In silico design of knowledge-based *Plasmodium falciparum* epitope ensemble vaccines

Shymaa Abdullah Damfo[1],

Pedro Reche[2], Derek Gatherer[3],

& Darren R Flower[1,*]

[1]School of Life and Health Sciences, Aston University, Aston Triangle, Birmingham, United Kingdom, B4 7ET.

[2] Immunomedicine Group, Facultad de Medicina, Departamento de Microbiologia I, Universidad Complutense de Madrid, Madrid, Spain.

[3] Division of Biomedical & Life Sciences, Faculty of Health & Medicine, Lancaster University, Lancaster LA1 4YW, UK.

* Corresponding author

Abstract

Malaria is a global health burden, and a major cause of mortality and morbidity in Africa. Here we design a putative malaria epitope ensemble vaccine by selecting an optimal set of pathogen epitopes. From the IEDB database, 584 experimentally-verified CD8+ epitopes and 483 experimentally-verified CD4+ epitopes were collected; 89% of which were found in 8 proteins. Using the PVS server, highly conserved epitopes were identified from variability analysis of multiple alignments of *Plasmodium falciparum* protein sequences. The allele-dependent binding of epitopes was then assessed using IEDB analysis tools, from which the population protection coverage of single and combined epitopes was estimated. Ten conserved epitopes from four well-studied antigens were found to have a coverage of 97.9% of the world population: 7 CD8+ T cell epitopes (LLMDCSGSI, FLIFFDLFLV, LLACAGLAYK, TPYAGEPAPF, LLACAGLAY, SLKKNSRSL, and NEVVVKEEY) and 3 CD4+ T cell epitopes (MRKLAILSVSSFLFV, KSKYKLATSVLAGLL and GLAYKFVVPGAATPYE). The addition of four heteroclitic peptides - single point mutated epitopes - increased HLA binding affinity and raised the predicted world population coverage above 99%.

Keywords

Vaccine design; MHC binding prediction; population coverage; malaria.

Highlights 4

We have designed an epitope ensemble vaccine targeting malaria.

We have extended previous work to include heteroclitic peptides in our epitope ensemble vaccine formulation.

We have estimated the level of population protection afforded by such an ensemble vaccine

We have promulgated a putative global malaria vaccine and ones specific to endemic regions of sub-Saharan Africa.

1 INTRODUCTION

Malaria is a mosquito-borne parasitic tropical disease. Ninety-one countries have continuing malaria transmission, with 212 million cases and 429,000 deaths in 2015. Over 92% of malaria cases occur in Africa and over 75% of malaria deaths in 13 countries within sub-Saharan Africa (WHO, 2016). *Plasmodium falciparum* is the principal malaria parasite in Africa. It has a three-stage lifecycle: exo-erythrocytic and erythrocytic stages in humans and a sporogonic cycle in mosquitoes. The female anopheline mosquito initiates infection by inoculating sporozoites during a blood meal. These invade hepatocytes, resulting in production and release of merozoites which then invade red blood cells. In non-immune adults, clinical symptoms of malaria typically appear 10-15 days after an infective mosquito bite. Symptoms range from mild fever to anaemia, impaired consciousness, severe respiratory distress, and death.

Individuals from endemic areas acquire a natural but incomplete protective immunity to *P. falciparum.* Cellular adaptive immunity is key to the response to *P. falciparum.* The CD8+ T cell-immunity has a role in the exoerythrocytic stage and in regulating immunosuppression during acute malaria (Tsuji, 2010). The transfer of CD8+ T cells clones into naïve mice confers protective immunity, while mice depleted of CD8+ T cells do not acquire protective immunity against malaria. Human CD8+ T cell responses have been detected against several *P. falciparum* proteins, including liver stage antigens. CD4+ T cell responses are needed to control parasite replication, yet cannot clear parasites from the blood. CD4+ T helper 1 cells (Th1) expressing IFN-γ play a role in acute infection, whereas antibody-helper T cells help prevent parasitaemia in the chronic phase (Perez-Mazliah and Langhorne, 2014).

Malaria eradication faces many challenges, such as the emerging resistance of *P. falciparum* to artemisinin and other anti-malarial drugs, resistance of anopheline mosquitoes to insecticides, the absence of treatment and prevention programs in many areas, and reduced global anti-malaria funding. Thus there remains a pressing need to develop effective vaccines, the most cost-effective public health intervention for infectious disease. The first successful human malaria vaccine consists of live, radiation-attenuated *P. falciparum* sporozoites (PfSPZ), which confers protective immunity by inducing a CD8+ T cell response (Epstein *et al.*, 2011; Weiss and Jiang, 2012). Despite such promise, live attenuated vaccines challenges (e.g. sterile manufacturing, extreme cold-chain requirements, and safety concerns) have prompted the development of subunit vaccines.

Current malaria subunit vaccines either target the parasite at the pre-erythrocytic stage or the asexual blood stage, or they act to block transmission. The selection of appropriate antigens as vaccine subunits is key to treating the different parasite stages. Vaccines targeting antigens expressed by gametes and zygotes, such as Pfs48/45, Pfs28, and Pfs25, may block transmission (Li *et al*., 2016) at the sexual and sporogonic stages. Although this would not prevent clinical disease, it may slow infection rates (Birkett *et al*., 2013). Pre-erythrocytic vaccines aim to prevent sporozoites entering the liver and developing into infective merozoites, targeting sporozoites' apical membrane antigens. However, such vaccines have failed to show significant efficacy against clinical malaria (Thera *et al*., 2011). Variant surface antigens expressed on the infected erythrocyte such as *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) elicit an immune response in humans, but high antigenic diversity may diminish vaccine efficacy (Barry and Arnott, 2014). Thus the highly conserved regions of *Plasmodium* antigens must be targeted.

Here, we design a putative peptide-based vaccine against malaria, comprising a small set of highly conserved epitopes of experimentally verified immunogenicity, which provide a wide population coverage. Our recent work has exemplified this approach against viruses: Hepatitis C (Molero-Abraham et al., 2013) and influenza (Sheikh et al., 2016). Since *P. falciparum* is likely to exhibit non-negligible allelic diversity in its antigenic proteins, we also considered epitope variants by introducing point mutations into the amino acid sequences of our selected epitopes. Such variants are commonly referred to as heteroclitic peptides, and increase MHC binding relative to the unsubstituted peptide, thereby potentiating intrinsic immunogenicity by acting as self-adjuvants (Adegoke et al., 2015; Adegoke and Grant, 2015)

2 METHODS

2.1 Collection of $CD8^+$ and $CD4^+$ T cell epitopes

P. falciparum epitopes recognised by the T cell-mediated immune response were obtained from the Immune Epitope Database (IEDB), accessible at URL: http://www.iedb.org (Vita et al., 2015). The search was restricted to positive T cell assays in humans.

2.2 Retrieval, clustering, and multiple sequence alignment generation for *P. falciparum* antigens

The complete sequence of each protein containing at least one downloaded epitope (termed antigen hereafter) was retrieved from the Universal Protein Resource, URL: http://www.uniprot.org/uniprot/. Each retrieved antigen was BLASTed against the non-redundant (NR) protein sequence database at the National Centre for Biotechnology information (NCBI), URL: https://blast.ncbi.nlm.nih.gov/Blast.cgi. Ten variant sequences corresponding to each identified *P. falciparum* antigen were selected. For the 8 antigens with the most epitopes a larger set containing more sequences was collected, and then clustered using CD-HIT (Fu et al., 2012), URL: http://weizhongli-lab.org/cdhit_suit. Sequences were clustered at identity thresholds of 90% and 98%. The two sets - highly homologous and CD-HIT clustered sequences - were submitted to the MUSCLE server http://www.ebi.ac.uk/Tools/msa/muscle/ (Edgar, 2004) to create multiple sequence alignments (MSA).

2.3 Calculating sequence variability of *P. falciparum* polyproteins

To identify conserved epitopes, we first analysed each MSAs using the Protein Variability Server (PVS), URL: http://imed.med.ucm.es/PVS/ (Garcia-Boronat et al., 2008), to separate conserved consensus sequences from sequences exhibiting variability (aligned positions with Shannon Entropy, H, above 0.5). Subsequently, conserved epitopes were identified as those epitopes falling within non-variable regions.

2.4 Estimation of MHC class I and class II binding

Affinities of conserved CD8+ T cell epitopes were predicted using the IEDB MHC-I binding prediction tool (http://tools.iedb.org/mhci/)(Kim et al., 2012). Epitopes with percentile rank of one or less for a given allele were considered to bind. Similarly, conserved CD4+ T cell epitopes were processed with the IEDB MHC-II binding prediction tool (http://tools.iedb.org/mhcii/), using the consensus method. Alleles in the top 10 percentile rank were considered to bind. This yielded distinct allele binding profiles for each epitope.

2.5 Calculation of predicted population coverage

Using the binding profile from 2.4 above, conserved epitopes were analysed for predicted population coverage (PPC) using the IEDB analysis tool, (http://tools.iedb.org/tools/population/iedb_input). This covers 78 populations across multiple geographical areas (Bui et al., 2006). PPC quantifies the percentage of a given population likely to produce an immune response to an epitope. PPC values for sub-Saharan Africa and globally were estimated for both individual epitopes and epitope combinations.

2.6 Generation of heteroclitic peptides and estimation their binding affinity and population coverage

Heteroclitic peptides were generated by making 19 virtual single amino acid substitutions at each position of each epitope sequence, giving a set of 19 x *l* variants, where *l* is the epitope length. Each generated variant peptide was analysed using the same methodology as in 2.4 and 2.5 above. The resulting predicted binding, allele binding profile, and PPC values were then compared to that of each progenitor epitope from which the heteroclitic variant set had been derived. This reduced the set of variant heteroclitic peptides to only those that had higher predicted affinity and higher predicted population coverage than the core epitope.

3 RESULTS

3.1. CD8+ T cell epitope component

We extracted 584 experimentally-confirmed CD8+ T cell epitopes from IEDB, deriving from 25 separate antigens. CD8+ epitope frequency varied considerably between these antigens, resulting in a long-tail distribution (Figure 1). Of the 25 IEDB epitope-bearing antigens, 8 contained 10 or more epitopes, giving 518 epitopes in total. As detailed in Figure 2, we analysed these epitopes using 4 different approaches.

First, for each of the 25 antigens, we created a separate multiple sequence alignment (MSA) comprising the 10 most similar but non-identical sequences found in the NCBI database. Variability analysis using the PVS server yielded 55 conserved epitopes deriving from 6 antigens: TRAP (22 epitopes), CSP (13 epitopes), AMA-1 (8 epitopes), LSA-3 (7 epitopes), EXP-1 (4 epitopes) and STARP (1 epitope).

Second, we analysed the MSAs corresponding to the 8 antigens bearing the highest number of epitopes. This yielded 47 conserved epitopes deriving from 4 proteins: TRAP (22 epitopes), CSP (13 epitopes), AMA-1(8 epitopes), and EXP-1(4 epitopes).

Third, we clustered at 90% similarity larger numbers of similar sequences corresponding to these 8 antigens. This yielded 34 conserved epitopes from the same 4 proteins.

Fourth, we repeated this process, again for the 8 antigens, but at 98% similarity. This yielded 28 conserved epitopes, again from the same 4 proteins. Using the IEDB-binding prediction tool, HLA binding profiles were calculated for the 4 sets of epitopes, and the population coverage (PPC) calculated using the IEDB population-coverage tool. No peptide had a PPC greater than 44.9%. (Table 1).

3.2. CD4+ T cell epitope component

We extracted 483 experimentally-confirmed CD4+ T cell epitopes in IEDB, deriving from 25 antigens. As detailed in Figure 3, we analysed CD4+ epitopes using 4 different approaches.

First, for each of these 25 antigens, we created a separate MSA, comprising the 10 most similar but non-identical sequences found in the NCBI database. Variability analysis using the PVS server yielded 191 conserved CD4+ T cell epitope, deriving from 19 antigens.

Second, 426 epitopes were collected from the 8 prioritised antigens shown in Figure 1. For consistency, we targeted these 8 due to the pre-eminence of CD8+ responses in protective immunity to malaria (Tsuji, 2010). Of the 426 epitopes, 153 were conserved and derived from

7 antigens: MSP-1 (68 epitopes), CSP (40 epitopes), DNAJ (15 epitopes), AMA-1 (13 epitopes), MSP-2 (7 epitopes), TRAP (6 epitopes) and EXP-1 (4 epitopes).

Third, for each of the 8 antigens, we clustered at 90% similarity larger numbers of similar sequences to generate a MSA. Variability analysis using PVS yielded 134 conserved CD4+ T cell epitopes.

Fourth, after clustering at 98%, 106 epitopes were conserved. Fourth, we repeated clustering, again for the 8 antigens, but at 98% similarity. This yielded 106 conserved CD4+ T cell epitopes. Using the IEDB-binding prediction tool, HLA binding profiles were calculated for the 4 sets of epitopes detailed above, and the population coverage (PPC) calculated using the IEDB population-coverage tool. No peptide had a PPC greater than 78.9%(Table 2).

3.3. Selecting high population coverage epitope combinations

For the 8 selected antigens, 8 CD8+ T cell epitopes (LLMDCSGSI, FLIFFDLFLV, LLACAGLAYK, TPYAGEPAPF, LLACAGLAY, LACAGLAYK, NEVVVKEEY and SLKKNSRSL) and 12 CD4+ epitopes (MRKLAILSVSSFLFV, SSVFNVVNSSIGLIM, KSKYKLATSVLAGLL, GLAYKFVVPGAATPY, AGLLGVVSTVLLGGV, KMEDYIKKNKTTIAN, DFNHYYTLKTGLEAD, HYYTLKTGLEADIKK, KYKIAGGIAGGLALL, KFSSSNNSVYNVQKL, DIEKKICKMEKCSSV, VKNVIGPFMKAVCVE) were common to the 3 data sets (10 most similar sequences, CD-Hit 90% and CD-Hit 98%). These provided a coherent basis for the construction of putative epitope ensemble vaccine designs. By the careful selection of conserved CD8+ and CD4+ epitopes it was possible to design a variety of different epitope collections, each with a combined PPC over 95% (Table 3).

By considering common conserved epitopes, with distinct HLA binding profiles, a putative optimal selection representing a minimally genetically-biased vaccine could be made: it comprised 7 CD8+ T cell epitopes (LLMDCSGSI, FLIFFDLFLV, LLACAGLAYK, TPYAGEPAPF, LLACAGLAY, NEVVVKEEY and SLKKNSRSL) and 3 CD4+ T cell epitopes (MRKLAILSVSSFLFV, KSKYKLATSVLAGLL and GLAYKFVVPGAATPY). This peptide set nominally covers 97.9% of the world population and over 85% of three African regions (East Africa: 88.2%; West Africa: 90.2%; Central Africa 86.3%; South Africa: 73.8%).

3.4. Selection of heteroclitic epitopes

Following Theiler *et al*. (2016), we sought to broaden and deepen the potential scope and efficacy of this putative vaccine peptide set by adding heteroclitic epitopes to the candidate

ensemble. Heteroclitic peptides should be self-adjuvanting, mimic natural variation exhibited by pathogen subspecies, but remain immunologically representative of a target epitope(Adegoke and Grant, 2015). For each epitope, a series of heteroclitic peptides was generated systematically by making 19 single amino acid substitutions at each position of the epitope sequence. For SLKKNSRSL this corresponded to 171 variants (19 x *l*, where *l* is the peptide length) and for GLAYKFVVPGAATPY this generated 285 variants.

After generating potential Heteroclitic epitopes, we assessed each variant for predicted binding affinity and calculated PPC. For example, LLMDCSGSI had 16 point-mutations with larger PPC values: the highest being the LLMDHSGSI epitope (60.0%); while the MRKLAILSASSFLFV variant bound 22 HLA molecules, allowing a single heteroclitic peptide to cover 81.8% of the world population. Heteroclitic epitopes had higher PPC values for all epitopes, except TPYAGEPAP and NEVVVKEEY (Table 4).

After assessment of binding profile and PPC of individual heteroclitic peptides, we then selected heteroclitic peptides for inclusion in proposed epitope ensemble vaccines. Addition of the heteroclitic peptide TFYAGEPAPF to the ensemble for example alone raised its world PPC to 98.8% and raised coverage in all four African regions over 80% (Table 5).

Adding FAIFFDLFLV (an analogue of FLIFFDLFLV mutated at position 2), MLACAGLAYK (an analogue of LLACAGLAYK mutated at position 1), TFYAGEPAPF (an analogue of TPYAGEPAPF mutated at position 2), and RLACAGLAY (an analogue of LLACAGLAY mutated at position 1) to our best combination, formed a final 14 epitope ensemble. The addition of heteroclitic peptides to this epitope combination raised the world PPC to 99.2%, East Africa to 94.6%, West Africa to 96.5%, Central Africa to 92.6%, and South Africa to 90.9%.

5 DISCUSSION

In the context of a worldwide vaccination programme, a broadly protective anti-malaria vaccine would assist malaria eradication both globally and in endemic areas. Natural human acquired immunity can generate antibodies against the blood stage antigens while immunisation with irradiated sporozoites can generate a cellular response preventing hepatocyte invasion (Doolan *et al*., 2014). Major challenges remain however, such as the multifactorial nature of protective immunity, the many epitopes recognised by the immune system, antigenic variation, and MHC allelic diversity. Subunit vaccines based on a single antigen may evoke too narrow an immune response and lack protection in genetically diverse populations. Thus it is important to prioritize both antigens and epitopes for vaccine design (Diez-Rivero and Reche, 2012; Molero-Abraham *et al*., 2015). The WHO has a strategic goal of developing a second-generation malaria vaccine with 75% or more efficacy by 2030 (WHO, 2016). An effective vaccine would likely include antigenic determinants inducing protective immunity not achieved naturally or following immunisation with an irradiated sporozoite. Here, we sought to identify highly conserved epitopes that could form the basis of a putative epitope ensemble vaccine against malaria, with a target population coverage of 95%.

One example of an epitope-based malaria vaccine - RTS,S - already exists. RTS,S is constructed from part of the CSP repeated region and C terminal region known to contain T cell epitopes (Cohen et al., 2010). Specifically, RTS include one conserved CD4+ T cell epitope and two polymorphic CD4+ and CD8+ T cell epitopes (Crompton *et al*., 2010). Although RTS,S demonstrates the feasibility of an epitope ensemble vaccine targeting malaria, in clinical trials following patients over 7 years, the efficacy of RTS,S is not maintained, being almost zero by the fourth year (Olotu et al., 2016). Our proposed epitope ensemble vaccine contains both conserved CD8+ and CD4+ T cell epitopes from 8 *P. falciparum* antigens with 95% population coverage.

In our designed putative vaccine, two epitopes were part of the CSP protein, expressed in large quantities by sporozoites (Florens et al., 2002), six were derived from the thrombospondin-related anonymous protein (TRAP), also known as sporozoite surface protein 2 (SSP2). CSP and TRAP/SSP2 immunogenicity has been demonstrated previously and it has been suggested that joint immunisation with both TRAP and CSP could be more efficacious than separate vaccines (Hodgson *et al*., 2015). Finally, our proposed peptide set

contains epitopes from AMA-1 and EXP-1, as well as CSP and TRAP. A multi-epitope vaccine targeting many antigens should induce a more robust immune response. Doolan *et al*. (2003) have shown that irradiated sporozoite-induced protective immunity following immunisation is dispersed over a large number of antigens. While previous studies show that sustained complete protection may be induced by irradiated sporozoites, no current vaccine yields such sustained immunity.

Nine of ten epitopes in our final ensemble were associated with the pre-erythrocytic (hepatic) stage antigens; and it is this stage which will induce cell-mediated responses. Many studies have shown hepatic stage antigens to induce CD8+ T cell response correlated to sterile protection (Ewer *et al*., 2013). One MHC-I epitope included in our epitope ensemble vaccine was derived from the AMA-1 antigen expressed in both pre-erythrocytic and erythrocytic stages. This AMA-1 derived epitope - NEVVVKEEY - shows binding affinity to the HLA-B44 molecule. No epitope in our final ensemble was derived from sexual stage antigens, since vaccines targeting sexual stage antigens are mainly mediated by antibodies.

Our final combination contains both CD8+ and CD4+ epitopes. All of the CD4+ epitopes were also independently reported by Doolan et al. (2000) as experimentally-confirmed, highly-conserved, cross-reactive T cell epitopes. All epitopes bound extensively to the most prevalent DRB1 alleles, and gave good recall of lymphoproliferative and cytokine (IFN-γ and IL-10) responses in Kenyan residents and irradiated sporozoite-immunised individuals. Doolan et al. (2000) estimated population coverage of these epitopes in five ethnicities (Caucasian, Hispanic, North American Black, Japanese and Chinese) with a higher minimum coverage of (72.0%) compared to the current study (60.3%). Eight conserved CD8+ epitopes showed binding affinities to four HLA-A alleles (A*01, A*02, A*03 and A*11) and seven HLA-B alleles (B*07, B*08, B*15, B*35, B*44 and B*53). Combining these epitopes provide a universal PPC value of 88.4%. According to Sette and Sidney (1999), epitopes binding to six HLA I allele super-types (A*01, A*02, A*03, B*07, A*24, and B*44) are needed to cover 99.3% of the world population. Thus, a CD8+ epitope that binds to HLA-A*24 is required to achieve optimal coverage.

Peptide-based vaccines have several advantages, including safety, the possibility of modifying peptide sequences to improve immunogenicity, and the ability to obtain focused immune responses for selected antigens(Purcell et al., 2007). Prior to this, exploratory work is also needed on the best means to deliver the set of peptides, whether as a cocktail of peptide

fragments, a single polypeptide, as part of an expression vector-based vaccine, as part of a DNA vaccine, or as part of a fusion vaccine with a carrier virus.

One of the significant advances over our previous work (Molero-Abraham et al., 2013; Sheikh et al., 2016) is the addition of carefully selected heteroclitic peptides. A well-known drawback of epitope ensemble vaccines is poor immunogenicity, usually necessitating the use of suitable adjuvants (Bayry et al., 2009). In this context, adding heteroclitic peptides should foment self-adjuvanting effects by augmenting the immune responses of individual epitopes and of the ensemble as a whole. For example, in the immunotherapy of human HIV, where heteroclitic peptides can raise CD8+ T cell proliferation by 20% relative to reference peptides (Adegoke et al., 2015). Adding variant sequences close to those of recognition epitopes is also known experimentally to broaden the breadth and depth of vaccine immune response in animal models (Abdul-Jawad et al., 2016; Barouch et al., 2010; Santra et al., 2010). Thus, the capacity of a heteroclitic peptide to enhance T cell function has important implications for chronic infection, where T cell exhaustion is more likely. Moreover, heteroclitic peptides bind to additional HLA molecules providing higher population coverage compared to native epitopes.

In our designed putative vaccine, single amino acid substitutions increased the population coverage by 1.6% to 12.1% for a single peptide. Thus, combining several variant peptides with core epitopes increases overall coverage to above 99%. Several heteroclitic peptides bind to HLA-A*30 and HLA-A*23, which are frequent alleles in African populations, and thus allow coverage to rise over 90% for African regions.

However, individual and combined epitopes provided lower coverage of Africa compared to the global population. The epitope TPYAGEPAPF binds to the HLA-B allele (B*53:01), which is common in the West African population (Spinola et al., 2011). Importantly, HLA-B*53 has been associated with the reduction in life-threatening malaria (Hill et al., 1991). Moreover, other epitopes did not bind to alleles with high frequencies in African regions. Thus, to reach nearly 95% coverage, other epitopes restricted by the prevalent super-types (HLA-A*23, HLA-A*30, HLA-B*58, and HLA-C*06) were required (Peterson et al., 2014; Spinola et al., 2011). In TPYAGEPAPF, substitution of proline (P) with phenylalanine (F) resulted in TFYAGEPAPF, which is restricted by HLA-A*23. Addition of this heteroclitic epitope to our vaccine ensemble provides higher coverage in Africa. MHC-II epitopes bind to many prevalent alleles in Africa, such as DRB3*02:02:01, DRB1*11:01:02, DRB4* 01:01:01, DRB5*01:01:01. Combination of

MHC-II epitopes failed to reach 95% coverage for African regions. Other epitopes restricted to allelic supertypes including DQA1*01:02:01, DQB1*03:01:01, DPA1*01:03:01, and DPB1*01:01:01 (Peterson et al., 2014) would be necessary to immunise most subpopulations in Africa.

6 Conclusion

Leveraging comparative sequence analysis, we used immunoinformatics to select an optimal set of experimentally-tested epitopes known to be immunogenic as a designed genetically-unbiased pre-validated epitope-based synthetic malaria vaccine. We selected an epitope set comprising LLMDCSGSI, FLIFFDLFLV, LLACAGLAYK, TPYAGEPAPF, LLACAGLAY, NEVVVKEEY, SLKKNSRSL, MRKLAILSVSSFLFV, KSKYKLATSVLAGLL and GLAYKFVVPGAATPY as the best candidate for a multi-epitope vaccine targeting malaria. Together, these epitopes are predicted to provide high universal coverage (97.9%). Both HLA class I and II epitopes are required for a broadly effective malaria vaccine. Our proposed epitope combination included 4 heteroclitic peptides selected for having increased HLA presentation profiles. Targeting a specific population is more difficult, highlighting the role that heteroclitic peptides might play in achieving high subpopulation coverage. Heteroclitic peptides with single amino acids substitution have different HLA binding profiles, raising the overall population coverage of our ensemble. In-vitro and in vivo testing of our ensemble vaccine for protective immune responses is now needed, including a thorough evaluation of peptide properties (such as solubility), their toxicity, heteroclitic peptides immunogenicity, and a rigorous examination of their efficacy against malaria.

ACKNOWLEDGEMENTS

Funding: This research was supported by Aston University. PAR is supported by grants BIO2014:54164-R and Inmunotek S.L.

REFERENCES

Abdul-Jawad, S., Ondondo, B., van Hateren, A., Gardner, A., Elliott, T., Korber, B., Hanke, T., 2016. Increased Valency of Conserved-mosaic Vaccines Enhances the Breadth and Depth of Epitope Recognition. Mol Ther 24, 375-84, doi:10.1038/mt.2015.210.

Adegoke, A., Gladney, K., Gallant, M., Grant, M., 2015. Heteroclitic Peptides Increase Proliferation and Reduce Evidence of Human Immunodeficiency Virus-Specific CD8(+) T Cell Dysfunction. Viral Immunol 28, 455-63, doi:10.1089/vim.2015.0036.

Adegoke, A. O., Grant, M. D., 2015. Enhancing Human Immunodeficiency Virus-Specific CD8(+) T Cell Responses with Heteroclitic Peptides. Front Immunol 6, 377, doi:10.3389/fimmu.2015.00377.

Ahouidi, A. D., Amambua-Ngwa, A., Awandare, G. A., Bei, A. K., Conway, D. J., Diakite, M., Duraisingh, M. T., Rayner, J. C., Zenonos, Z. A., 2016. Malaria Vaccine Development: Focusing Field Erythrocyte Invasion Studies on Phenotypic Diversity: The West African Merozoite Invasion Network (WAMIN). Trends Parasitol 32, 274-83, doi:10.1016/j.pt.2015.11.009.

Barouch, D. H., O'Brien, K. L., Simmons, N. L., King, S. L., Abbink, P., Maxfield, L. F., Sun, Y. H., La Porte, A., Riggs, A. M., Lynch, D. M., Clark, S. L., Backus, K., Perry, J. R., Seaman, M. S., Carville, A., Mansfield, K. G., Szinger, J. J., Fischer, W., Muldoon, M., Korber, B., 2010. Mosaic HIV-1 vaccines expand the breadth and depth of cellular immune responses in rhesus monkeys. Nat Med 16, 319-23, doi:10.1038/nm.2089.

Barry, A. E., Arnott, A., 2014. Strategies for designing and monitoring malaria vaccines targeting diverse antigens. Front Immunol 5, 359, doi:10.3389/fimmu.2014.00359.

Bayry, J., Lacroix-Desmazes, S., Kaveri, S. V., 2009. Novel therapeutic strategies for multiple sclerosis: potential of intravenous immunoglobulin. Nat Rev Drug Discov 8, 594, doi:10.1038/nrd2358-c1.

Birkett, A. J., Moorthy, V. S., Loucq, C., Chitnis, C. E., Kaslow, D. C., 2013. Malaria vaccine R&D in the Decade of Vaccines: breakthroughs, challenges and opportunities. Vaccine 31 Suppl 2, B233-43, doi:10.1016/j.vaccine.2013.02.040.

Bui, H. H., Schiewe, A. J., von Grafenstein, H., Haworth, I. S., 2006. Structural prediction of peptides binding to MHC class I molecules. Proteins 63, 43-52, doi:10.1002/prot.20870.

Cohen, T., Moise, L., Ardito, M., Martin, W., De Groot, A. S., 2010. A method for individualizing the prediction of immunogenicity of protein vaccines and biologic therapeutics: individualized T cell epitope measure (iTEM). J Biomed Biotechnol 2010, doi:10.1155/2010/961752.

Crompton, P. D., Pierce, S. K., Miller, L. H., 2010. Advances and challenges in malaria vaccine development. J Clin Invest 120, 4168-78, doi:10.1172/JCI44423.

Diez-Rivero, C. M., Reche, P. A., 2012. CD8 T cell epitope distribution in viruses reveals patterns of protein biosynthesis. PLoS One 7, e43674, doi:10.1371/journal.pone.0043674.

Doolan, D. L., Apte, S. H., Proietti, C., 2014. Genome-based vaccine design: the promise for malaria and other infectious diseases. Int J Parasitol 44, 901-13, doi:10.1016/j.ijpara.2014.07.010.

Doolan, D. L., Aguiar, J. C., Weiss, W. R., Sette, A., Felgner, P. L., Regis, D. P., Quinones-Casas, P., Yates, J. R., 3rd, Blair, P. L., Richie, T. L., Hoffman, S. L., Carucci, D. J., 2003. Utilization of genomic sequence information to develop malaria vaccines. J Exp Biol 206, 3789-802.

Doolan, D. L., Southwood, S., Chesnut, R., Appella, E., Gomez, E., Richards, A., Higashimoto, Y. I., Maewal, A., Sidney, J., Gramzinski, R. A., Mason, C., Koech, D., Hoffman, S. L., Sette, A., 2000. HLA-DR-promiscuous T cell epitopes from Plasmodium falciparum pre-erythrocytic-stage antigens restricted by multiple HLA class II alleles. J Immunol 165, 1123-37.

Edgar, R. C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32, 1792-7, doi:10.1093/nar/gkh340.

Epstein, J. E., Tewari, K., Lyke, K. E., Sim, B. K., Billingsley, P. F., Laurens, M. B., Gunasekera, A., Chakravarty, S., James, E. R., Sedegah, M., Richman, A., Velmurugan, S., Reyes, S., Li, M., Tucker, K., Ahumada, A., Ruben, A. J., Li, T., Stafford, R., Eappen, A. G., Tamminga, C., Bennett, J. W., Ockenhouse, C. F., Murphy, J. R., Komisar, J., Thomas, N., Loyevsky, M., Birkett, A., Plowe, C. V., Loucq, C., Edelman, R., Richie, T. L., Seder, R. A., Hoffman, S. L., 2011. Live

attenuated malaria vaccine designed to protect through hepatic CD8(+) T cell immunity. Science 334, 475-80, doi:10.1126/science.1211548.

Ewer, K. J., O'Hara, G. A., Duncan, C. J., Collins, K. A., Sheehy, S. H., Reyes-Sandoval, A., Goodman, A. L., Edwards, N. J., Elias, S. C., Halstead, F. D., Longley, R. J., Rowland, R., Poulton, I. D., Draper, S. J., Blagborough, A. M., Berrie, E., Moyle, S., Williams, N., Siani, L., Folgori, A., Colloca, S., Sinden, R. E., Lawrie, A. M., Cortese, R., Gilbert, S. C., Nicosia, A., Hill, A. V., 2013. Protective CD8+ T-cell immunity to human malaria induced by chimpanzee adenovirus-MVA immunisation. Nat Commun 4, 2836, doi:10.1038/ncomms3836.

Ferraro, B., Talbott, K. T., Balakrishnan, A., Cisper, N., Morrow, M. P., Hutnick, N. A., Myles, D. J., Shedlock, D. J., Obeng-Adjei, N., Yan, J., Kayatani, A. K., Richie, N., Cabrera, W., Shiver, R., Khan, A. S., Brown, A. S., Yang, M., Wille-Reece, U., Birkett, A. J., Sardesai, N. Y., Weiner, D. B., 2013. Inducing humoral and cellular responses to multiple sporozoite and liver-stage malaria antigens using exogenous plasmid DNA. Infect Immun 81, 3709-20, doi:10.1128/IAI.00180-13.

Florens, L., Washburn, M. P., Raine, J. D., Anthony, R. M., Grainger, M., Haynes, J. D., Moch, J. K., Muster, N., Sacci, J. B., Tabb, D. L., Witney, A. A., Wolters, D., Wu, Y., Gardner, M. J., Holder, A. A., Sinden, R. E., Yates, J. R., Carucci, D. J., 2002. A proteomic view of the Plasmodium falciparum life cycle. Nature 419, 520-6, doi:10.1038/nature01107.

Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150-2, doi:10.1093/bioinformatics/bts565.

Garcia-Boronat, M., Diez-Rivero, C. M., Reinherz, E. L., Reche, P. A., 2008. PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. Nucleic Acids Res 36, W35-41, doi:10.1093/nar/gkn211.

Ghosh, S., Kennedy, K., Sanders, P., Matthews, K., Ralph, S. A., Counihan, N. A., de Koning-Ward, T. F., 2017. The Plasmodium rhoptry associated protein complex is important for parasitophorous vacuole membrane structure and intraerythrocytic parasite growth. Cell Microbiol 19, doi:10.1111/cmi.12733.

Hill, A. V., Allsopp, C. E., Kwiatkowski, D., Anstey, N. M., Twumasi, P., Rowe, P. A., Bennett, S., Brewster, D., McMichael, A. J., Greenwood, B. M., 1991. Common west African HLA antigens are associated with protection from severe malaria. Nature 352, 595-600, doi:10.1038/352595a0.

Hodgson, S. H., Ewer, K. J., Bliss, C. M., Edwards, N. J., Rampling, T., Anagnostou, N. A., de Barra, E., Havelock, T., Bowyer, G., Poulton, I. D., de Cassan, S., Longley, R., Illingworth, J. J., Douglas, A. D., Mange, P. B., Collins, K. A., Roberts, R., Gerry, S., Berrie, E., Moyle, S., Colloca, S., Cortese, R., Sinden, R. E., Gilbert, S. C., Bejon, P., Lawrie, A. M., Nicosia, A., Faust, S. N., Hill, A. V., 2015. Evaluation of the efficacy of ChAd63-MVA vectored vaccines expressing circumsporozoite protein and ME-TRAP against controlled human malaria infection in malaria-naive individuals. J Infect Dis 211, 1076-86, doi:10.1093/infdis/jiu579.

John, C. C., Zickafoose, J. S., Sumba, P. O., King, C. L., Kazura, J. W., 2003. Antibodies to the Plasmodium falciparum antigens circumsporozoite protein, thrombospondin-related adhesive protein, and liver-stage antigen 1 vary by ages of subjects and by season in a highland area of Kenya. Infect Immun 71, 4320-5.

Kim, H., Certa, U., Dobeli, H., Jakob, P., Hol, W. G., 1998. Crystal structure of fructose-1,6-bisphosphate aldolase from the human malaria parasite Plasmodium falciparum. Biochemistry 37, 4388-96, doi:10.1021/bi972233h.

Kim, Y., Ponomarenko, J., Zhu, Z., Tamang, D., Wang, P., Greenbaum, J., Lundegaard, C., Sette, A., Lund, O., Bourne, P. E., Nielsen, M., Peters, B., 2012. Immune epitope database analysis resource. Nucleic Acids Res 40, W525-30, doi:10.1093/nar/gks438.

Li, Y., Leneghan, D. B., Miura, K., Nikolaeva, D., Brian, I. J., Dicks, M. D., Fyfe, A. J., Zakutansky, S. E., de Cassan, S., Long, C. A., Draper, S. J., Hill, A. V., Hill, F., Biswas, S., 2016. Enhancing immunogenicity and transmission-blocking activity of malaria vaccines by fusing Pfs25 to IMX313 multimerization technology. Sci Rep 6, 18848, doi:10.1038/srep18848.

Molero-Abraham, M., Lafuente, E. M., Flower, D. R., Reche, P. A., 2013. Selection of conserved epitopes from hepatitis C virus for pan-populational stimulation of T-cell responses. Clin Dev Immunol 2013, 601943, doi:10.1155/2013/601943.

Molero-Abraham, M., Glutting, J. P., Flower, D. R., Lafuente, E. M., Reche, P. A., 2015. EPIPOX: Immunoinformatic Characterization of the Shared T-Cell Epitome between Variola Virus and Related Pathogenic Orthopoxviruses. J Immunol Res 2015, 738020, doi:10.1155/2015/738020.

Nilsson, S., Moll, K., Angeletti, D., Albrecht, L., Kursula, I., Jiang, N., Sun, X., Berzins, K., Wahlgren, M., Chen, Q., 2011. Characterization of the Duffy-Binding-Like Domain of Plasmodium falciparum Blood-Stage Antigen 332. Malar Res Treat 2011, 671439, doi:10.4061/2011/671439.

Okenu, D. M., Riley, E. M., Bickle, Q. D., Agomo, P. U., Barbosa, A., Daugherty, J. R., Lanar, D. E., Conway, D. J., 2000. Analysis of human antibodies to erythrocyte binding antigen 175 of Plasmodium falciparum. Infect Immun 68, 5559-66.

Olotu, A., Fegan, G., Wambua, J., Nyangweso, G., Leach, A., Lievens, M., Kaslow, D. C., Njuguna, P., Marsh, K., Bejon, P., 2016. Seven-Year Efficacy of RTS,S/AS01 Malaria Vaccine among Young African Children. N Engl J Med 374, 2519-29, doi:10.1056/NEJMoa1515257.

Owalla, T. J., Palacpac, N. M., Shirai, H., Horii, T., Egwang, T. G., 2013. Association of naturally acquired IgG antibodies against Plasmodium falciparum serine repeat antigen-5 with reduced placental parasitemia and normal birth weight in pregnant Ugandan women: a pilot study. Parasitol Int 62, 237-9, doi:10.1016/j.parint.2013.01.006.

Perez-Mazliah, D., Langhorne, J., 2014. CD4 T-cell subsets in malaria: TH1/TH2 revisited. Front Immunol 5, 671, doi:10.3389/fimmu.2014.00671.

Peterson, E. C., Hambuchen, M. D., Tawney, R. L., Gunnell, M. G., Cowell, J. L., Lay, J. O., Jr., Blough, B. E., Carroll, F. I., Owens, S. M., 2014. Simple radiometric method for accurately quantitating epitope densities of hapten-protein conjugates with sulfhydryl linkages. Bioconjug Chem 25, 2112-5, doi:10.1021/bc500456z.

Pooe, O. J., Kollisch, G., Heine, H., Shonhai, A., 2017. Plasmodium falciparum Heat Shock Protein 70 Lacks Immune Modulatory Activity. Protein Pept Lett 24, 503-510, doi:10.2174/0929866524666170214141909.

Purcell, A. W., McCluskey, J., Rossjohn, J., 2007. More than one reason to rethink the use of peptides in vaccine design. Nat Rev Drug Discov 6, 404-14, doi:10.1038/nrd2224.

Read, D., Lensen, A. H., Begarnie, S., Haley, S., Raza, A., Carter, R., 1994. Transmission-blocking antibodies against multiple, non-variant target epitopes of the Plasmodium falciparum gamete surface antigen Pfs230 are all complement-fixing. Parasite Immunol 16, 511-9.

Santra, S., Liao, H. X., Zhang, R., Muldoon, M., Watson, S., Fischer, W., Theiler, J., Szinger, J., Balachandran, H., Buzby, A., Quinn, D., Parks, R. J., Tsao, C. Y., Carville, A., Mansfield, K. G., Pavlakis, G. N., Felber, B. K., Haynes, B. F., Korber, B. T., Letvin, N. L., 2010. Mosaic vaccines elicit CD8+ T lymphocyte responses that confer enhanced immune coverage of diverse HIV strains in monkeys. Nat Med 16, 324-8, doi:10.1038/nm.2108.

Schussek, S., Trieu, A., Apte, S. H., Sidney, J., Sette, A., Doolan, D. L., 2013. Immunization with apical membrane antigen 1 confers sterile infection-blocking immunity against Plasmodium sporozoite challenge in a rodent model. Infect Immun 81, 3586-99, doi:10.1128/IAI.00544-13.

Sette, A., Sidney, J., 1999. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. Immunogenetics 50, 201-12.

Sheikh, Q. M., Gatherer, D., Reche, P. A., Flower, D. R., 2016. Towards the knowledge-based design of universal influenza epitope ensemble vaccines. Bioinformatics 32, 3233-3239, doi:10.1093/bioinformatics/btw399.

Spinola, H., Couto, A. R., Peixoto, M. J., Anagnostou, P., Destro-Bisol, G., Spedini, G., Lopez-Larrea, C., Bruges-Armas, J., 2011. HLA class-I diversity in Cameroon: evidence for a north-south structure of genetic variation and relationships with African populations. Ann Hum Genet 75, 665-77, doi:10.1111/j.1469-1809.2011.00672.x.

Suwancharoen, C., Putaporntip, C., Rungruang, T., Jongwutiwes, S., 2011. Naturally acquired IgG antibodies against the C-terminal part of Plasmodium falciparum sporozoite threonine-asparagine-rich protein in a low endemic area. Parasitol Res 109, 315-20, doi:10.1007/s00436-011-2257-z.

Theiler, J., Yoon, H., Yusim, K., Picker, L. J., Fruh, K., Korber, B., 2016. Epigraph: A Vaccine Design Tool Applied to an HIV Therapeutic Vaccine and a Pan-Filovirus Vaccine. Sci Rep 6, 33987, doi:10.1038/srep33987.

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., Diallo, D. A., Diarra, I., Daou, M., Dolo, A., Tolo, Y., Sissoko, M. S., Niangaly, A., Sissoko, M., Takala-Harrison, S., Lyke, K. E., Wu, Y., Blackwelder, W. C., Godeaux, O., Vekemans, J., Dubois, M. C., Ballou, W. R., Cohen, J., Thompson, D., Dube, T., Soisson, L., Diggs, C. L., House, B., Lanar, D. E., Dutta, S., Heppner, D. G., Jr., Plowe, C. V., 2011. A field trial to assess a blood-stage malaria vaccine. N Engl J Med 365, 1004-13, doi:10.1056/NEJMoa1008115.

Tsuji, M., 2010. A retrospective evaluation of the role of T cells in the development of malaria vaccine. Exp Parasitol 126, 421-5, doi:10.1016/j.exppara.2009.11.009.

Vargas-Serrato, E., Barnwell, J. W., Ingravallo, P., Perler, F. B., Galinski, M. R., 2002. Merozoite surface protein-9 of Plasmodium vivax and related simian malaria parasites is orthologous to p101/ABRA of P. falciparum. Mol Biochem Parasitol 120, 41-52.

Vita, R., Overton, J. A., Greenbaum, J. A., Ponomarenko, J., Clark, J. D., Cantrell, J. R., Wheeler, D. K., Gabbard, J. L., Hix, D., Sette, A., Peters, B., 2015. The immune epitope database (IEDB) 3.0. Nucleic Acids Res 43, D405-12, doi:10.1093/nar/gku938.

Wahlgren, M., Bejarano, M. T., Troye-Blomberg, M., Perlmann, P., Riley, E., Greenwood, B. M., Patarroyo, M. E., Gonzales, C. I., Martinez, A., 1991. Epitopes of the Plasmodium falciparum clustered-asparagine-rich protein (CARP) recognized by human T-cells and antibodies. Parasite Immunol 13, 681-94.

Weiss, W. R., Jiang, C. G., 2012. Protective CD8+ T lymphocytes in primates immunized with malaria sporozoites. PLoS One 7, e31247, doi:10.1371/journal.pone.0031247.

Weng, H. B., Guo, X. H., Papoin, J., Wang, J., Coppel, R., Mohandas, N., An, X. L., 2014. Interaction of Plasmodium falciparum knob-associated histidine-rich protein (KAHRP) with erythrocyte ankyrin R is required for its attachment to the erythrocyte membrane. Biochimica Et Biophysica Acta-Biomembranes 1838, 185-192, doi:10.1016/j.bbamem.2013.09.014.

WHO, 2016. Malaria: Fact Sheet - http://www.who.int/mediacentre/factsheets/fs094/en/. Vol. 2017.

FIGURE 1.

CD8+ Epitope Frequency among *P. falciparum* Antigens

P. falciparum proteins containing 1 or more CD8+ epitopes, sorted by prevalence. **CSP:** Circumsporozoite protein (Hodgson et al., 2015), UniProt: P08307, CSP_PLAFW. **MSP:** Merozoite surface protein (Ahouidi et al., 2016), UniProt: P04933, MSP1_PLAFW. **SSP2/TRAP:** Sporozoite surface protein 2/Thrombospondin-related anonymous protein (John et al., 2003), UniProt: Q9GTW7, Q9GTW7_PLAFA. **AMA-1**: Apical membrane antigen 1 (Schussek et al., 2013), UniProt: Q94661, Q94661_PLAFA. **Pf332*:*** *P. falciparum* antigen 332, DBL-like protein (Nilsson et al., 2011), UniProt: Q7JPX6, Q7JPX6_PLAFA. **EXP-1**: Exported protein-1/Circumsporozoite-related antigen (John et al., 2003), UniProt: Q548U2, Q548U2_PLAFA. **SERA**: Serine-repeat antigen (Owalla et al., 2013) UniProt: Q9TY95, SERA_PLAF7. **Pfg27/25:** *P. falciparum* gamete antigen 27/25 (Read et al., 1994), UniProt: Q8IEU2, Q8IEU2_PLAF7. **P101:** 101 kDa malaria antigen (Vargas-Serrato et al., 2002), UniProt: Q8I5D2, ABRA_PLAF7. **PfHsp70:** *P. falciparum* Heat shock 70 kDa protein (Pooe et al., 2017), UniProt: P11144, HSP70_PLAFA. **LSA:** Liver stage antigen (Ferraro et al., 2013), UniProt: Q25893, Q25893_PLAFA. **EBA175:** Erythrocyte binding antigen175 kDa (Okenu et al., 2000), UniProt: Q95VT1, Q95VT1_PLAFA. **CARP**: Clustered-asparagine-rich protein (Wahlgren et al., 1991), UniProt: P13824, ARP2_PLAFA. **KAHRP:** Knob-associated histidine-rich protein (Weng et al., 2014), UniProt: P13817, KNOB_PLAFA. **41 kDa antigen:** Fructose-bisphosphate aldolase (Kim et al., 1998), UniProt: P14223, ALF_PLAFA. **RAP:** Rhoptry-associated protein (Ghosh et al., 2017), UniProt: Q25730, Q25730_PLAFA. **STRAP:** Sporozoite threonine-asparagine-rich protein (Suwancharoen et al., 2011), UniProt: B0Z8E6, B0Z8E6_PLAFA.

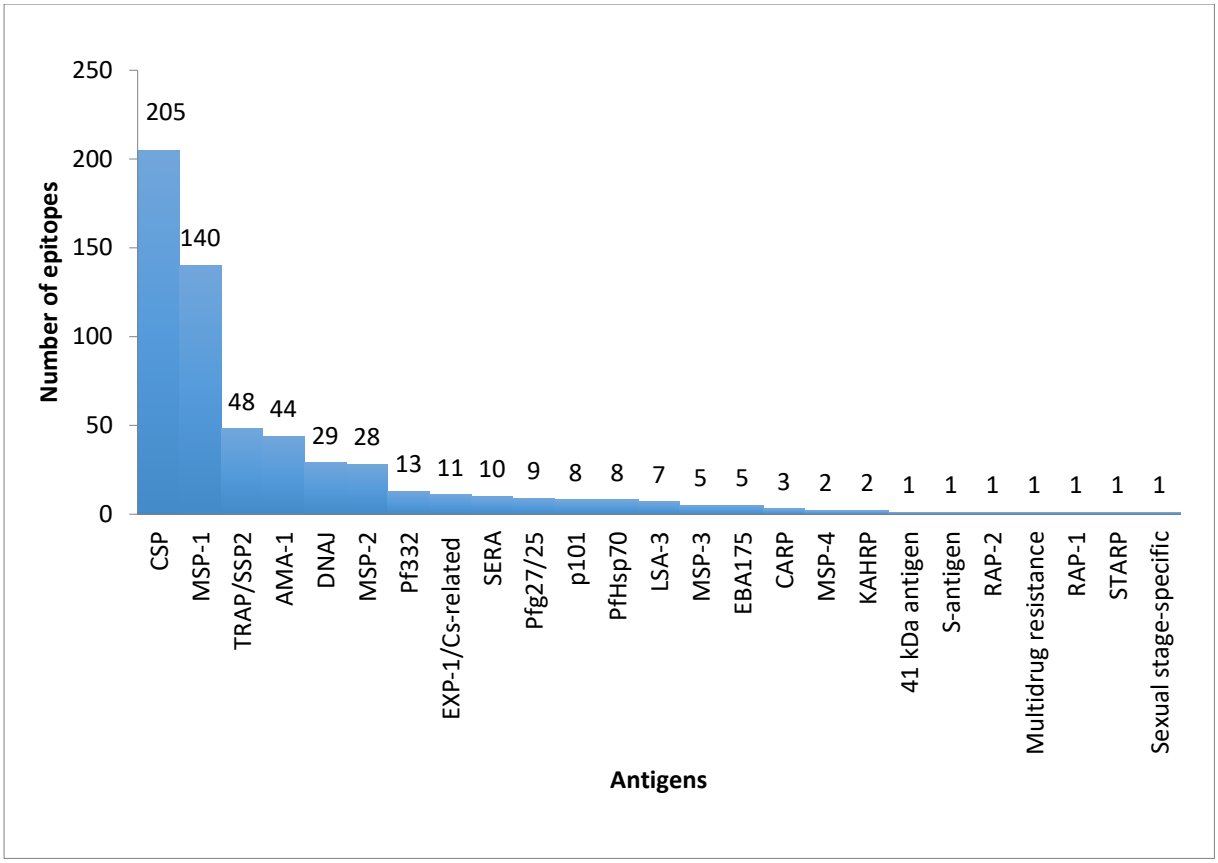
Number of epitopes
250
200
150
100
50
0
205
140
48
44
29
28
13
11
10
9
8
8
7
5
5
3
2
2
1
1
1
1
1
1
1
CSP
MSP-1
TRAP/SSP2
AMA-1
DNAJ
MSP-2
Pf332
EXP-1/Cs-related
SERA
Pfg27/25
p101
PfHsp70
LSA-3
MSP-3
EBA175
CARP
MSP-4
KAHRP
41 kDa antigen
S-antigen
RAP-2
Multidrug resistance
RAP-1
STARP
Sexual stage-specific
Antigens

FIGURE 2.

CD8 T cell Epitope selection overview.

The figures compares number of HLA I epitopes that were obtained from the primary results (in Red boxes) and epitopes after clustering protein sequences with the CD-Hit tool (in Blue boxes). (A) The number of conserved HLA I epitopes from all analysed sets. (B) The number of epitopes that show binding affinity toward HLA class I molecules.

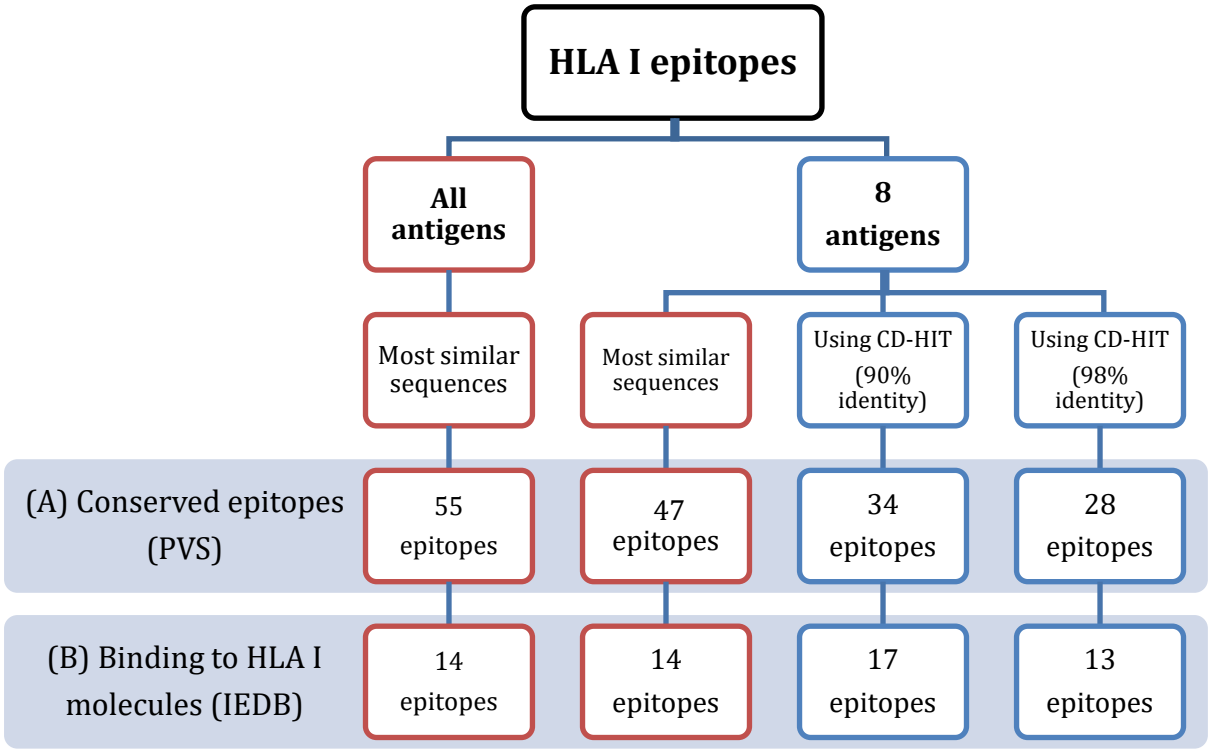

FIGURE 3.

CD4 T cell Epitope selection overview

This figure compares the number of HLA II epitopes obtained from the primary results (Red boxes) and epitopes after clustering protein sequences with the CD-Hit tool (Blue boxes). (A) The number of conserved HLA II epitopes from all analysed sets. (B) The number of epitopes that show affinity toward HLA II molecules.

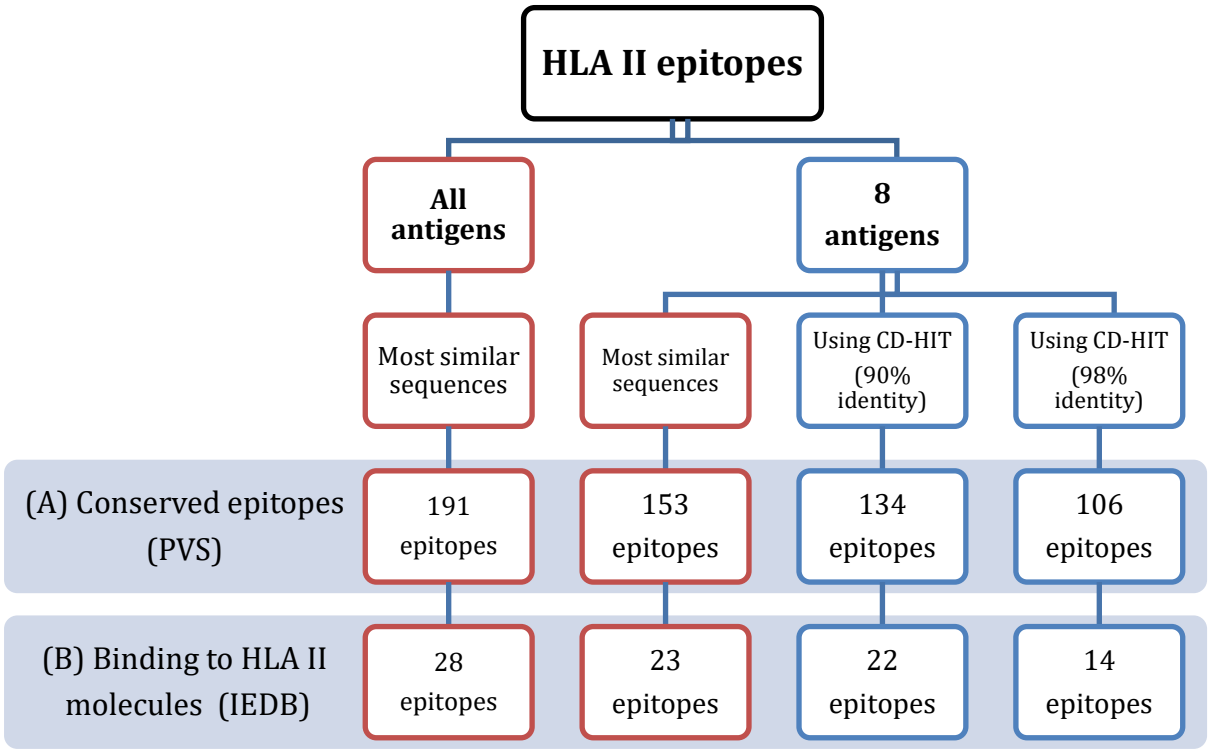

TABLE 1.

Conserved CD8+ T cell epitopes from *P. falciparum*

	HLA class I epitopes	Antigen	HLA I binding profile	% PPC	
				World	Africa*
1	LLMDCSGSI	TRAP	A*02:01, A*02:03, B*15:01	44.92	10.71
2	FLIFFDLFLV	TRAP	A*02:01, A*02:03, A*02:06	41.35	10.05
3	ALFFIIFNK	EXP-1	A*03:01, A*11:01, A*31:01, A*68:01	40.03	9.85
4	LLACAGLAYK	TRAP	A*03:01, A*11:01	30.92	5.91
5	TPYAGEPAPF	TRAP	B*07:02, B*35:01, B*51:01, B*53:01	29.39	24.00
6	LLACAGLAY	TRAP	A*01:01, B*15:01,	24.32	4.96
7	FEFTYMINF	AMA-1	B*40:01, B*44:02, B*44:03	20.88	5.38
8	LACAGLAYK	TRAP	A*11:01	15.53	1.18
9	NEVVVKEEY	AMA-1	B*44:02, B*44:03	13.63	4.85
10	KSHGKGYNW	AMA-1	A*32:01, B*57:01, B*58:01,	11.53	7.70
11	KNKEKALI	TRAP	B*08:01	10.55	4.64
12	SLKKNSRSL	CSP	B*08:01	10.55	4.64
13	YKSHGKGYNW	AMA-1	B*58:01-	3.42	5.18
14	GLIMVLSFL	CSP	A*02:03	0.97	0.08

Projected population coverage (PPC) values of conserved class I epitopes. PPC quantifies the percentage of a given population group who are estimated to produce some immune response to that antigen. The 14 epitopes shown were derived from both the 25 and 8 antigen sets. The eight epitopes also present in the CD-HIT clustering are shown in bold.

* Average coverage for East, West, South, and Central Africa.

TABLE 2.

Conserved CD4 T cell epitopes from *P. falciparum*

	HLA II epitopes	Antigen	HLA II binding profile	% PPC world	% PPC Africa*
1	MRKLAILSVSSFLFV	CSP	DRB1*01:01, DRB1*03:01 DRB1*04:01, DRB1*04:05 DRB1*07:01, DRB1*08:02 DRB1*12:01, DRB1*13:02 DRB1*11:01, DRB1*15:01 DRB4*01:01, DRB5*01:01 DPA1*01:03/DPB1*02:01 DPA1*01/DPB1*04:01 DPA1*02:01/DPB1*01:01 DPA1*02:01/DPB1*05:01 DPA1*02:01/DPB1*14:01 DPA1*03:01/DPB1*04:02	78.93	31.75
2	SSVFNVVNSSIGLIM	CSP	DRB1*03:01, DRB1*04:01 DRB1*04:05, DRB1*07:01 DRB1*08:02, DRB1*09:01 DRB1*11:01, DRB1*13:02 DRB1*15:01, DRB3*02:02 DRB5*01:01	74.13	30.27
3	SLRWIFKHVAKTHLK	DNAJ	DRB1*03:01, DRB1*04:01 DRB1*04:05, DRB1*07:01 DRB1*08:02, DRB1*09:01 DRB1*11:01, DRB1*12:01 DRB1*15:01, DRB3*02:02 DRB4*01:01, DRB5*01:01	72.96	27.90
4	NFFIFVTFNIKNESK	MSP-2	DRB1*03:01, DRB1*04:01 DRB1*04:05, DRB1*07:01 DRB1*08:02, DRB1*11:01 DRB1*12:01, DRB1*15:01 DRB3*01:01, DRB5*01:01 DPA1*01/DPB1*04:01 DPA1*01:03/DPB1*02:01 DPA1*02:01/DPB1*01:01 DPA1*03:01/DPB1*04:02 DPA1*02:01/DPB1*05:01 DPA1*02:01/DPB1*14:01	69.47	26.00
5	KSKYKLATSVLAGLL	EXP-1	DRB1*01:01, DRB1*04:01 DRB1*04:05, DRB1*07:01 DRB1*08:02, DRB1*09:01 DRB1*11:01, DRB1*15:01 DRB3*02:02, DRB5*01:01 DPA1*01/DPB1*04:01 DPA1*02:01/DPB1*01:01 DPA1*02:01/DPB1*05:01	66.73	20.25

			DPA1*02:01/DPB1*14:01 DPA1*03:01/DPB1*04:02 DQA1*01:02/DQB1*06:02		
6	GLAYKFVVPGAATPY	TRAP	DRB1*01:01, DRB1*03:01 DRB1*07:01, DRB1*08:02 DRB1*09:01, DRB1*11:01 DRB1*12:01, DRB5*01:01 DQA1*05:01/DQB1*03:01	60.28	27.05
7	LLSAFEFTYMINFGR	AMA-1	DRB1*04:01, DRB1*04:05 DRB1*07:01, DRB1*11:01 DRB1*15:01, DRB3*01:01 DRB5*01:01 DPA1*01:03/DPB1*02:01 DPA1*01/DPB1*04:01 DPA1*02:01/DPB1*01:01 DPA1*02:01/DPB1*05:01 DPA1*02:01/DPB1*14:01 DPA1*03:01/DPB1*04:02 DQA1*01:01/DQB1*05:01	53.70	16.83
8	AGLLGVVSTVLLGGV	EXP-1	DRB1*03:01, DRB1*04:01 DRB1*04:05, DRB1*11:01 DRB1*15:01, DRB5*01:01	53.40	17.45
9	KMEDYIKKNKTTIAN	MSP-1	DRB1*08:02, DRB1*09:01 DRB1*13:02, DRB1*04:01 DRB1*03:01, DRB1*11:01 DRB4*01:01	48.70	23.39
10	LKKLSSIMERYAGGK	DNAJ	DRB1*04:05, DRB1*08:02 DRB1*11:01, DRB1*04:01 DRB1*15:01, DRB1*09:01 DRB5*01:01	46.43	12.68
11	YKAYVSYKKRKAQEK	P101	DRB1*04:01, DRB1*08:02, DRB1*09:01, DRB1*11:01, DRB1*15:01, DRB3*02:02 DRB5*01:01,	44.18	11.29
12	GDIIKKMQTLWDEIM	DNAJ	DRB1*03:01, DRB1*04:01 DRB1*04:05, DRB1*08:02 DRB1*11:01, DRB3*01:01 DRB4*01:01 DQA1*01:01/DQB1*05:01	41.08	17.03
13	LTGYSLFQKEKMVLN	MSP-1	DRB1*03:01, DRB1*04:01 DRB1*08:02, DRB1*11:01 DRB5*01:01 DPA1*01/DPB1*04:01 DPA1*01:03/DPB1*02:01 DPA1*02:01/DPB1*01:01 DPA1*02:01/DPB1*05:01 DPA1*02:01/DPB1*14:01	38.72	15.64
14	DFNHYYTLKTGLEAD	MSP-1	DRB1*04:01, DRB1*04:05 DRB1*09:01, DRB1*11:01 DRB5*01:01	29.38	11.99

15	IRANELDVLKKLVFG	MSP-1	DRB1*03:01, DRB1*11:01 DPA1*03:01/DPB1*04:02 DQA1*01:01/DQB1*05:01	27.37	15.09
16	HYYTLKTGLEADIKK	MSP-1	DRB1*04:05, DRB1*08:02 DRB1*09:01, DRB1*11:01	21.43	11.64
17	KYKIAGGIAGGLALL	TRAP	DRB1*09:01, DRB1*11:01 DRB5*01:01 DQA1*05:01/DQB1*03:01	16.59	10.11
18	GSSPMEFLQIIEDYG	SERA	DRB1*04:05, DRB4*01:01, DRB1*11:01 DQA1*03:01/DQB1*03:02, DQA1*04:01/DQB1*04:02, DQA1*05:01/DQB1*02:01,	13.39	9.21
19	KFSSSNNSVYNVQKL	MSP-1	DRB1*04:01	11.21	0.54
20	IQNSLSTEWSPCSVT	CSP	DRB1*04:01	11.21	0.54
21	DIEKKICKMEKCSSV	CSP	DRB1*11:01	10.54	7.70
22	SNYPYNYVKVGEQCP	SERA	DRB1*11:01, DRB5*01:01	10.54	7.70
23	VKNVIGPFMKAVCVE	TRAP	DRB1*08:02	2.33	0.12
24	NNFMNRNMKNKNMNN	CARP	DRB5*01:01, DRB1*08:02	2.33	0.12
25	KLQKTYSQYKVQYDM	DNAJ	DRB1*04:05, DRB5*01:01	3.02	1.68
26	EEHVEEPASDVQQTS	DNAJ	DQA1*04:01/DQB1*04:02	0.00	0.00
27	MQTLWDEIMDINKRK	DNAJ	DRB3*01:01, DRB5*01:01, DQA1*01:01/DQB1*05:01	0.00	0.00
28	WMNLWDNGKILHNKN	SERA	DRB3*01:01	0.00	0.00

Projected population coverage (PPC) values of conserved HLA II-restricted epitopes. PPC quantifies the percentage of a given population group who are estimated to produce some immune response to that antigen. All 28 epitopes were derived from the set contained 25 antigens. 23 epitopes were derived from the set of 8 selected antigens; epitopes 11, 18, 22, 24 and 28 were derived from proteins other than the eight selected antigens. Epitopes shown in bold are also common to the CD-HIT clustered data. * Average coverage for East, West, South, and Central Africa.

TABLE 3.

Constituents of proposed minimal malaria Epitope Ensemble vaccines

Epitope Combination	% PPC				
	World	East Africa	West Africa	Central Africa	South Africa
LLMDCSGSI, ALFFIIFNK TPYAGEPAPF, MRKLAILSVSSFLFV	95.95	83.93	86.56	82.99	62.64
FLIFFDLFLV, LLACAGLAY, LLACAGLAYK, MRKLAILSVSSFLFV KSKYKLATSVLAGLL	95.73	80.63	78.95	73.93	55.71
FLIFFDLFLV, LLACAGLAY LLACAGLAYK, MRKLAILSVSSFLFV SSVFNVVNSSIGLIM	95.73	80.63	78.95	73.93	55.71
FLIFFDLFLV, LLACAGLAY, LLACAGLAYK, GLAYKFVVPGAATPY SSVFNVVNSSIGLIM	95.73	80.63	78.95%	73.93	55.71
FLIFFDLFLV, LLACAGLAY, LLACAGLAYK, SSVFNVVNSSIGLIM KSKYKLATSVLAGLL	95.27	76.88	78.06	70.68	55.71
LLMDCSGSI, LLACAGLAYK TPYAGEPAPF, MRKLAILSVSSFLFV	95.07	82.98	84.74	82.04	56.66
FLIFFDLFLV, LLACAGLAYK LLACAGLAY, MRKLAILSVSSFLFV	95.6	79.58	75.20	73.03	54.74

FLIFFDLFLV, LLACAGLAY LLACAGLAYK, KSKYKLATSVLAGLL GLAYKFVVPGAATPY	95.02	75.82	74.16	65.08	52.43

Composition of peptide sets corresponding to proposed epitope ensemble vaccines. A summary of the combination of epitopes (HLA I and HLA II epitopes) that were derived from the 8 selected antigens, able to achieve over 95% world population coverage.

TABLE 4.

Comparison of the binding profile and calculated PPC of heteroclitic peptides versus those of the original progenitor epitope.

Core peptide	HLA-profile	% PPC	Heteroclitic peptide	HLA binding profile	% PPC
LLMDCSGSI	A*02:01, A*02:03, B*15:01	44.92	**M**LMDCSGSI	A*02:01, A*02:03, A*02:06, B*15:01,	46.30
FLIFFDLFLV	A*02:01, A*02:03, A*02:06	41.3	F**A**IFFDLFLV	A*02:01, A*02:06, A*68:02, B*51:01	46.80
LLACAGLAYK	A*03:01, A*11:01	30.92	**M**LACAGLAYK	A*03:01, A*68:01, A*11:01	35.75
TPYAGEPAPF	B*07:02, B*35:01, B*51:01, B*53:01	29.39	T**F**YAGEPAPF	A*23:01, A*24:02	26.18
LLACAGLAY	A*01:01, B*15:01,	24.32	**R**LACAGLAY	B*15:01, A*03:01, A*30:02	25.87
SLKKNSRSL	B*08:01	10.55	**R**LKKNSRSL	B*08:01, B*07:02	22.61
MRKLAILSVSS FLFV	DRB1*01:01, DRB1*03:01, DRB1*04:01, DRB1*04:05, DRB1*07:01, DRB1*08:02, DRB1*12:01, DRB1*13:02, DRB1*11:01, DRB1*15:01, DRB4*01:01, DRB5*01:01, DPA1*01:03/DPB1*02:01, DPA1*01/DPB1*04:01,	78.93	MRKLAILS**A**SS FLFV	DRB1*04:01, DRB1*15:01, DRB1*01:01, DRB1*03:01, DRB1*12:01, DRB4*01:01, DRB1*04:05, DRB5*01:01, DRB1*07:01, DRB3*02:02, DRB1*09:01, DRB1*13:02, DRB3*01:01, DPA1*01:03/DPB1*02:01, DPA1*01/DPB1	81.81

	DPA1*02:01/DPB1*01:01, DPA1*02:01/DPB1*05:01, DPA1*02:01/DPB1*14:01, DPA1*03:01/DPB1*04:02			*04:01, DRB1*11:01, DPA1*02:01/DPB1*01:01, DRB1*08:02, DPA1*02:01/DPB1*14:01, DQA1*01:02/DQB1*06:02, DPA1*02:01/DPB1*05:01, DPA1*03:01/DPB1*04:02	
KSKYKLATSVLAGLL	DRB1*01:01, DRB1*04:01, DRB1*04:05, DRB1*07:01, DRB1*08:02, DRB1*09:01, DRB1*11:01, DRB1*15:01, DRB3*02:02, DRB5*01:01, DPA1*01/DPB1*04:01, DPA1*02:01/DPB1*01:01, DPA1*02:01/DPB1*05:01, DPA1*02:01/DPB1*14:01, DPA1*03:01/DPB1*04:02, DQA1*01:02/DQB1*06:02	66.73	K**I**KYKLATSVLAGLL	DRB1*09:01, DRB3*02:02, DRB1*01:01, DRB1*04:01, DPA1*02:01/DPB1*14:01, DRB1*08:02, DRB1*11:01, DRB1*04:05, DRB1*07:01, DPA1*03:01/DPB1*04:02, DQA1*01:02/DQB1*06:02, DPA1*01/DPB1*04:01, DRB5*01:01, DPA1*02:01/DPB1*01:01, DRB1*12:01, DRB1*15:01, DRB4*01:01, DRB1*13:02, DPA1*02:01/DPB1*05:01	72.95
GLAYKFVVPGAATPY	DRB1*01:01, DRB1*03:01, DRB1*07:01, DRB1*08:02, DRB1*09:01, DRB1*11:01, DRB1*12:01, DRB5*01:01, DQA1*05:01/DQB1*03:01	60.28	GL**R**YKFVVPGAATPY	DRB1*09:01, DRB5*01:01, DRB1*01:01, DRB1*11:01, DQA1*05:01/DQB1*03:01, DRB1*08:02, DRB1*12:01, DRB1*04:01, DRB1*07:01, DRB1*03:01	67.23

Comparison of native epitopes to heteroclitic variants, illustrating their PPC values. Substitution site are highlighted and underlined. Heteroclitic peptides have larger PPC values compared with core epitopes except TPYAGEPAP epitope. All possible mutations within NEVVVKEEY epitope show neither higher PPC value nor distinct HLA binding profile.

TABLE 5

Final proposed malaria epitope ensemble vaccine including one or several heteroclitic peptides.

Epitope combination	World %PPC	East Africa %PPC	West Africa %PPC	Central Africa %PPC	South Africa %PPC
LLMDCSGSI, FLIFFDLFLV, LLACAGLAYK, TPYAGEPAPF, LLACAGLAY, NEVVVKEEY, SLKKNSRSL, MRKLAILSVSSFLFV, KSKYKLATSVLAGLL, GLAYKFVVPGAATPY	97.89	88.15	90.23	86.31	73.83
LLMDCSGSI, FLIFFDLFLV, LLACAGLAYK, TPYAGEPAPF, LLACAGLAY, NEVVVKEEY, SLKKNSRSL, MRKLAILSVSSFLFV, KSKYKLATSVLAGLL, GLAYKFVVPGAATPY, **TFYAGEPAPF**	98.90	90.69	94.06	89.51	81.94
LLMDCSGSI, FLIFFDLFLV, LLACAGLAYK, TPYAGEPAPF, LLACAGLAY, NEVVVKEEY, SLKKNSRSL, MRKLAILSVSSFLFV, KSKYKLATSVLAGLL, GLAYKFVVPGAATPY, **MLACAGLAYK**	98.13	88.55	91.26	86.86	76.35
LLMDCSGSI, FLIFFDLFLV, LLACAGLAYK, TPYAGEPAPF, LLACAGLAY, NEVVVKEEY, SLKKNSRSL, MRKLAILSVSSFLFV, KSKYKLATSVLAGLL, GLAYKFVVPGAATPY, **FAIFFDLFLV**	98.00	90.39	90.92	87.68	78.81
LLMDCSGSI, FLIFFDLFLV, LLACAGLAYK, TPYAGEPAPF, LLACAGLAY, NEVVVKEEY, SLKKNSRSL, MRKLAILSVSSFLFV, KSKYKLATSVLAGLL, GLAYKFVVPGAATPY, **RLACAGLAY**	97.99	90.24	91.86	88.07	78.31

LLMDCSGSI, FLIFFDLFLV, LLACAGLAYK, TPYAGEPAPF, LLACAGLAY, NEVVVKEEY, SLKKNSRSL, MRKLAILSVSSFLFV, KSKYKLATSVLAGLL, GLAYKFVVPGAATPY, **MLMDCSGSI**	97.89	88.15	90.23	86.31	73.83
LLMDCSGSI, FLIFFDLFLV, LLACAGLAYK, TPYAGEPAPF, LLACAGLAY, NEVVVKEEY, SLKKNSRSL, MRKLAILSVSSFLFV, KSKYKLATSVLAGLL, GLAYKFVVPGAATPY, **RLKKNSRSL**	97.89	88.15	90.23	86.31	73.83
LLMDCSGSI, FLIFFDLFLV, LLACAGLAYK, TPYAGEPAPF, LLACAGLAY, NEVVVKEEY, SLKKNSRSL, MRKLAILSVSSFLFV, KSKYKLATSVLAGLL, GLAYKFVVPGAATPY, **MRKLAILSASSFLFV**	97.89	88.15	90.23	86.31	73.83
LLMDCSGSI, FLIFFDLFLV, LLACAGLAYK, TPYAGEPAPF, LLACAGLAY, NEVVVKEEY, SLKKNSRSL, MRKLAILSVSSFLFV, KSKYKLATSVLAGLL, GLAYKFVVPGAATPY, **KIKYKLATSVLAGLL**	97.89	88.15	90.23	86.31	73.83
LLMDCSGSI, FLIFFDLFLV, LLACAGLAYK, TPYAGEPAPF, LLACAGLAY, NEVVVKEEY, SLKKNSRSL, MRKLAILSVSSFLFV, KSKYKLATSVLAGLL, GLAYKFVVPGAATPY, **GLRYKFVVPGAATPY**	97.89	88.15	90.23	86.31	73.83
LLMDCSGSI, FLIFFDLFLV, LLACAGLAYK, TPYAGEPAPF, LLACAGLAY, NEVVVKEEY, SLKKNSRSL, MRKLAILSVSSFLFV, KSKYKLATSVLAGLL, GLAYKFVVPGAATPY, **FAIFFDLFLV, TFYAGEPAPF**	98.98	92.66	94.60	90.71	86.03
LLMDCSGSI, FLIFFDLFLV, LLACAGLAYK, TPYAGEPAPF, LLACAGLAY, NEVVVKEEY, SLKKNSRSL, MRKLAILSVSSFLFV,	99.05	94.28	95.79	92.15	89.25

KSKYKLATSVLAGLL, GLAYKFVVPGAATPY, **FAIFFDLFLV, TFYAGEPAPF, RLACAGLAY**					
LLMDCSGSI, FLIFFDLFLV, LLACAGLAYK, TPYAGEPAPF, LLACAGLAY, NEVVVKEEY, SLKKNSRSL, MRKLAILSVSSFLFV, KSKYKLATSVLAGLL, GLAYKFVVPGAATPY, **FAIFFDLFLV, MLACAGLAYK, TFYAGEPAPF, RLACAGLAY**	99.21	94.55	96.46	92.56	90.85

The combination of suggested vaccine candidates with single or multiple heteroclitic peptides (underlined in bold) as identified in Table 4, including the original optimal epitope ensemble for comparison. A minimum of four heteroclitic peptides are required to achieve >90% for all African regions.