba, B.; Diallo, D.A.; et al. Epitope-based sieve analysis of Plasmodium falciparum sequences from a FMP2.1/AS02A vaccine trial is consistent with differential vaccine efficacy against immunologically relevant AMA1 variants. *Vaccine* **2020**, *38*, 5700–5706. [CrossRef]

61. Polley, S.D.; Tetteh, K.K.; Lloyd, J.M.; Akpogheneta, O.J.; Greenwood, B.M.; Bojang, K.A.; Conway, D.J. Plasmodium falciparum merozoite surface protein 3 is a target of allele-specific immunity and alleles are maintained by natural selection. *J. Infect. Dis.* **2007**, *195*, 279–287. [CrossRef] [PubMed]

62. Kang, J.M.; Moon, S.U.; Kim, J.Y.; Cho, S.H.; Lin, K.; Sohn, W.M.; Kim, T.S.; Na, B.K. Genetic polymorphism of merozoite surface protein-1 and merozoite surface protein-2 in Plasmodium falciparum field isolates from Myanmar. *Malar. J.* **2010**, *9*, 131. [CrossRef] [PubMed]

63. Richter, S.S.; Diekema, D.J.; Heilmann, K.P.; Dohrn, C.L.; Riahi, F.; Doern, G.V. Changes in pneumococcal serotypes and antimicrobial resistance after introduction of the 13-valent conjugate vaccine in the United States. *Antimicrob. Agents Chemother.* **2014**, *58*, 6484–6489. [CrossRef] [PubMed]

64. Lyke, K.E.; Diallo, D.A.; Dicko, A.; Kone, A.; Coulibaly, D.; Guindo, A.; Cissoko, Y.; Sangare, L.; Coulibaly, S.; Dakouo, B.; et al. Association of intraleukocytic Plasmodium falciparum malaria pigment with disease severity, clinical manifestations, and prognosis in severe malaria. *Am. J. Trop. Med. Hyg.* **2003**, *69*, 253–259. [CrossRef]

65. A global network for investigating the genomic epidemiology of malaria. *Nature* **2008**, *456*, 732–737. [CrossRef]

66. Duffy, P.E.; Sahu, T.; Akue, A.; Milman, N.; Anderson, C. Pre-erythrocytic malaria vaccines: Identifying the targets. *Expert Rev. Vaccines* **2012**, *11*, 1261–1280. [CrossRef]

67. Taylor-Robinson, A.W. Immunity to liver stage malaria: Considerations for vaccine design. *Immunol. Res.* **2003**, *27*, 53–70. [CrossRef]

68. Sauerwein, R.W. The first malaria vaccine: A significant milestone. *Ned. Tijdschr. Geneeskd.* **2015**, *159*, A9730.

69. Mahmoudi, S.; Keshavarz, H. Efficacy of Phase 3 Trial of RTS, S/AS01 Malaria Vaccine in infants: A systematic review and meta-analysis. *Hum. Vaccines Immunother.* **2017**, *11*, 261. [CrossRef]

70. Rosenthal, P.J. The RTS,S/AS01 vaccine continues to show modest protection against malaria in African infants and children. *Evid. Based Med.* **2015**, *20*, 179. [CrossRef]

71. Rampling, T.; Ewer, K.J.; Bowyer, G.; Edwards, N.J.; Wright, D.; Sridhar, S.; Payne, R.; Powlson, J.; Bliss, C.; Venkatraman, N.; et al. Safety and efficacy of novel malaria vaccine regimens of RTS,S/AS01B alone, or with concomitant ChAd63-MVA-vectored vaccines expressing ME-TRAP. *NPJ Vaccines* **2018**, *3*, 49. [CrossRef] [PubMed]

72. Bowyer, G.; Grobbelaar, A.; Rampling, T.; Venkatraman, N.; Morelle, D.; Ballou, R.W.; Hill, A.V.S.; Ewer, K.J. CXCR3(+) T Follicular Helper Cells Induced by Co-Administration of RTS,S/AS01B and Viral-Vectored Vaccines Are Associated With Reduced Immunogenicity and Efficacy Against Malaria. *Front. Immunol.* **2018**, *9*, 660. [CrossRef] [PubMed]

73. Collins, K.A.; Snaith, R.; Cottingham, M.G.; Gilbert, S.C.; Hill, A.V.S. Enhancing protective immunity to malaria with a highly immunogenic virus-like particle vaccine. *Sci. Rep.* **2017**, *7*, 46621. [CrossRef] [PubMed]

74. Halbroth, B.R.; Sebastian, S.; Salman, A.M.; Ulaszewska, M.; Gola, A.; Longley, R.J.; Janse, C.J.; Khan, S.M.; Hill, A.V.S.; Spencer, A.J. Preclinical Development and Assessment of Viral Vectors Expressing a Fusion Antigen of Plasmodium falciparum LSA1 and LSAP2 for Efficacy against Liver-Stage Malaria. *Infect. Immun.* **2020**, *88*, e00573-19. [CrossRef] [PubMed]

75. Longley, R.J.; Halbroth, B.R.; Salman, A.M.; Ewer, K.J.; Hodgson, S.H.; Janse, C.J.; Khan, S.M.; Hill, A.V.S.; Spencer, A.J. Assessment of the Plasmodium falciparum Preerythrocytic Antigen UIS3 as a Potential Candidate for a Malaria Vaccine. *Infect. Immun.* **2017**, *85*, e00641-16. [CrossRef]

76. Kimani, D.; Jagne, Y.J.; Cox, M.; Kimani, E.; Bliss, C.M.; Gitau, E.; Ogwang, C.; Afolabi, M.O.; Bowyer, G.; Collins, K.A.; et al. Translating the immunogenicity of prime-boost immunization with ChAd63 and MVA ME-TRAP from malaria naive to malaria-endemic populations. *Mol. Ther.* **2014**, *22*, 1992–2003. [CrossRef]

77. Beeson, J.G.; Kurtovic, L.; Dobano, C.; Opi, D.H.; Chan, J.A.; Feng, G.; Good, M.F.; Reiling, L.; Boyle, M.J. Challenges and strategies for developing efficacious and long-lasting malaria vaccines. *Sci. Transl. Med.* **2019**, *11*, eaau1458. [CrossRef]

78. Patarroyo, M.E.; Arevalo-Pinzon, G.; Reyes, C.; Moreno-Vranich, A.; Patarroyo, M.A. Malaria Parasite Survival Depends on Conserved Binding Peptides' Critical Biological Functions. *Curr. Issues Mol. Biol.* **2016**, *18*, 57–78.

79. Mackellar, D.C.; O'Neill, M.T.; Aly, A.S.; Sacci, J.B.; Cowman, A.F., Jr.; Kappe, S.H. Plasmodium falciparum PF10_0164 (ETRAMP10.3) is an essential parasitophorous vacuole and exported protein in blood stages. *Eukaryot. Cell* **2010**, *9*, 784–794. [CrossRef]

80. Ajibola, O.; Diop, M.F.; Ghansah, A.; Amenga-Etego, L.; Golassa, L.; Apinjoh, T.; Randrianarivelojosia, M.; Maiga-Ascofare, O.; Yavo, W.; Bouyou-Akotet, M.; et al. In silico characterisation of putative Plasmodium falciparum vaccine candidates in African malaria populations. *Sci. Rep.* **2021**, *11*, 16215. [CrossRef]

81. Manske, M.; Miotto, O.; Campino, S.; Auburn, S.; Almagro-Garcia, J.; Maslen, G.; O'Brien, J.; Djimde, A.; Doumbo, O.; Zongo, I.; et al. Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. *Nature* **2012**, *487*, 375–379. [CrossRef] [PubMed]

82. WHO. *World Malaria Report 2021*; WHO: Geneva, Switzerland, 2021.

83. Hughes, A.L.; Verra, F. Very large long-term effective population size in the virulent human malaria parasite Plasmodium falciparum. *Proc. Biol. Sci.* **2001**, *268*, 1855–1860. [CrossRef] [PubMed]

84. Sondo, P.; Derra, K.; Lefevre, T.; Diallo-Nakanabo, S.; Tarnagda, Z.; Zampa, O.; Kazienga, A.; Valea, I.; Sorgho, H.; Ouedraogo, J.B.; et al. Genetically diverse Plasmodium falciparum infections, within-host competition and symptomatic malaria in humans. *Sci. Rep.* **2019**, *9*, 127. [CrossRef] [PubMed]

85. Abukari, Z.; Okonu, R.; Nyarko, S.B.; Lo, A.C.; Dieng, C.C.; Salifu, S.P.; Gyan, B.A.; Lo, E.; Amoah, L.E. The Diversity, Multiplicity of Infection and Population Structure of P. falciparum Parasites Circulating in Asymptomatic Carriers Living in High and Low Malaria Transmission Settings of Ghana. *Genes* **2019**, *10*, 434. [CrossRef]

86. Nascutiu, A.M. "Hide-and-seek" as modus vivendi of malaria parasites. *Rom. Arch. Microbiol. Immunol.* **2002**, *61*, 301–314.

87. Park, A.W.; Daly, J.M.; Lewis, N.S.; Smith, D.J.; Wood, J.L.; Grenfell, B.T. Quantifying the impact of immune escape on transmission dynamics of influenza. *Science* **2009**, *326*, 726–728. [CrossRef]

88. Campeotto, I.; Goldenzweig, A.; Davey, J.; Barfod, L.; Marshall, J.M.; Silk, S.E.; Wright, K.E.; Draper, S.J.; Higgins, M.K.; Fleishman, S.J. One-step design of a stable variant of the malaria invasion protein RH5 for use as a vaccine immunogen. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 998–1002. [CrossRef]

89. Patarroyo, M.E.; Cifuentes, G.; Bermudez, A.; Patarroyo, M.A. Strategies for developing multi-epitope, subunit-based, chemically synthesized anti-malarial vaccines. *J. Cell. Mol. Med.* **2008**, *12*, 1915–1935. [CrossRef]

90. Lee, S.M.; Wu, C.K.; Plieskatt, J.; McAdams, D.H.; Miura, K.; Ockenhouse, C.; King, C.R. Assessment of Pfs25 expressed from multiple soluble expression platforms for use as transmission-blocking vaccine candidates. *Malar. J.* **2016**, *15*, 405. [CrossRef]

91. Thompson, C.M.; Petiot, E.; Mullick, A.; Aucoin, M.G.; Henry, O.; Kamen, A.A. Critical assessment of influenza VLP production in Sf9 and HEK293 expression systems. *BMC Biotechnol.* **2015**, *15*, 31. [CrossRef]

92. Doolan, D.L.; Hoffman, S.L. The complexity of protective immunity against liver-stage malaria. *J. Immunol.* **2000**, *165*, 1453–1462. [CrossRef] [PubMed]

93. Carvalho, L.H.; Sano, G.; Hafalla, J.C.; Morrot, A.; Curotto de Lafaille, M.A.; Zavala, F. IL-4-secreting CD4+ T cells are crucial to the development of CD8+ T-cell responses against malaria liver stages. *Nat. Med.* **2002**, *8*, 166–170. [CrossRef] [PubMed]

94. Overstreet, M.G.; Chen, Y.C.; Cockburn, I.A.; Tse, S.W.; Zavala, F. CD4+ T cells modulate expansion and survival but not functional properties of effector and memory CD8+ T cells induced by malaria sporozoites. *PLoS ONE* **2011**, *6*, e15948. [CrossRef] [PubMed]

95. Schofield, L.; Villaquiran, J.; Ferreira, A.; Schellekens, H.; Nussenzweig, R.; Nussenzweig, V. Gamma interferon, CD8+ T cells and antibodies required for immunity to malaria sporozoites. *Nature* **1987**, *330*, 664–666. [CrossRef]

96. Helg, A.; Mueller, M.S.; Joss, A.; Poltl-Frank, F.; Stuart, F.; Robinson, J.A.; Pluschke, G. Comparison of analytical methods for the evaluation of antibody responses against epitopes of polymorphic protein antigens. *J. Immunol. Methods* **2003**, *276*, 19–31. [CrossRef]

97. Terasaki, P.I. A brief history of HLA. *Immunol. Res.* **2007**, *38*, 139–148. [CrossRef]

98. Zhao, W.; Sher, X. Systematically benchmarking peptide-MHC binding predictors: From synthetic to naturally processed epitopes. *PLoS Comput. Biol.* **2018**, *14*, e1006457. [CrossRef]

99. Sanchez-Trincado, J.L.; Gomez-Perosanz, M.; Reche, P.A. Fundamentals and Methods for T- and B-Cell Epitope Prediction. *J. Immunol. Res.* **2017**, *2017*, 2680160. [CrossRef]