

Protein interaction analysis of *Plasmodium falciparum* circumsporozoite protein variants with human immunoproteins explains RTS,S vaccine efficacy in Ghana

Cheikh Cambel Dieng^{1,*}, Colby T. Ford^{2,3,*}, Anita Lerch⁴, Dickson Donu⁵, Jennifer Huynh¹, Kovidh Vegesna³, Jun-tao Guo², Daniel A. Janies², Linda Amoah^{5,6}, Yaw Afrane⁷, and Eugenia Lo^{1,3,✉}

¹Department of Biological Sciences, University of North Carolina at Charlotte, Charlotte, NC, USA

²Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC, USA

³School of Data Science, University of North Carolina at Charlotte, Charlotte, NC USA

⁴Department of Biological Sciences, University of Notre Dame, Notre Dame, IN, USA

⁵Department of Immunology, Noguchi Memorial Institute for Medical Research, University of Ghana, Accra, Ghana

⁶West Africa Center for Cell Biology of Infectious Pathogens, University of Ghana, Accra, Ghana

⁷Department of Microbiology, University of Ghana Medical School, Accra, Ghana

*These authors contributed equally.

Background. The world's first malaria vaccine RTS,S provides only partial protection against *Plasmodium falciparum* infections. The explanation for such low efficacy is unclear. This study examined the associations of parasite genetic variations with binding affinity to human immunological proteins including human leukocyte antigen (HLA) and T cell receptors (TCR) involved in RTS,S-induced immune responses.

Methods. Multiplicity of infections was determined by amplicon deep sequencing of merozoite surface protein 1 (*PfMSP1*). Genetic variations in the C-terminal of circumsporozoite protein (*PfCSP*) gene were examined across 88 samples of *P. falciparum* collected from high and low transmission settings of Ghana. Binding interactions of *PfMSP1* variants and HLA/TCR were analyzed using NetChop and HADDOCK predictions. Anti-CSP IgG levels were measured by ELISA in a subset of 10 samples.

Findings. High polyclonality was detected among *P. falciparum* infections. A total 27 CSP haplotypes were detected among samples. A significant correlation was detected between the CSP and MSP multiplicity of infection (MOI). No clear clustering of haplotypes was observed by geographic regions. The number of genetic differences in *PfCSP* between 3D7 and non-3D7 variants does not influence binding interactions to HLA/T cells nor anti-CSP IgG levels. Nevertheless, *PfCSP* peptide length significantly affects its molecular weight and binding affinity to the HLA.

Interpretations. The presence of multiple non-3D7 strains among *P. falciparum* infections in Ghana impact the effectiveness of RTS,S. Longer *PfCSP* peptides may elicit a stronger immune response and should be considered in future version RTS,S. The molecular mechanisms of RTS,S cell-mediated immune responses related to longer CSP peptides warrants further investigations.

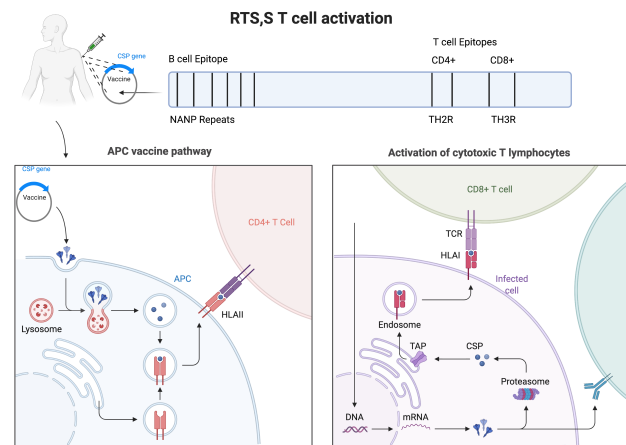


Fig. 1. RTS,S T cell Activation.

Introduction

Plasmodium falciparum remains the most dominant malaria parasite in Sub-Saharan Africa that accounts for almost all mortality in malaria (1). Despite the many interventions in place, malaria burden remains high over years and mortality is the highest in young children (2). RTS,S is a pre-erythrocytic vaccine that aims to trigger the human immune system to defend against *P. falciparum* malaria (3). RTS,S is designed based on a portion of the central Asn-Ala-Asn-Pro (NANP) amino acid repeats (B-cell epitopes) and the C-terminal (T cell epitopes) of the circumsporozoite protein (CSP) of the 3D7 lab strain citeRN13. The C-terminal, precisely the Th2R and Th3R regions, encode epitopes recognized by the CD8+ and CD4+ T cells (Figure 1). The Th2R and Th3R portions (RTS) fuse to the native hepatitis B surface antigen (S) and self-assemble into virus-like particles exposed on the surface of the CSP protein (4–7). Phase III trials conducted at 11 sites across seven African countries indicated a 36–55% efficacy against clinical malaria in 5 to 17 month-old children who received three primary doses of RTS,S plus a booster at 20 months (3, 8–12). The explanation for such low efficacy is unclear.

High genetic diversity was reported in the *PfCSP* C-terminal as malaria transmission increases across geographic

Bioinformatics | *Plasmodium falciparum* | Malaria | Vaccine Evasion | Haplotyping | Genetic Diversity | Circumsporozoite Protein | Ghana

Correspondence: cdieng@uncc.edu, colby.ford@uncc.edu, or eugenia.lo@uncc.edu

areas (13, 14). In high transmission areas, multiple genetically diverse parasite strains are common within the same host and in the community (15). Diverse parasite strains could harbor different CSP phenotypes contributing to variations in the invasion and or evasion mechanism of the parasites. Antibody-mediated immunity against blood-stage *P. falciparum* was strain-specific (16, 17). Functional responses elicited against a particular parasite strain do not yield comparable levels of inhibition against heterologous strains (18, 19). RTS,S efficacy against non-3D7 parasite strains was lower than that against 3D7 strain (20), further suggesting that allele-specific immunity is important in eliciting protection. Vaccine efficacy reduced from 50.3% for parasites with a perfect match to *PfCSP* 3D7 C-terminal sequence to 33.4% for parasites with any amino acid mismatch. Besides, vaccine efficacy declined as the number of amino acid differences increased (20). Mutational changes in amino acid sequence could alter CSP protein conformation and binding reaction of the CSP peptide with human leukocyte antigens (HLA), which is vital to the induction of T cell immune response, such as production and secretion of cytokines and the capacity to mediate cytotoxicity (21, 22). The recognition of *P. falciparum* by CD8+ and CD4+ T cell requires the intervention of antigen presenting cells (APCs) that display polypeptide fragments of *PfCSP* on their surfaces (Figure 1). The interaction between T cell and the CSP peptide is mediated by the HLA, which is expressed at the surface of APCs (23). The antigenic peptide recognition depends on the degradation of proteins by the 26S proteasome or lysosomal proteolysis, which generates appropriate peptide size to dock into the groove of HLA molecules (24–27). It is not clear if non-3D7 *PfCSP* variants exhibit different binding affinities to HLA and impact T cell recognition. Computational analyses using a machine learning-based model to predict protein structure, folding (28), macromolecules interactions and/or binding affinities have been the forefront of protein-ligand complexes studies (29). Such mechanistic insights would shed light on low RTS,S efficacy.

In Ghana, malaria accounts for 30% of outpatient attendances and 23% inpatient admissions in 2017(30). The country is characterized by heterogeneous transmission patterns driven by different climate and landscape features. Transmission is seasonal in the northern savannah, and perennial in the central forest and southern coastal areas (2). Ghana is one of the three African countries where RTS,S is being implemented along with Kenya and Malawi (2, 11). In this study, the multiplicity of infections (MOI) and CSP diversity among *P. falciparum* samples from different transmission settings of Ghana were compared. We aim to (1) determine the within and between host genetic diversity of *P. falciparum* across transmission settings; (2) clarify the effect of polyclonal infections on vaccine efficacy; (3) use protein predicted structure to identify mutational changes that may alter the overall conformation of the CSP protein and the binding affinity to HLA/T cell receptors. Our results have important implications on the development of next generation RTS,S. Namely, that CSP sequences that result in peptides of specific lengths

may contribute to an increase in binding affinity between the HLA and T cell receptors.

Methods

Study sites and sample collection. Five sites were selected from three ecological zones of Ghana (31) including Pagaza (PZ) in Tamale Municipality and Kpalsogou (KG) in Kumbungu district in the northern savannah region; Duase (KD) in Konongo and Seikwa district in the central forest region; and Ada (AD) and Dodowa (DO) in the southern coastal region. Sample collection was conducted during June-July of 2019. About 200 μ L capillary blood samples were collected from 88 subjects who were diagnosed as malaria-positive by Access Bio CareStartTM malaria rapid diagnostic test (RDT). DNA was extracted using Quick-DNA kit (Zymo Research) following the manufacturer's protocol and stored in -20°C. *Plasmodium* species identification and DNA quantification was performed by real-time quantitative PCR of the 18S rRNA gene based on the published protocols (31).

Amplicon deep sequencing and analysis. The 395-bp *PfMSP1* and the 321-bp C-terminal of *PfCSP* containing the Th2R and Th3R regions were amplified using previously published primers (32, 33). The Illumina partial adapters (34) were added to the purified PCR products prior to library constructions (See Table 1). Multiple samples were pooled and paired-end sequenced on the Illumina HiSeq 2500 (35). Coverage of at least 50,000 reads were obtained for each sample amplicon (sequences and sample information are available on NCBI under BioProject [PRJNA783000](#)).

The HaplotypR package was used for SNP and haplotype detection based on *PfCSP* and *PfMSP1* sequences (36). Given that *PfMSP1* is length polymorphic, haplotypes were generated based on the size of the amplicon (Table 1), which were determined by iterative testing of various thresholds to distinguish between real haplotypes and random noise. MOI was calculated as the number of genetically distinct parasite clones co-infecting a single host with cut-off settings for haplotype calling requiring a haplotype to be supported by ≥ 3 reads in ≥ 2 samples (including independent replicates of the same sample) (36). A haplotype network was generated with the TCS for multiple sequence alignment evaluation and phylogenetic reconstruction (37). The construction of the haplotype network involves extracting the sequences two by two to build a T-Coffee library. The library was then used to obtain a Transitive Consistency Score (TCS) of every pair of aligned residues. Then, by averaging the pair-wise TCS score over all pairs of aligned residues in each target, SequenceTCS and the AlignmentTCS metrics were derived (38). The network was visualized using a TCS web-based program `tcsBU` (39).

CSP peptide prediction and protein interaction modeling. To predict the resulting polypeptide expressed after proteolytic degradation of the CSP protein, the complete *PfCSP* haplotype sequences were scored using NetChop

Primer set	Name	Sequence	PCR Conditions	Product length (bp)
PfCSP	Forward	5'-AATCAAGGTAATAATGGACAAGG-3'	95 °C for 4 min, followed by 40 cycles of 94 °C (1 min), 50 °C (1 min) and 72 °C (1 min) and final extension at 72 °C for 4 min	321
	Reverse	5'-CTAATTAAGGAACAAGAAGG-3'		
PfMSP1	Forward	5'-GAAGCTTTAGAAGATGCAGTATTGA-3'	94 °C (3 min), followed by 35 cycles at 94 °C (30 s), 55 °C (30 s) and 72 °C (1 min), final extension at 72 °C (6 min).	395
	Reverse	5'-CAAAGAGTTCGGGATATTGAG-3'		
Illumina Partial Adapters	Forward	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'		
	Reverse	5'-GACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'		

Table 1. Sequences of the primers used in the study and corresponding references.

Setting	CSP	MSP1
Method	SNP-based	Length Polymorphic
Minimum Mismatch Rate (for SNPs)	0.50	NA
Minimum Coverage	3	3
Detection Limit	1/50	1/100
Minimum Occurrence	3	2
Minimum Coverage per Sample	25	25

Table 2. HaplotypR settings for generating CSP and MSP1 haplotypes.

3.0 (40). NetChop provides a prediction score for each residue as to its probability of being expressed in a short peptide (~8 amino acids in length) yielded from the proteasomal degradation process. Residues with an expression probability $\geq 50\%$ in the Th2R and Th3R regions of the protein were selected. These short peptide sequences were then assembled as 3D structures using the Peptide Builder package in Biopython (41). A PDB format file of each peptide was produced and used in subsequent interaction modeling. For comparison, we analysed the sequences through netMHC 1 and 2 for TH2R and TH3R, respectively, to provide binding affinity predictions by position in the region. In netMHC, peptides are predicted if they were among the highest prediction score for HLA class I affinity, TAP transport efficiency, and C-terminal proteasomal cleavage as a weighted sum of the three individual prediction scores at the set threshold (42).

The interaction between CSP, HLA, and T cell receptor proteins is complex in that it requires the intervention of different components of the immune system (Figure 1). In humans, there are many different alleles of HLA-I and HLA-II molecules. However, these variants are expressed in distinct frequencies in different ethnic groups (43). According to the Allele Frequency Net Database, the HLA locus A (HLA-A) and HLA-DRB1 are the most frequent in Ghana and confer protection against severe malaria (44, 45).

Using the generated CSP peptide structure along with reference structures for the Human Leukocyte Antigen 2A (HLA-2A) and T Cell Receptor (TCR), we predicted the interaction of these three structures using HADDOCK 2.4 (29). Specifically, we used PDB: 6TRN (chain A) for the HLA-2A structure and PDB: 6PY2 (chains D and E) for the TCR structure, renumbering residues according to HADDOCK's requirements. We specified the interacting residues between the HLA structure, the CSP peptide, and the TCR structure by first docking the CSP peptide in the protein-binding groove of the HLA specifically between the $\alpha 1$ and $\beta 1$ subunits (46). This HLA-CSP peptide complex was then docked at the antigen-binding site of the TCR on top of the α and β loci (46). The prediction of these interactions was derived from various biophysical factors including van der Waals energy, electrostatic energy, desolvation energy, and restraints violation energy, which were collectively used to derive a

HADDOCK score to quantify changes in protein-protein interaction resulting from CSP peptide variations (47, 48).

Measurement of anti-CSP antibody levels. Plasma from participant blood samples were examined for antibody titers against PfCSP by standardized enzyme-linked immunosorbent assays (ELISA) following a previous protocol (49). Recombinant IgG (BP055) was used as a standard calibrator for total IgG measurements. The PfCSP antigen was diluted to a concentration of 1 $\mu\text{g/ml}$ in carbonate buffer (pH 9.0) and coated at 100 $\mu\text{l/well}$ onto ELISA plates at 4°C overnight, followed by washing with 1xPBS/Tween20 for four times. Plasma from malaria naïve individuals from Denmark was used as negative control samples. Absorbance was measured using the ELx808 Absorbance Reader (Biotek, USA) and resulting optical density values were converted to IgG concentration using ADAMSEL (Ed Remarque, BPRC). This process workflow is depicted in Figure 2.

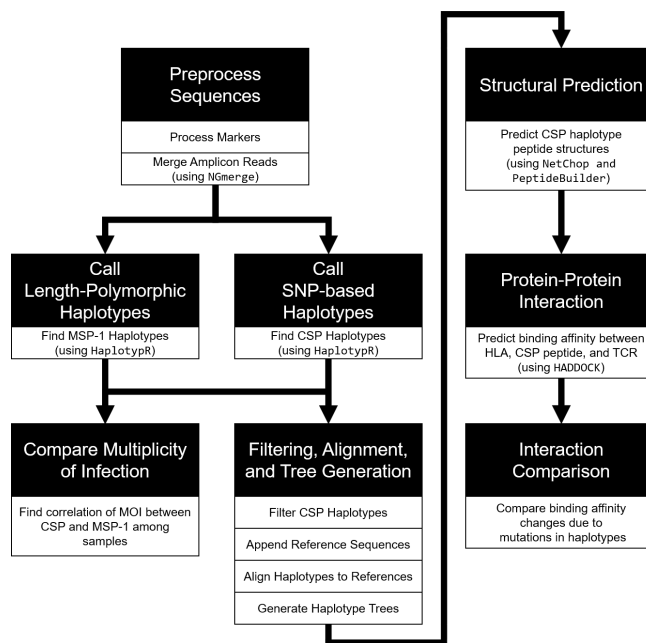


Fig. 2. Process workflow for generating and comparing CSP haplotypes and assessing human immunoprotein interaction.

Results

Multiplicity of infection (MOI) and Sequence Comparison. Based on PfMSP1, the proportion of polyclonal samples among the study sites was 76%, with the mean MOI of 2.63 among the 88 samples (Table 3). A total of 30 haplotypes were detected, ranging from 1 to 6 distinct clones within a single host (Figure 3). Considerable size polymor-

phisms were observed among different the *PfMSP1* haplotypes. Polyclonal infections were detected in 80% of the samples from the North and all samples in the central and southern regions. The mean MOI was highest in central followed by the north and south Ghana, but such difference was not significant (Table 3)

For *PfCSP*, 27 haplotypes were detected among samples, ranging from 1 to 8 haplotypes/clones within a host (Figure 3). Nucleotide diversity (Pi) and haplotype diversity (Hd) were similar among the three studied regions (Table 3). Though the Pi values were found to be less than 0.02 in each region, Hd values ranged from 0.938–0.951, indicating a high clonal diversity in the parasite populations. For all samples, MOI values estimated from the *PfMSP1* and *PfCSP* were significantly correlated with one another ($R = 0.33$; p -value = 0.005). When samples were analyzed by regions, only those in the north showed a significant correlation ($R = 0.41$; p -value = 0.039). The northern, central, and southern populations shared similar *PfCSP* haplotypes, and no clear clustering of haplotypes was observed by geographic regions. Of the 27 *PfCSP* haplotypes, haplotype IDs CSP-10 and CSP-31 were most closely related to the 3D7 haplotype with only one SNP difference (Figure 4). By contrast, haplotypes with IDs CSP-15, CSP-117, and CSP-22 were the most distant from 3D7, each with 11-12 nucleotide differences. For CSP-117, mutations observed at the Th2R and Th3R regions were non synonymous and thus, this haplotype has identical amino acid sequence as 3D7. (Figure 5)

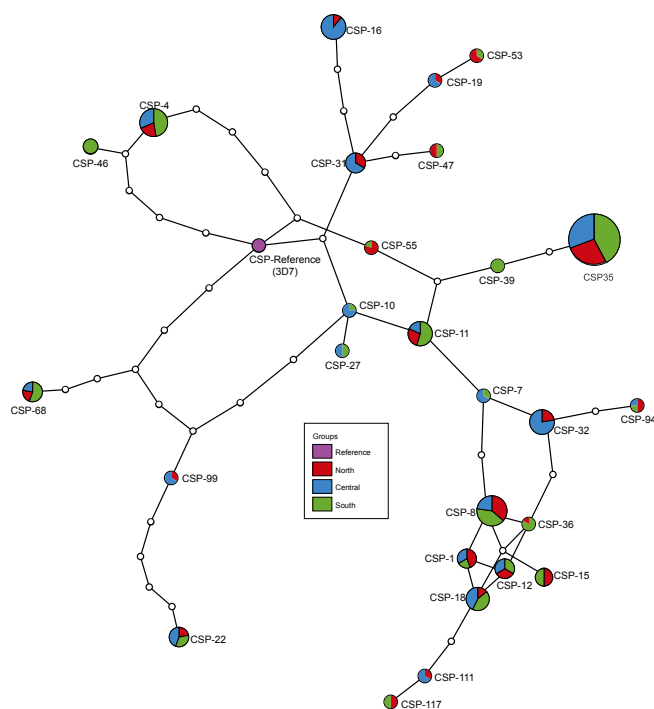


Fig. 4. Haplotype network of the 27 CSP haplotypes. This was generated using TCS (50) and then colored with region information using tcsBU (51).

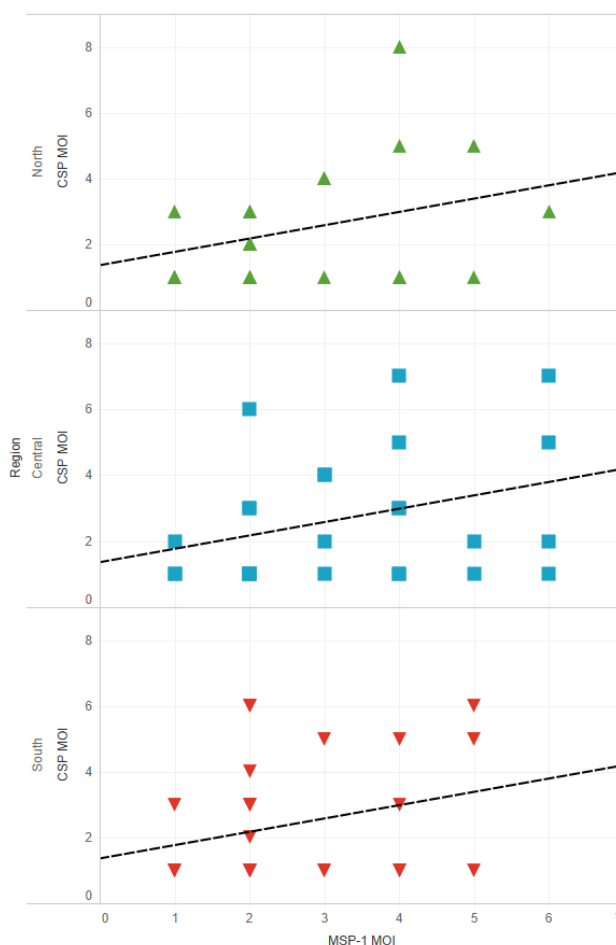


Fig. 3. Relationship of CSP and MSP1 minimum multiplicity of infection by geographic region. (Overall $R = 0.3197$, p -Value: 0.0052)

Binding affinity of *PfCSP* variants. The C-terminal of the *PfCSP* sequences that contained the Th2R and Th3R regions (known as T cell epitope regions (52)) were analyzed separately for binding affinity. In the Th3R region, four peptides (VIELYE, VIQLYA, VIELYA, VIQLYE) were detected based on six amino acid residues (Figures in Table 8). HADDOCK scores of these peptides with the HLA-TCR complex were very similar without clear differences (Table 4).

Compared to Th3R, the Th2R region of *PfCSP* was more variable and contained 17 distinct peptide sequences (Figures in Table 7). The binding peptides of Th2R predicted by NetChop ranged from 4 to 9 residues in length and all begin with a glutamate (D) residue. As shown in Figure 6, there is a significant correlation between a peptide's molecular weight and the interaction between the HLA and CSP peptide. Longer peptides have a better binding affinity within the HLA groove ($R = -0.9415$; p -value = <0.0001).

However, these trends were not observed when we included the T cell structure to the to the HLA-CSP peptide ($R = 0.2658$; p -value = 0.3025). The HADDOCK scores of Th2R peptides were compared among haplotypes that were closely related and distant from the 3D7 reference. For instance, CSP-15 one of the most distant haplotypes, that provides a 5-residue peptide, showed the lowest HADDOCK

	Total Samples	<i>PfMSP1</i>		<i>PfCSP</i>	
		% Polyclonality	Multiplicity of Infections (mean, [range])	Nucleotide Diversity	Haplotype Diversity (Hd)
North	27	80	2.9, [1-6]	0.017	0.951 (-0.011)
Central	27	100	3.1, [1-6]	0.017	0.945 (-0.009)
South	29	100	2, [0-4]	0.018	0.938 (-0.009)

Table 3. Comparison of multiplicity of infections (MOI) and percentage of polyclonal samples based on *PfMSP1*, as well as nucleotide (Pi) and haplotype (Hd) diversity based on *PfCSP* among samples from north, central, and south Ghana. (Note: This only represents 83 of the 88 samples where we have both CSP and MSP-1 sequences.)

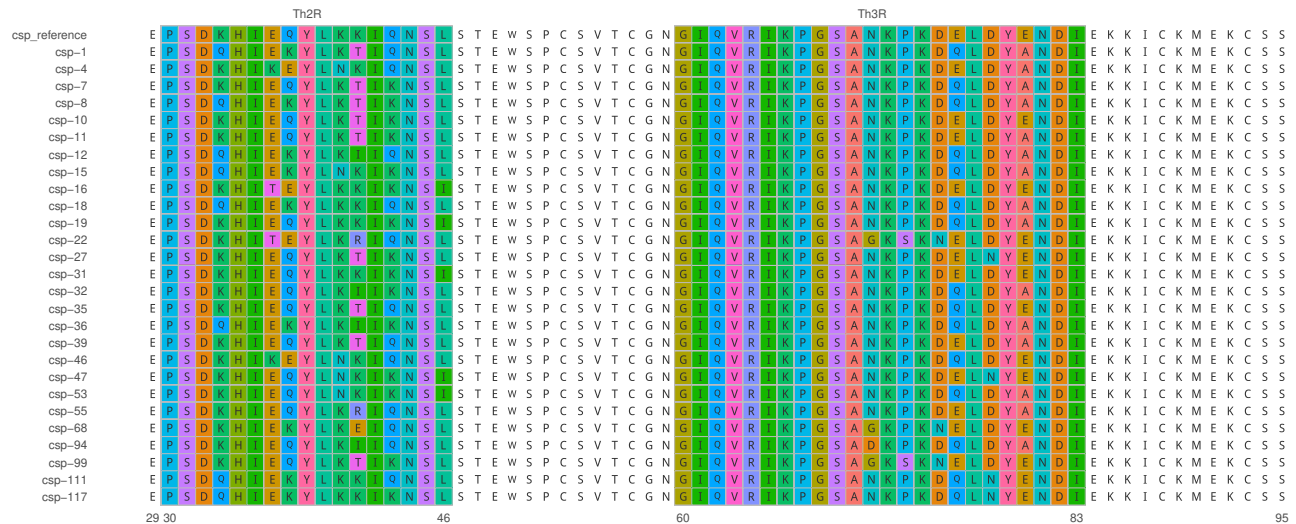


Fig. 5. Sequences of the 27 CSP haplotypes found in this study. The Th2R and Th3R regions are highlighted to show variations from the 3D7 reference sequence.

T Cell Type / Region	CSP Haplotype	CSP Peptide Sequence	CSP Peptide Length	HLA-CSPpép HADDOCK score	HLA-CSPpép-TCR HADDOCK score	Peptide Molecular Weight (g/mol)
CD4 / Th2R	csp-reference	DIEIQN	6	-62.5	-124.5	730.77
	csp-1, 18, 111	DELIQN	6	-70.6	-101.3	730.77
	csp-4, 46	DIEYIQN	7	-76.1	-113.8	893.95
	csp-7, 10, 11, 27, 99	DIELIN	6	-66.3	-100	715.8
	csp-8, 36	DELIN	5	-58.7	-88.3	602.64
	csp-12	DELIQN	7	-82	-104.3	843.93
	csp-15	DEYLI	5	-62.8	-132.1	651.71
	csp-16	DITELI	6	-70.9	-102.6	702.8
	csp-19, 31	DIEI	4	-48.3	-91.3	488.54
	csp-22	DITELRIQN	9	-108.9	-76.1	1101.23
	csp-32	DIEYLIN	7	-86	-96.7	878.98
	csp-35, 39	DIELIQN	7	-75	-103.2	843.93
	csp-47, 53	DIEQYI	6	-69.8	-100.9	779.85
	csp-55	DIEYLRIQN	9	-98	-118	1163.3
	csp-68	DELEIN	6	-79.5	-98.8	731.76
csp-94	DIEYLIQN	9	-92.4	-89.7	1120.27	
csp-117	DELI	4	-41.3	-128.6	488.54	
CD8 / Th3R	csp-reference, 10, 16, 22, 27, 31, 47, 68, 99	VIELYE	6	-76.6	-108.7	764.87
	csp-1, 7, 8, 12, 15, 18, 19, 32, 36, 53, 94	VIQLYA	6	-69.3	-118.4	705.85
	csp-4, 11, 39, 55	VIELYA	6	-75.5	-125.5	706.84
	csp-35, 46, 111, 117	VIQLYE	6	-77.7	-110.7	763.89

Table 4. Summary of the interaction of the different CSP peptides and both HLA molecules and T cell structures.

score of -132.1 (the highest binding affinity; 6.1% higher than 3D7). Similarly, CSP-117, which provides a 4-residue peptide, showed a HADDOCK score of -128.6 (3.3% better than 3D7), the second-best interaction with the HLA and T cell complex. By contrast, of all haplotypes, CSP-22 and CSP-8 showed the lowest binding affinity. CSP-22, which provides a 9-residue peptide, showed a HADDOCK score of -76.1 (the lowest binding affinity; 38.9% less than 3D7). CSP-8, which provides a 5-residue peptide, showed a HADDOCK score of -80.3 (35.5% less than 3D7). While CSP-10 and

CSP-31 were the two most closely related haplotypes to 3D7, their HADDOCK scores were markedly different. CSP-10, which provides a 6-residue peptide (DIELIN), has a similar HADDOCK score as 3D7, but CSP-31, which contains a 4-residue peptide (DIEI), has one of the lowest HADDOCK scores. See Table 4. These findings indicated that the number of genetic variations from 3D7 does not correlate with the binding affinity with the HLA and T cell complex.

To further elucidate if residues with larger side chains, such as tyrosine or glutamine, may cause the CSP peptide to



Fig. 6. Correlation plots of HADDOCK scores by Th2R CSP haplotype peptide molecular weight.

block the HLA groove, thereby reducing the binding interaction, we compared the HADDOCK scores and other evaluation metrics for peptides containing and lacking tyrosine or glutamine. The χ^2 -statistic indicated no significant difference among these peptides. (Table 5; Figure 7)

HLA-CSP peptide-TCR Metric	Peptide Contains	χ^2 Statistic	p-value
Restrains violation energy	Glutamine	2.676	0.102
Restrains violation energy	Tyrosine	0.091	0.763
Desolvation energy	Glutamine	3.177	0.075
Desolvation energy	Tyrosine	0.041	0.840
Electrostatic energy	Glutamine	1.815	0.178
Electrostatic energy	Tyrosine	1.222	0.269
van der Waals energy	Glutamine	1.335	0.248
van der Waals energy	Tyrosine	0.205	0.650
HADDOCK score	Glutamine	0.231	0.630
HADDOCK score	Tyrosine	0.364	0.546

Table 5. Kruskal-Wallis test results indicating the chi-square and p-values of binding affinity matrices between CSP peptides that contain or lack tyrosine or glutamine. No significant correlations below the $\alpha = 0.05$ level were found.

Association of peptide length with binding affinity. To determine if peptide molecular weight (based on peptide length) of CSP influences the binding affinity of HLA and T cell complex, we created correlation plots between HADDOCK score and various peptide molecular weight derived from lysosomal degradation (Figure 6). A strong negative correlation was detected between the HADDOCK score of the HLA-CSP peptide and the molecular weight of the CSP peptide ($R = -0.94$; p -value = <0.001), suggesting that longer CSP peptides confer to stronger binding affinity. However, when TCR was added to the HLA-CSP complex, the HADDOCK score was not significantly correlated with the molecular weight of the complex ($R = 0.2658$; p -value = 0.3025).

These findings suggest that longer peptides may bind better within the HLA groove and the size of the *PfCSP* peptide predicted based on proteolysis may play role in recognition by T cell receptors. Other binding metrics were also assessed in Table 6.

Association of *PfCSP* antibody levels with ligand-peptide binding affinity. For a subset of samples, *PfCSP* antibody levels were measured to determine if a binding affinity correlation exists between CSP peptide and HLA/TCR complex. No significant correlation was found between *PfCSP* antibody level and the binding affinity of CSP peptide with HLA-CSP ($R = 0.61$; p -value = 0.059) nor HLA-CSP TCR complex ($R = 0.23$; p -value = 0.52). The antibody level was also not associated with qPCR-based parasitemia among the infected samples ($R = 0.58$; p -value = 0.082 ; Figure 8)

Discussion

Plasmodium falciparum infections in highly endemic areas are usually complex and polyclonal, which explains the great genetic polymorphism within the species (53). Such diversity could impact different antimalarial interventions such as the emergence of resistant parasite strains and breakthrough infections of malaria vaccines. Here, we predicted that the presence of multi-clone infections and high prevalence of the non-3D7 *PfCSP* variants may allow the parasites to escape RTS,S-induced immunity and explain the relatively low efficacy observed in phase 3 trial (9). Consistent with other studies, *PfMSP1* is more variable than *PfCSP* in distinguishing parasite clones within hosts (36).

The positive correlation of *PfCSP* MOI and *PfMSP1* MOI

T cell Type and CSP Region	Interaction	Peptide Feature	Metric	Correlation	p-value
CD4+ / Th2R	HLA-CSP peptide	Molecular Weight	HADDOCK score	0.941	<0.0001***
			Electrostatic energy	0.097	0.711
			Desolvation energy	0.681	0.003**
			Restraints violation energy	0.614	0.009**
			van der Waals energy	0.907	<0.0001***
	HLA-CSP peptide-TCR	Molecular Weight	HADDOCK score	0.266	0.302
			Electrostatic energy	-0.368	0.146
			Desolvation energy	0.686	0.002**
			Restraints violation energy	0.530	0.029*
			van der Waals energy	0.281	0.275
CD8+ / Th3R	HLA-CSP peptide	Molecular Weight	HADDOCK score	-0.736	0.264
			Electrostatic energy	-0.767	0.233
			Desolvation energy	0.428	0.572
			Restraints violation energy	0.661	0.339
			van der Waals energy	0.757	0.243
	HLA-CSP peptide-TCR	Molecular Weight	HADDOCK score	0.917	0.083
			Electrostatic energy	0.766	0.234
			Desolvation energy	0.252	0.748
			Restraints violation energy	0.541	0.459
			van der Waals energy	0.035	0.965

Table 6. Correlations of peptide features against docking metrics by interaction and T cell type and *PfCSP* region. Significant correlations below the $\alpha = 0.05$ level are denoted in red and with one or more asterisks (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).

values suggested that individuals who were infected with a higher number of parasite clones harbored multiple *PfCSP* variants that could impact immune responses induced by RTS,S. Among the Ghanaian samples, the genetic heterogeneity of the *PfCSP* was relatively high, with up to six distinct haplotypes within a sample. While three of the samples had identical Th3R sequences as the 3D7 reference, no samples had the 3D7-type Th2R sequences. This result agrees with a previous study that found no polymorphism in Th3R, suggesting that this region may be under balancing selection (22, 54). The four Th3R peptides (VIELYE, VIQLYA, VIELYA, VIQLYE) had similar HADDOCK scores and conceivably similar binding affinity.

In this study, a clear correlation was detected between the molecular weight of CSP peptides and binding affinity when the peptide was docked in the protein-binding groove between the $\alpha 1$ - and $\beta 1$ - subunits of HLA. However, when the HLA-CSP complex was docked at the antigen-binding site of the TCR structure, the molecular weight of the peptide was not significantly associated to binding affinity. It is possible that longer peptides with exposed side chains allow for a better interaction with the HLA receptor, but the size of the entire CSP-HLA complex does not impact its interaction with T cell receptors (26, 55). The peptide-binding groove of HLA-I is known to contain deep binding pockets surrounded by polymorphic sidechains, which imposes tight constraints on the residues and overall peptide length to achieve binding. By contrast, the binding groove of HLA-II is open, allowing peptides of various lengths to fit into the binding groove (56). This explains the challenge in predicting interactions of the HLA-II-peptide compared to HLA-I-peptide complexes. Peptide length has been shown to affect the interactions with HLA molecules. For instance, a study of CD8+ T cell response to the BZLF1 protein of the Epstein-Barr virus showed that individuals with HLA-B*18:01 re-

sponded strongly and exclusively to an octamer peptide and those with HLA-B*44:03+ responded to a dodecamer peptide (57), highlighting the importance of peptide length for T cell response (58).

Previous studies have shown that vaccine efficacy declined as the number of amino acid differences increased, suggesting that allele-specific immunity is important in eliciting protection (20). In Ghana, multiple CSP haplotypes are present within a polyclonal sample and most CSP haplotypes are markedly different from the 3D7 strain at the Th2R region. Haplotypes that were most closely related to 3D7 could have peptide length and/or molecular weight different from 3D7 after proteolysis and exhibit completely different binding affinity with HLA. This finding suggests that the CSP peptide length rather than amino acid differences alone determine binding affinity and downstream immune responses. Further, residues with larger side chains such as tyrosine or glutamine in the CSP peptide have been shown to block the HLA groove (59), thereby reducing binding interaction. Nevertheless, our analyses showed no significant effect of tyrosine or glutamine on the docking of the CSP peptides.

Our results indicated no significant correlation between anti-CSP antibodies and binding affinity for both HLA-CSP and HLA-CSP-TCR complexes. However, only the most prevalent haplotype within the sample was investigated and the presence of other variants may affect antibody production. A recent study showed that high proportions of strain-specific antibody responses are likely to be elicited during high transmission season due to increased MOI and in turn, a higher preponderance of cross-reactive antibodies (18).

Limitations. It is important to note that this study has a few limitations. First, our samples may represent only a portion of CSP variants present in Africa. None of the haplotypes in our samples were similar to the 3D7 reference. Including samples with a high prevalence of 3D7-matching haplotypes

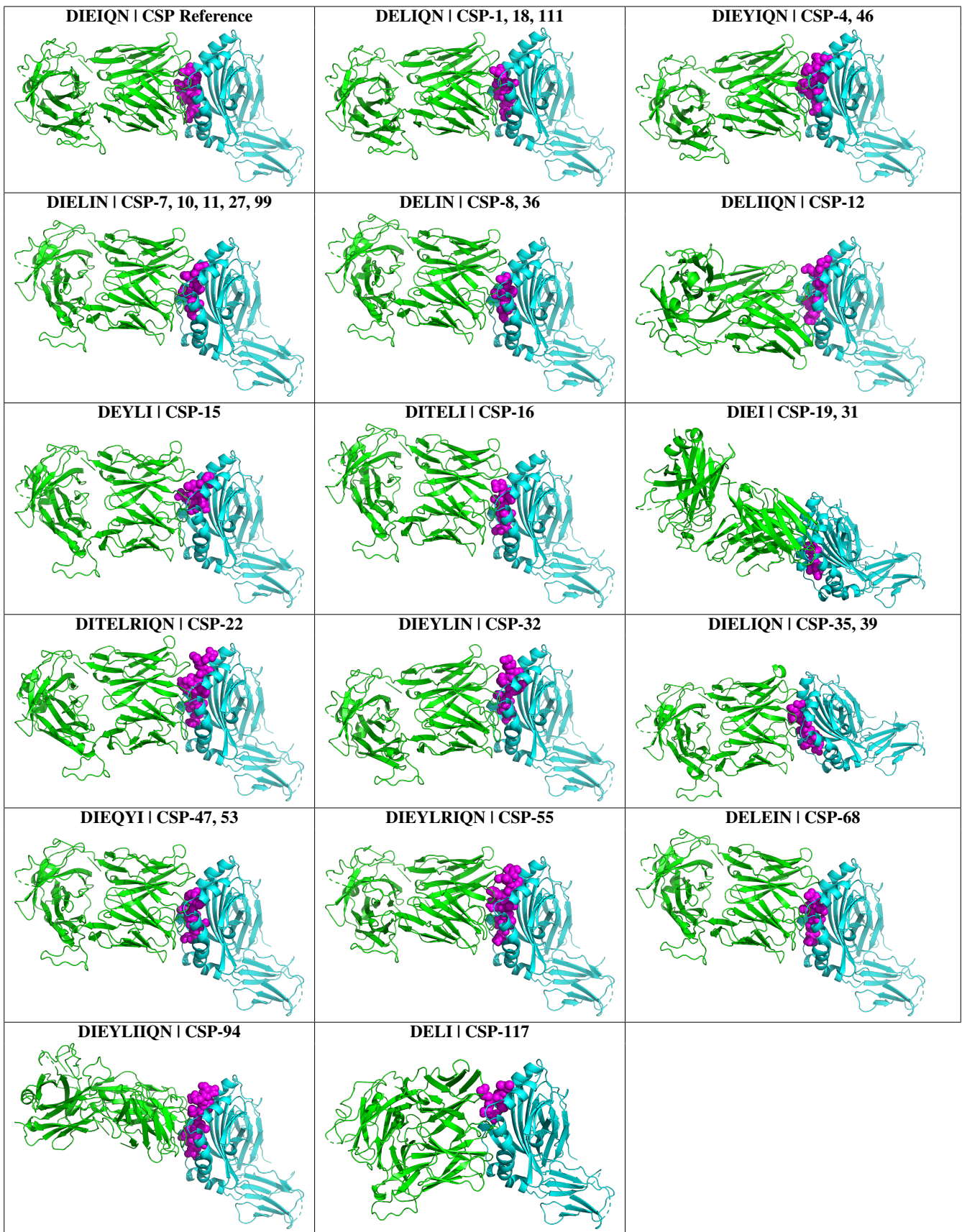


Table 7. Th2R Region HADDOCK Protein Interaction Predictions using MHC Class II HLA-DR (PDB: 6V1A) and T cell receptor T594 (PDB: 6PY2). Note that the TCR structure is shown in green, the CSP peptide is shown in magenta, and the HLA structure is shown in cyan.

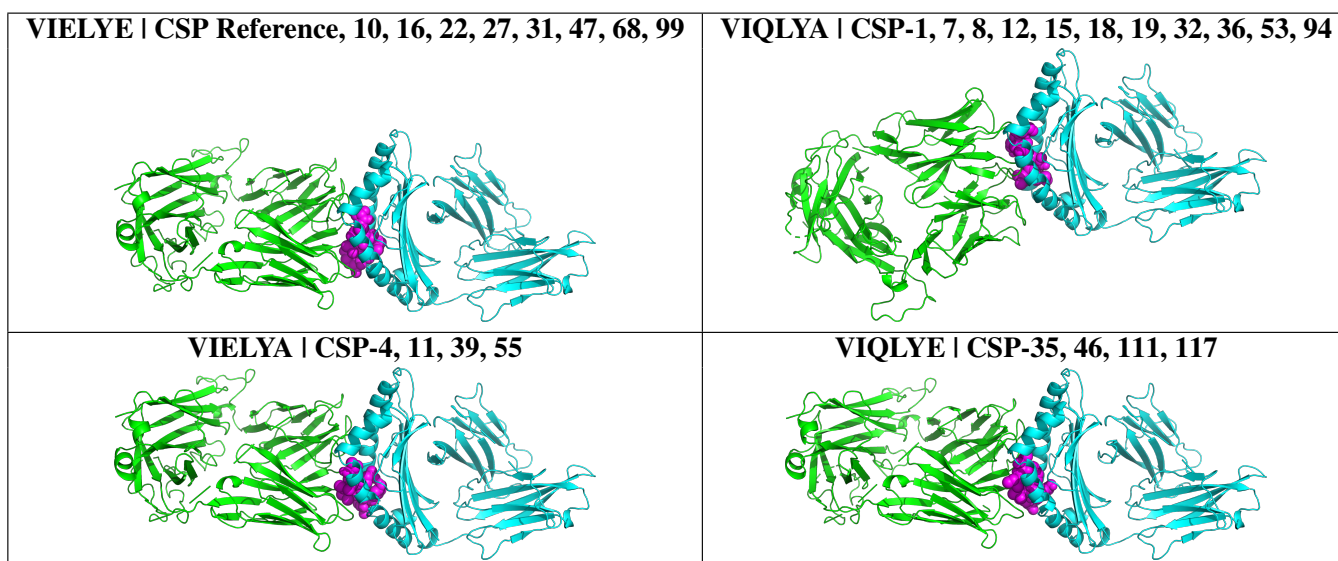


Table 8. Th3R Region HADDOCK Protein Interaction Predictions using MHC Class I HLA-A2*01 (PDB: 6TRN) and T cell receptor T594 (PDB: 6PY2). Note that the TCR structure is shown in green, the CSP peptide is shown in magenta, and the HLA structure is shown in cyan.

at the Th2R and Th3R regions would allow us to compare docking results at the T cell epitopes and vaccine efficacy with samples of mostly non-3D7 haplotypes as in the present study. RTS,S immune responses induced by other variants merits further investigations. Second, the HLA-2A and HLA-DR1 were selected in the protein modeling analysis by their high prevalence in Ghana and potential protection against severe malaria (60), even though there are several classes of HLA-I and HLA-II molecules in other ethnic groups that may exhibit different binding affinities (61). Similarly, the choice of the T cell receptor T594 was limited by the lack of representative T cell receptor structures bound to a *Plasmodium* protein in the Protein Data Bank. Third, only Th2R and Th3R regions of PfCSP were examined but epitopes in the N-terminal and NANP repeat regions may also affect binding interactions with HLA/TCR and corresponding immune responses. Future studies should validate the *in silico*-derived candidate sequences *in vitro* using custom-designed assays to investigate the HLA and TCR interactions. Also, the impact of PfCSP diversity in the N-terminal and central repeat regions on B-epitope recognition and vaccine efficacy remains unclear.

Conclusions

This study reveals a high level of polyclonality in *P. falciparum* infections across different transmission settings of Ghana. Mutations were found predominantly in the Th2R region of PfCSP and the most prevalent PfCSP haplotypes are not closely related to the 3D7 reference. The length of the Th2R peptide is significantly correlated with binding affinity with the HLA molecules but genetic relatedness to 3D7 does not impact binding interactions. These findings provide new insights into low vaccine efficacy and inform development of next generation RTS,S.

Future Work

In our upcoming research, we plan to validate the *in silico*-derived candidate sequences *in vitro* using custom-designed assays to test HLA and TCR interaction. Similar analyses are needed to further investigate the binding interactions and impact of other potential CSP haplotypes on vaccine design in other parts of Africa and the world using our methodologies described in this work.

Supplementary Materials

All PfMSP1 and PfCSP amplicon reads were uploaded to the National Center for Biotechnology Information Short Read Archive (NCBI SRA) under BioProject number: [PRJNA783000](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA783000).

All data, scripts, and results from this work are available at [GitHub.com/colbyford/Ghana_CSP_Haplotypes](https://github.com/colbyford/Ghana_CSP_Haplotypes).

ACKNOWLEDGEMENTS

We thank the field team from The University of Ghana for their technical assistance in sample collection; the communities and hospitals, for their support and willingness to participate in this research; and undergraduates at The University of North Carolina Charlotte who were involved in sample processing.

FUNDING

This research was funded by the 2019 University of North Carolina at Charlotte Faculty Research Grant. We also acknowledge funding and logistical support from several entities of the University of the North Carolina at Charlotte including: The Department of Bioinformatics and Genomics, the Bioinformatics Research Center, University Research Computing, the College of Computing and Informatics and the College of Liberal Arts and Sciences. We are grateful for funding from the Belk Family.

COMPETING FINANCIAL INTERESTS

The authors declare no conflicts of interest.

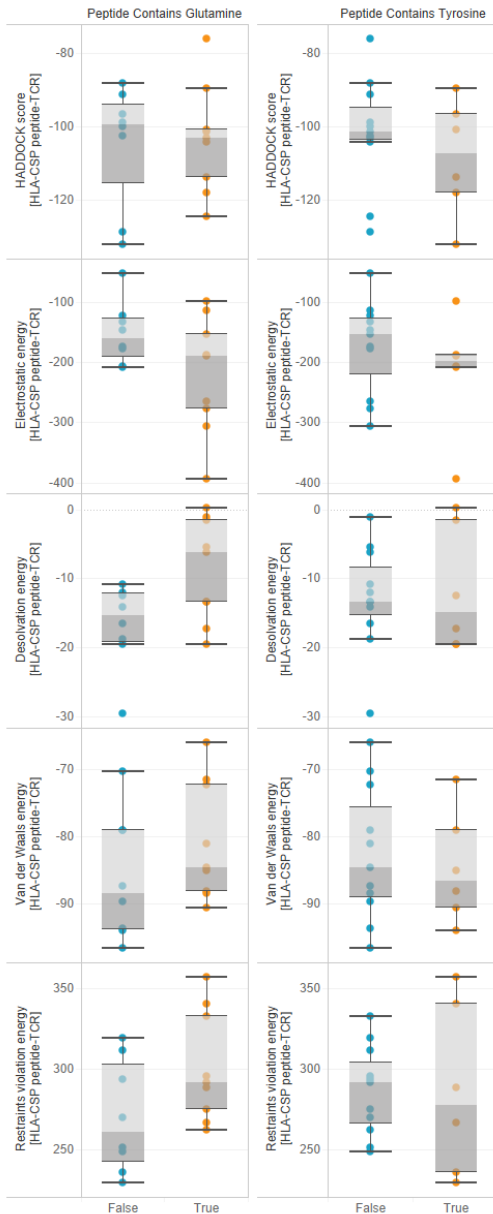


Fig. 7. Boxplots breakdown of statistics of Th2R CSP peptides by the presence or absence of glutamine or tyrosine.

Bibliography

1. Amambua-Ngwa, L. Amenga-Etego, E. Kamau, R. Amato, A. Ghansah, L. Golassa, M. Randrianavelojosia, D. Ishengoma, T. Apinjoh, O. Maïga-Ascofaré, B. Andagalu, W. Yavo, M. Bouyou-Akotet, O. Kolapo, K. Mane, A. Worwui, D. Jeffries, V. Simpson, U. D'Alessandro, D. Kwiatkowski, and A. A. Djimde. Major subpopulations of plasmodium falciparum in sub-saharan africa. *Science*, 365(6455):813–816, 2019. ISSN 0036-8075. doi: 10.1126/science.aav5427.
2. T. Awine, K. Malm, C. Bart-Plange, and S. P. Silal. Towards malaria control and elimination in ghana: challenges and decision making tools to guide planning. *Glob Health Action*, 10(1): 1381471, 2017. ISSN 1654-9716 (Print) 1654-9880. doi: 10.1080/16549716.2017.1381471.
3. S. T. Agnandji, B. Lell, J. F. Fernandes, B. P. Abossolo, B. G. Methogo, A. L. Kabwende, A. A. Adegnikla, B. Mordmüller, S. Issifou, P. G. Kremsner, J. Sacarlal, P. Aide, M. Lanaspá, J. J. Aponte, S. Machevo, S. Acacio, H. Bulo, B. Sigauque, E. Macete, P. Alonso, S. Abdulla, N. Salim, R. Minja, M. Mpina, S. Ahmed, A. M. Ali, A. T. Mtoro, A. S. Hamad, P. Mutani, M. Tanner, H. Tinto, U. D'Alessandro, H. Sorgho, I. Valea, B. Bihoun, I. Guiraud, K. Kaboré, O. Sombié, R. T. Guiguemdé, J. B. Ouédraogo, M. J. Hamel, S. Kariuki, M. Oneko, C. Odero, K. Otieno, N. Awino, M. McMorrow, V. Muturi-Kioi, K. F. Laserson, L. Slutsker, W. Otieno, L. Otieno, N. Otsyula, S. Gondi, A. Otieno, V. Owira, E. Oguk, G. Odongo, J. B. Woods, B. Ogutu, P. Njuguna, R. Chiliengi, P. Akoo, C. Kerubo, C. Maingi, T. Lang, A. Olotu, P. Bejon, K. Marsh, G. Mwambingu, S. Owusu-Agyei, K. P. Asante, K. Osei-Kwakye, O. Boahen, D. Dosoo, I. Asante, G. Adjei, E. Kwara, D. Chandramohan, B. Greenwood, J. Lusingu, S. Gesase, A. Malabeja, O. Abdul, C. Mahende, E. Liheluka, L. Malle, M. Lemnge, T. G. The-

ander, C. Drakeley, D. Ansong, T. Agbenyega, S. Adjei, H. O. Boateng, T. Rettig, J. Bawa, J. Sylverken, D. Sambian, A. Sarfo, A. Agyekum, et al. A phase 3 trial of rts, s/as01 malaria vaccine in african infants. *N Engl J Med*, 367(24):2284–95, 2012. ISSN 0028-4793. doi: 10.1056/NEJMoa1208394.

4. S. Casares, T. D. Brumeanu, and T. L. Richie. The rts, s malaria vaccine. *Vaccine*, 28(31): 4880–94, 2010. ISSN 0264-410x. doi: 10.1016/j.vaccine.2010.05.033.
5. J. Cohen, V. Nussenzweig, R. Nussenzweig, J. Vekemans, and A. Leach. From the circumsporozoite protein to the rts, s/as candidate vaccine. *Hum Vaccin*, 6(1):90–6, 2010. ISSN 1554-8600. doi: 10.4161/hv.6.1.9677.
6. A. M. Didierlaurent, C. Collignon, P. Bourguignon, S. Wouters, K. Fierens, M. Fochesato, N. Dendouga, C. Langlet, B. Malissen, B. N. Lambrecht, N. Garçon, M. Van Mechelen, and S. Morel. Enhancement of adaptive immunity by the human vaccine adjuvant as01 depends on activated dendritic cells. *J Immunol*, 193(4):1920–30, 2014. ISSN 0022-1767. doi: 10.4049/jimmunol.1400948.
7. D. M. Gordon, T. W. McGovern, U. Krzych, J. C. Cohen, I. Schneider, R. LaChance, D. G. Heppner, G. Yuan, M. Hollingdale, M. Slaoui, and et al. Safety, immunogenicity, and efficacy of a recombinantly produced plasmodium falciparum circumsporozoite protein-hepatitis b surface antigen subunit vaccine. *J Infect Dis*, 171(6):1576–85, 1995. ISSN 0022-1899 (Print) 0022-1899. doi: 10.1093/infdis/171.6.1576.
8. S Clinical Trials Partnership (2014) The RTS. Efficacy and safety of the rts, s/as01 malaria vaccine during 18 months after vaccination: a phase 3 randomized, controlled trial in children and young infants at 11 african sites. *PLoS Med*, 11(7):e1001685, 2014. ISSN 1549-1277 (Print), 1549-1277. doi: 10.1371/journal.pmed.1001685.
9. S. T. Agnandji, B. Lell, S. S. Soulanoudjingar, J. F. Fernandes, B. P. Abossolo, C. Conzelmann, B. G. Methogo, Y. Doucka, A. Flamen, B. Mordmüller, S. Issifou, P. G. Kremsner, J. Sacarlal, P. Aide, M. Lanaspá, J. J. Aponte, A. Nhamuave, D. Quelhas, O. Bassat, S. Mandjate, E. Macete, P. Alonso, S. Abdulla, N. Salim, O. Juma, M. Shomari, K. Shubis, F. Machera, A. S. Hamad, R. Minja, A. Mtoro, A. Sykes, S. Ahmed, A. M. Urassa, A. M. Ali, G. Mwanga, M. Tanner, H. Tinto, U. D'Alessandro, H. Sorgho, I. Valea, M. C. Tahita, W. Kaboré, S. Ouédraogo, Y. Sandrine, R. T. Guiguemdé, J. B. Ouédraogo, M. J. Hamel, S. Kariuki, C. Odero, M. Oneko, K. Otieno, N. Awino, J. Omoto, J. Williamson, V. Muturi-Kioi, K. F. Laserson, L. Slutsker, W. Otieno, L. Otieno, O. Nekoye, S. Gondi, A. Otieno, B. Ogutu, R. Wasuna, V. Owira, D. Jones, A. A. Onyango, P. Njuguna, R. Chiliengi, P. Akoo, C. Kerubo, J. Gitaka, C. Maingi, T. Lang, A. Olotu, B. Tsafa, P. Bejon, N. Peshu, K. Marsh, S. Owusu-Agyei, K. P. Asante, K. Osei-Kwakye, O. Boahen, S. Ayamba, K. Kayan, R. Owusu-Ofori, D. Dosoo, I. Asante, G. Adjei, G. Adjei, D. Chandramohan, B. Greenwood, J. Lusingu, S. Gesase, A. Malabeja, O. Abdul, H. Kilavo, C. Mahende, E. Liheluka, et al. First results of phase 3 trial of rts, s/as01 malaria vaccine in african children. *N Engl J Med*, 365(20): 1863–75, 2011. ISSN 0028-4793. doi: 10.1056/NEJMoa1102287.
10. C. J. Duncan and A. V. Hill. What is the efficacy of the rts, s malaria vaccine? *Bmj*, 343: d7278, 2011. ISSN 0959-8138 (Print) 0959-8138. doi: 10.1136/bmj.d7278.
11. M. B. Laurens. Rts, s/as01 vaccine (mosquirix™): an overview. *Hum Vaccin Immunother*, 16(3):480–489, 2020. ISSN 2164-5515 (Print) 2164-5515. doi: 10.1080/21645515.2019.1669415.
12. A. Olotu, G. Fegan, J. Wambua, G. Nyangweso, K. O. Awuondo, A. Leach, M. Lievens, D. Lebouilleux, P. Njuguna, N. Peshu, K. Marsh, and P. Bejon. Four-year efficacy of rts, s/as01e and its interaction with malaria exposure. *N Engl J Med*, 368(12):1111–20, 2013. ISSN 0028-4793 (Print) 0028-4793. doi: 10.1056/NEJMoa1207564.
13. S. Jongwutiwes, C. Putaporntip, and A. L. Hughes. Bottleneck effects on vaccine-candidate antigen diversity of malaria parasites in thailand. *Vaccine*, 28(18):3112–7, 2010. ISSN 0264-410X (Print) 0264-410x. doi: 10.1016/j.vaccine.2010.02.062.
14. G. D. Weedall, B. M. Preston, A. W. Thomas, C. J. Sutherland, and D. J. Conway. Differential evidence of natural selection on two leading sporozoite stage malaria vaccine candidate antigens. *Int J Parasitol*, 37(1):77–85, 2007. ISSN 0020-7519 (Print) 0020-7519. doi: 10.1016/j.ijpara.2006.09.001.
15. D. J. Conway, C. Roper, A. M. Oduola, D. E. Arnot, P. G. Kremsner, M. P. Grobusch, C. F. Curtis, and B. M. Greenwood. High recombination rate in natural populations of plasmodium falciparum. *Proc Natl Acad Sci U S A*, 96(8):4506–11, 1999. ISSN 0027-8424 (Print) 0027-8424. doi: 10.1073/pnas.96.8.4506.
16. S. Dutta, S. Y. Lee, A. H. Batchelor, and D. E. Lanar. Structural basis of antigenic escape of a malaria vaccine candidate. *Proc Natl Acad Sci U S A*, 104(30):12488–93, 2007. ISSN 0027-8424 (Print) 0027-8424. doi: 10.1073/pnas.0701464104.
17. K. A. Kusi, B. W. Faber, A. W. Thomas, and E. J. Remarque. Humoral immune response to mixed pfama1 alleles; multivalent pfama1 vaccines induce broad specificity. *PLoS One*, 4(12):e8110, 2009. ISSN 1932-6203. doi: 10.1371/journal.pone.0008110.
18. K. A. Kusi, E. A. Manu, T. Manful Gwira, E. Kyei-Baafour, E. K. Dickson, J. A. Amponsah, E. J. Remarque, B. W. Faber, C. H. M. Kocken, D. Dodo, B. A. Gyan, G. A. Awandare, F. Atuguba, A. R. Oduro, and K. A. Koram. Variations in the quality of malaria-specific antibodies with transmission intensity in a seasonal malaria transmission area of northern ghana. *PLoS One*, 12(9):e0185303, 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0185303.
19. P. G. McQueen and F. E. McKenzie. Host control of malaria infections: constraints on immune and erythropoietic response kinetics. *PLoS Comput Biol*, 4(8):e1000149, 2008. ISSN 1553-734X (Print) 1553-734x. doi: 10.1371/journal.pcbi.1000149.
20. D. E. Neafsey, M. Juraska, T. Bedford, D. Benkeser, C. Valim, A. Griggs, M. Lievens, S. Abdulla, S. Adjei, T. Agbenyega, S. T. Agnandji, P. Aide, S. Anderson, D. Ansong, J. J. Aponte, K. P. Asante, P. Bejon, A. J. Birkett, M. Bruls, K. M. Connolly, U. D'Alessandro, C. Dobaño, S. Gesase, B. Greenwood, J. Grimsby, H. Tinto, M. J. Hamel, I. Hoffman, P. Kamthunzi, S. Kariuki, P. G. Kremsner, A. Leach, B. Lell, N. J. Lennon, J. Lusingu, K. Marsh, F. Martinson, J. T. Molel, E. L. Moss, P. Njuguna, C. F. Ockenhouse, B. R. Ogutu, W. Otieno, L. Otieno, K. Otieno, S. Owusu-Agyei, D. J. Park, K. Pellé, D. Robbins, C. Russ, E. M. Ryan, J. Sacarlal, B. Sogoloff, H. Sorgho, M. Tanner, T. Theander, I. Valea, S. K. Volkman, Q. Yu, D. Lapierre, B. W. Birren, P. B. Gilbert, and D. F. Wirth. Genetic diversity and protective efficacy of the rts, s/as01 malaria vaccine. *N Engl J Med*, 373(21):2025–2037, 2015. ISSN 0028-4793 (Print) 0028-4793. doi: 10.1056/NEJMoa1505819.
21. P. Moris, E. Jongert, and R. G. van der Most. Characterization of t-cell immune responses in

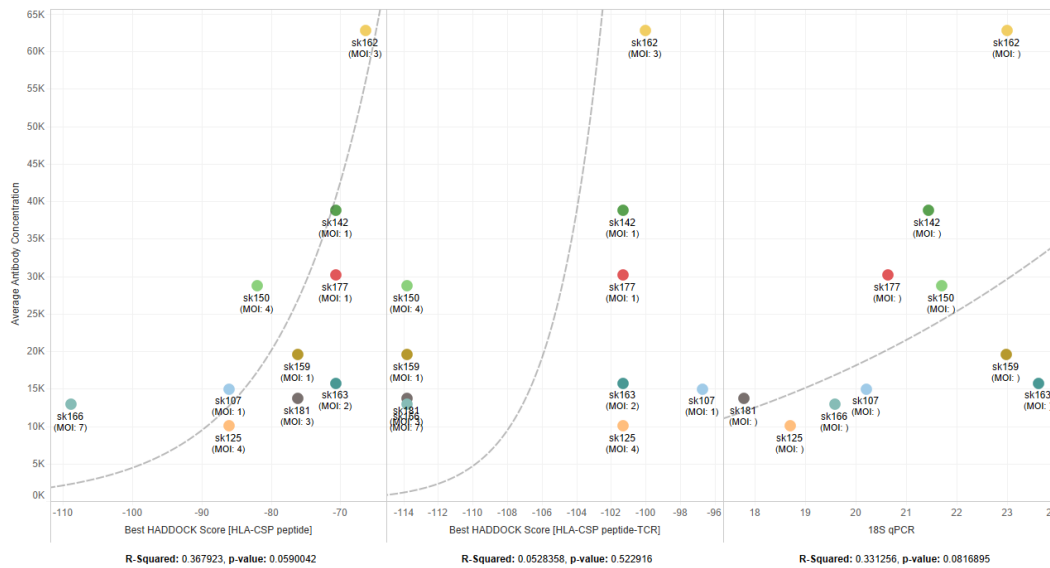


Fig. 8. Correlation plots showing a potential, though not significant, logarithmic relationship between antibody concentration and binding affinity.

clinical trials of the candidate rts,s malaria vaccine. *Hum Vaccin Immunother*, 14(1):17–27, 2018. ISSN 2164-5515 (Print) 2164-5515. doi: 10.1080/21645515.2017.1381809.

22. M. N. Wykes, J. M. Horne-Debets, C. Y. Leow, and D. S. Karunaratne. Malaria drives t cells to exhaustion. *Front Microbiol*, 5:249, 2014. ISSN 1664-302X (Print) 1664-302X. doi: 10.3389/fmicb.2014.00249.

23. Q. J. Liu and B. Gao. Manipulation of mhc-ii/ctr interaction for immune therapy. *Cell Mol Immunol*, 5(3):171–82, 2008. ISSN 1672-7681 (Print) 1672-7681. doi: 10.1038/cmi.2008.21.

24. E. J. Sijts and P. M. Kloetzel. The role of the proteasome in the generation of mhc class i ligands and immune responses. *Cell Mol Life Sci*, 68(9):1491–502, 2011. ISSN 1420-682X (Print) 1420-682X. doi: 10.1007/s00018-011-0657-y.

25. J. A. Speir, J. Stevens, E. Joly, G. W. Butcher, and I. A. Wilson. Two different, highly exposed, bulged structures for an unusually long peptide bound to rat mhc class i rt1-aa. *Immunity*, 14(1):81–92, 2001. ISSN 1074-7613 (Print) 1074-7613. doi: 10.1016/s1074-7613(01)00091-7.

26. L. J. Stern and D. C. Wiley. Antigenic peptide binding by class i and class ii histocompatibility proteins. *Structure*, 2(4):245–51, 1994. ISSN 0969-2126 (Print) 0969-2126. doi: 10.1016/s0969-2126(00)00026-5.

27. F. E. Tynan, N. A. Borg, J. J. Miles, T. Beddoe, D. El-Hassen, S. L. Silins, W. J. van Zuylen, A. W. Purcell, L. Kjer-Nielsen, J. McCluskey, S. R. Burrows, and J. Rossjohn. High resolution structures of highly bulged viral epitopes bound to major histocompatibility complex class i. implications for t-cell receptor engagement and t-cell immunodominance. *J Biol Chem*, 280(25):23900–9, 2005. ISSN 0021-9258 (Print) 0021-9258. doi: 10.1074/jbc.M503060200.

28. W. Zhang, J. Yang, B. He, S. E. Walker, H. Zhang, B. Govindarajoo, J. Virtanen, Z. Xue, H. B. Shen, and Y. Zhang. Integration of quark and i-tasser for ab initio protein structure prediction in casp11. *Proteins*, 84 Suppl 1(Suppl 1):76–86, 2016. ISSN 0887-3585 (Print) 0887-3585. doi: 10.1002/prot.24930.

29. S. J. de Vries, M. van Dijk, and A. M. Bonvin. The haddock web server for data-driven biomolecular docking. *Nat Protoc*, 5(5):883–97, 2010. ISSN 1750-2799. doi: 10.1038/nprot.2010.32.

30. R. Shretta, S. P. Silal, K. Malm, W. Mohammed, J. Narh, D. Piccinini, K. Bertram, J. Rockwood, and M. Lynch. Estimating the risk of declining funding for malaria in ghana: the case for continued investment in the malaria response. *Malar J*, 19(1):196, 2020. ISSN 1475-2875. doi: 10.1186/s12936-020-03267-9.

31. C. C. Dieng, L. Gonzalez, K. Pestana, S. B. Dhikrullahi, L. E. Amoah, Y. A. Afrane, and E. Lo. Contrasting asymptomatic and drug resistance gene prevalence of plasmodium falciparum in ghana: Implications on seasonal malaria chemoprevention. *Genes (Basel)*, 10(7), 2019. ISSN 2073-4425. doi: 10.3390/genes10070538.

32. A. Jalloh, M. Jalloh, and H. Matsuoaka. T-cell epitope polymorphisms of the plasmodium falciparum circumsporozoite protein among field isolates from sierra leone: age-dependent haplotype distribution? *Malar J*, 8:120, 2009. ISSN 1475-2875. doi: 10.1186/1475-2875-8-120.

33. D. Zhong, E. Lo, X. Wang, D. Yewhalaw, G. Zhou, H. E. Atieli, A. Githeko, E. Hemming-Schroeder, M. C. Lee, Y. Afrane, and G. Yan. Multiplicity and molecular epidemiology of plasmodium vivax and plasmodium falciparum infections in east africa. *Malar J*, 17(1):185, 2018. ISSN 1475-2875. doi: 10.1186/s12936-018-2337-y.

34. T. C. Glenn, R. A. Nilsen, T. J. Kieran, J. G. Sanders, N. J. Bayona-Vásquez, J. W. Finger, T. W. Pierson, K. E. Bentley, S. L. Hoffberg, S. Louha, F. J. Garcia-De Leon, M. A. Del Rio Portilla, K. D. Reed, J. L. Anderson, J. K. Meece, S. E. Aggrey, R. Rekaya, M. Alabady, M. Belanger, K. Winker, and B. C. Faircloth. Adapterama i: universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed illumina libraries (itru & inext). *PeerJ*, 7:e7755, 2019. ISSN 2167-8359 (Print) 2167-8359. doi: 10.7717/peerj.7755.

35. J. B. Holm, M. S. Humphrys, C. K. Robinson, M. L. Settles, S. Ott, L. Fu, H. Yang, P. Gajer, X. He, E. McComb, P. E. Gravit, K. G. Ghanem, R. M. Brotman, and J. Ravel. Ultrahigh-throughput multiplexing and sequencing of >500-base-pair amplicon regions on the illumina hiseq 2500 platform. *mSystems*, 4(1), 2019. ISSN 2379-5077 (Print) 2379-5077. doi: 10.1128/mSystems.00029-19.

36. A. Lerch, C. Koepfli, N. E. Hofmann, C. Messerli, S. Wilcox, J. H. Kattenberg, I. Betuela, L. O’Connor, I. Mueller, and I. Felger. Development of amplicon deep sequencing markers and data analysis pipeline for genotyping multi-clonal malaria infections. *BMC Genomics*, 18(1):864, 2017. ISSN 1471-2164. doi: 10.1186/s12864-017-4260-y.

37. A. R. Templeton, K. A. Crandall, and C. F. Sing. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and dna sequence data. iii. cladogram estimation. *Genetics*, 132(2):619–33, 1992. ISSN 0016-6731 (Print) 0016-6731.

38. J. M. Chang, P. Di Tommaso, V. Lefort, O. Gascuel, and C. Notredame. Tcs: a web server for multiple sequence alignment evaluation and phylogenetic reconstruction. *Nucleic Acids Res*, 43(W1):W3–6, 2015. ISSN 0305-1048 (Print) 0305-1048. doi: 10.1093/nar/gkv310.

39. A. Múrias dos Santos, M. P. Cabezas, A. I. Tavares, R. Xavier, and M. Branco. tcsbu: a tool to extend tcs network layout and visualization. *Bioinformatics*, 32(4):627–8, 2016. ISSN 1367-4803. doi: 10.1093/bioinformatics/btv636.

40. C. Keşmir, A. K. Nussbaum, H. Schild, V. Detours, and S. Brunak. Prediction of proteasome cleavage motifs by neural networks. *Protein Eng*, 15(4):287–96, 2002. ISSN 0269-2139 (Print) 0269-2139. doi: 10.1093/protein/15.4.287.

41. M. Z. Tien, D. K. Sydykova, A. G. Meyer, and C. O. Wilke. Peptidebuilder: A simple python library to generate model peptides. *PeerJ*, 1:e80, 2013. ISSN 2167-8359 (Print) 2167-8359. doi: 10.7717/peerj.80.

42. Claus Lundegaard, Kasper Lamberth, Mikkel Harndahl, Søren Buus, Ole Lund, and Morten Nielsen. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. *Nucleic Acids Research*, 36(suppl_2):W509–W512, 05 2008. ISSN 0305-1048. doi: 10.1093/nar/gkn202.

43. Jose L. Sanchez-Trincado, Marta Gomez-Perosanz, and Pedro A. Reche. Fundamentals and methods for t- and b-cell epitope prediction. *Journal of Immunology Research*, 2017: 2680160, 2017. ISSN 2314-8861. doi: 10.1155/2017/2680160.

44. A. V. Hill, C. E. Allsopp, D. Kwiatkowski, N. M. Anstey, P. Twumasi, P. A. Rowe, S. Bennett, D. Brewster, A. J. McMichael, and B. M. Greenwood. Common west african hla antigens are associated with protection from severe malaria. *Nature*, 352(6336):595–600, 1991. ISSN 0028-0836 (Print) 0028-0836. doi: 10.1038/352595a0.

45. L. R. Pratt-Riccio, J. C. Lima-Junior, L. J. Carvalho, M. Theisen, E. C. Espindola-Mendes, F. Santos, J. Oliveira-Ferreira, A. C. Goldberg, C. T. Daniel-Ribeiro, and D. M. Banic. Antibody response profiles induced by plasmodium falciparum glutamate-rich protein in naturally exposed individuals from a brazilian area endemic for malaria. *Am J Trop Med Hyg*, 73(6):1096–103, 2005. ISSN 0002-9637 (Print) 0002-9637.

46. K. W. Wucherpfennig, E. Gagnon, M. J. Call, E. S. Huseby, and M. E. Call. Structural biology of the t-cell receptor: insights into receptor assembly, ligand recognition, and initiation of signaling. *Cold Spring Harb Perspect Biol*, 2(4):a005140, 2010. ISSN 1943-0264. doi: 10.1101/cshperspect.a005140.

47. P. L. Kastirits, J. P. Rodrigues, and A. M. Bonvin. Haddock(2p2): a biophysical model for predicting the binding affinity of protein-protein interaction inhibitors. *J Chem Inf Model*, 54(3):826–36, 2014. ISSN 1549-9596 (Print) 1549-9596. doi: 10.1021/ci4005332.

48. A. Vangone, J. P. Rodrigues, L. C. Xue, G. C. van Zundert, C. Geng, Z. Kurkuoglu, M. Nellen, S. Narasimhan, E. Karaca, M. van Dijk, A. S. Melquiond, K. M. Visscher, M. Trellet, P. L. Kastirits, and A. M. Bonvin. Sense and simplicity in haddock scoring: Lessons from casp-capri round 1. *Proteins*, 85(3):417–423, 2017. ISSN 0887-3585 (Print) 0887-3585. doi: 10.1002/prot.25198.

49. L. E. Amoah, H. B. Abagna, R. Ayanful-Torgby, S. O. Blankson, and N. A. Aryee. Diversity and immune responses against plasmodium falciparum gametocytes in non-febrile school children living in southern ghana. *Malar J*, 18(1):265, 2019. ISSN 1475-2875. doi: 10.1186/s12936-019-2895-7.

50. M. Clement, D. Posada, and K. A. Crandall. Tcs: a computer program to estimate gene genealogies. *Molecular Ecology*, 9(10):1657–1659, 2000. doi: <https://doi.org/10.1046/j.1365-294x.2000.01020.x>.

51. António Múrias dos Santos, Maria Pilar Cabezas, Ana Isabel Tavares, Raquel Xavier, and

- Madalena Branco. tcsBU: a tool to extend TCS network layout and visualization. *Bioinformatics*, 32(4):627–628, 10 2015. ISSN 1367-4803. doi: 10.1093/bioinformatics/btv636.
52. M. J. Lockyer, K. Marsh, and C. I. Newbold. Wild isolates of *Plasmodium falciparum* show extensive polymorphism in t cell epitopes of the circumsporozoite protein. *Mol Biochem Parasitol*, 37(2):275–80, 1989. ISSN 0166-6851 (Print) 0166-6851. doi: 10.1016/0166-6851(89)90159-x.
 53. L. Rénia and Y. S. Goh. Malaria parasites: The great escape. *Front Immunol*, 7:463, 2016. ISSN 1664-3224 (Print) 1664-3224. doi: 10.3389/fimmu.2016.00463.
 54. N. K. Almeida-de Oliveira, R. Abreu-Fernandes, L. Lima-Cury, A. R. Lavigne, A. Pina-Costa, D. S. Perce-da Silva, M. Catanho, A. D. Rossi, P. Brasil, C. Tadeu Daniel-Ribeiro, and M. F. Ferreira-da Cruz. Balancing selection and high genetic diversity of *Plasmodium vivax* circumsporozoite central region in parasites from Brazilian Amazon and Rio de Janeiro Atlantic forest. *PLoS One*, 15(11):e0241426, 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0241426.
 55. F. E. Tynan, N. A. Borg, J. J. Miles, T. Beddoe, D. El-Hassen, S. L. Silins, W. J. van Zuylen, A. W. Purcell, L. Kjer-Nielsen, J. McCluskey, S. R. Burrows, and J. Rossjohn. High resolution structures of highly bulged viral epitopes bound to major histocompatibility complex class I: implications for t-cell receptor engagement and t-cell immunodominance. *J Biol Chem*, 280(25):23900–9, 2005. ISSN 0021-9258 (Print) 0021-9258. doi: 10.1074/jbc.M503060200.
 56. J. S. Brugge and R. L. Erikson. Identification of a transformation-specific antigen induced by an avian sarcoma virus. *Nature*, 269(5626):346–8, 1977. ISSN 0028-0836 (Print) 0028-0836. doi: 10.1038/269346a0.
 57. M. J. Rist, A. Theodossis, N. P. Croft, M. A. Neller, A. Welland, Z. Chen, L. C. Sullivan, J. M. Burrows, J. J. Miles, R. M. Brennan, S. Gras, R. Khanna, A. G. Brooks, J. McCluskey, A. W. Purcell, J. Rossjohn, and S. R. Burrows. HLA peptide length preferences control CD8+ T cell responses. *J Immunol*, 191(2):561–71, 2013. ISSN 0022-1767. doi: 10.4049/jimmunol.1300292.
 58. X. Y. Meng, H. X. Zhang, M. Mezei, and M. Cui. Molecular docking: a powerful approach for structure-based drug discovery. *Curr Comput Aided Drug Des*, 7(2):146–57, 2011. ISSN 1573-4099 (Print) 1573-4099. doi: 10.2174/157340911795677602.
 59. S. G. Remesh, M. Andreatta, G. Ying, T. Kaever, M. Nielsen, C. McMurtrey, W. Hildebrand, B. Peters, and D. M. Zajonc. Unconventional peptide presentation by major histocompatibility complex (MHC) class I allele HLA-A*02:01: Breaking confinement. *J Biol Chem*, 292(13):5262–5270, 2017. ISSN 0021-9258 (Print) 0021-9258. doi: 10.1074/jbc.M117.776542.
 60. A. V. Hill, C. E. Allsopp, D. Kwiatkowski, N. M. Anstey, P. Twumasi, P. A. Rowe, S. Bennett, D. Brewster, A. J. McMichael, and B. M. Greenwood. Common West African HLA antigens are associated with protection from severe malaria. *Nature*, 352(6336):595–600, 1991. ISSN 0028-0836 (Print) 0028-0836. doi: 10.1038/352595a0.
 61. H. Hananantachai, J. Patarapotikul, J. Ohashi, I. Naka, S. Looreesuwan, and K. Tokunaga. Polymorphisms of the HLA-B and HLA-DRB1 genes in Thai malaria patients. *Jpn J Infect Dis*, 58(1):25–8, 2005. ISSN 1344-6304 (Print) 1344-6304.