Effect of circulating *PfCSP* variants in Ghana and the implication on vaccine development

Amegashie Elikplim[1], Brodsky Elia[2]; Ghansah Anita [1]

1. Noguchi Memorial Institute for Medical Research, College of Health Sciences, University of Ghana
2. Pine Biotech, Inc.,1441 Canal St., New Orleans, LA, USA 70112

UNIVERSITY OF GHANA
Noguchi Memorial Institute for Medical Research

Pine Biotech
Bioinformatics Training and Research Collaborations

OmicsLogic Bioinformatics for Infectious Diseases

Introduction

- A major challenge in the development of successful malaria vaccines is the antigenic diversity of vaccine candidate genes in natural parasite populations leading to immune escape.
- The most advanced vaccine is the RTS,S/ASO1 vaccine designed based on the immunogenic regions of the 3D7 laboratory strain of *Plasmodium falciparum* Circumsporozoite protein (*PfCSP*) (Fig.1)
- The degree of malaria transmission intensity affects the extent of genetic diversity of PfCSP whereby diverse *PfCSP* non-vaccine haplotypes in specific immune epitopes of the PfCSP antigen were detected reflecting in low levels of the 3D7 strain in Ghanaian natural parasite populations regions (Amegashie et al., 2020)
- Poor vaccine efficacy against non-vaccine parasite strains has been observed (Neafsey et al., 2015)
- However, the effect of these changes on protein structure and interactions is poorly understood. Therefore, we studied the genetic diversity of *PfCSP* between two distinct ecological regions in Ghana and the impact of amino acid mutations specifically in the immunodominant T cell epitopes on the protein structure, stability and potentially its effect on T cell mediated immunity .
- PfCSP amino acid data were obtained from two variable malaria transmission sites, Cape Coast and Navrongo in Ghana (Fig.2). Cape Coast has an Entomological inoculation rate (EIR) of 50 infective mosquito bites per year and Navrongo has EIR of 1132 infective mosquito bites per year.

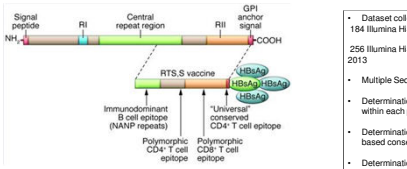

Fig.1: Circumsporozoite protein (CSP) and RTS,S vaccine construct

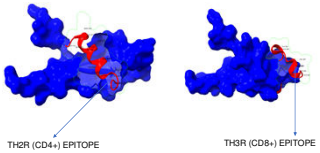

Fig.2: Reference Protein structure (3VDJ.pdb) of PfCSP

Methods

- Dataset collection and Sequencing:
  184 Illumina HiSeq sequence data was collected from Cape Coast in 2013.
  256 Illumina HiSeq sequence data was collected from Navrongo in 2010, 2011 and 2013
- Multiple Sequence Alignment and Variant Analysis with ClustalW
- Determination of the patterns of amino acid sequence substitutions frequency within each population in R
- Determination of amino acid variations between both populations, using entropy-based conservation analysis in ChimeraX.
- Determination of BLOSUM 62 scores and phytochemical properties of amino acid substitutions
- Prediction of the impact of mutations at the T cell immune epitopes upon protein stability and flexibility with DynaMut tool

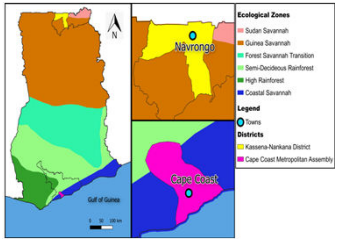

Fig.3: Eco-Epidemiological Zones in Ghana

Results

Distribution of TH2R (CD4+ epitope) and TH3R (CD8+ epitope) Amino acids:

- The distribution graph (Fig.4) revealed the two geographically distinct Ghanaian parasite populations had varying levels of polymorphisms at similar positions with the same types of amino acids that populate these sites appearing to be conserved between them.
- These similar mutable positions 314, 321, 324, 327 at the TH2R epitope , and 352, 359 at the TH3R epitope were seen in both populations.
- Common mutations seen in both populations include: (K314Q, L320I, Q324R, L327I, N352D, D359N) but at different frequencies in the population (Table 1 &2).
- K314Q, L327I, D359N is more prevalent in Navrongo than in Cape Coast ($p = 0.000227$*, 0.000000161*,0.009522* Pearson's Chi square correlation co-efficient respectively)
- N352D is more prevalent in Cape Coast than in Navrongo ($p = 0.02905$* Pearson's Chi square correlation co-efficient respectively)
- Amino acid residues reflecting geographical difference are:
  - N321H seen only in Cape Coast
  - N321K, N321Q, N321T, P354S, Q357E, D359N seen only in Navrongo (Fig.4,Table1&2)
- In Navrongo, some mutations had higher frequencies than the reference amino acid residue
- K317E had the highest mutation rate observed in 78.1% (202/256) in the population followed by N321K seen in 77.4% (199/256) of the population (Table 1)

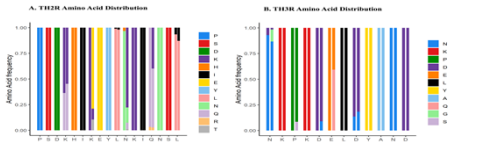

Fig.4: The amino acid column represents the frequency of amino acids seen in isolates from Cape Coast (left half of the column) and Navrongo (right half of the column)

Table 1: TH3R Amino Acid frequency in Cape Coast and Navrongo

Cape Coast	Pos	Ref	Variant	Ref freq	Variant freq
TH2R (311-327)	314	K	Q	117 (63.8 %)	67 (36.2%)
	320	L	I	182 (98.9 %)	2 (1.1%)
	321	N	H	178 (96.8 %)	6 (3.2%)
	324	Q	R	178 (96.8 %)	6 (3.2%)
	327	L	I	172 (93.5 %)	12 (6.5%)

Navrongo	Pos	Ref	Variant	Ref freq	Variant freq
TH2R (311-327)	314	K	Q	140 (54.9%	116 (45.1%)
	317	K	E T	27 (10.9%)	202 (78.1%) 27 (10.5%)
	320	L	I	251 (98.1%)	5 (1.9%)
	321	N	K Q T	38 (15.2%)	199 (77.4%) 16 (6.2%) 13 (1.2%)
	324	Q	K R	147 (57.6%)	102 (39.7%) 7 (2.7%)
	327	L	I	223 (87.2%)	33 (12.8%)

Table 2: TH3R Amino Acid frequency in Cape Coast and Navrongo

Cape Coast	Pos	Ref	Variant	Ref freq	Variant freq
TH3R (352 – 363)	352	N	D	171 (93.0 %)	13 (7.0%)
	359	D	N	159 (86.5)	25 (13.5)

Navrongo	Pos	Ref	Variant	Ref freq	Variant freq
TH3R (352 – 363)	352	N	G D	222 (86.8%)	30 (11.7%) 4 (1.6%)
	354	P	S	234 (91.4%)	22 (8.6%)
	356	D	N	233 (91.1%)	23 (8.9%)
	357	E	Q	151 (59.1%)	105 (40.9%)
	359	D	N	209 (81.7%)	47 (18.3%)

Results

Protein Sequence and Structure: Amino Acid Conservation

- The entropy based conservation analysis of amino acid revealed variable positions at the TH2R and TH3R immunodominant epitopes in both populations (Fig.5).
- High variability was observed Navrongo compared to Cape Coast. These variations are likely to affect T cell recognition and binding
- The positive BLOSUM score revealed that most of these mutations are likely to substitute frequently at those positions except for K317E and P354S which had negative score meaning they occurs rarely
- Amino acid residues are grouped based on their phytochemical properties in terms of volume, hydropathy and their chemistry
- The phytochemical changes includes changes from hydrophobic to hydrophilic residue, basic to acidic, basic to amide, amide to acidic, small to medium, and very small to small (Table 3)
- To determine the likelihood of these phytochemical changes to preserve the structure of the protein, mutation prediction analysis was done
- Mutation prediction analysis revealed 8 mutations predicted to be destabilizing (Table 4)

Fig. 5: Position based conservation analysis of amino acid of Cape Coast

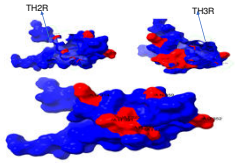

Fig.6: Position based conservation analysis of amino acid of Navrongo

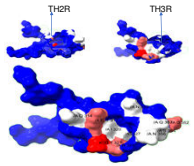

Table 3: Blosum62 scores, Physiochemical properties and mutation effect prediction

Cape Coast	Position	Variant	Type of substitution	Blosum62 score
TH2R (311-327)	314	K > Q	Basic to amide Large to medium Hydrophilic	1
	320	L > I	Hydrophobic Large Aliphatic	2
	321	N > H	Amide to basic Small to medium Hydrophilic	1
	324	Q > R	Hydrophilic Amide to Basic Medium to large	1
	327	L > I	Hydrophobic Aliphatic Large	2
TH3R (352 – 363)	352	N > D	Hydrophilic Amide to acidic small	1
	359	D > N	Hydrophilic Acidic to amide Small	1

Navrongo	Position	Variant	Type of Substitution	Blosum62 score
TH2R (311-327)	314	K > Q	Hydrophilic Large to medium	1
	317	K > E, T	Basic to acidic K>E Basic to hydroxyl K>T Large to medium K >T Large to small K>T Hydrophilic	1, -1
	320	L > I	Hydrophobic Large Aliphatic	2
	321	N > K,Q,T	Amide to Basic (N>K), small to large (N>K) Amide (N>Q), small to medium (N>Q) Amide to Hydroxyl (N>T) Hydrophilic	0, 0, 0
	324	Q > R	Amide to basic Medium to large Hydrophilic	1
	327	L > I	Hydrophobic Large Aliphatic	2
TH3R (352 – 363)	352	N > D	Hydrophilic Amide to acidic Small	1
	354	P > S	Hydrophobic to hydrophilic Aliphatic to hydroxyl Very small to small	-1
	356	D > N	Hydrophilic Amide to acidic Small	1
	357	E > Q	Acidic to amide Hydrophilic medium	2
	359	D > N	Hydrophilic Acidic to amide Small	1

Table 4: Mutation effect prediction

Mutation	Prediction outcome (Stability effect ΔΔG, kcal/mol)	Δ Vibrational Entropy Energy between Wild-type and Mutant, kcal.mol-1. K-1)
K317E	-0.058 - Destabilizing	0.433 - Increase of molecule flexibility
K317T	-1.043 - Destabilizing	0.718 - Increase of molecule flexibility
N321Q	-0.096 - Destabilizing	0.037 - Increase of molecule flexibility
N321H	-0.048 - Destabilizing	0.036 - Increase of molecule flexibility
L327I	-0.151 - Destabilizing	0.009 - Decrease of molecule flexibility
N352D	-0.107 - Destabilizing	0.045 - Increase of molecule flexibility
D356N	-0.036 - Destabilizing	-0.016 - Decrease of molecule flexibility
E357Q	-0.239 - Destabilizing	0.004 - Increase of molecule flexibility

Conclusion

- The amino acid mutations at the T cell epitopes may alter structural conformation, consequently compromising the host immune recognition by T cells (Zeeshan et al., 2012)
- Vaccine efficacy tended to decrease with the number of mismatches with 3D7 at these seven amino acid positions at 299, 301, 317, 354, 356, 359 and 361 (Neafsey et al., 2015), in this study mutations were observed at some of these loci in the Ghanaian population (301,317,354,356, and 359)
- Isolates in Navrongo revealed increased mutations (11 mutable sites in both TH2R and TH3R epitopes) than in Cape Coast (7 mutable sites) and this is likely attributed to higher malaria transmission patterns (EIR =1132) in Navrongo as compared to Cape Coast (EIR = 50).
- Geographical differences in amino acids even within the same country provides insight into the need to design location specific *PfCSP* vaccines
- Radical replacement of an amino acid was observed whereby an amino acid is exchanged into another with different physiochemical properties and this may lead to changes in protein structure with implication of reducing immune recognition by T-cells
- There were radical replacement of amino acid residues in terms of**: Hydropathy**; hydrophobic to hydrophilic residues (P354S), **Chemistry;** basic to acidic (K317E), basic to amide (K314Q), basic to hydroxyl (K317T), amide to basic (N321K, N321H, Q324R), amide to acidic (N352D, D356N), acidic to amide (E357Q, D359N), **Volume;** small to medium (N321H), very small to small (P354S), small to medium (N321Q),small to large (N321K), medium to large (Q324R), large to medium (K314Q, K317T), large to small (K317T).
- These changes may alter the conformation of the protein structure with a potential of affecting its stability and interaction with T-cell receptors and reduce epitope recognition, leading to reduced vaccine efficacy
- Mutation effect prediction analysis revealed 8 mutations predicted to have destabilizing effect on the protein (Table 4) with 3 located in positions previously observed to reduce vaccine efficacy (K317E , K317T and D356N) (Neafsey et al, 2015).
- In Navrongo, there were amino acids with negative BLOSUM scores in positions K317T and P354S suggesting that these amino acids substitute rarely and may negatively be selected against.
- Overall, this study reveals mutations which are likely to occur frequently at specific positions and may affect the vaccine efficacy, therefore aside from the wild-type, these mutants could also be included in the vaccine component.
- Considering the extent of diversity in circulating parasites from different transmission settings even within a country can provide us with insight into how well RTS,S/AS01 may perform if implemented on a large scale in that country

References

Amegashie, E. A., Amenga-Etego, L., Adobor, C., Ogoti, P., Mbogo, K., Amambua-Ngwa, A., & Ghansah, A. (2020). Population genetic analysis of the Plasmodium falciparum circumsporozoite protein in two distinct ecological regions in Ghana. Malaria journal, 19(1), 1-14.

Neafsey DE, Juraska M, Bedford T, Benkeser D, Valim C, Griggs A, et al. Genetic diversity and protective efficacy of the RTS,S/AS01 malaria vaccine. N Engl J Med. 2015;373:2025–37

Zeeshan, M., Alam, M. T., Vinayak, S., Bora, H., et al., (2012). Genetic variation in the plasmodium falciparum circumsporozoite protein in india and its relevance to RTS,S malaria vaccine. PLoS ONE, 7:1–10.

Further information:amegashie.elikplim@students.jkuat.ac.ke