BIOINFORMATION
Discovery at the interface of physical and biological sciences

BIOMEDICAL INFORMATICS

Open access

www.bioinformation.net
Volume 12(3)

Hypothesis

Implications from predicted B-cell and T-cell epitopes of *Plasmodium falciparum* merozoite proteins EBA175-RII and Rh5

Kevin Kariuki Wamae[1,2], & Lynette Isabella Ochola-Oyier*[1,2]

[1]Centre for Biotechnology and Bioinformatics, University of Nairobi, Kenya; [2]KEMRI-Wellcome Trust Collaborative Programme, Kilifi, Kenya; P.O. Box 230, Kilifi – 80108, Kenya; Email: kwamae@kemri-wellcome.org; Fax:+ 254 20 2711673; *Corresponding author

Received February 05, 2016; Revised March 21, 2016; Accepted March 25, 2016; Published June 15, 2016

Abstract:
The leading circumsporozoite protein (CSP) based malaria vaccine, RTS,S, though promising, has shown limited efficacy in field studies. There is therefore, still a need to identify other malaria vaccine targets. Merozoite antigens are potential vaccine candidates, since naturally acquired antibodies generated against them inhibit erythrocyte invasion and in some cases result in the clinical protection from disease. We thus used in silico tools (BCPreds, NetMHCcons and NetMHCIIpan 3.0) to predict B-cell epitopes (BCEs) and T-cell epitopes (TCEs) in two merozoite invasion proteins, EBA175-RII and Rh5. Initially, we validated these tools using CSP to determine whether the algorithms could predict the epitopes in the RTS,S vaccine. In EBA175-RII, we prioritised three BCEs ^{15}REKRKGMKWDCKKKNDRSNY^{34}, ^{420}SNRKLVGKINTNSNYVHRNKQ^{440} and ^{528}WISKKKEEYNKQAKQYQEYQ^{547}, a CD8+ epitope ^{553}KMYSEFKSI^{561} and a CD4+ epitope ^{440}QNDKLFRDEWWK VIKKD^{456}. Three Rh5 epitopes were prioritised, a BCE ^{344}SCYNNNFCNTNGIRYHYDEY^{363}, a CD8+ epitope ^{198}STYGKCIAV^{206} and a Rh5 CD4+ epitope 180TFLDYYKHLSYNSIYHKSSTY^{200}. All these epitopes are in the region involved in the proteins' interaction with their erythrocyte receptors, thus enabling erythrocyte invasion. Therefore, upon validation of their immunogenicity, by ELISA using serum from a malaria endemic population, antibodies to these epitopes may inhibit erythrocyte invasion. All the epitopes we predicted in EBA175-RII and Rh5 are novel. We also identified polymorphic epitopes that may escape host immunity, as some variants were not predicted as epitopes, suggesting that they may not be immunogenic regions. We present a set of epitopes that following *in vitro* validation provide a set of molecules to screen as potential vaccine candidates.

Keywords: Malaria, polymorphism, epitope, vaccine

Background:
Malaria is caused by the unicellular protozoan parasite, *Plasmodium falciparum*, that remains an important public health concern due to the high rates of mortality and morbidity in children under 5 years of age **[1].** The resistance of the parasite to the current first line antimalarial drug, artemisinin, in South East Asia **[2–4]** and mosquito resistance to pyrethroids **[5–8]**, highlight that malaria control is yet to be achieved. Additionally, the malaria vaccine candidate (MosquirixTM), RTS,S, based on the circumsporozoite protein (CSP), was approved for use by European regulators in July 2015, however it has shown limited success and waning efficacy over time **[9].** Hence, there is a need to identify novel vaccine targets. Previous studies have shown that in silico tools can identify B-cell epitopes (BCE) and T-cell epitopes (TCE) **[10, 11],** making this approach a quicker way to prioritise potential immunogenic targets for in vitro validation. A prime target for the design of a malaria vaccine is the invasive blood-stage form of the parasite, the merozoite, which invades red blood cells (RBCs) initiating the blood stage infection and the clinical symptoms of disease **[12].**

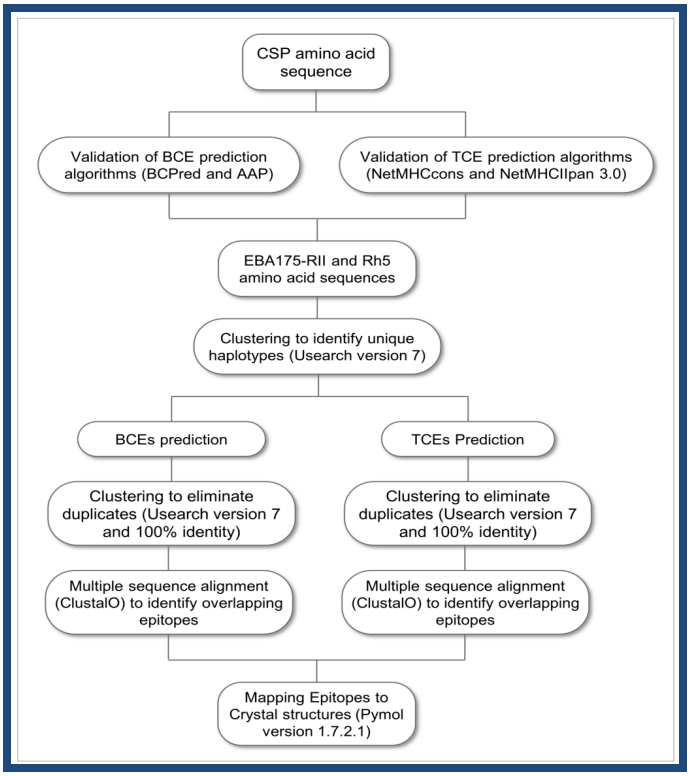

Figure 1: Flowchart showing the epitope prediction pipeline starting with a validation of the algorithm using CSP as a control, followed by BCE and TCE predictions in EBA175-RII and Rh5.

The mechanism by which the merozoite selects and successfully invades a RBC is complex, involving various receptor-ligand interactions **[13].** The Duffy binding ligands (DBLs) and reticulocyte binding-like homologues (Rhs) located in the micronemes and rhoptries, respectively, are two main families of proteins thought to play key roles in the invasion process **[14].** DBL molecules are thought to be essential in the formation of the tight junction, which precedes entry into the RBCs **[15].** The first merozoite ligand identified to bind to RBCs was erythrocyte binding antigen-175 (EBA175) **[16].** EBA175 interacts with glycophorin A (GypA) on the RBC surface via its erythrocyte binding domain (EBD) or region II (RII). EBA175-RII is a target for invasion inhibitory antibodies **[17–21]** and the EBA175-GypA interaction is a major RBC invasion pathway **[19]. It** has also become a leading malaria vaccine candidate **[22, 23],** thus immunogenic epitopes within EBA175-RII should be exploited as potential vaccine candidates.

The Rh family includes Rh1, 2a, 2b, 3, 4 and 5, only the latter two have defined RBC receptors, complement receptor 1 **[24]** and basigin **[25],** respectively. Rh5 has recently become a leading malaria vaccine candidate **[26],** due to evidence from previous studies, which have shown it has a limited number of single nucleotide polymorphisms (SNPs), only five non-synonymous SNPs **[27].** There has been no demonstration of Rh5 allele-specific immunity **[28],** additionally Rh5 antibodies were shown to inhibit RBC invasion **[29–31]** and have been associated with protection against malaria **[32]**. The crystal structures for both EBA175-RII (PDB code 1ZRL) **[33]** and Rh5 (PDB code 4U0Q) **[34]** have been published and the residues involved in binding to their respective RBC receptors identified. This therefore makes these two proteins, ideal candidates for the *in silico* discovery of vaccine targets.

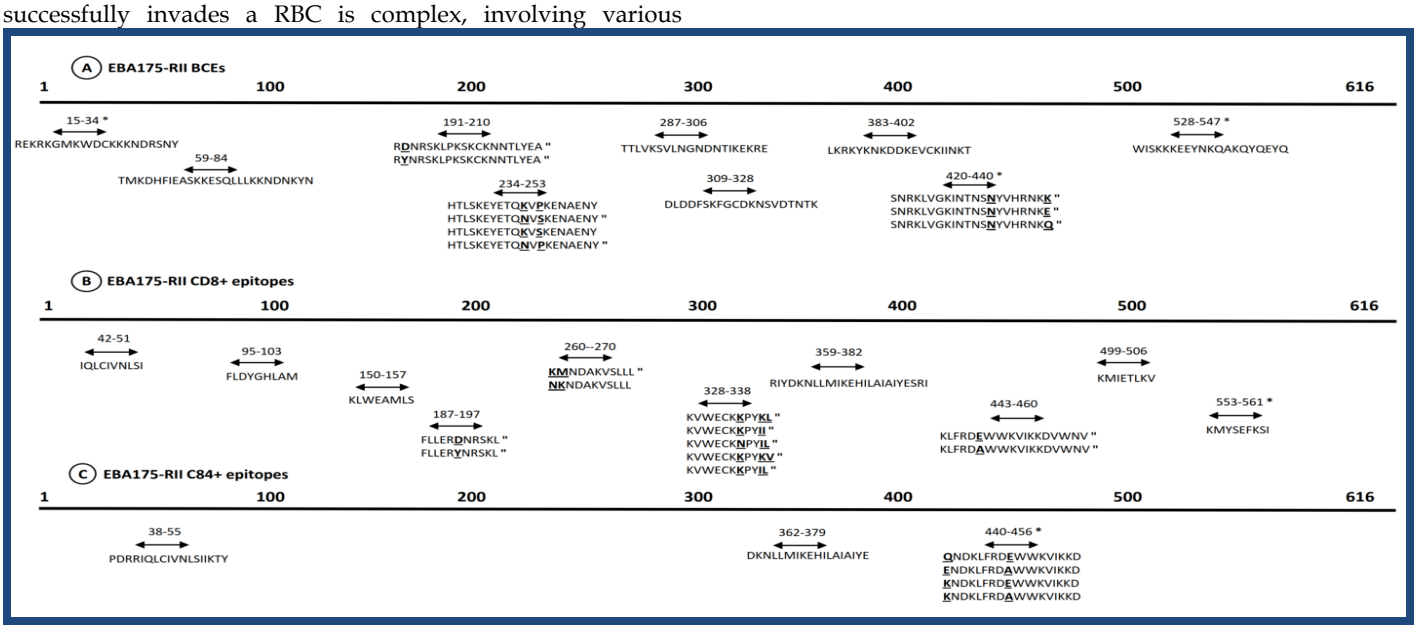

Figure 2: A schematic view of the EBA175-RII predicted epitopes mapped to the full EBA175-RII sequence, not drawn to scale. (A) EBA175-RII predicted BCEs. (B) EBA175-RII predicted CD8+ epitopes. (C) EBA175-RII predicted CD4+ epitopes. The numbers displayed above each epitope and separated by hyphens represent the amino acid regions that each epitope encompasses. The residues in bold and underlined represent polymorphic sites within the respective epitopes. Of the polymorphic epitopes, those marked with " represents the epitopes that were predicted. The epitope positions marked with * represent epitopes that overlap with residues involved in binding to GypA.

ISSN 0973-2063 (online) 0973-8894 (print)

The aim of this study was to predict BCEs and TCEs in EBA175-RII and Rh5 and to map them back to their crystal structures to determine their location in the tertiary protein. We validated the prediction tools using immunogenic, in vitro verified circumsporozoite protein (CSP) epitopes that included: a central NANP (NANP3 or NANPNANPNANP) repeat region that represents the BCE **[35],** three T-helper epitopes, CS.T3 ^{363}DIEKKICKMEKCSSV^{377}, Th2R ^{311}PSDKHIKEYLNKIQNSL^{327} and Th3R ^{341}GIQVRIKPGSANKPKDELDYANDI^{364} [36] as well as three cytotoxic TCEs located in the immuno-dominant C-terminal region of CSP including ^{336}VTCGNGIQVR^{345}, ^{386}GLIMVLSFL^{394} and ^{353}KPKDELDYANDIEKKICKMEKCS^{375} **[37,38].** All the above listed epitopes are components of the RTS,S vaccine.

Methodology

Protein Sequence Data Sets

The *P. falciparum* 3D7 laboratory isolate protein sequences were obtained for Rh5 (PF3D7_0424100), EBA175-RII (PF3D7_0731500) and CSP (PF3D7_0304600) from PlasmoDB (http://PlasmoDB .org). 19 EBA175-RII sequences and 52 Rh5 sequences (750bp region after the intron) were obtained by capillary sequencing of field isolates from Kilifi County, Kenya between 2007 and 2009 (Ochola-Oyier *et al.* 2016; GenBank accession numbers EBA175 KU526236-KU526265 and Rh5 KU525880-KU525986). The amino acid sequences were clustered at a 100% identity using Usearch version 7.0.1001 **[39],** to obtain unique haplotypes **(Figure 1**).

Validation of B-cell Epitope Prediction Algorithms

The selection of servers, epitope length and antigenic score cut offs were based on previous studies. We used the BCPREDS server (http://ailab.ist.psu.edu/bcpreds/index.html) for the prediction of BCEs **[40,41]** and two algorithms were selected, AAP and BCPred **Figure 1**. The CSP sequence was submitted to the server with the default parameters and BCE lengths of 20mers **[42].** Predicted BCEs with an antigenic score of >0.8 were selected **[40,41]** and included CSP BCEs identified from both algorithms after clustering them at 100% identity to exclude duplicates. The final predicted epitopes were then clustered at 50% identity with the *in vitro* verified $NANP_3$ BCE to determine epitopes that were similar. This criterion lowers the stringency and identifies a larger number of epitopes, taking into account any limitations in the tools to predict epitopes.

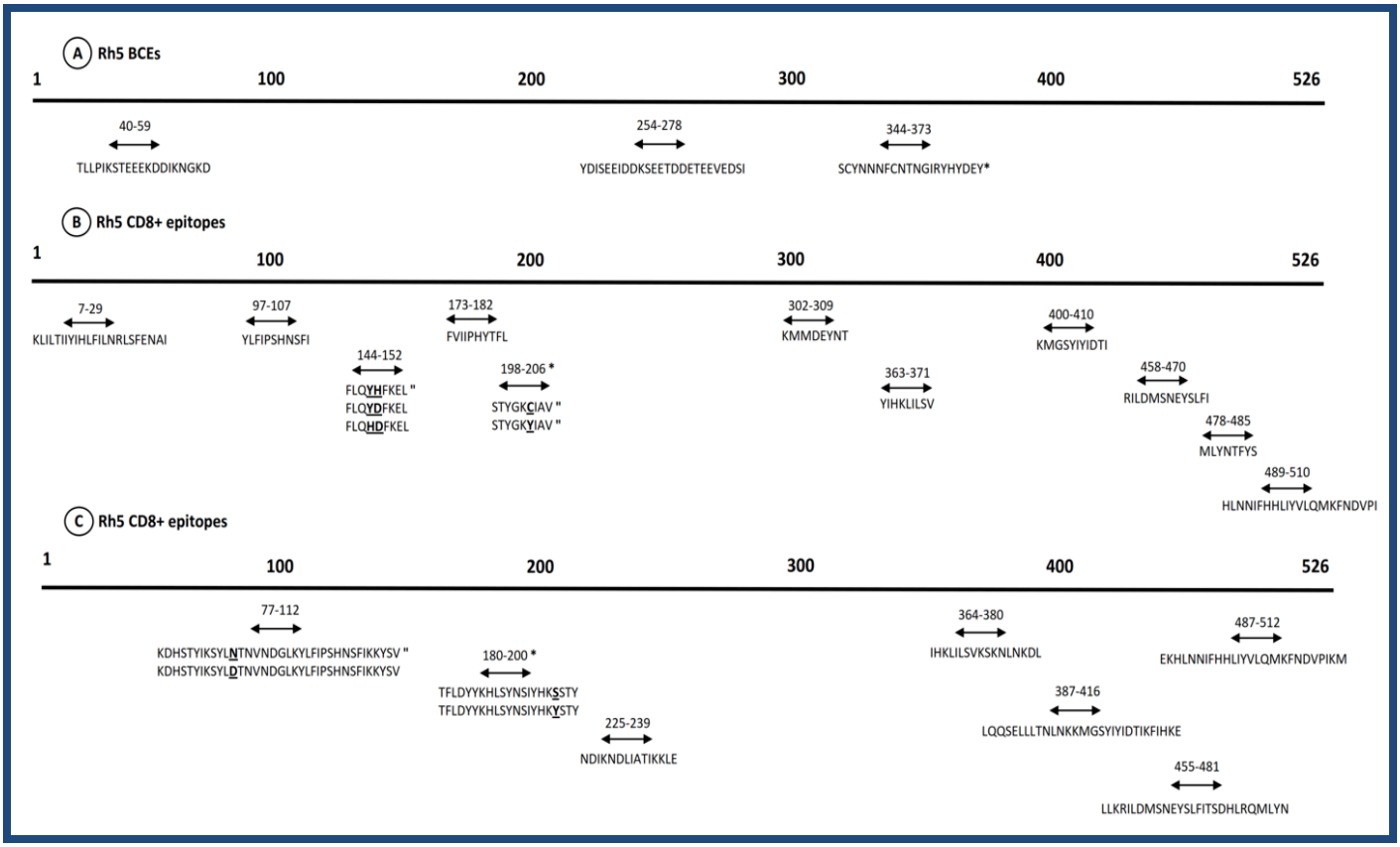

Figure 3: A schematic view of the Rh5 predicted epitopes mapped to the full Rh5 sequence, not drawn to scale: **A)** Rh5 predicted BCEs; **B)** Rh5 predicted CD8+ epitopes; **C)** Rh5 predicted CD4+ epitopes. The numbers displayed above each epitope and separated by hyphens represent the amino acid regions that each epitope encompasses. The residues in bold and underlined represent polymorphic sites within the respective epitopes. Of the polymorphic epitopes, those marked with " represents the epitopes that were predicted. The epitope positions marked with * represent epitopes that overlap with residues involved in binding to basigin

Validation of T-cell Epitope Prediction Algorithms

We selected HLA alleles that were common globally, from malaria endemic areas and those associated with resistance to malaria infection **(Table 1).** The HLAs that confer protection against malaria were obtained from malaria endemic regions in Africa and Asia, with the rationale that individuals expressing these alleles are likely to generate an immune response during an infection. We therefore selected 6 class I HLA alleles for cytotoxic T-cell lymphocytes (CD8+), HLA-A02:01, HLA-A02:04, HLA-A02:11, HLA-B15:13, HLA-B27:05, HLA-B35:01nd HLA-B53:01 and 9 class II HLA alleles for helper T-cells (CD4+), DRB1*01:01, DRB1*04:01, DRB1*11:01, DRB1*11:08, DRB1*13:02, DRB1*13:16, DRB1*14:21, DRB1*15:03 and HLAQA1*01:02-DQB1*05:01. NetMHCcons **[43]** and NetMHCIIpan 3.0 **[44]** algorithms were selected for the MHC class I and class II binding predictions, respectively.

We then determined if these tools could predict the *in vitro* verified TCEs directly from the full CSP protein sequence (**Figure 1).** The CSP sequence was submitted to the NetMHCcons server with the default parameters, a peptide length of 8-11mers and the HLA class I alleles mentioned earlier were selected. For NetMHCIIpan 3.0, the parameters for the CSP sequence were similar to those of NetMHCcons except for the HLA class II alleles and the 15mer epitope length. We then determined how well the prediction algorithms identified the experimentally verified CSP TCEs, firstly, by identifying promiscuous epitopes (those that bound to multiple HLA alleles) then clustering them against the experimentally verified CSP TCEs to identify overlaps at a threshold of 50%. This took into consideration the limited number of HLA alleles used and excluded any duplicated epitopes.

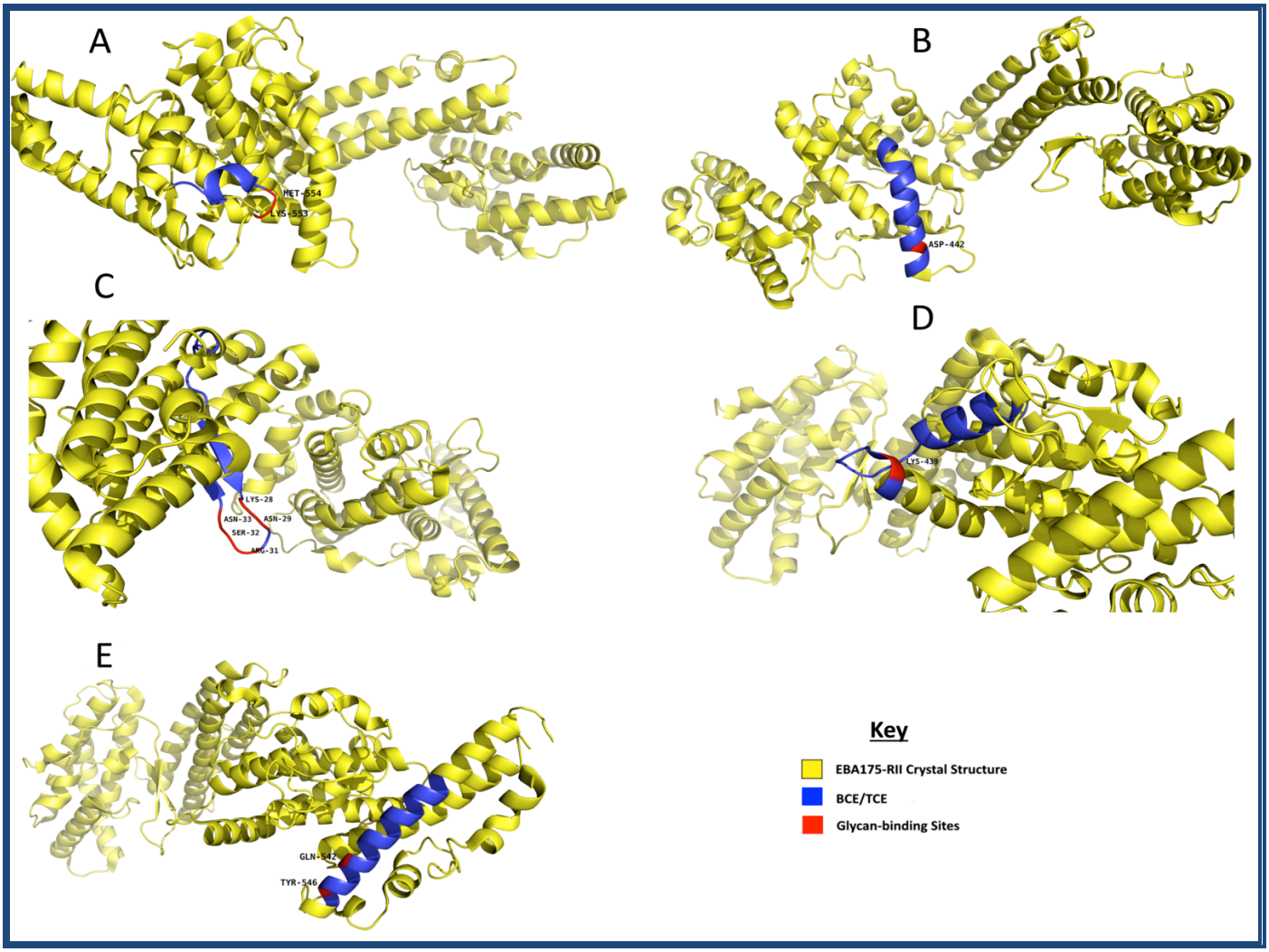

Figure 4: The crystal structure of EBA175-RII showing (A) the overlap between the predicted CD8+ epitope (aa 553-561) and the glycan binding sites at residues Lys-553 and Met-554, (B) the overlap between the predicted CD4+ epitope (aa 440-456) and the glycan binding sites at residue Asp-442, (C) the overlap between the predicted BCE (aa 15-34) and the glycan binding sites at residues Lys-28, Asn-29, Arg-31, Ser-32 and Asn-33, (D) the overlap between the predicted BCE (aa 420-440) and the glycan binding sites at residue Lys-439 and (E) the overlap between the predicted BCE (aa 528-547) and the glycan binding sites at residues Gln-542 and Tyr-546.

ISSN 0973-2063 (online) 0973-8894 (print)

BIOINFORMATION 12(3): 82-91 (2016)

EBA175-RII and Rh5 BCE and TCE predictions

The selected parameters and cut-offs that gave suitable results for CSP were used to predict BCEs and TCEs in both EBA175-RII and Rh5 **(Figure 1).** Since the epitopes were generated from the haplotype sequences for both EBA175-RII and Rh5, we aligned them to their respective 3D7 lab isolate sequence to identify the polymorphic epitopes. We considered as one epitope, multiple epitopes aligning to the same loci. We prioritised the number of predicted EBA175-RII and Rh5 epitopes for in vitro validation by clustering them at 100% identity to eliminate duplicates. We then identified epitopes in the regions that are involved in the EBA175-GypA and Rh5-basigin interactions. These epitopes were mapped onto the published Rh5 and EBA175-RII crystal structures using Pymol Version 1.7.2.1 to identify their locations in the folded protein.

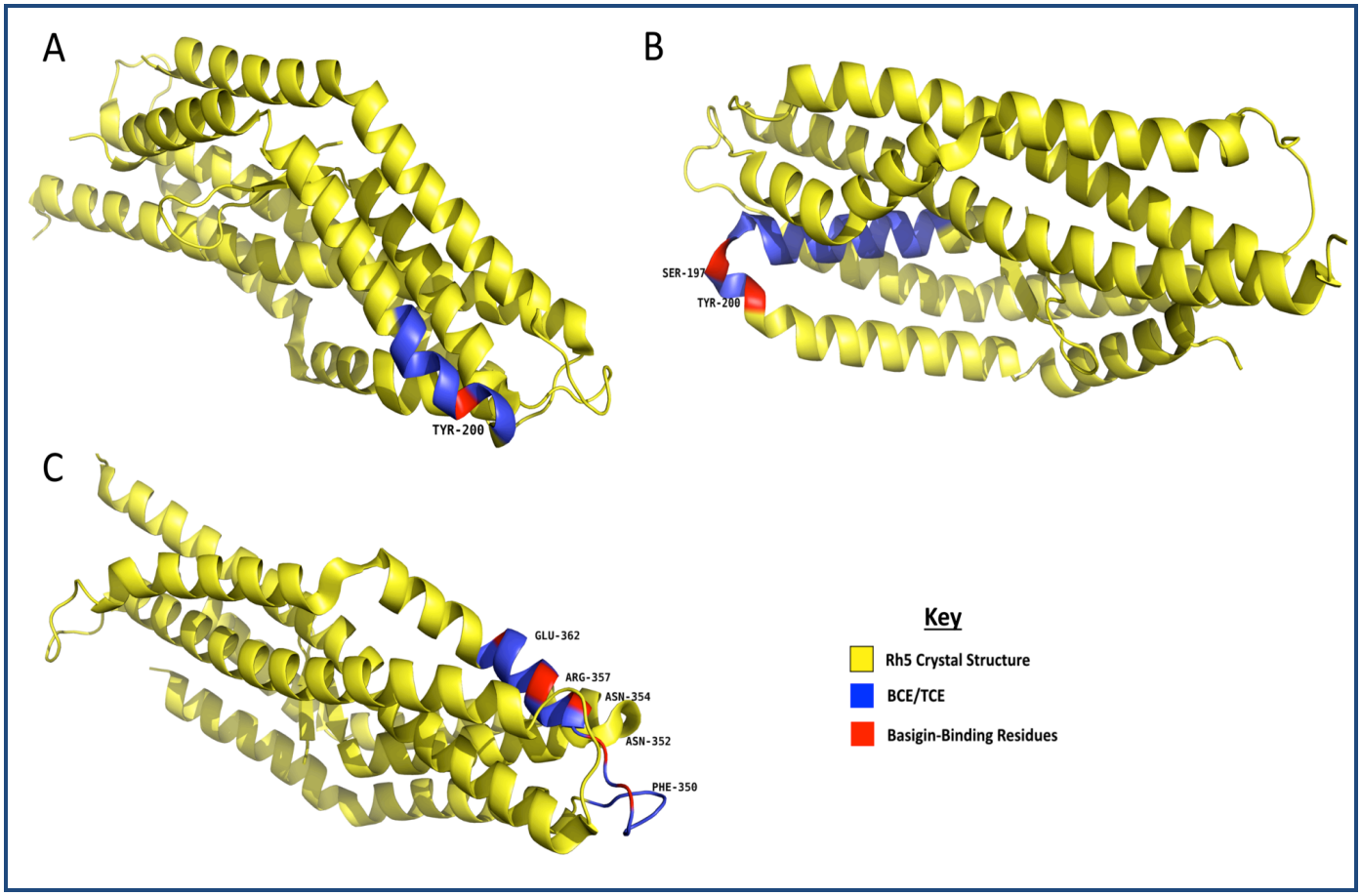

Figure 5: The crystal structure of Rh5 showing **(A)** the overlap between the predicted CD8+ epitope (aa 198-206) and the residue Tyr-200 involved in binding to basigin, **(B)** the overlap between the predicted CD4+ epitope (aa 180-200) and the residues Ser-197 and Tyr-200 involved in binding to basigin and **(C)** the overlap between the predicted BCE (aa 344-363) and the residues Phe-350, Asn-352, Asn-354, Arg-357 and Glu-362 involved in binding to basigin.

Results:

Prediction of BCEs and TCEs in CSP

After clustering the 22 predicted BCEs, we remained with 18 unique epitopes **(Table 2)** of which 7 contained the CSP BCE, $NANP_3$. Since all the predicted CSP BCEs had antigenic scores of 1, we used this value in our selection of EBA175-RII and Rh5 BCEs.

From the 84 predicted CSP CD8+ epitopes (**Table 3),** 7 span two of the *in vitro* verified CSP CD8+ epitopes, ^{336}VTCGNGIQVR^{345} and ^{386}GLIMVLSFL^{394}. None of the epitopes bound to all the 6 HLA class I alleles. Two similar epitopes (^{386}GLIMVLSFL^{394} [9mer] and ^{386}GLIMVLSFLFL^{396} [11mer]) that span the *in vitro* verified CSP epitope ^{386}GLIMVLSFL^{394} bound to a maximum of 3 class I HLA alleles. Therefore, selected EBA-175-RII and Rh5 class I epitopes had to bind to a minimum of 3 HLA alleles.

28 of the 121 predicted CSP CD4+ epitopes **(Table 4**) span the three in vitro verified CSP CD4+ epitopes (CS.T3 ^{363}DIEKKICKMEKCSSV^{377}, Th2R ^{311}PSDKHIKEYLNKIQNSL^{327} and Th3R ^{341}GIQVRIKPGSANKPKDELDYANDI^{364}). Two similar epitopes, ^{317}KEYLNKIQNSLSTEW^{331} and ^{316}IKEYLNKIQNSLSTE^{330}, which span the CSP epitope 311-327 bound to 8 of the 9 class II HLA alleles. The selected EBA-175-RII and Rh5 class II epitopes therefore had to bind to a minimum of 8 HLA alleles.

ISSN 0973-2063 (online) 0973-8894 (print)

BIOINFORMATION 12(3): 82-91 (2016)

Table 1: Common HLA alleles and HLA alleles associated with resistance to malaria

Reference Population	Reference	HLA Alleles
World+	[1]	A*02:01, A*02:04, B*27:05, DRB1*01:01, DRB1*04:01
Kenya (Africa)+	[2]	DRB1*1101, DRB1*1503, DRB1*1302, DRB1*1108, DRB1*1316 and DRB1*1421
Mali (Africa)+	[3]	B*53
Gambia (Africa)*	[4]	B*53:01
New Delhi, India (Asia)*	[5]	A*02:11
Malaysia (Asia)*	[6]	B*15:13
Gambia (Africa)*	[7]	B*53:01 and DRB1*13:02
West Africa*	[8]	DRB1*01:01
Kenya, Uganda & Tanzania East Africa*	[9]	DQA1*01:02-DQB1*05:01

+ HLA alleles that are common within the respective populations; * HLA alleles that have been shown to confer resistance to malaria

EBA175-RII Epitope Predictions

The twelve haplotypes for both EBA175-RII and Rh5 **(Table 5)** revealed that both proteins contained indels and SNPs and we examined their impact on epitope prediction.

Nine EBA175-RII BCEs were predicted **Figure 2A** 6 conserved

([15]REKRKGMKWDCKKKNDRSNY[34],
[59]TMKDHFIEASKKESQLLLKKNDNKYN[84],
[287]TTLVKSVLNGNDNTIKEKRE[306],
[309]DLDDFSKFGCDKNSVDTNTK[328],
[383]LKRKYKNKDDKEVCKIINKT[402] and
[528]WISKKKEEYNKQAKQYQEYQ[547])

and

3 polymorphic

([190]ERDNRSKLPKSKCKNNTLYEA[210],
[234]HTLSKEYETQKVPKENAENY[253] and
[420]SNRKLVGKINTNSNYVHRNKQ[440]).

Of the 4 variants (KP, KS, NP or NS at codons 244 and 246, respectively) in the polymorphic epitope 234-253, only the NS and NP variants were predicted as epitopes. In the other 2 polymorphic epitopes, all the variants were predicted as epitopes. Three of the conserved EBA175-RII BCEs (15-34, 420-440 and 528-547) overlapped with residues, K28, N29, R31, S32 and N33, K439, Q542 and Y546, that have been shown to interact with GypA.

Ten EBA175-RII CD8+ epitopes were predicted **(Figure 2B),** which can be summarized as 6 conserved ([42]IQLCIVNLSI[51], [95]FLDYGHLAM[103], [150]KLWEAMLS[157], [359]RIYDKNLLMIKEHILAIAIYESRI[382], [499]KMIETLKV[506] and [553]KMYSEFKSI[561]) and 4 polymorphic ([187]FLLERDNRSKL[197], [260]NKNDAKVSLLL[270], [328]KVWECKKPYKL[338] and [443]KLFRDEWWKVIKKDVWNV[460]) epitopes. Within the polymorphic epitope 260-270, at codons 260 and 261, variants NK and KM were identified and only the KM variant was predicted as an epitope. All the variants were predicted as epitopes in the other polymorphic regions. The conserved epitope 553-561 overlapped with residues, K553 and M554, which are involved in binding to GypA. Three EBA175-RII CD4+ epitopes were predicted (**Figure 2C),** of which 2 were conserved ([38]PDRRIQLCIVNLSIIKTY[55] and [362]DKNLLMIKEHI LAIAIYE[379]) and 1 was polymorphic ([440]QNDKLFRDEWWKVIKKD[456]). The four variants KA, EA, QE and KE at codons 440 and 448, respectively were identified in the polymorphic epitope and only the KA and EA variants were predicted as epitopes. This epitope also included residue D442 that interacts with GypA.

Rh5 Epitope Predictions

The 3 predicted Rh5 BCEs (**Figure 3A),** [40]TLLPIKST EEEKDDIKNGKD[59], [254]YDISEEIDDKSEETDDETEEVEDSI[278] and [344]SCYNNNFCNTNGIRYHYDEY[363], were all conserved and epitope 344-363 is in the region shown to interact with basigin that includes residues F350, N352, N354, R357 and E362. Eleven Rh5 CD8+ epitopes were predicted **Figure 3B),** 9 of which were conserved ([7]KLILTIIYIHLFILNRLSFENAI[29], [97]YLFIPSHNSFI[107], [173]FVIIPHYTFL[182], [302]KMMDEYNT[309], [363]YIHKLILSV[371], [400]KMGSYIYIDTI[410], [458]RILDMSNEYSLFI[470], [478]MLYNTFYS[485] and [489]HLNNIFHHLIYVLQMKFNDVPI[510]) and 2 were polymorphic ([144]FLQYHFKEL[152] and [198]STYGKCIAV[206]). The polymorphic epitope 144-152 contained the variants YH, YD and HD at codons 147 and 148, respectively, and only the YH variant was predicted as an epitope. In the other polymorphic epitope, both variants of C203Y were predicted as epitopes and it included residue Y200 that interacts with basigin.

Of the 7 Rh5 CD4+ epitopes predicted (**Figure 3C),** 5 were conserved ([225]NDIKNDLIATIKKLE[239], [364]IHKLILSVKSKNL NKDL[380], [387]LQQSELLLTNLNKKMGSYIYIDTIKFIHKE[416], [455]LLKRILDMSNEYSLFITSDHLRQMLYN[481] and [487]EKHLNNIFHHLIYVLQMKFNDVPIKM[512]) and 2 were polymorphic ([77]KDHSTYIKSYLNTNVNDGLKYL FIPSHNS FIKKYSV[112] and [180]TFLDYYKHLSYNSIYHKSSTY[200]). Within the polymorphic epitope 77-112, codon 88 was a singleton SNP and consisted of a D or N and only the N variant was predicted as an epitope. Both variants (codon S197Y) in the other polymorphic epitope were predicted and it also included residue Y200 that interacts with basigin.

Mapping of candidate epitopes to their respective crystal structures

For purposes of selecting candidate epitopes for *in vitro* validation, we considered the epitopes located in regions previously described as being involved in ligand-receptor interactions. We mapped these epitopes onto the protein tertiary structures to determine their spatial positioning within the erythrocyte binding domains. They included EBA175-RII CD8+ epitope 553-561 **(Figure 4A),** CD4+ epitope 440-456 **(Figure 4B)** and three EBA175-RII BCEs including 15-34, 420-440 and 528-547 **(Figure 4C, 4D & 4E).** The Rh5 epitopes included a CD8+ epitope 198-206 **(Figure 5A),** a CD4+ epitope 180-200 **(Figure 5B)** and a BCE 344-363 **(Figure 5C).**

ISSN 0973-2063 (online) 0973-8894 (print)

Table 2: The BCEs predicted from CSP

Amino Acid Position	Predicted Circumsporozoite Protein (CSP) BCE	Prediction Score	*Epitopes overlapping with $NANP_3$
27-46	GSSSNTRVLNELNYDNAGTN	0.982	
83-12	NEKLRKPKHKKLKQPADGNP	1	
85-14	KLRKPKHKKLKQPADGNPDP	1	
99-118	NPNVDPNANPNVDPNANPNV	1	
122-141	ANPNVDPNANPNVDPNANPN	1	
124-143	PNVDPNANPNANPNANPNAN	1	*
128-147	PNANPNANPNANPNANPNAN	1	*
13-149	ANPNANPNANPNANPNANPN	1	*
131-15	NPNANPNANPNANPNANPNA	1	*
133-152	NANPNANPNANPNANPNANP	1	*
195-214	NPNVDPNANPNANPNANPNA	1	*
258-277	ANPNANPNANPNANPNKNNQ	1	*
274-293	KNNQGNGQGHNMPNDPNRNV	1	
289-38	PNRNVDENANANSAVKNNNN	1	
297-316	ANANSAVKNNNNEEPSDKHI	1	
326-354	SLSTEWSPCSVTCGNGIQVR	1	
328-347	STEWSPCSVTCGNGIQVRIK	1	
349-368	GSANKPKDELDYANDIEKKI	1	

(*) The underlined epitopes highlight the predicted peptides that contained the in vitro verified CSP epitope (NANPNANPNANP). The prediction scores ranged from 1 (most antigenic) to 0 (least antigenic).

Table 3: An extract showing the top predicted CD8+ epitopes from CSP. The table shows the predicted epitopes (peptides) as well as the HLA class I alleles to which they bound to HLA-epitope binding prediction values are given in nanomolar (nM) inhibitory concentration 50 (IC50) values, where values less than 50nM and less than 500nM represent strong and weak binders, respectively. Above this values, epitopes are regarded as poor binders. We, therefore, selected only the epitopes with less than or equal to 500nM values for validation of the NetMHCcons algorithm.

Pos	Peptide	HLA-A02:01 nM	HLA-A02:04 nM	HLA-A02:11 nM	HLA-B15:13 nM	HLA-B27:05 nM	HLA-B53:01 nM	Binders
388	MVLSFLFL	33.3	155.1	25.97	1452.7	11355.63	1139.41	3
387	IMVLSFLFL	99.33	311.18	100.95	2192.24	9499	12315.54	3
386	LIMVLSFL	81.75	288.22	40.03	2932.65	15967.2	8297.38	3
386	LIMVLSFLFL	82.64	320.14	126.02	2783.33	12183	3932.89	3
385	GLIMVLSFL	40.68	288.74	6.97	8870.49	14963.56	31740.46	3
385	GLIMVLSFLFL	56.28	407.06	25.14	8315.53	12517.05	29906.73	3
369	KMEKCSSVFNV	59.09	235.8	4.17	14707.63	14643.23	26265.23	3
318	YLNKIQNSL	53.03	409.47	5.35	3033.35	11922.2	17130.58	3
8	SVSSFLFV	41.57	200.45	31.55	7565.59	22818.9	18985.07	3
6	ILSVSSFLFV	16.66	112.31	9.14	7999.42	21852.38	19932.31	3
5	AILSVSSFL	80	388.8	14.01	6191.57	18578.65	22090.1	3
5	AILSVSSFLFV	13.86	77.22	4.88	13703.2	18985.07	25426.37	3
3	KLAILSVSSFL	55.38	139.95	17.88	7412.19	12249.09	25842.4	3
0	MMRKLAILSV	135.94	275.89	8.34	1530.44	5774.65	17695.75	3
386	LIMVLSFLF	2766.91	1713.98	1682.07	320.98	8571.13	301.1	2
384	IGLIMVLSF	20478.82	24790.42	20589.91	390.33	14252.45	731.18	2
379	VVNSSIGLIMV	439.71	1217.85	10.63	13259.87	26550.96	22330.41	2
369	KMEKCSSV	398.91	1050.31	8.9	18665.63	18378.72	31569.21	2
325	SLSTEWSPCSV	61.04	604.46	6.79	12258.96	23699.6	27131.77	2
318	YLNKIQNSLST	444.5	2415.14	39.17	13745.59	17888.25	28640	2
52	MNYYGKQENW	34798.02	36917.57	30893.41	107.44	22330.41	107.14	2
33	VLNELNYDNA	268.76	1145.48	10.93	29318.68	28485.48	31912.64	2
12	FLFVEALF	801.61	1350.13	195.32	192.06	9812.39	2812.19	2

Discussion:

In this study, we demonstrated that *in silico* tools can predict *in vitro* verified BCEs and TCEs in CSP, the protein used in the RTS,S subunit malaria vaccine. This technique may prove to be a useful way to rapidly prioritize potential vaccine targets, especially when coupled with *in vitro* validation experiments. Sedegah et al., (2013) used ELISpot assays and *in silico* prediction to identify novel CD8+ epitopes in CSP. Of the 5 *in vitro* verified CD8+ epitopes, 4 overlapped with our predicted CSP CD8+ epitopes, ^{387}LIMVLSFLF^{39}, ^{13}FLFVEALFQE^{22}, ^{376}SVFNVVNSSI^{385} and ^{12}SFLFVEALF^{20}. Rodrigues-da-Silva *et al.* (2016) also combined *in silico* prediction and *in vitro* validation to identify a candidate BCE in *P. vivax* merozoite surface protein (MSP) 9. The *in vitro* verified CSP BCE ($NANP_3$) was predicted with an antigenic score of 1, the highest possible score for a predicted epitope. This suggests that these epitopes are likely to be the most antigenic in comparison to all other predicted epitopes. We did not predict all the *in vitro* verified CSP TCEs, perhaps due to the limited panel of 15 class I and II HLA alleles selected. We also did not predicted epitopes shown in previous studies to potently inhibit invasion. For instance, two BCEs 332-344 and 410-422 mapped by Ambroggio et al. (2013) **[45],** which encompass the previously reported monoclonal antibodies R215, R217 and R256 **[20],** fall within the region involved in binding to GypA. Moreover, the monoclonal antibody to ^{28}AIKK^{31} identified by Ord *et al.* (2014)

[
46] was not predicted. However, despite these limitations, our observations with the *in vitro* verified CSP epitope predictions formed the basis of our selection criteria for EBA175 RII and Rh5 BCE and TCE predictions. We do however note that a larger number of HLA alleles can be used in the predictions, as demonstrated by a similar study which used 34 HLA alleles to predict CSP CD8+ epitopes **[10].**

Table 4: An extract showing the top predicted CD4+ epitopes from CSP. The table shows the predicted epitopes (peptides) as well as the HLA class-II alleles to which they bound to. HLA-epitope binding prediction values are given in nanomolar (nM) inhibitory concentration 50 (IC50) values, where values less than 50nM and less than 500nM represent strong and weak binders, respectively. Above this values, epitopes are regarded as poor binders. We, therefore, selected only the epitopes with less than or equal to 500nM values for validation of the NetMHCIIpan 3.0 algorithm.

Pos	Peptide	DRB1*01:01	DRB1*04:01	DRB1*11:01	DRB1*11:08	DRB1*13:02	DRB1*13:16	DRB1*14:21	DRB1*15:03	HLA-DQA1*01:02-DQB1*05:01	Binders
		nM	nM	nM	nM	nM	nM	nM	nM	nM	
376	VFNVVNSSIGLIMVL	26.25	146.58	342.02	84.63	121.82	192.98	12.45	252.06	4236.33	8
375	SVFNVVNSSIGLIMV	20.46	83.51	196.14	65.37	112.67	154.77	12.84	207.6	4000.1	8
374	SSVFNVVNSSIGLIM	13.81	51.36	119.31	42.17	118.22	139.9	10.32	146.52	4577.66	8
373	CSSVFNVVNSSIGLI	21	74.16	197.04	58.87	120.2	136.23	12.02	189.9	3826.61	8
372	KCSSVFNVVNSSIGL	40.8	118.85	363.2	98.41	188.7	212.12	17.17	399.72	5254.22	8
316	KEYLNKIQNSLSTEW	21.84	110.32	142.35	19.2	288.54	282.79	7.15	455.95	4468.24	8
315	IKEYLNKIQNSLSTE	35.44	162.15	194.91	30.51	415.49	491.47	11.13	422.86	6690.59	8
377	FNVVNSSIGLIMVLS	40.98	240.34	552.46	121.52	188.56	306.43	16.35	427.96	3884.55	7
314	HIKEYLNKIQNSLST	49.64	203.67	217.06	36.8	439.63	635.44	12.7	343.24	5954.27	7
3	KLAILSVSSFLFVEA	40.59	483.17	410.08	86.56	2490.13	4272.18	72.58	191.37	298.3	7
2	RKLAILSVSSFLFVE	31.39	285.06	280.81	50.66	1515.59	3153.05	41.08	162.31	419.19	7
371	EKCSSVFNVVNSSIG	126.27	218.01	784.65	176.64	240.65	306.91	27.94	760.37	5577.59	6
370	MEKCSSVFNVVNSSI	316.86	470.09	1455.69	448.1	250.45	340.35	53.7	972.57	5084.76	6
317	EYLNKIQNSLSTEWS	38.19	189.76	299.27	34.58	461.59	502.9	11.47	749.09	4722.96	6
313	KHIKEYLNKIQNSLS	89.3	296.82	219.98	51.13	593.17	834.12	14.12	349.69	5097.87	6
58	QENWYSLKKNSRSLG	36.68	437.42	25.15	28.13	1331.35	1562.82	16.06	461.29	14240.43	6
11	SFLFVEALFQEYQCY	112.56	1129.53	422.28	91.26	4407.55	5287.98	139.93	293.54	92.71	6
10	SSFLFVEALFQEYQC	77.3	1020.31	342.54	77.18	5192.95	6345.46	126.91	290.58	84.75	6
9	VSSFLFVEALFQEYQ	52.74	964.15	270.89	66.04	5056.45	6597	119.79	291.15	71.84	6
8	SVSSFLFVEALFQEY	53.02	1123.79	334.05	76.83	5571.38	7059	164.14	343.19	68.52	6
1	MRKLAILSVSSFLFV	21.95	174.1	215.66	34.57	893.13	1689.38	25.67	121.32	677.45	6
0	MMRKLAILSVSSFLF	14.22	125.39	138.93	25.71	790.61	1451.34	14.77	82.29	704.86	6
365	KKICKMEKCSSVFNV	211.46	659.03	247.84	124.06	402.21	568.36	10.61	669.23	10099.23	5

Table 5: Unique Haplotypes generated from EBA175-RII and Rh5 isolates

(i) Rh5 Isolates		(ii) EBA175-RII Isolates	
Isolate	**Haplotypes**	**Isolate**	**Haplotypes**
Rh5 3D7	N-----YHSCK	EBA175 3D7 (RII)	KEKKDKPISENKKKLNQERE
ID01	N-----YHSYK	ID01	KKEKDNSISKNKNILNKESE
ID02	N-----HDSYK	ID02	EEEEDNS--KKMKKLNEASK
ID03	N-----HDSCK	ID03	KKEKDNS--KKMKKVNEASK
ID04	N-----YDSYK	ID04	KKEKYNSISENKKKLNEASE
ID05	N-----HDSCN	ID05	KKEKDKSISENKKKVNQERE
ID06	N-----YHYYK	ID06	KEEKDKPISENKKILKQESE
ID07	NDYKNVYHSYK	ID07	KEEKDKPISENKKIINKESE
ID08	N-----YDSYN	ID08	KKEEDNS--KKMKKLNKASK
ID09	N-----YHSCN	ID09	KEKEDKPISENKKKLNQERE
ID10	D-----HDSYK	ID10	KEEKDNPISKNKKKLNQERK
ID11	N-----YHSYN	ID11	KEKEDNS--KKMKKVNQERE

We determined the impact of polymorphisms on epitope prediction in EBA175-RII and Rh5. Fewer BCEs and TCEs were predicted in the polymorphic regions than in the conserved regions and some variants were not predicted as epitopes. For instance, the polymorphic codons 147 and 148 in the Rh5 CD8+ polymorphic epitope ^{144}FLQYHFKEL^{152}, consisted of three variants, YH, YD and HD, and the YD and HD variants were not predicted as epitopes. It appears that in this *in silico* analysis, particular amino acid combinations escape prediction as immunogenic epitopes. The polymorphisms in *P. falciparum* merozoite antigens are thought to be the result of immune selection, thus allowing the parasites to escape detection by host immune responses. In natural infections, immune escape has been demonstrated in polymorphic antigens MSP2 and apical membrane antigen 1 (AMA1), as allele-specific immunity **[47, 48].** Subsequently, immune responses generated to one allele of AMA1 or MSP2 only protects against the same allele and not a different allele. Perhaps, *in silico* tools could indicate

ISSN 0973-2063 (online) 0973-8894 (print)

potential variant epitopes that may escape immunity. Allele-specific immunity has not been described for either EBA175-RII or Rh5 and more recently a study by Gandhi *et al.* (2014) **[49]** found no evidence of allele-specific immunity in CSP. Nevertheless, we hypothesize that the polymorphisms in these antigens may be driven by host immunity, resulting in allele-specific immunity or escape from immune detection or a redirection of the immune response away from important functional regions, such as those involved in allowing the antigen to bind the RBC receptor. In the case of Rh5, it appears that the polymorphic codons 147 and 148 fall outside the region required for the interaction with basigin. *in vitro* validation is required to test these assumptions.

All BCEs and TCEs predicted for EBA175-RII and Rh5, both polymorphic and conserved are novel. However, to prioritise epitopes for *in vitro* validation we focused on epitopes that would interfere with the functional roles of EBA175-RII and Rh5 in erythrocyte invasion. We rationalized that if we target regions of the proteins that can inhibit ligand-receptor interactions, these molecules if immunogenic may be effective in preventing parasite invasion and ultimately malaria pathology.

We prioritized 8 epitopes for *in vitro* validation,

three EBA175-RII BCEs,

^{15}REKRKGMKWDCKKKNDRSNY^{34},
^{420}SNRKLVGKINTNSNYVHRNKQ^{440} and
^{528} WISKKKEEYNKQAKQYQEYQ^{547},

a EBA175-RII CD8+ epitope ^{553}KMYSEFKSI^{561},
a EBA175-RII CD4+ epitope ^{440}QNDKLFRDEWWKVIKKD^{456},
a Rh5 BCE ^{344}SCYNNNFCNTNGIRYHYDEY^{363},
a Rh5 CD8+ epitope ^{198}STYGKCIAV^{206} and
one Rh5 CD4+ epitope ^{180}TFLDYYKHLSYNSIYHKSSTY^{200}.

These epitopes cover both conserved and polymorphic regions, since we recognize that a combination of both regions is likely to be more effective in inhibiting RBC invasion. We recommend the aforementioned epitopes for *in vitro* validation, by testing their immunogenicity using sera from a malaria endemic population. In particular, the TCEs are of interest, since to the best of our knowledge no study has evaluated T-cell responses to Rh5 and only Malhotra *et al.* (2005) **[50]** have evaluated T-cell responses to EBA175-RII, but the epitopes were not mapped.

Conclusion:
The BCE and TCE prediction algorithms resulted in multiple putative epitopes. This can be attributed to a lack of sufficient training data to further benchmark these tools and improve their performance. It also highlights the need to couple the use of *in silico* epitope prediction tools with *in vitro* validation of predicted epitopes to improve the accuracy of the pipeline and provide the training data required. Nonetheless, *in silico* tools provide a quick way to identify potential vaccine targets that can then be screened *in vitro* to determine their immunogenicity and viability as possible malaria vaccine candidates.

Conflict of interest statement:
The authors have no conflicts of interest to report.

Author Contributions:
L.I.O-O and K.K.W were involved equally in conception, design and undertaking of the study as well as drafting and revision of the article.

Acknowledgement:
We acknowledge the Malaria Capacity Development Consortium (MCDC) re-entry grant to L.I.O-O which funded the generation of the EBA175 and Rh5 sequence data from the Kilifi isolates. We thank Prof. James Ochanda and Dr. George Obiero the previous and current CEBIB Director, respectively for supporting this work. We thank the Director of the Kenya Medical Research Institute for permission to publish the article.

References:
[1] WHO. World Malaria Report 2015. World Health Organization, Geneva.; 2015.
[2] Noedl H *et al. N Engl J Med*. 2008 **359**: 2619 [PMID: 19064625]
[3] Dondorp AM *et al. N Engl J Med*. 2009 **36**: 455 [PMID: 19641202]
[4] Amaratunga C *et al. Lancet Infect Dis*. 2012 **12**: 851 [PMID: 22940027]
[5] Hargreaves K *et al. Med Vet Entomol*. 2000 **14**: 181 [PMID: 10872862]
[6] Casimiro S *et al. J Med Entomol*. 2006 **43**: 276 [PMID: 16619611]
[7] Protopopoff N *et al. Trop Med Int Heal*. 2008 **13**: 1479 [PMID: 18983277]
[8] Ranson H *et al. Malar J*. 2009 8: 299 [PMID: 20015411]
[9] Agnandji ST *et al. N Engl J Med*. 2012 **367**: 2284 [PMID: 23136909]
[10] Sedegah M *et al. Malar J*. 2013 **12**: 185 [PMID: 23738590]
[11] Rodrigues-da-Silva RN *et al.* Braga ÉM, editor. *PLoS One*. 2016 **11**: e0146951 [PMID: 26788998]
[12] Miller LH *et al. Science*. 1994 **264**: 1878 [PMID: 8009217]
[13] Cowman AF *et al. Cell*. 2006 **124**: 755 [PMID: 16497586]
[14] Duraisingh MT *et al. Subcell Biochem*. Springer New York; 2008 **47** :46 [PMID: 18512340]
[15] Adams JH *et al. Cell*. 1990 **63**: 141 [PMID: 2170017]
[16] Camus D *et al. Science*. 1985 **230**: 4725. [PMID: 3901257]
[17] Narum DL *et al. Infect Immun*. 2000 **68**: 1964 [PMID: 10722589]
[18] Pandey K *et al. Mol Biochem Parasitol*. 2002 **123**: 23 [PMID: 12165386]
[19] Jiang L *et al. Proc Natl Acad Sci U S A*. 2011 **108**: 7553 [PMID: 21502513]
[20] Sim BKL *et al. PLoS One*. 2011 **6**: e18393. [PMID: 21533224]
[21] Ord RL *et al. PLoS One*. 2012 **7**: e30251 [PMID: 22253925]
[22] Jones TR *et al. J Infect Dis*. 2001 **183**: 303 [PMID: 11110648]
[23] El Sahly HM et al. *Clin Vaccine Immunol*. 2010 **17**: 1552 [PMID: 20702657]
[24] Tham W-H *et al. Proc Natl Acad Sci U S A*. 2010 **107**: 17327 [PMID: 20855594]
[25] Crosnier C *et al. Nature*. 2011 **480**: 534 [PMID: 22080952]
[26] Douglas AD et al. *Cell Host Microbe*. 2015 **17**: 130 [PMID: 25590760]
[27] Manske M et al. *Nature*. Nature Publishing Group; 2012 **487**: 375

[PMID: 22722859]
[28] Bustamante LY *et al. Vaccine*. Elsevier Ltd. 2013 **31**: 373 [PMID: 23146673]
[29] Douglas AD *et al. Nat Commun*. 2011 **2**: 601 [PMID: 22186897]
[30] Williams AR *et al. PLoS Pathog*. 2012 **8** :e1002991 [PMID: 23144611]
[31] Patel SD *et al. J Infect Dis*. 2013 **208**: 1679 [PMID: 23904294]
[32] Tran TM *et al. J Infect Dis*. 2014 **209**: 789 [PMID: 24133188]
[33] Tolia NH *et al. Cell*. 2005 **122**: 183 [PMID: 16051144]
[34] Wright KE *et al. Nature*. 2014 515: 427 [PMID: 25132548]
[35] Zavala F *et al. Science*. 1985 **228**: 1436 [PMID: 2409595]
[36] Zeeshan M *et al. PLoS One*. 2012 **7**: e43430 [PMID: 22912873]
[37] Malik A *et al. Proc Natl Acad Sci U S A*. 1991 **88**: 3300 [PMID: 1707538]
[38] Doolan DL *et al. Immunity*. 1997 **7**: 97 [PMID: 9252123]
[39] Edgar RC. *Bioinformatics*. 2010 **26**; 2460 [PMID: 20709691]
[40] Barh D *et al. Bioinformation*. 2010 **5**:77 [PMID: 21346868]
[41] Oprea M *et al. Biologicals*. Elsevier Ltd. 2013 **41**: 148 [PMID: 23582120]
[42] Douglas AD *et al. J Immunol*. 2014 **92**: 245 [PMID: 24293631]
[43] Karosiene E *et al. Immunogenetics*. 2012 **64**: 177 [PMID: 22009319]
[44] Karosiene E *et al. Immunogenetics*. 201365: 711 [PMID: 23900783]
[45] Ambroggio X *et al.* Sinnis P, editor. *PLoS One*. 2013 **8**: e56326 [PMID: 23457550]
[46] Ord RL *et al. Malar J*. 2014 **13**: 326 [PMID: 25135070]
[47] Genton B *et al. J Infect Dis*. 2002 **185**: 820 [PMID: 11920300]
[48] Ouattara A *et al. J Infect Dis*. 2013 **207**: 511 [PMID: 23204168]
[49] Gandhi K *et al.* Tetteh KKA, editor. *PLoS One*. 2014 **9**: e101783 [PMID: 24992338]
[50] Malhotra I *et al. Infect Immun*. 2005 **73**: 3462 [PMID: 15908375]

Edited by P Kangueane
Citation: **Wamae** & **Ochola-Oyier,** Bioinformation 12(3): 82-91 (2016)