SUMMARY OF THE FINAL REPORT OF THE PROJECT:

The malaria parasite has a complex, multistage life cycle occurring within two living beings, the vector mosquitoes and the vertebrate hosts. During its complex life cycle in the human host, *Plasmodium falciparum* merozoite invades erythrocyte through multiple protein-protein interactions. Among these, interactions between RBC glycophorins with parasite Erythrocyte Binding Antigens are important. Since antigenic variation in parasite proteins is a major barrier to the development of effective vaccine, we explored the pattern of sequence diversity in two vaccine candidates (EBA-175/140) in *P. falciparum* isolates of Kolkata. The objectives are to understand the genetic and population levels forces responsible for shaping the diversity of *eba-175* and *eba-140* genes and to characterize repercussion of genetic diversity in the F2 region of EBA-ligands as they interact with host glycophorins. Venous blood samples (n=92) from *P. falciparum* infected malaria patients were used to sequence RII regions of *Pfeba-175* and *Pfeba-140* genes. Sequence divergence of parasite proteins were subjected to appropriate statistical and phylogenetic analyses using DnaSP and MEGA tools. Strength of host-parasite interactions between parasite proteins (mutant and prototype Pf3D7) and RBC glycophorin were compared using a novel molecular docking approach. *P. falciparum* Kolkata isolates experienced a recent population expansion as documented by negative value of Tajima's D, Fu & Li's statistics, unimodal mismatch distribution and star-like median-joining network for both *Pfeba* loci. Positive selection seemed to play a major role in shaping the diversity of *Pfeba-175* (d_N/d_S=2.45, and McDonald-Kreitman P-value= 0.04). *In silico* molecular docking demonstrated that high frequency haplotypes of *eba-175* were capable of engaging the parasite ligand into energetically more favorable interaction with GPA. A contrasting pattern of evolution and protein-protein interaction was noted for PfEBA-140. Together, this study provides a firm genetic support favoring a dominant role of EBA-175 in erythrocyte invasion and underscores the need to consider the observed F2-variants while designing vaccines that target EBA-175 parasite antigen.

Final Report of the work done on the Major Research Project

"Decoding the Genetics of Host-Parasite Interaction during Erythrocyte Invasion by *Plasmodium falciparum* in Human malaria".

Submitted to

UNIVERSITY GRANTS COMMISSION
BAHADUR SHAH ZAFAR MARG
NEW DELHI – 110 002.

By
Dr. Sanghamitra Sengupta
Associate Professor, Department of Biochemistry.
University of Calcutta.
35, Ballygunge Circular Road
Kolkata-700 019

UNIVERSITY GRANTS COMMISSION
BAHADUR SHAH ZAFAR MARG
NEW DELHI – 110 002.

Annual/Final Report of the work done on the Major Research Project

1.	Project report No. 1st /2nd /3rd/Final:	Final.
2.	UGC Reference No:	F.No.- 43-346/2014(SR); Dated: 09.09.2015
3.	Period of report:	01/07/2015 to 30/06/2018 (3 years).
4.	Title of research project	Decoding the Genetics of Host-Parasite Interaction during Erythrocyte Invasion by *Plasmodium falciparum* in Human malaria.
5.	Name of the Principal Investigator	Dr. Sanghamitra Sengupta
6.	Department	Department of Biochemistry.
7.	University/College where work has progressed	University of Calcutta.
8.	Effective date of starting of the project	01/07/2015 (The grant has been received by the University on 24.11.2015)
9.	Total amount approved	Rs. 14, 79, 600.00
10.	Total grant received	Rs. 12, 93,789.00
11.	Total expenditure	Rs.14, 23, 215 .00

c. Report of the work done: Attached in a separate sheet.

Report of the work done

TITLE OF THE PROJECT: Decoding the Genetics of Host-Parasite Interaction during Erythrocyte Invasion by *Plasmodium falciparum* in Human malaria.

Introduction:

Malaria is a devastating infectious disease particularly in tropical and sub-tropical countries belonging to Africa and South-East Asia. As per the World malaria report (November 2018), 219 and 217 million cases of malaria were registered in 2017 and 2016, respectively. The estimated number of malaria deaths stood at 435 000 in 2017, a similar number to the previous year. Pregnant women, infants and those over 60 years old are at greatest risk (27th March, 2019) [1].

Human malaria may be caused by five species. Of these, *Plasmodium falciparum* is the most prevalent and account for the majority of the global malaria cases [2]. *P. falciparum* undergoes a complex life cycle requiring both a human and a mosquito host. The disease is mostly caused by the asexual blood stage involving Erythrocyte invasion. This stage consists of multiple receptor–ligand interactions between merozoite and erythrocyte surface leading to attachment, apical reorientation, tight-junction formation, and finally entry of the parasite into the erythrocyte (Figure 1A). Manifestation of symptoms and disease burden are associated with erythrocyte invasion rendering this process a hotspot of malaria research. A variety of merozoite proteins interact with each other and the host receptors present on RBC for successful invasion (Figure 1B). Many of these such as Merozoite surface Protein (MSPs), Apical membrane antigen-1 (AMA-1), Circumsporozoite protein (CSP) and *Plasmodium falciparum* erythrocyte membrane protein -1 (PfEMP-1) are vaccine candidates and well-known for their genetic diversity, one of the challenges in malaria vaccine development. Several strategies for vaccine development continue to be hindered by a poor understanding of antimalarial immunity and the genetic diversity of malaria parasites [3]. Nevertheless, there is a paucity of genetic data for two major merozoite protein family, namely Duffy binding-like (DBL) family and reticulocyte binding-like (RBL) family which play central role in apical reorientation of merozoite and rhoptry discharge.

Recent successes in malaria control are threatened by drug-resistant *Plasmodium* parasites and insecticide-resistant Anopheles mosquitoes, and first generation vaccines offer only partial protection. Given this, our research attempted to connect the genetic findings from *Plasmodium* clinical isolates to the parasite infection biology and identified a greater biological importance of EBA-175 in the erythrocyte invasion, providing a firm justification for the molecule to be a therapeutic target.

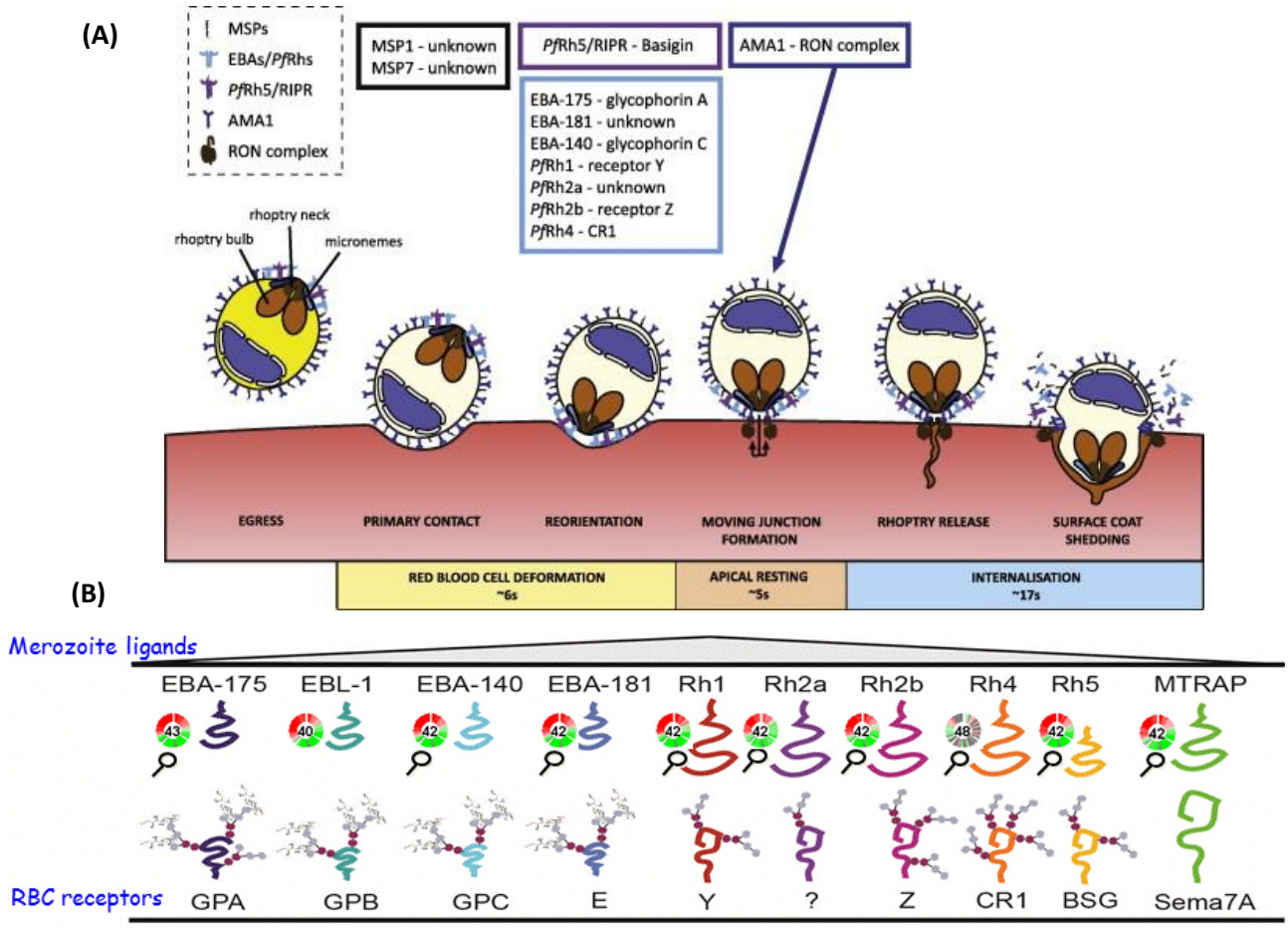

Figure 1: (A) Erythrocyte Invasion is a tightly regulated, multistep process involving many Ligands (Harvey et al; 2012). (B) Parasite invasion ligands and human erythrocyte receptors (http://mpmp.huji.ac.il/maps/ligandAndreceptor.html).

(i) OBJECTIVES OF THE PROJECT:

Since antigenic diversity in malaria-endemic population is a major barrier for effective vaccine development, this study focuses on exploring the impact of genetic diversity of the parasite ligands on the host-pathogen interaction. Of the several ligands we prioritized to sample sequence variations of EBA-175 and EBA-140. Our specific objectives as per the original proposal were

1. To study the genetic diversity of region II of genes encoding two parasite invasion ligands namely *EBA-175* and *EBA-140*.
2. To study the association of Glycophorin A gene polymorphisms with *P. falciparum* blood infection level.
3. To infer the host-parasite interaction (sequence-driven) in terms of *in silico* assessment of protein-protein interaction.

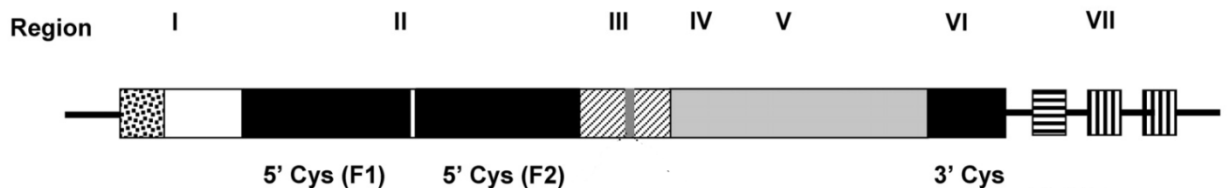

Figure 2: Overall structure of the *EBA-175 and EBA-140* gene that encodes region I-VII.

PfEBA-175 and PfEBA-140 are composed of a functionally important cysteine-rich region at their N-terminal, termed as region II (RII), which is composed of two related domains, F1 and F2, responsible for receptor binding [4,5]. EBA-175 interacts with host RBC receptor glycophorin A (GPA) through multiple glycans of GPA in a sialic acid dependent manner [6, 7, 8] whereas, DBL domain of EBA-140 interacts with glycophorin C (GPC) through its O-linked and N-linked sialylated glycans (Figure 2) [9,10,11]. To gain an understanding of host-parasite (*P. falciparum*) interactions given the genetic variability of parasite field isolates, we explored the sequence divergence of two parasite antigens, *Pfeba-175* and *Pfeba-140* that interact with glycophorins in host RBC during blood stage of malaria [12].

(ii) METHODOLOGY:

Studying the sequence diversity of *Pfeba-175 and Pfeba-140:*

Total genomic DNA was extracted from *P. falciparum* infected malaria patient's blood sample and for each isolate, F1 (*eba-175*: 825 bp; *eba-140*: 840 bp), F2 (*eba-175*: 852 bp; *eba-140*: 882 bp) and a small linker region (*eba-175*: 105 bp; *eba-140*: 72 bp) of RII segment of *eba-175* and *eba-140* genes were amplified using multiple overlapping primers those have been designed using *P. falciparum* reference sequence (PF3D7_0731500) retrieved from plasmoDB database (Table 1).

Table 1: Nucleotide sequences of oligonucleotide primers

Sl. No.	Gene Name	Primer IDs and Sequences
1	*Pfeba-175*	*eba-175* (FP1) : 5'-GGAAAGGAATGAAATGGGATTG-3'
2		*eba-175* (RP1) : 5'-ACGCTTTTAACGAGAGTAGTAG-3'
3		*eba-175* (FP2) : 5'-CGTTGATACAAACACAAAGGTGT-3'
4		*eba-175* (RP2) : 5'-TACTTCTGGACACATCGTAC-3'
5		*eba-175* (FP3) : 5'-GAATGGCATACGTTATCGAAAG-3'
6		*eba-175* (RP3) : 5'-CATTCCAATAATCAGTACCTCC-3'
7	*Pfeba-140*	*eba-140* (FP1) : 5'-GGATATGTCGAGAAAACAGTGAG-3'
8		*eba-140* (RP1) : 5'-ATACAAGTTCGTCAAAACCTCC-3'
9		*eba-140* (FP2) : 5'-TGTGATTGCAGATATACTGCTACT-3'
10		*eba-140* (RP2) : 5'-ACATTCTCCTTGTTTTCATCCCAC-3'
11		*eba-140* (FP3) : 5'-GTGGGATGAAAACAAGGAGAATGT-3'
12		*eba-140* (RP3) : 5'-AATACTTATTGGCGTTCCATCACC-3'
13		*eba-140* (FP4) : 5'-GGAGGTTTTGACGAACTTGTAT-3'
14		*eba-140* (RP4): 5'-AAAGTTTGTCTTCTGGGGGGTT-3'

Amplification was done using following protocol: 5 mins at 94°C, followed by 40 cycles of denaturation at 94°C for 30 s, annealing at 58°C for 45 s, elongation at 72°C for 1 min and a final extension at 72°C for 5 min. PCR amplicons were analyzed by agarose gel electrophoresis and purified by QIAquick Gel Extraction kit (Qiagen, Hilden, Germany). The purified amplicons were sequenced under the cycling program of initial denaturation at 94°C for 30s, 25 cycles of denaturation at 94°C for 10s, 5 s at 50°C and 4 min at 60°C, followed by holding at 4°C. Sequencing was performed in both directions, using the forward and reverse primers using Big Dye ver. 3.1 dye terminator technology and ran on ABI Prism 3500 Genetic Analyzer (Applied Biosystems, Foster City, CA). Forward and reverse sequence reads were stitched to obtain the entire F1 or F2 region. Sequence identity was validated using NCBI BLAST (http://blast.ncbi.nlm.nih.gov/). Reads showing variations from the 3D7 reference were resequenced to confirm the variants. Nucleotide sequences of F1 and F2 of *eba-175* and *eba-140* genes reported in this study were submitted to GenBank.

Sequence alignment and data analysis:

The sequences were aligned with the corresponding prototypic Pf3D7 reference sequence to identify nucleotide and the corresponding amino acid changes and the extent of genetic variability was also compared with 3D7 using MEGA6 tool (http://www.megasoftware.net/). Nucleotide mismatches and statistical analyses were carried out using DnaSP version 5.1 [13]. Variant sites with quality score >30 in both forward and reverse sequences were included in the analyses. The number of segregating sites (S), haplotypes (H), haplotype diversity (Hd), nucleotide diversity (π, θ) and the average number of pair-wise nucleotide differences, raggedness index (r), minimum number of recombination event (Rm) were estimated using the DnaSP 5.10. The phylogenetic and molecular evolutionary analyses were conducted using MEGA version 6.0. Tajima's D [14] and Fu & Li's [15] D & F statistics were applied to test the neutral theory of evolution. Nucleotide diversity (π) was estimated from total number of segregating sites and average number of mismatches between two sequences was calculated for all segregating sites as well as for parsimony informative sites. Positive values for Tajima's D correspond to positive selection whereas negative values indicate negative or purifying selection. The effect of natural selection was assessed by estimating the rate of the synonymous and nonsynonymous substitutions following Nei and Gojobori's [16] method with Jukes and Cantor correction using MEGA V 6.0 [17]. The rates of d_N and d_S were compared using the Z test ($P < 0.05$) in MEGA6. The ratio of non-synonymous to synonymous substitution (d_N/d_S) is widely used as an indicator of the action of natural selection in gene sequences. An excess of non-synonymous relative to synonymous mutation is a clear signal of positive selection whereas a lack of non-synonymous relative to synonymous polymorphisms suggests negative or purifying selection. Haplotype network was constructed using Median-Joining method in Network 5.0.0.1 [18] to comprehend the genetic and population level forces. McDonald-Kreitman (MK) test [19] was performed to compare inter- and intra-specific nucleotide changes in RII of *eba-175* and *eba-140* using the orthologus sequences from *Plasmodium reichenowi* (AJ251848; *eba-175* and

AJ438829, AY572433; *eba-140*) and *Plasmodium gaboni* (XM_018785148; eba-175). A comparative evaluation of different population genetic parameters based on RII data of *eba-175* and *eba-140* from different geographical regions was studied. Fisher's exact test was applied to test for significant departure from neutrality using DnaSP v5.10.

Genotyping of Glycophorin A:

Purified genomic DNAs from 50 randomly chosen *P. falciparum* infected patients samples were used as templates for amplification of a 1253 bp genomic region of GYPA covering exon 3 (96 bp) and exon 4 (39 bp). GYPA exon 3 and exon 4 specific amplification primers designed from GYPA (GC04M144109) reference sequence (Table 2). Each amplicon was purified and sequenced using Big Dye ver 3.1 and ran on ABI Prism 3500 Genetic Analyzer. Sequence identity was validated using NCBI BLAST and DNA sequence data were submitted in GenBank. On other hand, two SNPs (rs12645789 and rs6857303) were selected for examining the association of Glycophorin A with blood parasite level. Genotype assignment was carried out based on size discrimination of PCR products digested with *ScaI* (rs12645789) and *SpeI* (rs6857303) (Thermo Fisher scientific) at 37°C according to manufacturer's protocol. The primer pairs used in amplification reactions were listed in the Table 2. PCR was carried out using following protocol: denaturation at 94°C for 5 mins, followed by 40 cycles of denaturation at 94°C for 30 sec, annealing at 60°C for 20 sec, extension at 72°C for 30 sec, and then completed with a final extension at 72°C for 5 mins using a thermal cycler (Applied Biosystems® GeneAmp® PCR System 9700). The allele and genotype frequencies for GYPA SNPs were estimated by gene counting and Hardy–Weinberg equilibrium was evaluated by $\chi 2$ test using Haploview (http://www.broad.mit.edu/mpg/haploview/). Association between allele/genotype frequency and blood parasitemia was computed using SPSS version 10.0 and the result was considered "statistically significant" if P-value was <0.05.

Table 2: Nucleotide sequences of oligonucleotide primers:

Sl. No.	Gene Name	Primer IDs and Sequences
1	GYPA	GYPAex3fwd : 5'-GAGAGTTTGTCTTTCATAATACGCT-3';
2		GYPAex3rev : 5'-TACATCATGAAGCTCTCTTACTACC-3'
3		GYPAex4fwd : 5'-GGTAGTAAGAGAGCTTCATGATGTA-3'
4		GYPAex4rev : 5'-ACATATGCTCTTCTAAGATAGACAC-3'
5		GYPA_FP_ rs12645789 : 5'TCCTGGTCAACATGATGCAGAG-3'
6		GYPA_RP_ rs12645789:5'-AGGAGAATCTCACTATGCTGCAG-3'
7		GYPA_FP_ rs6857303: 5'-CAAAATCTCTGCAAAGGGGATT-3'
8		GYPA_RP_ rs6857303: 5'-AGGAGAATCTCACTATGCTGCAG-3'

Ligand –receptor interaction study and homology modeling:

Construction of an appropriate interaction model between EBA-175 and Glycophorin A receptor was based on the seminal work of Tolia et al. [20] who proposed that crystallographic structure of RII (PDB ID: 1ZRO) reveals the interaction is mediated through two symmetrical RII molecules (chain A and B) positioned in an antiparallel fashion and binding sites (o-linked glycans) of the sialylated receptor glycophorin A were located on two channels created by dimeric organization of RII. These channels were created using 340-554 and 28-33 amino acid residues of EBA-175 (PF3D7_0731500.1) and PfEBA-175 engages with GPA through five O-glycans (32, 36, 43, 46, 49 residues) present in a 37 amino acids stretch coded by exon 3 (24–57 residues) and part of exon 4 (58–60 residues) [21,22]. Although five crystal structures of GPA (1AFO, 5EH4, 5EH6, 2KPE, 2KPF) are available in RCSB database (https://www.rcsb.org/), none of them include extracellular domain responsible for interaction with RII. Since 37 residues long peptide of GPA assumes a random coil structure as predicted by GOR4 tool (Figure 6A), multiple conformers could be generated using the extracellular domain of the protein. To reduce the number of possible conformers, certain criteria were imposed while docking the host receptor with parasite dimers. First, human GPA binds to PfEBA-175 through the glycosylated side-chains and the potential glycan binding residues in the ligand were 340–343, 415, 417, 422, 429, 439, 442, 542, 546, 550–554 of chain A and 28–33 of chain B. Glycosylated residues at positions 43, 46, 49 on GYPA were considered to be essential for forming the interaction complex (Salinas et al., 2014). To initiate the model building, backbone atoms of 37 residues coiled peptide was retrieved from RCSB protein databank randomly and receptor file was constructed by sequentially incorporating appropriate residues. The coordinates of the side chains of each residue were generated using AutoDock (autodock.scripps.edu/). Initially, a set of 30 conformers of the GPA extra cellular part were created using Accelrys-DS viewer 2.0 (accelrys.com) which approximated two feasible interaction models as proposed by Tolia et al. [20]. After docking the receptor onto the ligand using Accelrys-DS viewer 2.0, further refinement of the complex was brought about by rotating selected residues about φ and Ψ angles using viewerLite (https://windowsfileviewer.com) and DS-viewer 2.0. Torsional strain in the resulting models was minimized using Accelry-DS-viewer 2.0 through the process of energy minimization until the energy between two successive cycles is less that 50 kcal/mol. Three best models describing GYPA- PfEBA-175 complex satisfied all these conformation adjustment rules. The suitability of each models was examined by estimating residue-wise interaction energy ($E^{Interaction}=E^{electrostatic}+E^{van\ der\ Waals}$) based on CHARMM parameters that considered electrostatic and van der Waals interactions within a distance cutoff of 10.0 Å between the atoms of the receptor and the ligand. The $E^{interaction}$ estimates for interaction between GYPA and F2KH1/F2KH3 were also computed following the above pipeline. To build the model between PfEBA-140 and GPC, crystallographic structure of EBA-140 monomer (PDB ID: 4JNO) was retrieved from RCSB Protein Data Bank. Due to unavailability of crystal structure of GPC and any hypothesis mimicking PfEBA-140-GPC interaction, here the complex was build up based on PfEBA-175 and GPA interaction model. From the 3D model of PfEBA-175-GPA already

constructed, the side-chains of the EBA-175 residues were replaced using target residues from EBA-140 followed by addition of side chains using AutoDock. In a similar fashion the sequences of the GPA and GPC were superimposed to achieve maximum alignment of the glycans. The residues of the GPA were then modified using AutoDock to resemble that of GPC. Following the same principle, complex between mutant EBA-140 and GPC was generated. To study the effect of *Pfeba-175* and *Pfeba-140* variants prevalent in Kolkata parasite population on host parasite interaction, 3D structures of complex between GPA and EBA-175 and GPC and EBA-140 were generated following the structurally and energetically most favorable model. AutoDock and Accelrys-DS-viewer 2.0 were employed to generate mutated side chains along with structural refinement by successive energy minimization. $E^{Interaction}$ estimates for two and one predominant haplotypes of *Pfeba-175* and *Pfeba-140* respectively were computed following the above protocol. The more negative is the interaction energy the stronger is the binding strength between the two molecules. For homology modeling, EBA-175 (PDB ID: 1ZRL: Chain A) protein file of *P. falciparum* strain 3D7 was extracted from RCSB protein data bank. PDB files were generated for two prevalent haplotypes (F2KH1 and F2KH3) originated from F2 domain sequences of EBA-175 in Kolkata parasite population by incorporating mutant amino acids in protein chain using SwissModel Workspace. Three dimensional structures of mutant EBA-175 polypeptides having variant amino acids in F2 domain were generated using Swiss pdb-Viewer software and compared with respect to the wild type *P. falciparum* 3D7 EBA-175 protein structure. The root mean square deviation (RMSD) and surface charge distribution pattern of the wild type and variant amino acids residue in two haplotypes were estimated by the SPDB-viewer and GRASP2 software.

(iii) WORK DONE

Sequence diversity and population structure of region II and F2 domain of *eba-175* and *eba-140* genes:

An assessment of 48 parasite sequences representing region II of *Pfeba-175* and 42 sequences representing RII of *Pfeba-140*, identified 49 and 18 variant sites respectively among Kolkata clinical isolates. Of these, the numbers of parsimony informative sites were turned out to be 27 for *Pfeba-175* and 9 for *Pfeba-140* loci. Estimates of Watterson's θ and nucleotide diversity (π) were 0.0071 ± 0.001 and 0.0048 ± 0.0004 respectively for *eba-175* loci. The abundance of low frequency variants gave rise to negative values of Tajima's D (−1.116) and statistically significant values of Fu & Li's D* (−2.56) and F* (−2.466) (Table 3). The distribution of pairwise nucleotide differences was plotted to yield a unimodal pattern with a raggedness index of 0.0053 (Figure 3). Both recombination (Rm = 7) and mutation were responsible for generating high haplotype diversity (0.983 ± 0.01) in the RII of *Pfeba-175*. The median joining network connecting the parsimony informative haplotypes generated a cobweb phylogeny, characteristic of population expansion (Figure 3). Comparison of 42 sequences representing RII of *Pfeba-140* revealed that average nucleotide diversity indices were 0.0024 ± 0.0006 (θ) and 0.00126 ± 0.0002 (π) respectively. Tajima's D, Fu and Li′s D* and F* were −1.569, −1.79 and −2.032

respectively (Table 3). Pairwise mismatch distribution pattern was unimodal and median-joining network assumed a star-like architecture (Figure 3). The ratio of nonsynonymous substitutions over number of nonsynonymous sites (d_N= 0.0054 ± 0.0011 for *eba-175* and 0.0014 ± 0.0005 for *eba-140*) was higher than synonymous substitutions over number of synonymous sites (d_S = 0.0022 ± 0.0008 for *eba-175* and 0.0008 ± 0.0004 for *eba- 140*) for both loci. This, in effect, resulted in statistically significant higher value of d_N/d_S ratio of 2.45 (Z-test: $P < 0.05$) for *eba-175* locus only indicating the possibility that positive selection shaped the pattern of *eba-175* diversity (Table 3). As an additional support, MK-test was applied. Due to an excess of intra-specific nonsynonymous polymorphic changes (49 nonsynonymous and 10 synonymous amino acid changes) relative to nonsynonymous divergence between *P. falciparum* and *P. reichenowi* eba-175 sequences (77 nonsynonymous and 36 synonymous amino acid changes), a significant departure from neutrality was observed ($P = 0.04$). In contrast, the comparison of the level of polymorphic (14 nonsynonymous and 4 synonymous amino acid changes) and fixed differences (46 nonsynonymous and 15 synonymous amino acid changes) recorded in RII of *Pfeba-140* did not vary significantly (Table 4). Comparative evaluation of different population genetic parameters based on RII data of *eba-175* and *eba-140* from different geographical regions and this study revealed distinguishing pattern of parasite evolution [23]. Although haplotype and nucleotide diversities (Hd and π) of both loci displayed similar levels variation across regions, the nucleotide diversity estimates of *eba-175* were higher than those of *eba-140*. A striking difference in the distribution of Tajima's D and Fu & Li′s D* statistics across countries was noted (Figure 4). F2 region (852 bp) of *Pfeba-175* was thus sequenced in a greater depth (n = 92). A total of 18 polymorphic sites were detected, of which 16 polymorphic sites corresponded to nonsynonymous amino acid changes and one site was found to be triallelic (A/G/C). The present study identified 15 haplotypes with a haplotype diversity of 0.681 ± 0.047. Nucleotide diversity θ and π were 0.0059 ± 0.002 and 0.0028 ± 0.0003 respectively with Tajima's D estimate being −1.52 (Table 3). The mismatch distribution pattern of F2 displayed a unimodal characteristic with raggedness index of 0.0103 (Figure 5), reminiscent of a recent population expansion with overabundance of few advantageous alleles as revealed by a significant rise of d_N/d_S ratio (5.5) (Table 3, Figure 5). An inspection of median-joining network formed among the Kolkata haplotypes using F2 variants (n= 15) displayed two clusters centered by F2KH1 (0.556) and F2KH3 (0.163) from which emanated several low frequency haplotypes (Figure 5). These two haplotypes were derived from Pf3D7 sequence through accumulation of 2 (at positions 584th and 664th in F2KH3) and 3 (at positions 584th, 592th and 664th in F2KH1) mutations with frequencies 0.315 (Q584K), 0.956(R664S) and 0.641 (Q584E), 0.641 (E592A), and 0.956 (R664S) respectively. In contrast to the above observation, median joining network of *Pfeba-140* haplotypes yielded star like phylogeny where the central and most predominant node represented Pf3D7 prototype. The pairwise mismatch distribution plot of *Pfeba-140* sequences followed that expected under conditions of neutral evolution (Figure 5). This contrasting pattern of population structures at *Pfeba-175* and *Pfeba-140* genes in the Kolkata isolates motivate us to probe the possible biological underpinning of higher prevalence of F2KH1 and F2KH3 haplotypes in Kolkata *P. falciparum* samples using a novel rationale based manual docking

experiment that compare the efficacies of interaction between EBA proteins (Pf3D7 and mutant ligands) and host receptors.

Genetic epidemiology: To investigate if the frequency distribution of alleles encoding GYPA protein harbored any differences between the *P. falciparum* infected individuals with high and low blood parasite counts, RFLP based genetic association was carried out in 150 patient samples using rs12645789 & rs6857303. Our study did not find any significant differences of allele or genotype frequencies between the groups with low and high parasite indices (Table 5). Besides, GYPA exon 3 and part of exon 4 did not harbor any variants in the P. falciparum infected patients under study compared to the reference sequence (NCBI GenBank accession no: GYPA: MF447757-MF447806). Therefore, models evaluating protein-protein interactions between host and parasite was constructed using GYPA reference sequence (UniProtKB: P02724) retrieved from UniProt (http://www.uniprot.org).

Table 3: Genetic variation in region II and F2 domain of *P. falciparum eba-175* and *eba-140* genes.

Gene name	***Pfeba-175***		***Pfeba-140***	
Domain	Region II	F2	Region II	F2
No of samples (n)	48	92	42	42
Length sequenced (bp)	1560	594	1716	882
Polymorphic sites	49	18	18	9
Singleton variable sites	22	11	9	6
Parsimony informative sites	27	7	9	3
No of haplotypes (H)	37	15	20	9
Haplotype diversity (Hd±SD)	0.983 ±0.01	0.681 ±0.047	0.898 ±0.033	0.542 ±0.087
Average number of pairwise differences (K)	7.486	1.668	2.157	0.787
Raggedness index	0.0053	0.0103	0.0497	0.0803
Recombination event (R_m)	7	-	3	1
Nucleotide diversity (π±SD)	0.0048 ±0.0004	0.0028 ±0.0003	0.00126 ±0.0002	0.00089 ±0.00021
Watterson's θ±SD	0.0071± 0.001	0.0059± 0.002	0.0024± 0.0006	0.0024± 0.00079
Tajima's D	-1.116	-1.52	-1.569	-1.80449*
Fu & Li's *D**	-2.56*	-3.392*	-1.79	-2.44491
Fu & Li's *F**	-2.466*	-3.301*	-2.032	-2.63191*
Average number of non-synonymous sites, d_N (SE)	0.0054 ±0.0011	0.0033 ±0.002	0.0014 ±0.0005	0.0009 ±0.0004
Average number of synonymous sites, d_S (SE)	0.0022 ±0.0008	0.0006 ±0.00006	0.0008 ±0.0004	0.0011 ±0.0006
d_N/d_S	2.45*	5.5*	1.75	0.82

** P-value significant (p<0.05).*

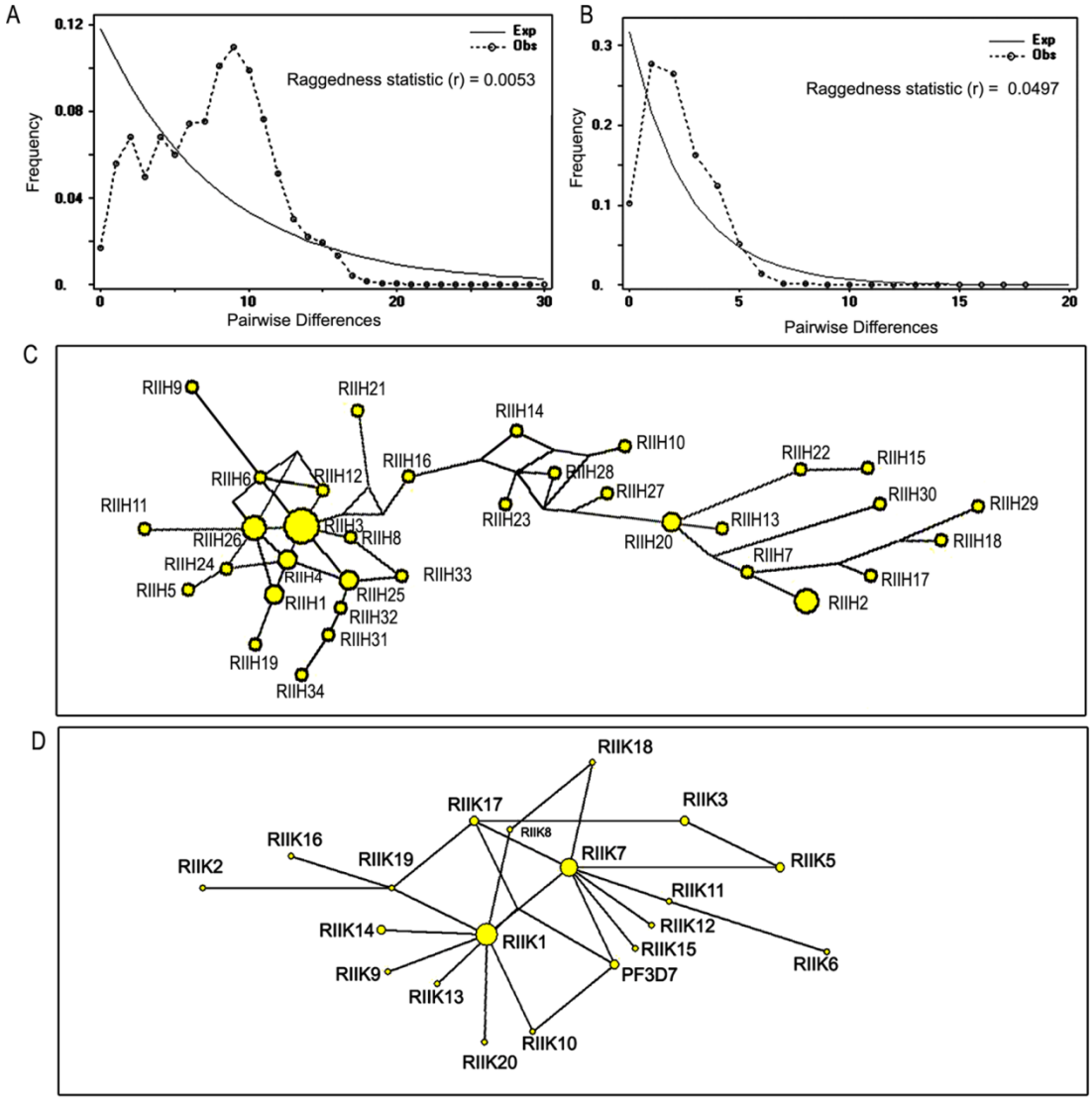

Figure 3: Mismatch distribution pattern and median-joining network of RII in Kolkata clinical isolates. (A and B) The pattern of pairwise nucleotide differences in the Pfeba-175 and Pfeba-140 genes respectively. The black dotted line indicates the observed distribution and the solid grey colored line indicates mismatch distribution pattern under neutral expectation. (C and D) Network analyses depicting the genetic relation of Pfeba-175 and Pfeba-140 haplotypes respectively. Each yellow circle represents a haplotype and the size of the circle is proportional to its frequency, a line joining haplotypes is proportional to distance between the nodes.

Table 4: McDonald-Kreitman test for selective neutrality for RII of Pfeba-175 and Pfeba-140 genes.

Gene region	Pair of parasite species Under study	Nucleotide differences				Fisher's exact test
		Inter-specific fixed		Intra-specific polymorphic		
		Syn	Nonsyn	Syn	Nonsyn	P-value
RII	*Pfeba-175 & Preba-175*	36	77	10	49	0.04*
	Pfeba-175 & Pgeba-175	63	165	10	49	0.09
	Pfeba-140 & Preba-140	15	46	4	14	1.00

Pf, P. falciparum; Pr, P. reichenowi; Pg, P.gaboni. Syn, synonymous; Nsyn, nonsynonymous substitution.

** indicates P value (two tailed) significant (<0.05).*

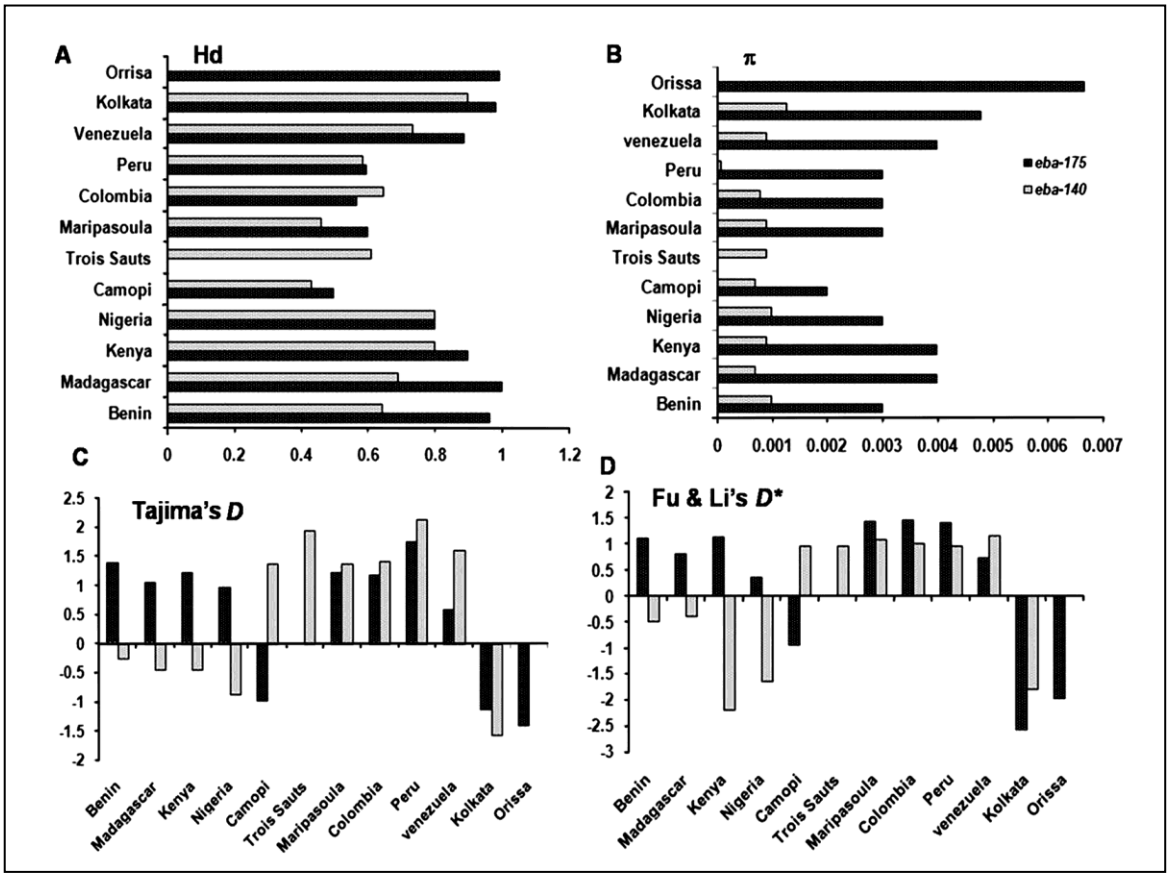

Figure 4: Comparison of parasite genetic parameters from regions belonging to Africa, South America and India. (A) Haplotype diversity (Hd) (B) Average nucleotide diversity (□) (C) Tajima's D index (D) Fu & Li's D* index. Values are estimated for two genes, eba-175 (black bars) and eba-140 (grey bars).

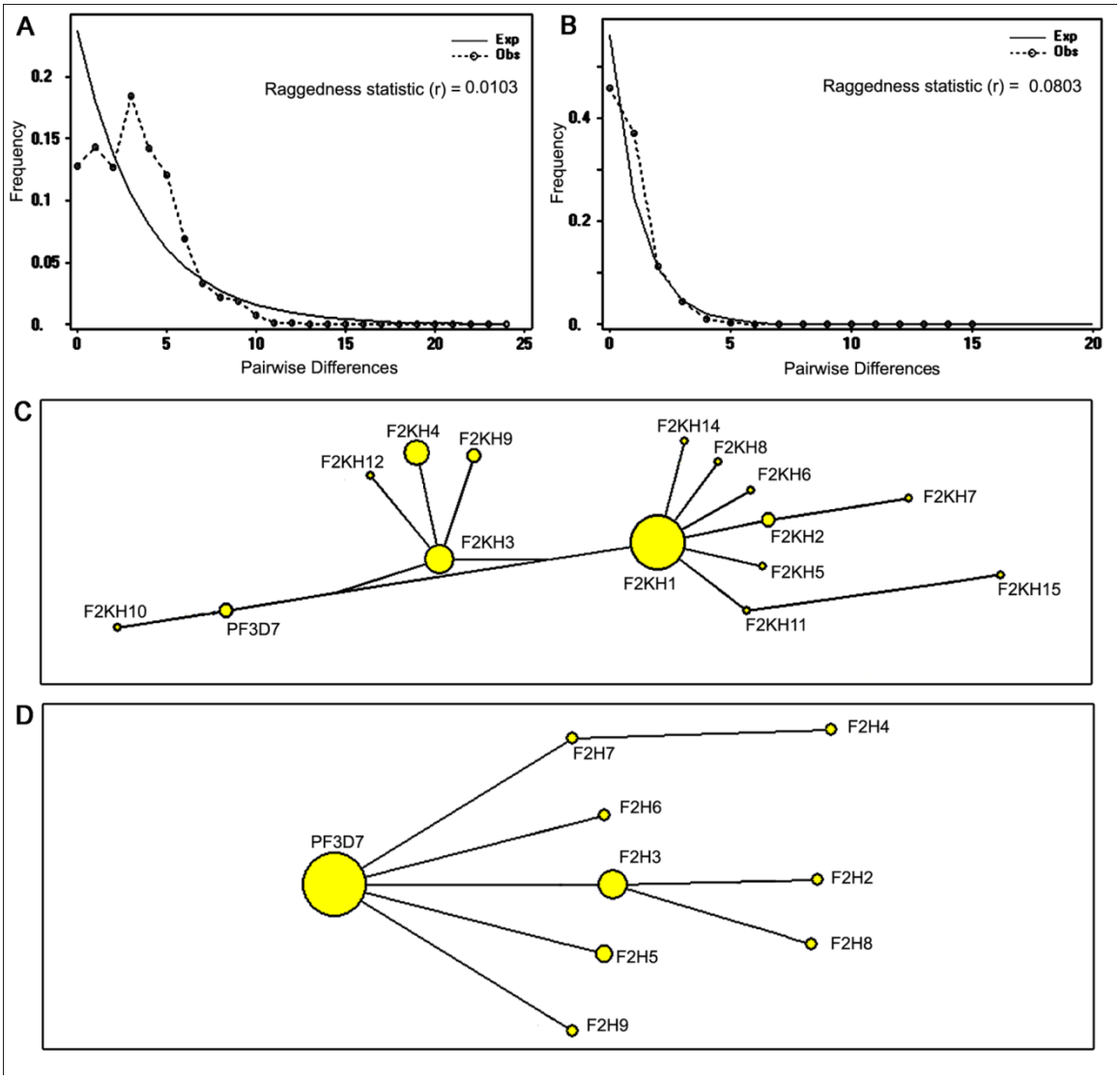

Figure 5: Mismatch distribution pattern and median-joining network based on F2 region of genes encoding erythrocyte binding proteins in Kolkata parasite population. (A and B) The pattern of pairwise nucleotide differences in the Pfeba-175 and Pfeba-140 genes respectively. The black dotted line indicates the observed distribution and the solid grey line indicates mismatch distribution pattern expected under neutral theory. (C and D) Genetic relation of Pfeba-175 and Pfeba-140 haplotypes respectively shown in the form of network. Both median-joining networks give rise to star-like appearance. Yellow circles represent a haplotype and the size of the circle is proportional to its frequency, a line joining haplotypes is proportional to connection limits between the nodes.

Table 5: Mapping of interaction between host receptor and parasite ligand:

SNPs ID	Genotype frequency		Allele frequency	
	High/Low Parasitemia	χ2 (P value)	High/Low Parasitemia	χ2 (P value)
rs12645789 G>A	0.81/0.17/0.02	0.996(0.608)	0.89/0.11	0.341(0.559)
	0.75/0.23/0.02		0.87/0.13	
rs6857303 C>T	0.24/0.67/0.09	7.23(0.027)*	0.57/0.43	3.14(0.076)
	0.45/0.44/0.11		0.68/0.32	

Possible interaction models between relevant part of EBA-175 (3D7) and glycophorin protein were constructed based on assumptions discussed in the Methodology section. Three models satisfied the criteria imposed by this study. In both model 1 and model 2, extracellular part of GPA traverses through the channels of EBA-175 dimer (Figure 6; A, B) However in case of model 1 the receptor assumed an upward turn while in model 2 the receptor followed a downward direction once it emerged from the cavity. On the contrary, GPA wrapped EBA-175 dimer externally in model 3 (Figure 6; C). Among these three, model 1 was selected for further study for following reasons. First, in model 1, in addition to three residues (at positions 43, 46 and 49) two other glycosylated moieties (at positions 32 and 36) of GPA showed favorable interactions with the ligand (Table 6). Second, Interaction energy between receptor and Pf3D7-F2 polypeptides was found to be more favorable for model 1 (-304.74 kcal/mol) which was remarkably lower from that of model 2 (−214.23 kcal/mol) and model 3 (−168.83 kcal/mol). Next, in the model 1, Pf3D7-F2 was replaced by F2KH1 and F2KH3 mutant peptides and corresponding interaction energies were imputed. Mutant peptides F2KH1 and F2KH3 yielded $E^{Interaction}$ estimates of −351.45 kcal/mol and −346.19 kcal/mol respectively, which were much lower than $E^{Interaction}$ (−304.74 kcal/mol) between Pf3D7 and GPA (Table 7, Figure 7B). The surface electrostatic potential and root mean square deviation based on c-α (0.43A°) and backbone atoms (0.49A°) of the mutants were found to differ from that of Pf3D7 sequence (Figure 7C). On the contrary, interaction energies of GPC with PfEBA-140 (−528 kcal/mol) and GPC with mutant ligand (−546 kcal/mol), harboring amino acid changes at positions N239S & K261T, did not differ significantly (Figure 8).

A closer look in the Wireframe interaction models showed that two mutants (at positions Q584E/K and E592A) that resulted in difference of electrostatic distribution in the F2KH1 and F2KH3 polypeptides were located in the inner wall of the EBA-channel (Figure 7A). Thus, it may be assumed that the nonsynonymous amino acid changes in the Pfeba-175 gene imparted selective advantage for PfEBA175-GPA interaction during erythrocyte invasion. Altogether employing structural modeling with genetic data our study for the first time demonstrates the importance of two EBA-variants in mediating efficient interaction between EBA-175 and glycophorin A.

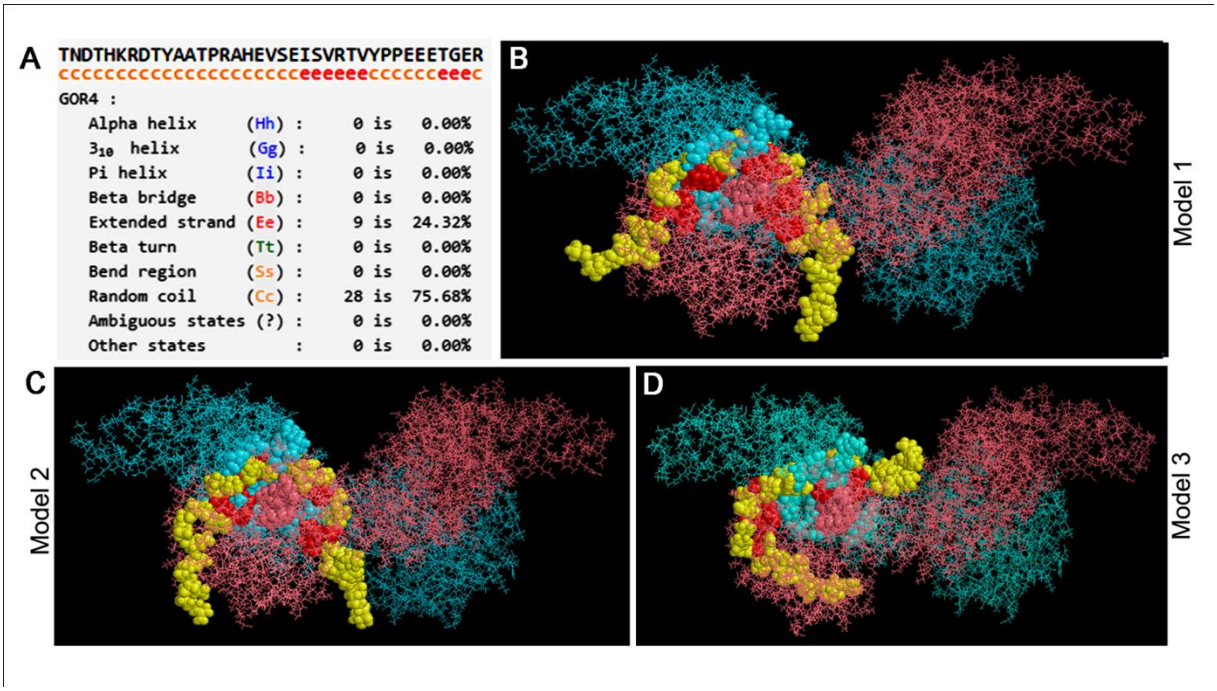

Figure 6: Prediction of secondary structure of the interacting segment of GPA and probable models depicting between PfEBA-175 (RII) and GPA. (A) The extracellular domain of GPA receptor is predicted to form a coiled structure as suggested by secondary structure prediction tool, GOR 4. (B) In model 1, the extracellular domain of GPA traverses through the cavity of EBA-175 dimer and takes a upward turn. (C) In model 2, after passing through the cavity of dimeric ligand, GPA receptor follows a downward direction. (D) In the model 3, GPA is shown to wrap the ligand from outside without entering through the cavity. Monomers of EBA-175, RII are shown in pink and cyan. GPA is shown in yellow space-filled residues and five glycosylated (O-glycans) residues are marked with red color.

Table 6

GPA residues		Estimated interaction Energy ($E^{interaction}$) in Kcal/mol								
		Model 1			Model 2			Model 3		
#	ID	Van der Waals	Electro static	Total	Van der Waals	Electro static	Total	Van der Waals	Electro static	Total
24	T	0	0.17	0.17	0.00	0.00	0.00	-1.36	-0.39	-1.75
25	N	-0.03	-0.36	-0.39	0.00	0.00	0.00	-2.07	-0.74	-2.81
26	D	-0.03	-2.21	-2.24	0.00	-0.12	-0.12	-0.23	6.09	5.86
27	T	-0.50	-0.40	-0.90	-0.02	-0.03	-0.05	-0.42	0.83	0.41
28	H	-0.51	8.49	7.98	0.00	1.31	1.31	-5.14	2.81	-2.33
29	K	-2.31	-34.04	-36.35	-0.05	4.17	4.12	-1.97	-6.46	-8.43
30	R	-5.84	-32.51	-38.35	-0.37	3.37	3.00	-2.28	-42.07	-44.35
31	D	-7.17	-36.55	-43.72	-0.98	-10.47	-11.45	-2.71	17.90	15.19
32	T	-11.41	-20.88	-32.29	-4.36	-5.17	-9.53	-16.10	21.53	5.43
33	Y	-5.49	6.35	0.86	-6.43	2.72	-3.71	-3.74	7.71	3.97
34	A	-2.30	3.88	1.58	-1.61	0.03	-1.58	-2.62	-1.77	-4.39
35	A	-1.33	0.14	-1.19	-1.66	2.73	1.07	-3.15	13.14	9.99
36	T	-4.14	-4.38	-8.52	-4.27	-28.54	-32.81	-9.84	4.09	-5.75
37	P	-2.40	3.69	1.29	-0.11	-0.64	-0.75	-1.14	-3.05	-4.19
38	R	-2.04	-9.21	-11.25	-0.55	11.18	10.63	-5.10	-20.49	-25.59
39	A	-1.64	-4.69	-6.33	-1.51	-1.42	-2.93	-0.15	-2.22	-2.37
40	H	-4.72	-0.91	-5.63	-1.70	-19.36	-21.06	-3.38	2.20	-1.18
41	E	-4.77	1.29	-3.48	-2.44	-80.32	-82.76	-2.33	-20.82	-23.15
42	V	-2.75	-2.36	-5.11	-1.94	-6.46	-8.40	-2.35	-3.17	-5.52
43	S	-7.32	-23.19	-30.51	-5.51	8.12	2.61	-5.62	3.13	-2.49
44	E	-1.32	-50.83	-52.15	-2.29	-0.18	-2.47	-2.05	-3.71	-5.76
45	I	-1.12	-12.83	-13.95	-1.02	-2.63	-3.65	-1.07	-4.07	-5.14
46	S	-10.88	6.28	-4.60	-6.89	28.49	21.6	-6.70	21.04	14.34
47	V	-1.36	0.66	-0.7	-1.52	-1.49	-3.01	-1.47	-1.54	-3.01
48	R	-2.09	22.32	20.23	-2.15	16.63	14.48	-2.16	12.67	10.51
49	T	-5.32	-10.64	-15.96	-9.89	13.79	3.90	-9.76	9.40	-0.36
50	V	-0.16	1.79	1.63	-2.21	4.20	1.99	-2.20	3.18	0.98
51	Y	-1.61	4.66	3.05	-5.35	-9.83	-15.18	-5.28	-3.92	-9.20
52	P	-1.31	-3.16	-4.47	-1.90	-5.23	-7.13	-2.25	-3.88	-6.13
53	P	-0.29	-0.79	-1.08	-3.98	-3.00	-6.98	-4.32	-0.10	-4.42
54	E	-2.10	-20.69	-22.79	-2.80	-17.10	-19.90	-2.79	-10.56	-13.35
55	E	-0.87	1.90	1.03	-6.01	-20.48	-26.49	-5.72	-24.84	-30.56
56	E	-0.09	-3.56	-3.65	-5.47	1.09	-4.38	-5.33	-1.31	-6.64
57	T	-0.03	2.13	2.10	-4.03	-8.33	-12.36	-3.85	-9.58	-13.43
58	G	0.00	0.71	0.71	-1.24	-5.78	-7.02	-1.24	-6.34	-7.58
59	E	0.00	0.24	0.24	-0.40	-2.06	-2.46	-0.39	1.02	0.63
60	R	0.00	0.00	0.00	-0.08	7.32	7.24	-0.08	3.82	3.74
$E^{interaction}$				-304.74 kcal/mol			-214.23 kcal/mol			-168.83 kcal/mol

Table 7

GPA amino acid residues		Estimated interaction Energy ($E^{interaction}$) in Kcal/mol					
		F2KH1			F2KH3		
#	Name	Van der Waals	Electrostatic	Total	Van der Waals	Electrostatic	Total
24	T	0.00	-0.72	-0.72	0.00	-0.71	-0.71
25	N	-0.03	-0.42	-0.45	-0.03	-1.13	-1.16
26	D	-0.03	-2.93	-2.96	-0.03	-2.81	-2.84
27	T	-0.50	-1.11	-1.61	-0.50	-2.07	-2.57
28	H	-0.51	8.88	8.37	-0.51	9.20	8.69
29	K	-2.31	-33.46	-35.77	-2.31	-33.24	-35.55
30	R	-5.83	-30.74	-36.57	-5.83	-32.21	-38.04
31	D	-7.13	-37.75	-44.88	-7.15	-38.15	-45.30
32	T	-11.52	-23.83	-35.35	-11.48	-22.96	-34.44
33	Y	-5.44	6.61	1.17	-5.43	6.04	0.61
34	A	-2.28	-2.39	-4.67	-2.31	-0.26	-2.57
35	A	-1.31	-0.09	-1.40	-1.32	-0.20	-1.52
36	T	-4.05	-6.05	-10.10	-4.09	-6.61	-10.7
37	P	-2.45	2.97	0.52	-2.45	3.94	1.49
38	R	-1.99	-5.01	-7.00	-2.01	-7.38	-9.39
39	A	-1.64	-4.96	-6.60	-1.65	-4.61	-6.26
40	H	-5.02	8.12	3.10	-5.01	6.46	1.45
41	E	-4.74	4.70	-0.04	-4.75	3.25	-1.5
42	V	-2.75	-3.18	-5.93	-2.74	-2.19	-4.93
43	S	-7.32	-18.70	-26.02	-7.32	-18.10	-25.42
44	E	-1.30	-51.43	-52.73	-1.33	-53.18	-54.51
45	I	-1.13	-7.84	-8.97	-1.13	-7.41	-8.54
46	S	-10.63	11.99	1.36	-10.71	7.97	-2.74
47	V	-1.63	-0.53	-2.16	-1.72	3.05	1.33
48	R	-0.68	-41.04	-41.72	-2.31	5.76	3.45
49	T	-4.95	-2.80	-7.75	-5.09	-32.55	-37.64
50	V	-0.15	0.06	-0.09	-0.14	2.19	2.05
51	Y	-1.45	4.27	2.82	-1.39	5.08	3.69
52	P	-1.24	-4.85	-6.09	-1.04	-4.65	-5.69
53	P	-0.29	-0.13	-0.42	-0.28	-1.36	-1.64
54	E	-2.12	-24.73	-26.85	-2.07	-22.73	-24.80
55	E	-0.81	-0.65	-1.46	-0.83	-6.30	-7.13
56	E	-0.09	-2.23	-2.32	-0.09	-5.69	-5.78
57	T	-0.03	0.91	0.88	-0.03	2.03	2.00
58	G	0.00	0.72	0.72	0.00	0.48	0.48
59	E	0.00	0.24	0.24	0.00	-0.06	-0.06
60	R	0.00	0.00	0.00	0.00	0.00	0.00
$E^{interaction}$				-351.45 kcal/mol			-346.19 kcal/mol

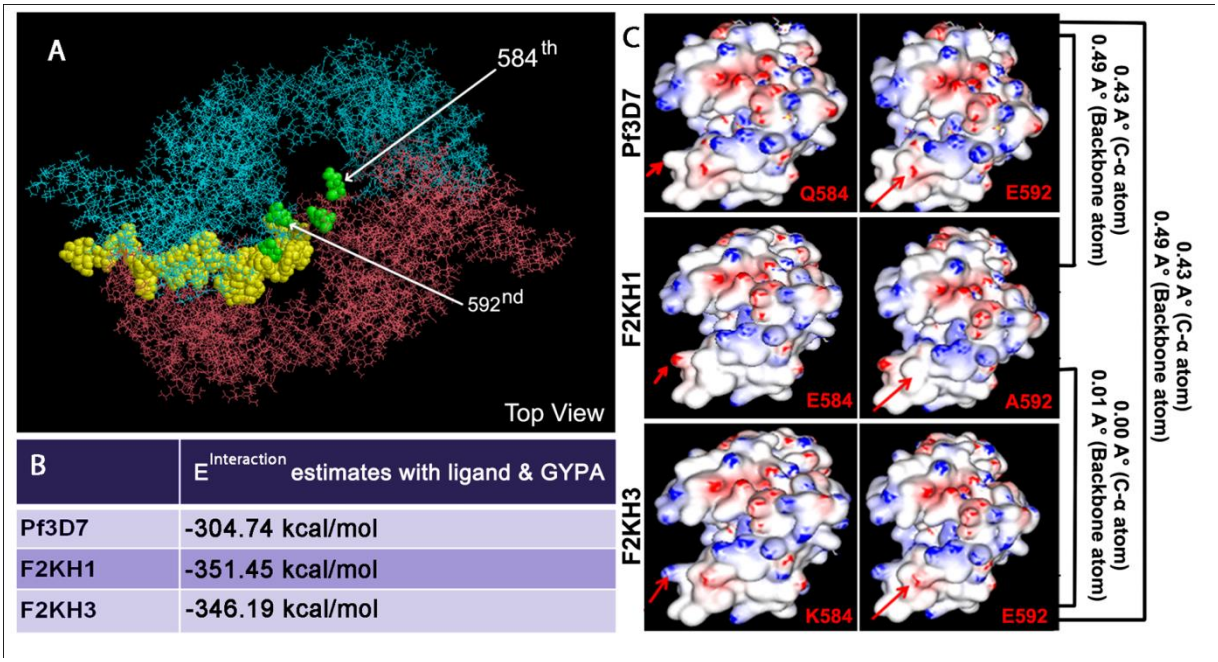

Figure 7: . Predicted interaction model between parasite ligand and host receptor, their estimated EInteraction and electrostatic redistribution as per homology modeling considering the prevalent mutations in the EBA-175 molecule. (A) Top view of receptor-ligand complex showing that the mutant residues at position 584 (green in color), 592 (green in color) are situated at the inner wall of the cavity formed by EBA-175 dimer. (B) Comparison of the computed interaction energies ($E^{Interaction}$) between the reference polypeptide of receptor and parasite ligands encoded by two prevalent haplotypes with respect

to Pf3D7. (C) Homology modeling and surface charge distribution pattern of wild-type and variant PfEBA-175 polypeptides.

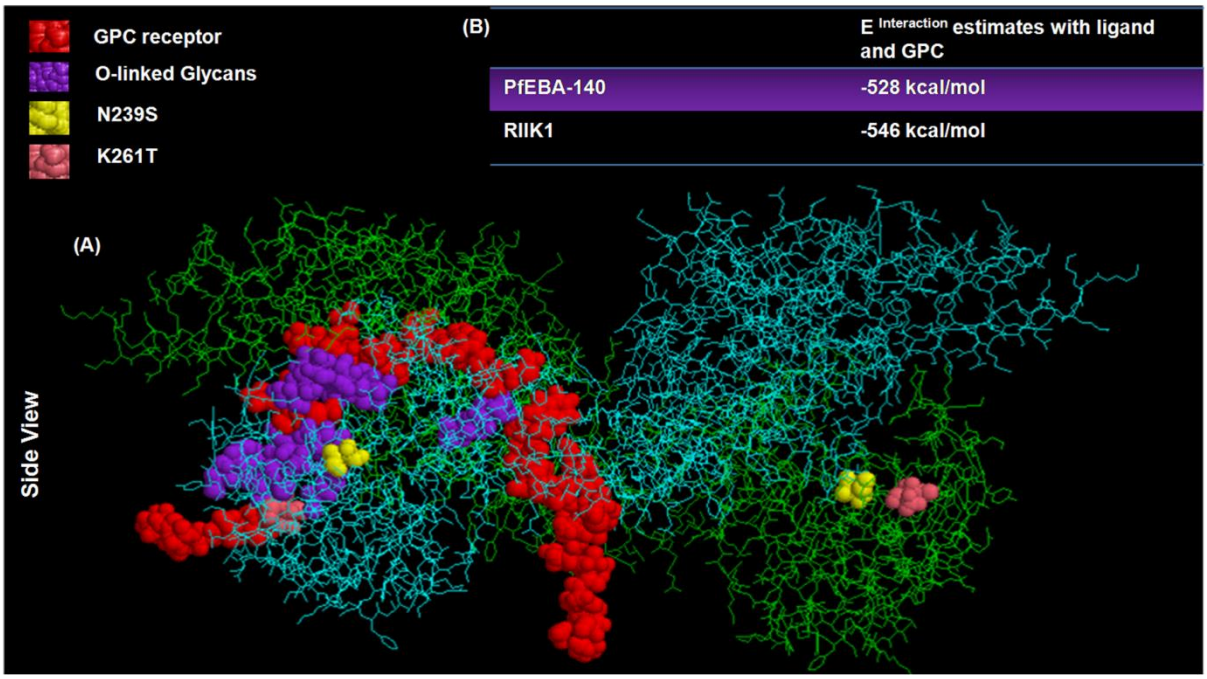

Figure 8: Interaction model between parasite ligand (EBA-140) and host receptor (GPC), their estimated interaction energies (EInteraction) between wild type and mutant ligand with GPC receptor. (A) Side view of receptor-ligand complex showing GPC in red space-filled residues and glycosylated (O-glycans) residues are marked with purple color. The mutant residues at position 239 (yellow in color), 261 (pink in color) are situated in the close proximity of the receptor (B) Comparison of the computed interaction energies (EInteraction) between the reference polypeptide of receptor and parasite ligands encoded by one prevalent haplotype with respect to Pf3D7.

Significance

Fluctuation of parasite genetic diversity reported from different geographical regions is believed to be one of the major obstacles in developing a successful vaccine against *P. falciparum* mediated human malaria. This population-based analysis of sequence diversity and establishment of a logical model mimicking host-parasite interaction will provide clues to predict the functional repercussion of parasite mutations in RBC invasion. The knowledge derived from this study is anticipated to be utilized for rational and efficacious vaccine development.

References :

1. WHO World Malaria Report, 2019 (http://www.who.int/malaria/publications/world-malaria-report-2015/report/en/).

2. Dayananda KK, Achur RN, Gowda DC. Epidemiology, drug resistance, and pathophysiology of Plasmodium vivax malaria (2018). J Vector Borne Dis. Jan-Mar;55(1):1-8.

3. Reiman JM, Kumar S, Rodriguez IB, Gnidehou S, Ito K, Stanisic DI, Lee M, McPhun V, Majam V, Willemsen NM, Batzloff MR, Raja AI, Dooley B, Hoffman SL, Yanow SK, Good MF. Induction of immunity following vaccination with a chemically attenuated malaria vaccine correlates with persistent antigenic stimulation.(2018); Clin Transl Immunology. Apr 11;7(4):e1015.

4. Baum J, Thomas AW, Conway DJ. Evidence for diversifying selection on erythrocyte-binding antigens of Plasmodium falciparum and P. vivax. (2003) Genetics. Apr;163(4):1327-36.

5. Maier, A.G., Baum, J., Smith, B., Conway, D.J., Cowman, A.F., 2009. Polymorphisms in erythrocyte binding antigens 140 and 181 affect function and binding but not receptor specificity in Plasmodium falciparum. Infect. Immun. 77 (4), 1689–1699.

6. Sim, B.K., Carter, J.M., Deal, C.D., Holland, C., Haynes, J.D., Gross, M., 1994. Plasmodium falciparum: further characterization of a functionally active region of the merozoite invasion ligand EBA-175. Exp. Parasitol. 78 (3), 259–268.

7. Salinas, N.D., Paing, M.M., Tolia, N.H., 2014. Critical glycosylated residues in exon three of erythrocyte glycophorin A engage Plasmodium falciparum EBA-175 and define receptor specificity. MBio 5 (5), e01606–14.

8. Pandey, K.C., Singh, S., Pattnaik, P., Pillai, C.R., Pillai, U., Lynn, A., Jain, S.K., Chitnis, C.E., 2002. Bacterially expressed and refolded receptor binding domain of Plasmodium falciparum EBA-175 elicits invasion inhibitory antibodies. Mol. Biochem. Parasitol. 123 (1), 23–33.

9. Camus, D., Hadley, T.J., 1985. A Plasmodium falciparum antigen that binds to host erythrocytes and merozoites. Science 230 (4725), 553–556.

10. Joshua-Tor, L., Tolia, N.H., Enemark, E.J., Sim, B.K.L., 2005. Structural basis for the EBA-175 erythrocyte invasion pathway of the malaria parasite Plasmodium falciparum. Cell 122, 183–193.

11. Lin, D.H., Malpede, B.M., Batchelor, J.D., Tolia, N.H., 2012. Crystal and solution structures of Plasmodium falciparum erythrocyte-binding antigen 140 reveal determinants of receptor specificity during erythrocyte invasion. J. Biol. Chem. 287 (44), 36830–36836.

12. Chowdhury P., Sen S., Kanjilal SD., Sengupta S., Genetic structure of two erythrocyte binding antigens of *Plasmodium falciparum* reveals a contrasting pattern of selection, 2018, *Infect Genet Evol*. 2018 Jan;57:64-74.

13. Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X., Rozas, R., 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19 (18), 2496–2497.

14. Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123 (3), 585–595.

15. Fu, Y.X., Li, W.H., 1993. Statistical tests of neutrality of mutations. Genetics 133 (3), 693–709.

16. Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3 (5), 418–426

17. Jukes, T., Cantor, C., 1969. Mammalian protein metabolism. In: HN, M. (Ed.), Evolution of Protein Molecules. Academic Press, New York, pp. 21–132.

18. Bandelt, H.J., Forster, P., Rohl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16 (1), 37–48.

19. McDonald, J.H., Kreitman, M., 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351 (6328), 652–654.

20. Tolia, N.H., Enemark, E.J., Sim, B.K., Joshua-Tor, L., 2005. Structural basis for the EBA-175 erythrocyte invasion pathway of the malaria parasite Plasmodium falciparum. Cell 122 (2), 183–193.

21. Salinas, N.D., Paing, M.M., Tolia, N.H., 2014. Critical glycosylated residues in exon three of erythrocyte glycophorin A engage Plasmodium falciparum EBA-175 and define receptor specificity. MBio 5 (5), e01606–14.

22. Baum, J., Ward, R.H., Conway, D.J., 2002. Natural selection on the erythrocyte surface. Mol. Biol. Evol. 19 (3), 223–229.

23. Yalcindag, E., Rougeron, V., Elguero, E., Arnathau, C., Durand, P., Brisse, S., Diancourt, L., Aubouy, A., Becquart, P., D'Alessandro, U., Fontenille, D., Gamboa, D., Maestre, A., Ménard, D., Musset, L., Noya, O., Veron, V., Wide, A., Carme, B., Legrand, E., Chevillon, C., Ayala, F.J., Renaud, F., Prugnolle, F., 2014. Patterns of selection on Plasmodium falciparum erythrocyte-binding antigens after the colonization of the new world. Mol. Ecol. 23 (8), 1979–1993.

(iii) SUMMARY OF THE FINDINGS:

The malaria parasite has a complex, multistage life cycle occurring within two living beings, the vector mosquitoes and the vertebrate hosts. During its complex life cycle in the human host, *Plasmodium falciparum* merozoite invades erythrocyte through multiple protein-protein interactions. Among these, interactions between RBC glycophorins with parasite Erythrocyte Binding Antigens are important. Since antigenic variation in parasite proteins is a major barrier to the development of effective vaccine, we explored the pattern of sequence diversity in two vaccine candidates (EBA-175/140) in *P. falciparum* isolates of Kolkata. The objectives are to understand the genetic and population levels forces responsible for shaping the diversity of *eba-175* and *eba-140* genes and to characterize repercussion of genetic diversity in the F2 region of EBA-ligands as they interact with host glycophorins. Venous blood samples (n=92) from *P. falciparum* infected malaria patients were used to sequence RII regions of *Pfeba-175* and *Pfeba-140* genes. Sequence divergence of parasite proteins were subjected to appropriate statistical and phylogenetic analyses using DnaSP and MEGA tools. Strength of host-parasite interactions between parasite proteins (mutant and prototype Pf3D7) and RBC glycophorin were compared using a novel molecular docking approach. *P. falciparum* Kolkata isolates experienced a recent population expansion as documented by negative value of Tajima's D, Fu & Li's statistics, unimodal mismatch distribution and star-like median-joining network for both *Pfeba* loci. Positive selection seemed to play a major role in shaping the diversity of *Pfeba-175* (d_N/d_S=2.45, and McDonald-Kreitman P-value= 0.04). *In silico* molecular docking demonstrated that high frequency haplotypes of *eba-175* were capable of engaging the parasite ligand into energetically more favorable interaction with GPA. A contrasting pattern of evolution and protein-protein interaction was noted for PfEBA-140. Together, this study provides a firm genetic support favoring a dominant role of EBA-175 in erythrocyte invasion and underscores the need to consider the observed F2-variants while designing vaccines that target EBA-175 parasite antigen.

(iv) PUBLICATION OUT OF THE PROJECT:

Chowdhury P., Sen S., Kanjilal SD., Sengupta S., Genetic structure of two erythrocyte binding antigens of *Plasmodium falciparum* reveals a contrasting pattern of selection, 2018, *Infect Genet Evol*. 2018 Jan;57:64-74.(PMID: 29128519).

(v) WHETHER OBJECTIVES WERE ACHIEVED:

Yes, the progress has been made according to original plan of work and towards achieving the objective.

(vi) WHETHER ANY PH.D. ENROLLED/PRODUCED OUT OF THE PROJECT:

The Project Fellow Pramita Chowdhury has been registered for her Ph. D. under the supervision of Dr. Sanghamitra Sengupta, Associate Professor, Department of Biochemistry, University of Calcutta.

(vii) Abstracts published in proceedings of Seminars/Conferences attended:

1. 2018: (25th -27th November): Attended 87th Annual Conference (Society of Biological Chemists, India) on " Genome Biology in Health and Disease" organized by School of Life Sciences, Manipal academy of Higher Education, Manipal and presented a poster entitled" **Genetic analysis of two erythrocyte binding ligands of *Plasmodium falciparum* to evaluate direction of natural selection and strength of host-parasite interaction**".

2. 2017: (17th March): Attended a one day Symposium on "Emerging Trends in Biology" organized by Department of Biochemistry, University of Calcutta and presented a poster entitled "**Genetic footprint of Selective Sweep in PfEBA-175, a mojor vaccine candidate in *P. falciparum* malaria".**

3. 2016: (4th October): Attended National Seminar on "Frontiers in Biotechnology" organized by Department of Biotechnology, St. Xavier's College, Kolkata and presented a poster entitled "Genetic diversity of *Plasmodium falciparum* EBA-175 in West Bengal".

A summary of the work done for the period (Annual basis)

Work done	Month/Year
1. Acquisition of equipment and consumables.	1/1
2. Sample collection.	4/1
3. Standardization of protocols .	1/1
4. Sequencing of EBA-175 and genetic diversity analysis	6/1
5. Genetic epidemiology (GYPA- tagSNP genotyping, statistical analysis and resequencing)	1/2
6. Ligand binding region of GYPA (exon 3 and exon 4) 0s+equencing.	3/2
7. Collation of host and parasite data, *in silico* homology modeling and molecular docking.	8/2
8. Sequencing of EBA-140 and genetic diversity Analysis.	6/3
9. Collation of host and parasite data, Molecular Docking, other relevant Bioinformatic analyses	3/3
10. Final data analysis, data submission, report and manuscript preparation and publication	3/3

Sanghamitra Sengupta
2.7.19

SIGNATURE OF
THE PRINCIPAL INVESTIGATOR

02/07/2019

By **REGISTRAR/ PRINCIPAL (Seal)**

Deputy Registrar
University of Calcutta

DR. SANGHAMITRA SENGUPTA
Associate Professor
Dept. of Biochemistry
UNIVERSITY OF CALCUTTA

2/7/19

SECRETARY
Faculty Council for P.G.
Studies in Sc. Tech. & Engg.
Ag. & Vet. Sc.
Calcutta University

FORWARDED

Sanjay Ghosh

Prof. Sanjay Ghosh
Head
Department of Biochemistry
University of Calcutta