Comparative study of Plasmodium falciparum Erythrocyte Membrane Protein 1- DBLα domain variants with respect to antigenic variations and docking interaction analysis with glycosaminoglycans

Journal:	*Molecular BioSystems*
Manuscript ID:	MB-ART-05-2014-000274.R1
Article Type:	Paper
Date Submitted by the Author:	20-Jun-2014

Complete List of Authors:	Agrawal, Megha; University of Pune, Bioinformatics Centre Ozarkar, Aarti; University of Pune, Molecular Biology Research Laboratory, Department of Zoology, Centre for Advanced Studies Gupta, Shipra; University of Pune, Bioinformatics Centre Deobagkar, Dileep; University of Pune, Bioinformatics Centre; University of Pune, Molecular Biology Research Laboratory, Department of Zoology, Centre for Advanced Studies Deobagkar, Deepti; University of Pune, Bioinformatics Centre; University of Pune, Department of Zoology

Comparative study of *Plasmodium falciparum* Erythrocyte Membrane Protein 1- DBLα domain variants with respect to antigenic variations and docking interaction analysis with glycosaminoglycans

Author names and affiliations

Megha R Agrawal[a,#], Aarti D Ozarkar[b,#], Shipra Gupta[a], Dileep N. Deobagkar[a,b], Deepti D Deobagkar[a,b,*]

a. Bioinformatics Centre, University of Pune, Pune 411007, India

b. Molecular Biology Research Laboratory, Department of Zoology, Centre for Advanced Studies, University of Pune, Pune- 411007, India

 # Both the authors have equal contribution in this manuscript.

*** Corresponding author:**

Prof. (Dr.) Deepti D Deobagkar

Bioinformatics Centre & Department of Zoology,

Center of Advanced Studies,

University of Pune,

Pune 411007, India.

Ph No.: +91-9921184871

Email: deepti.deobagkar@gmail.com

The variant surface antigen PfEMP1 (*Plasmodium falciparum* erythrocyte membrane protein 1) encoded by the polymorphic multi-copy *var* gene family plays an important role in parasite biology and the host-parasite interactions. Sequestration and antigenic variation is an essential component in the survival and pathogenesis of *Plasmodium falciparum* and contributes to chronic infection. The DBLα domain of PfEMP1 is a potential target for immuno-epidemiological studies and has been visualized as a vaccine candidate against severe malaria. Specific host receptors like heparin, heparan sulphate, blood group A and complement receptor 1 have been reported to bind DBLα domain. Although heparin has been experimentally shown to disrupt the parasite-host interaction and effectively disrupt rosetting, the binding sites for the DBLα domain and mechanism behind heparin-mediated rosette inhibition have not been elucidated. In this study, 3D structures and epitopes of DBLα domain in 3D7 and in two Indian isolates have been predicted and compared. We have carried out docking studies on DBLα domains with human GAG receptors (heparin and heparan sulphate) to predict the strength of association between the protein–ligand interactions. The DBLα domain structures showed extensive diversity and polymorphism in their binding sites. The docking results indicate that heparin binds more effectively with high affinity as compared to heparan sulphate with some common interacting residues. These common residues can play an important role in rosetting and will aid in the designing of inhibitors specific to the interactions between DBLα and heparin or heparan sulphate would be important in malaria treatment. Thus it may lead to the development of novel interference strategies to block red blood cell invasion and provide protection against malaria.

1. Introduction

Malaria is endemic to around 100 countries containing half of the global population. Approximately two million people annually die of malaria, in Africa [1], Brazil [2], Indonesia [3], Tanzania [4], India [5, 6] and young children are most affected in these regions (www.who.int/). *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) is one of the important parasite protein which shows antigenic variation and has been implied in immune evasion and sequestration [7]. The severity of the disease is characterized by PfEMP1 mediated rosetting phenomenon, where infected red blood cells forms rosettes with uninfected RBCs [8]. The enhanced invasion of RBCs and sequestration of high rosette densities in the microvasculature of vital organ are probably the major factors in pathophysiology. Rosetting has been associated with malaria severity [9]. PfEMP1 is also associated through its interaction with variety of host-cell receptors thus acting as an adhesin allowing pRBC to avoid splenic clearance and this leads to manifestation of severe malaria through excessive sequestration [10].

PfEMP1 is a large multidomain protein (200-300 KD) encoded by multicopy (~60 copy per genome) *var* gene family and is expressed on the surface of pRBC in mutually exclusive fashion [11]. The parasite has been reported to express only one PfEMP1 at a time on the erythrocyte surface [12, 13]. The *var* gene consists of two exon structure. The first exon is large and codes for multiple extracellular domains like N-terminal segment (NTS), Duffy binding-like (DBL) domains, Cys rich inter-domain regions (CIDR), C2 domains, one trans-membrane region (TM). Seven types of DBL domains (α, α1, β, γ, δ, ε, and x) and four types of CIDR domains (α, α1, β, and γ) have been reported on the basis of sequence similarity. The second exon is small and codes for most conserved cytoplasmic tail the acidic terminal segment (ATS). Based on the 5' UTR sequence similarity, *var* genes are classified into different upstream sequence (UPS) groups namely UPSA, UPSB, UPSC and UPSE as well as some intermediate groups: UPSB/A and UPSB/C [14]. These UPS

groups are related with chromosomal position (subtelomeric or central regions) of the *var* genes, as well as domain complexity of the encoded PfEMP1[15]. DBLα domain was found to be more conserved than other domains and can be divided into three subdomains (SD1, SD2 and SD3) comprising of several conserved and partially conserved alpha helices as well as very few beta-sheets (Fig. 1). It has been reported that DBLα domain plays a direct role in the process of rosette formation and its interaction to several host receptors - heparan sulphate (HS), blood group A antigen and complement receptor 1 (CR1)[16]. Though, there is diversity among rosetting phenotypes, some sulphated glycosaminoglycans (GAG) such as heparin and heparan sulphate have been reported to disrupt rosettes[17]. Heparan sulphate (HS) is found on endothelial cells in the microvasculature [18, 19]. The parasite mediates its interaction with HS via the N-terminal portion of PfEMP1, and more precisely the DBL1α domain. Heparin is a naturally occurring GAG and is a potent anticoagulant. Heparin derivatives are used in treatment of severe malaria. GAG mimetic molecules like low anticoagulant heparins (LAH) are found to disrupt rosettes of fresh clinical isolates of malaria patients [20]. DBLα antibodies have been shown to disrupt rosettes and protect against sequestration of *Plasmodium falciparum* infected erythrocyte, suggesting that DBLα can be used as a vaccine against severe malaria [21, 22]. This is one of the crucial phenomena of natural inhibition of PfEMP1 DBLα domain mediated rosette formation which needs to be further analysed from the perspective of developing anti-rosetting treatment in malaria.

The role of PfEMP1 in disease severity has been complicated due to the *var* gene diversity. This diversity has been reported in field isolates of different geographical regions [23-28]. The great antigenic diversity and unique regulation of switching between different antigenic variants of PfEMP1 enables the parasite to evade the immune system and to maintain severe and chronic infection [7]. Several reports have stressed this molecule as a vaccine candidate [29] and anti-rosetting agent [30]. In view of the genetic polymorphism and antigenic variation in the DBLα domain variants

of PfEMP1, we have analysed the entire DBLα domain variants in relation to their structure, epitope prediction and interaction analysis with GAG receptors in 3D7, IGH*var* and RAJ116*var* (Indian isolates). We have used homology modelling technique to design 3D structures of ~59 DBLα domain variant of 3D7 reference isolate, 41 variants and 39 variants each of two Indian isolates *i.e.* IGH*var* and RAJ116*var* respectively. Based on best quality representative 3D models from each UPS groups, the DBLα domains of all three isolates were analysed using molecular dynamics simulation to stabilise their conformations. The energy of simulated proteins was verified with several tools to ensure their stability and to provide a valid input for docking interaction analysis. Our study distinguishes itself from the earlier studies [31] as it provides genome wide structure predictions of all DBLα domain variants with exhaustive validation and their comparative analysis, in order to get a better insight into the epitope prediction and the contribution of evolutionary conserved residues in host receptor binding interactions, antigenic site variations, occupancy of important protein–ligand interactions during simulation time, etc.

With the fundamental aim of structure prediction, this study focuses on interaction analysis of DBLα domain variants with heparin and heparan sulphate, which have ability to disrupt the rosettes. The extensive comparative analysis of interacting residues provides insights into identification of common heparin binding sites shared by almost all DBLα domain variants from all three isolates. This would provide the information for developing heparin analogs, which would have potential to inhibit rosetting phenomena. The discontinuous epitope prediction analysis within DBLα domain variants and identification of common heparin binding site provides valuable inputs in antigenic variation and designing anti-rosetting strategies. It may also comment on the probable role of PfEMP1-DBLα domain as either drug or vaccine target.

2. Material & Methods

2.1 Dataset

The 59 variants of 3D7 PfEMP1 protein sequences were downloaded from NCBI (http://www.ncbi.nlm.nih.gov). The contigs containing *var* gene sequence variants of two Indian isolate IGH*var* and RAJ116*var* were retrieved from the dataset provided by Broad Institute (http://www.broadinstitute.org/) and dataset provided by Rask et al, 2010 [32]. The contig sequences were transformed into the protein sequence using Expasy Translate Tool (http://web.expasy.org/translate/). As the scope of this study involves the analysis of structure as well as sequence variation of most conserved PfEMP1 protein domain *i.e.* DBLα, the sequence boundaries for this domain were gathered using VarDom 1.0 server for all variants of all three isolates [32]. The sequence boundaries were then used to extract exact domain sequences in a particular variant from corresponding whole PfEMP1 protein. In this way, we have collected dataset of 136 sequences for DBLα domain from all three, 3D7 (56), IGH*var* (41) and RAJ116*var* (39) *Plasmodium falciparum* parasite isolates.

2.2 Sequence Analysis

Multiple sequence alignment was carried out for the whole dataset comprising of DBLα domain sequences using ClustalW [33], Multalin [34], ESPpript [35] online servers for multiple sequence alignment. Phylogenetic tree construction was done using Molecular Evolutionary Genetics Analysis (MEGA6) software [36, 37]. The phylogenetic trees were refined with bootstrap-confirmed neighbor-joining trees (1000 replicates) option within MEGA6 using Jones-Taylor-Thornton (JTT) substitution model with rate of gamma distribution for all isolates and their subdomain wise distributed sequences. Trees were edited and visualized using FigTree v1.4.0.

2.3 Structure Prediction and Validation

To carry out large scale structural comparison analysis of PfEMP1 DBLα domain variants from all three isolates, homology models for all DBLα variants were generated by automated structure modelling tool of Modweb which is based on modeller algorithm [38]. The models were built using either of two templates of DBLα1 domain of the *Plasmodium falciparum* membrane protein 1 (PfEMP1) from the *Var*O strain (PDB: 2XU0, 2YK0) retrieved from protein databank. As a result, we had in total 136 predicted structures from retrieved sequence dataset. The structures were validated and characterised to verify the quality of generated 3D structure by Structural Analysis and Verification meta-Server (SAVESv4) using PROCHECK [39], Verify 3D [40, 41], Errat [42], Discovery studio 3.5 (http://accelrys.com).

2.4 Molecular Dynamics (MD) Simulation

The three sets comprising of best quality 3D predicted structures each from all three parasite isolates and UPSA, B and C group have been selected for molecular dynamics simulation studies. The MD simulations for these structures were performed using Gromacs 4.5.3 [43, 44] under OPLS 2005 atoms force field on BRAF facility provided by CDAC (http://bioinfo.cdac.in/hpc.xhtml). The 3D structures of DBLα domain were immersed in a cubic box of 1.0 nm and periodic boundary conditions were applied using editconf tool followed by addition of SPC water molecules according to system. System was made electrically neutral by adding Na^+ using the 'genion' tool. The system was first minimized for energy in 50000 steps by steepest descent method to remove excessive strain. The minimized system was then subjected to MD in two steps. Initially NVT ensemble (constant number of particles, volume, and temperature) was performed for 100 ps, followed NPT ensemble (constant number of particles, pressure, and temperature) for 100 ps. The well equilibrated system was then subjected to molecular MD for 20 ns. Temperature was kept constant at 310 K with Andersen thermostat, pressure coupling of 1bar with Berendsen algorithm and system was further allowed to undergo production runs. LINCS algorithm [45] was used to constrain

the lengths of all bonds while the waters molecules were restrained using the SETTLE algorithm [46]. The trajectory files were analysed by using g_rms utility of GROMACS to obtain the root-mean square deviation (RMSD) values. RMSD values for Cα atoms from the initial structure were considered as a necessary condition to determine the convergence of the proteins toward equilibrium and calculated by:

$$RMSD(t_1, t_2) = [\frac{1}{M}\sum_{i=1}^{N} mi\ ||r_1(t_1) - r_2(t_2)||^2]^{\frac{1}{2}}$$

Where $\sum_{i=1}^{N}$ and $r_i(t)$ is the position of atom i at time t. The shape of protein molecule at all instants of simulation is indicated through hydrodynamic radius obtained using radius of gyration calculated by:

$$R_g = (\frac{\sum_i ||r_i||^2 m_i}{\sum_i m_i})^{1/2}$$

Mi is a mass of atom i and ri position of atom i with respect to the centre of mass of the molecule [47].

2.5 B cell epitope prediction

There are reports that PfEMP1 protein invoked the immune response in host cell [7, 29]. In order to investigate the presence of antigenic sites on these PfEMP1-DBLα domains, we have carried out the B cell epitope prediction analysis for all the structures. The linear B cell epitope prediction was carried out using online BepiPred 1.0 software which uses a combination of a hidden Markov model and a propensity scale method for linear B cell epitope Prediction [48]. We have also subjected the structures to conformation based discontinuous B cell prediction which found more sensible and effective. The discontinuous B cell epitope was predicted using DiscoTope2.0 software [49]. DiscoTope uses a combination of amino acid statistics, spatial information, and surface exposure [50].

It is trained on a compiled data set of discontinuous epitopes from 76 X-ray structures of antibody/antigen protein complexes. We have used specific cut-off of 1.9 for DiscoTope predictions which specifies 0.95% specificity and 0.17 sensitivity whereas for BepiPred epitope prediction set cut-off was 0.7 [50].

2.6 Docking Interaction with heparin and heparan sulphate

In order to check the interaction of heparin and heparan sulphate with DBLα domain variants, heparin coordinates were taken from heparin complexes in the Protein Data Bank (1bfb [51], 1e0o [52], and 1hpn [53]) while heparan sulphate coordinates were taken from Pubchem chemical database (CID:53477714). The docking interaction analysis for all 136 structures with heparin was performed using PatchDock (http://bioinfo3d.cs.tau.ac.il/PatchDock/) online server with default parameters [54]. Further interaction energies of 10 best solutions were calculated using FireDock (http://bioinfo3d.cs.tau.ac.il/FireDock/) [55, 56]. The docking interactions of nine simulated structure were further validated using Autodock Vina [57]. All nine simulated structures which have been taken into consideration for docking using Autodock Vina, have been pre-processed and minimized by adding polar hydrogens and gasteiger charges using the AutodockTool (ADT), a free graphics user interface (GUI) of MGL-tools. The grid box parameters were set in such a way that the search will perform over entire protein surface. Default values were used for all other docking parameters. Both the ligands were prepared by AutodockTool (ADT) and then subjected to docking interactions using Autodock Vina 4.2. Simultaneous docking runs were performed using heparan sulphate for comparative analysis with heparin interactions. The interaction analysis was performed using Discovery studio 3.5. Structural visualizations and high-resolution images were generated using PyMol (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.).

3 Results

3.1 *Var* genes group analysis

The DBL α domain of PfEMP1 dataset comprising of all three isolates - 3D7 (56), IGH*var* (41) and RAJ116*var* (39) and templates were used for analysis. Multiple sequence alignment of whole dataset with templates exhibits variation among sequence length and composition. These sequences have an average of ~ 45-66 % amino acid similarity and the maximum similarity was less than 75% (Fig. 2). Two regions, each comprising of ~20 residues, one at 110-130 amino acid **(MD1: CTvLARSFADIGDIvRGkDly**) and second at 220-250 amino acid residues **(MD2: vptyfDYVPQylRWfeEWaeDfc**), were observed to be highly conserved throughout the DBLα sequences of all the three parasite isolates. The former conserved region is a part of αH3 of subdomain 2, while the later one is a part of αH7 of subdomain 3. In addition to these two highly conserved regions, there are three partially conserved regions PD1, PD2 and PD3 which showed variant degree of conservation with respect to UPS group wise alignment. The partially conserved region **(PD1: GACAP*RRLhlC**Nlexi**) which is part of αH1 of subdomain 1 seems to be more conserved in UPSB group sequences (shown in Fig. S1) as compared to UPSA and UPSC whereas residues in **(PD2: hDLLGNvLVtAKyEG*sIV** and **PD3: ny*kLREdWW*aNRdqVWkAiTC)** regions in αH2 and αH5 of subdomain 2 are seen to be more conserved in UPSA group sequences as compared to UPSB and UPSC as illustrated in Fig. S2, S3. The sequence logos were generated for all five highly as well as partially conserved regions to determine the propensity of a particular amino acid to appear at certain position in particular UPS group of *var* gene. The conservation of these domains in PfEMP1 across different isolates of *P. falciparum* parasite raises a possibility that it may play a crucial role in DBLα mediated rosetting phenomena. The multiple sequence alignment of all DBLα domain sequences indicated that insertions were found at the position of β2 sheet formation in subdomain 2 in some variants from RAJ116*var* and 3D7 thus leading to the

variable lengths of β2. These insertions in β2 sheets may affect the overall stability of domain architecture.

The phylogenetic analysis across all isolates showed that UPSA group sequences were highly evolved and comparatively more conserved than UPSB and UPSC. UPSA group sequences formed a single cluster without any interruption by other UPS group sequences whereas UPSB group sequences formed a cluster interspersed by UPSC group sequences. These observations suggest that UPSA is a unique most evolved group however UPSB and C are evolutionary related as shown in Fig. 3. The sequences from subdomains of all UPS groups *i.e.* SD1, SD2 and SD3 were aligned separately and the results suggested a similar picture as seen in Fig. 3. The sequences from UPSA group were conserved in all the subdomains *i.e.* SD1, SD2 and SD3. In the SD1 analysis, UPSA group was found to be fully conserved while the sequence (gi_124512768) from UPSB group was showing similarity with the members of the UPSA group sequences, which was the most conserved group of all the three groups (Fig. 4A). Similarly, sequences (gi_124512768) from UPSB and (gi_86171174) from UPSC were showing similarity with the members of the UPSA group in case of SD2 subdomain (Fig. 4B). Additionally, while analysing the tree of SD3 domain sequences, UPSB group sequences namely gi_124512768, gi_124512758 and UPSC group sequence Raj116*var*11 were also found to be conserved and clubbed in UPSA group sequences due to its high remarkable similarity with UPSA group sequences (Fig. 4C).

In subdomain wise analysis of all the UPS groups, only one partially conserved domain (PD1) was found in SD1 domain, which involves formation of disulphide bridges between two cysteine residues whereas in subdomain SD2 two partially conserved regions and one fully conserved region were present and only one cysteine residue of MD1 was involved in the formation of disulphide bridge network. Similarly, in case of SD3 there were eight cysteine residues present and they

formed one cysteine bridge under MD2 while additional disulphide bridges were involved in the cysteine bridge network formation given in Table 1. This part of the molecule is interesting since the number of cysteine residues present therein have been found associated with the rosetting phenotype of the parasite and the severity from malaria.

3.2 Modelling of DBLα domain variants

The sequence identity of all DBLα variants with corresponding PDB templates (2K0 and 2XU0) along with the Modweb z-dope score has been given in (Table 2, S1). Almost all DBLα domain targets exhibited sequence identity between 40-60%, which is quite more than 30% threshold for successful homology modelling. Further quality checks of the modelled structures were carried out using programs such as ERRAT, Verify 3D and their respective scores were calculated. The reliability of the backbone torsion angles Φ, ψ of the modelled proteins was examined by PROCHECK and the corresponding values and percentage for residues in core, allowed, generously allowed, and disallowed regions for all modelled structures were depicted in Table 3, S2. The percentage of residues in disallowed region range from 0 - 1.7 and none of these residues are a part of the functional site.

We have seen extensive variation among modelled structures of different variants of DBLα domains. This variability was reflected within the average main chain and c-α RMSD ~ 2.5Å when calculated from overall structure of corresponding template. Additionally, we have used a different approach where we have calculated a residue wise RMSD from corresponding template. The residues taken into consideration for RMSD calculation are the same conserved residues comprising of two of the most conserved regions mentioned above. The RMSD values for all the structures with their corresponding templates with respect to highly conserved domains are given in Table 4, S3. The results from residue wise RMSD calculation exhibited average RMSD for first

conserved domain in IGH*var* variants for both c-α and main chain to be ~2.0Å and for second conserved region ~1.9Å. This RMSD range was conserved in case of RAJ116*var* variants *i.e.* ~1.9Å for both the conserved regions. Unlike IGH*var* and RAJ116*var*, 3D7 DBLα domain variants showed somewhat higher range of average RMSD for both c-α and main chain *i.e.* ~2.2 for second conserved region and ~2.3 for first conserved region. Thus, the second conserved region is more conserved and assumes a stable conformation.

3.3 Epitope prediction of DBLα domain variants

To verify the potentiality of DBLα domain as a vaccine candidate, we have undertaken the B cell epitope prediction analysis to identify the presence of antigenic sites on DBLα domains. The linear as well as conformational epitope prediction analysis predicted several B cell epitopes at variable positions in different DBLα domain variants. The B cell prediction results given by BepiPred and DiscoTope for all variants in dataset are provided in supplementary files (S1, S2). The B cell epitope results have been predicted using very high threshold of >0.7 for BepiPred and >1.9 for DiscoTope specifying high specificity and low sensitivity. As shown in Fig. 5, 6A and 6B epitopes were predicted in the variable regions however few residues from partially conserved regions showed probability of being a B cell epitope.

The epitopes were not predicted in main domains in all structures across the different isolates. The sequence logos of the main domains *i.e.* MD1, MD2 and partially conserved domains *i.e.* PD1, PD2 and PD3 are shown in the Fig. 5. It was observed that main domains have more confidence to be conserved than the partially conserved domains. The same has been observed through the conformational epitope prediction analysis approach. The rationale behind this approach is that, although the residues are shown to be conserved throughout different isolates, there is extensive variation in other parts of the domain. This suggests a definite role in parasite antigenic domain

mediated virulence. Hence, if the conformations of these functional residues are found to be conserved, they could be considered for future therapeutic target in malaria treatment.

3.4 Molecular Dynamics Simulation

Further, to check whether, we could identify major interacting residues by considering these two fully conserved domains as potential sites for drug designing, we have selected three structures from each UPS groups on the basis of their identity and Z-dope score for further study (Table 2). As an indication of the degree of refinement during the simulations, the positional root mean square deviation (RMSD) from the respective original predicted structure after a least-squares best fit was calculated for each structure investigated (nine models, three each from three parasite isolates) as a function of the simulation time for whole protein. The RMSD values calculated after 20 ns of simulation for all structures are given in Table 5. It has been observed that there were notable differences in RMSD with respect to the initial predicted structures after 20 ns of simulation. In all cases RMSD has shown to be increased.

The analysis of occurrence of salt bridges before and after simulation showed that there were remarkable increases in number of salt bridges formed after 20ns simulation which ultimately confirm the overall stability of protein conformers. The quality of simulated structures was also confirmed by PROCHECK, Verify 3D and Errat (Table 3). The domain architecture of different DBLα variants from all three isolates also affected the stability time of protein confirmations where IGH*var*, RAJ116*var* and 3D7 DBLα domain variants showed stability at around 18ns. RMSD values for gi_86171174 increases from 0.08 nm to 0.3609 nm at 0.91 ns and it became constant (0.35 nm) till 5 ns followed by a sharp increase of 0.5032 nm at 11.39 ns followed by decrease of 0.435 nm till 12.472 ns and again it become constant 0.4379 nm till 20 ns. RMSD values for IGH*var*34 increased from 0.333 nm to 0.443 nm at 17.15 ns followed by atomic fluctuation within

protein that levelled off around 16.19 ns (0.38 nm) and then demonstrated a stable trajectory between 0.35 nm and 0.38 nm. Raj116*var*28, RMSD value has increased from 0 ns to 0.4394 nm at 6.75 ns and then becomes constant till 20 ns.

The RMSD value for IGH*var*05 started from 0.077 nm and increased to 0.37 nm around 10.358 ns followed by sharp increase of 0.47 nm at 15.5 ns and finally became stable after 18.566 ns (0.409nm). Similarly, RMSD value for RAJ116*var*06 started from 0.0831 nm and increased to 0.4181 nm around 7.36 ns and then became approximately constant till 14.0 ns followed by a sharp increase of 0.459 nm at 16.322 ns and decreased to 0.37nm finally became stable after 18.674 ns i.e. 0.378 nm and in case of gi_124512768, the RMSD value started from 0.079 nm and increases to 0.4677 nm at around 5.134 ns and decreased to 0.3899 nm till 8.882 ns and again increased to 0.522 nm at around 13.138 and finally became stable.

Subsequently, RMSD value for gi_124505159 started from 0.059 nm and increased to 0.3376 nm around 5.468 ns followed by a sharp decrease of 0.263 nm at 7.08 and finally become stable from 0.3189 nm to 0.325 nm at 13.87 ns to 17.75 ns. In case of RAJ116*var*19, RMSD value started from 0.417 nm at 0.01 ns and decreased sharply till 0.276 nm at 1.55 ns followed by sharp increase of 0.3427 nm at 7.156 ns and again decreased to 0.119 nm at 17.122 ns and finally became stable. Additionally, RMSD values in case of IGH*var*26 increases 0.25 nm at 8ns and finally become stable till 20ns (Fig. 7).

At the end of simulation for 20ns, we obtained the stable conformations of nine DBLα domain variants that served as optimized inputs for the docking algorithm. It was observed that beta sheets along with the salt bridges have increased in the conserved domains in most of the simulated structures which were responsible for stability of the structural domains.

3.5 DBLα domain variants interaction analysis

The docking interaction analysis of all DBLα domain variants with heparin and heparan sulphate exhibited that almost all the complexes were stabilized by variable number of hydrogen bonds in different DBLα domain variants likely with heparin RAJ116*var*19(9hb), RAJ116*var*06(9hb), RAJ116*var*28(13hb), IGH*var*05(11hb), IGH*var*26(8hb), IGH*var*34(15hb), gi_124512768(7hb), gi_124505159(11hb), gi_86171174(17hb), similarly, in case of heparan sulphate RAJ116*var*19(11hb), RAJ116*var*06(8hb), RAJ116*var*28(10hb), IGHb*var*05(7hb), IGH*var*26(8hb), IGH*var*34(14hb), gi_124512768(7hb), gi_124505159(7hb), gi_86171174(11hb), thus suggesting significant contribution of hydrogen bonds, attractive van der Waals forces in host-parasite domain interaction. The numbers of hydrogen bonds were more with heparin as compared to heparan sulphate. The binding energies for heparin and heparin sulphate were ranging from -78.04 to -38.00 kcal/mol and -78.04 to -22.61 kcal/mol, respectively. This indicates significant interaction between heparin and heparan sulphate with DBLα domain variants respectively. The repulsive van der Waals, atomic contact energies and global interaction energies for nine simulated proteins are calculated, while docking interactions energies of rest of the predicted structures of DBLα domain variants with heparin are given Table S4 and Table S5. Heparan sulphate molecule binds to PfEMP1 through interactions of negatively charged carboxyl and sulphate groups with basic amino acid residues. Heparin, which is highly sulphated, is likely to provide the crucial sulphate groups required to interact with key basic amino acids in a target protein that binds to heparan sulphate in a more selective manner. Almost all DBLα domain variants were shown to be interacting with highly conserved regions (MD1 and MD2) which signify that the two highly conserved regions of DBLα domain were considered as most probable site of interaction for host receptors heparin and heparan sulphate as shown in Fig. 8A, B. The common binding sites of heparin and heparan sulphate included hydrophilic polar and positively charged amino acids namely Thr, Ser, Tyr, Lys, Arg and His. Heparan sulphate binds strongly with Ser residue. The docking results were validated with

AutoDock vina and the same sites have been observed with heparin and heparan sulphate binding. The binding affinities for heparin are in the range of 7-8 kcal/mol whereas for heparan sulphate is in the range of 6-7 kcal/mol, thus, showing significant binding with heparin for anti-rosetting phenomena.

4. Discussion

The key findings of the present study suggest presence of structural and epitope variations among the different PfEMP1 variants of Indian isolates as compared to 3D7. Secondly, docking studies with heparin and heparan sulphate revealed that the active site lies between both main domains (MD1, MD2) of the DBLα. This signifies that PfEMP1 interacts with heparin and employs anti-rosetting whereas when heparan sulphate interacts with this complex, it starts the rosetting activity. These findings underline the importance of the MD1 and MD2 domain in molecular events linked to the rosetting and anti-rosetting phenomena. These events are crucial in host parasite interaction and influence the ability of parasite to persist and survive in the host.

The Indian isolates, IGH*var* and Raj116*var* showed diversity when compared with 3D7 with respect to its sequence, structure and epitopes. Amongst them Raj116*var* variants also showed maximum polymorphism. The length variations in the DBLα domains are due to the variable loops length and these Indian isolates demonstrate significant diversity and evolved more due to recombination events over the period and can cause severe disease conditions. Epitope prediction results revealed that Raj116*var* variants have fewer epitopes as compared to other isolates. The docking interaction studies revealed that heparin binds more effectively as compared to heparan sulphate.

Previously, it has been reported that UPSA group variants may have evolved specialized binding properties that contribute to preferential expression in severe malaria infections as compared to UPSB/C, those are associated with mild or uncomplicated malaria only [58, 59]. It has been reported that UPSA group parasites have an ability to form rosettes [60, 61] and are frequently expressed in severe malaria patients. Subsequent *in vitro* cultivation generates random switching to UPSB/C *var* genes [62-64], suggesting significant role for immune pressure [65]. Our findings suggested that UPSA

group genes are more evolutionary evolved and their pattern is more conserved as compared to UPSB/C as given in Fig. 3, 4. Subsequent analysis of all isolates indicated that Raj116*var* variants have more variation in sequence length and structure as compared to IGH*var* and 3D7. It was also reported that SD3 loop and ALNRKE motif of SD2 loop of DBLα is an important target for anti-rosetting activity [66]. In our study DBLα variants revealed that both SD3 and SD2 loop are involved in rosetting and anti-rosetting activities, however, the ALNRRE motif was not found to be interacting with heparin and heparin sulfate. Two fully conserved cysteine residues are found in SD2 domain while eight cysteine residues are present in SD3 domain. Further analysis of the fully conserved cysteine residues shows that only one cysteine residue lie in MD1 and MD2 domains of SD2 and SD3 loops, whereas rest of the conserved cysteine residues are involved in the cysteine–cysteine network formation (Table 1). Cys2 (DBLα-*Var*O strain) and Cys4 (most frequent in genome) conserved cysteine residues are correlated well with severe malaria and mild malaria, respectively [67]. The disulphide bonds play an important role in predicting how the DBL1α domain adapts to constant immune pressure. Cysteine residues have a significant role in keeping the protein scaffold stable and allowing the parasite to explore more surface area at the variable loops [68]. This molecular analysis of cysteine residues is indeed interesting since the number of cysteine residues present have been found associated with the rosetting and anti-rosetting phenotype of the parasite.

Our results are expressed in terms of the root-mean-square deviation (Cα-RMSD and main-chain-RMSD) of the alignment of predicted positions (MD1, MD2) of Cα and main-chain atoms with respect to the native structures. On comparing DBLα variant protein structures with main domains (MD1, MD2), the trend of RMSD values fluctuated mostly in UPSB group members (Table 4). With the use of B cell epitope prediction algorithms, we were able to map the epitopes within the subdomains. There are reports where the linear B cell epitopes were predicted using sequence, however this technique have not been very successful [49]. We have predicted B cell epitopes from

the primary structure of protein. Most of those epitopes are variable and buried on the inside of the protein and were found in the variable loop region as compared to MD1 and MD2 domains (Fig. 5, 6), and sequence logos of each UPS sub group showed the degree of conservation around 95% of the top (MD1, MD2) domains residues as compared with partially conserved domains sequences (PD1, PD2 and PD3). The same trend has been found in the sequences of all three isolates. However, Raj116*var* variants have less number of epitopes as compared to others. This suggests that in case of Raj116*var* variants probability of progression of infection to severe malaria is higher. In light of these findings, it was suggested that a drug can be designed by considering MD1 and MD2 domains, besides much variation in other part of protein domains which reflects a definite role in parasite antigenic domain mediated virulence. Further, the study was undertaken to investigate the correlation of DBLα domain of PfEMP1 variants of Indian isolates with epitopes and binding properties of GAG receptors. Three representatives of each subgroup *i.e.* UPSA/B/C were considered for simulation to characterise the binding interactions with heparin and heparan sulphate. After simulation, our results further depicted that optimizing the parameters for individual proteins leads to approx. 50% gains over the root-mean-square deviation as reported in Table 5. Although, PfEMP1 protein is a globular protein, in most of the cases protein become marginally stable after 15 ns because the free energy released when the protein folded into its native conformation is relatively small in all isolates Fig. 7. DBLα variants with a heparin or heparan sulphate shows binding property that contains clusters of positively charged and hydrophilic amino acid residues in the MD1 and MD2 domains. The study of the DBLα variants showed the presence of potential GAG-binding motifs in the sequences. It was observed that binding affinity is more influenced by the ionic interactions between the highly acidic sulphate groups and the basic side chains of arginine, lysine and histidine. Arginine and lysine binds more strongly and they provide structural clues about heparin binding sites. It was reported earlier that 12 mer oligomer docked well with the protein and YFR motif along with these residues are important for heparin binding

interactions [69]. Thus, these residues can facilitate the design of peptides that binds efficiently. Hydrophillic interactions also play an important role in heparin-protein interactions. In this study, heparin-binding variants showed that MD1 and MD2 domains are around 50-60% soluble and these interactions may influence some important processes in development of disease.

It has been reported that low anticoagulant heparin (LAH) - Sevuparin / DF02 can be used as an adjunct treatment of severe malaria (Patent No - WO2013095276A1). Thus, we performed docking experiments of this compound with the DBLα domain variants. It shows interactions at the same binding site as that of heparin and heparin sulphate. LAH deactivate heparin, displace heparan sulphate by competitive binding, thus, dissolving cell aggregates and leads to de-polymerization of the GAG. *In vitro* inhibition leads to blocking and reverse sequestration of pRBCs. This study is, thus, highly relevant for drug development and understanding disease pathogenesis.

Thus, the distribution of *var* genes in 3D7 and two Indian isolate genomes have been extensively studied; we found that Indian isolates varied with their structural conformations and B cell epitopes. However common binding sites in the DBLα domain involved in interacting with heparin and heparin sulphate, *i.e.* MD1 and MD2 were identified. Characterisation of the binding pattern of heparin and heparin sulphate with DBLα variants focuses on rosetting and anti-rosetting phenomena and this will enhance the study of the mechanism of erythrocyte invasion and pathogenesis and may form the basis for drug design and ligand blocking therapeutics for malaria.

5. Conclusion

Sequestration of pRBCs in the microvasculature is a relevant phenomenon in the disease pathology, with PfEMP1-mediated rosetting being one of the major contributors. An extensive comparative analysis of the antigenic epitope prediction and host receptor binding interactions was carried out in

order to gain an insight into the structural variations and to understand the interacting domains involved in heparin and heparin sulphate binding. The DBLα domain of PfEMP1 variants in 3D7 strain and Indian isolates shows diverse epitopes and structural variations due to which the parasite evades the host immune response. Predicted epitopes of these variants are located in SD2 and SD3 loop regions of the DBLα proteins and some of them were mapped in partially conserved subdomains. The surface expression of PfEMP1 influences the overall binding affinity of infected erythrocytes. The DBLα variant structures and comparative analysis provides new insights for robust understanding of erythrocyte invasion, pathogenesis and the prediction of heparin and heparan sulphate binding of DBLα variants to infected RBCs that will facilitate the drug design and ligand blocking therapeutics. The development of treatments that could diminish sequestration and disrupt rosetting is urgently needed to decrease the disease mortality. Thus, it can be predicted that although this domain may have limitations in its use as a vaccine candidate for malaria due to the vast repertoire they appear to be a very promising candidate for designing a therapeutic drug target.

Acknowledgments: The support provided by Department of Biotechnology- COE to Bioinformatics Center and Department of Science &Technology, Govt. of India are gratefully acknowledged. We thank to BRAF supercomputer facility of Centre for Development of Advanced Computing (C-DAC), Pune for providing the access for performing MD simulation.

Funding: Funding for this work was provided by the DBT-COE to Bioinformatics Center (http://dbtindia.nic.in/) and DST (SR/WOS-A/LS-146/2011, Aarti Ozarkar; http://www.dst.gov.in/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. T. P. Eisele, D. A. Larsen, N. Walker, R. E. Cibulskis, J. O. Yukich, C. M. Zikusooka and R. W. Steketee, *Malar J*, 2012, **11**, 93.

2. J. Oliveira-Ferreira, M. V. Lacerda, P. Brasil, J. L. Ladislau, P. L. Tauil and C. T. Daniel-Ribeiro, *Malar J*, 2010, **9**, 115.

3. I. R. Elyazar, S. I. Hay and J. K. Baird, *Adv Parasitol*, 2011, **74**, 41-175.

4. J. Mugasa, W. Qi, S. Rusch, M. Rottmann and H. P. Beck, *Malar J*, 2012, **11**, 230.

5. A. Das, A. R. Anvikar, L. J. Cator, R. C. Dhiman, A. Eapen, N. Mishra, B. N. Nagpal, N. Nanda, K. Raghavendra, A. F. Read, S. K. Sharma, O. P. Singh, V. Singh, P. Sinnis, H. C. Srivastava, S. A. Sullivan, P. L. Sutton, M. B. Thomas, J. M. Carlton and N. Valecha, *Acta Trop*, 2013, **121**, 267-273.

6. A. Kumar, L. Chery, C. Biswas, N. Dubhashi, P. Dutta, V. K. Dua, M. Kacchap, S. Kakati, A. Khandeparkar, D. Kour, S. N. Mahajan, A. Maji, P. Majumder, J. Mohanta, P. K. Mohapatra, K. Narayanasamy, K. Roy, J. Shastri, N. Valecha, R. Vikash, R. Wani, J. White and P. K. Rathod, *Acta Trop*, 2012, **121**, 246-255.

7. N. D. Pasternak and R. Dzikowski, *Int J Biochem Cell Biol*, 2009, **41**, 1463-1466.

8. A. Barragan, V. Fernandez, Q. Chen, A. von Euler, M. Wahlgren and D. Spillmann, *Blood*, 2000, **95**, 3594-3599.

9. A. Rowe, J. Obeiro, C. I. Newbold and K. Marsh, *Infect Immun*, 1995, **63**, 2323-2326.

10. S. M. Kraemer and J. D. Smith, *Curr Opin Microbiol*, 2006, **9**, 374-380.

11. J. D. Smith, B. Gamain, D. I. Baruch and S. Kyes, *Trends Parasitol*, 2001, **17**, 538-545.

12. A. Scherf, R. Hernandez-Rivas, P. Buffet, E. Bottius, C. Benatar, B. Pouvelle, J. Gysin and M. Lanzer, *EMBO J*, 1998, **17**, 5418-5426.

13. Q. Chen, V. Fernandez, A. Sundstrom, M. Schlichtherle, S. Datta, P. Hagblom and M. Wahlgren, *Nature*, 1998, **394**, 392-395.

14. T. Lavstsen, A. Salanti, A. T. Jensen, D. E. Arnot and T. G. Theander, *Malar J*, 2003, **2**, 27.

15. S. M. Kraemer and J. D. Smith, *Mol Microbiol*, 2003, **50**, 1527-1538.

16. K. Flick and Q. Chen, *Mol Biochem Parasitol*, 2004, **134**, 3-9.

17. U. Lindahl and L. Kjellen, *J Intern Med*, 2013, **273**, 555-571.

18. A. M. Vogt, A. Barragan, Q. Chen, F. Kironde, D. Spillmann and M. Wahlgren, *Blood*, 2003, **101**, 2405-2411.

19. N. S. Gandhi and R. L. Mancera, *Chem Biol Drug Des*, 2008, **72**, 455-482.

20. A. M. Leitgeb, K. Blomqvist, F. Cho-Ngwa, M. Samje, P. Nde, V. Titanji and M. Wahlgren, *Am J Trop Med Hyg*, 2011, **84**, 390-396.

21. Q. Chen, F. Pettersson, A. M. Vogt, B. Schmidt, S. Ahuja, P. Liljestrom and M. Wahlgren, *Vaccine*, 2004, **22**, 2701-2712.

22. D. Angeletti, L. Albrecht, M. Wahlgren and K. Moll, *Malar J*, 2013, **12**, 32.

23. D. S. Chen, A. E. Barry, A. Leliwa-Sytek, T. A. Smith, I. Peterson, S. M. Brown, F. Migot-Nabias, P. Deloron, M. M. Kortok, K. Marsh, J. P. Daily, D. Ndiaye, O. Sarr, S. Mboup and K. P. Day, *PLoS One*, 2011, **6**, e16629.

24. A. R. Trimnell, S. M. Kraemer, S. Mukherjee, D. J. Phippard, J. H. Janes, E. Flamoe, X. Z. Su, P. Awadalla and J. D. Smith, *Mol Biochem Parasitol*, 2006, **148**, 169-180.

25. D. Deobagkar, *Scholarly Research Exchange*, 2009, **2009**.

26. S. M. Kraemer, S. A. Kyes, G. Aggarwal, A. L. Springer, S. O. Nelson, Z. Christodoulou, L. M. Smith, W. Wang, E. Levin, C. I. Newbold, P. J. Myler and J. D. Smith, *BMC Genomics*, 2007, **8**, 45.

27. A. E. Barry, A. Leliwa-Sytek, L. Tavul, H. Imrie, F. Migot-Nabias, S. M. Brown, G. A. McVean and K. P. Day, *PLoS Pathog*, 2007, **3**, e34.

28. A. D. Ozarkar, D. Prakash, D. N. Deobagkar and D. D. Deobagkar, *Protein Pept Lett*, 2007, **14**, 528-530.

29. J. G. Beeson, J. A. Chan and F. J. Fowkes, *Expert Rev Vaccines*, 2013, **12**, 105-108.

30. Y. Adams and J. A. Rowe, *PLoS One*, 2013, **8**, e73999.

31. C. O. Buckee and M. Recker, *PLoS Comput Biol*, 2012, **8**, e1002451.

32. T. S. Rask, D. A. Hansen, T. G. Theander, A. Gorm Pedersen and T. Lavstsen, *PLoS Comput Biol*, 2010, **6**.

33. A. Aiyar, *Methods Mol Biol*, 2000, **132**, 221-241.

34. F. Corpet, *Nucleic Acids Res*, 1988, **16**, 10881-10890.

35. P. Gouet, X. Robert and E. Courcelle, *Nucleic Acids Res*, 2003, **31**, 3320-3323.

36. G. Stecher, L. Liu, M. Sanderford, D. Peterson, K. Tamura and S. Kumar, *Bioinformatics*, 2014.

37. K. Tamura, G. Stecher, D. Peterson, A. Filipski and S. Kumar, *Mol Biol Evol*, 2013, **30**, 2725-2729.

38. U. Pieper, B. M. Webb, G. Q. Dong, D. Schneidman-Duhovny, H. Fan, S. J. Kim, N. Khuri, Y. G. Spill, P. Weinkam, M. Hammel, J. A. Tainer, M. Nilges and A. Sali, *Nucleic Acids Res*, 2014, **42**, D336-346.

39. R. A. Laskowski, J. A. Rullmannn, M. W. MacArthur, R. Kaptein and J. M. Thornton, *J Biomol NMR*, 1996, **8**, 477-486.

40. J. U. Bowie, R. Luthy and D. Eisenberg, *Science*, 1991, **253**, 164-170.

41. R. Luthy, J. U. Bowie and D. Eisenberg, *Nature*, 1992, **356**, 83-85.

42. C. Colovos and T. O. Yeates, *Protein Sci*, 1993, **2**, 1511-1519.

43. H. J. C. Berendsen, D. van der Spoel and R. van Drunen, *Computer Physics Communications*, 1995, **91**, 43-56.

44. B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *Journal of Chemical Theory and Computation*, 2008, **4**, 435-447.

45. J. Y. Xie, G. H. Ding and M. Karttunen, *Biochim Biophys Acta*, 2013, **1838**, 994-1002.

46. S. Miyamoto and P. A. Kollman, *Journal of Computational Chemistry*, 1992, **13**, 952-962.

47. A. K. Verma, S. Gupta, S. Verma, A. Mishra, N. S. Nagpure, S. P. Singh, A. K. Pathak, U. K. Sarkar, M. Singh and P. K. Seth, *J Mol Model*, 2013, **19**, 1285-1294.

48. J. E. Larsen, O. Lund and M. Nielsen, *Immunome Res*, 2006, **2**, 2.

49. P. Haste Andersen, M. Nielsen and O. Lund, *Protein Sci*, 2006, **15**, 2558-2567.

50. J. V. Kringelum, C. Lundegaard, O. Lund and M. Nielsen, *PLoS Comput Biol*, 2012, **8**, e1002829.

51. S. Faham, R. E. Hileman, J. R. Fromm, R. J. Linhardt and D. C. Rees, *Science*, 1996, **271**, 1116-1120.

52. L. Pellegrini, D. F. Burke, F. von Delft, B. Mulloy and T. L. Blundell, *Nature*, 2000, **407**, 1029-1034.

53. B. Mulloy, M. J. Forster, C. Jones and D. B. Davies, *Biochem J*, 1993, **293 (Pt 3)**, 849-858.

54. D. Schneidman-Duhovny, Y. Inbar, R. Nussinov and H. J. Wolfson, *Nucleic Acids Res*, 2005, **33**, W363-367.

55. E. Mashiach, D. Schneidman-Duhovny, N. Andrusier, R. Nussinov and H. J. Wolfson, *Nucleic Acids Res*, 2008, **36**, W229-232.

56. N. Andrusier, R. Nussinov and H. J. Wolfson, *Proteins*, 2007, **69**, 139-159.

57. O. Trott and A. J. Olson, *J Comput Chem*, 2010, **31**, 455-461.

58. Y. Kalmbach, M. Rottmann, M. Kombila, P. G. Kremsner, H. P. Beck and J. F. Kun, *J Infect Dis*, 2010, **202**, 313-317.

59. J. H. Janes, C. P. Wang, E. Levin-Edens, I. Vigan-Womas, M. Guillotte, M. Melcher, O. Mercereau-Puijalon and J. D. Smith, *PLoS Pathog*, 2011, **7**, e1002032.

60. L. Albrecht, K. Moll, K. Blomqvist, J. Normark, Q. Chen and M. Wahlgren, *Malar J*, 2011, **10**, 17.

61. I. Vigan-Womas, M. Guillotte, C. Le Scanf, S. Igonet, S. Petres, A. Juillerat, C. Badaut, F. Nato, A. Schneider, A. Lavergne, H. Contamin, A. Tall, L. Baril, G. A. Bentley and O. Mercereau-Puijalon, *Infect Immun*, 2008, **76**, 5565-5580.

62. A. Bachmann, S. Predehl, J. May, S. Harder, G. D. Burchard, T. W. Gilberger, E. Tannich and I. Bruchhaus, *Cell Microbiol*, 2011, **13**, 1397-1409.

63. K. Blomqvist, J. Normark, D. Nilsson, U. Ribacke, J. Orikiriza, P. Trillkott, J. Byarugaba, T. G. Egwang, F. Kironde, B. Andersson and M. Wahlgren, *Mol Biochem Parasitol*, 2010, **170**, 74-83.

64. Q. Zhang, Y. Zhang, Y. Huang, X. Xue, H. Yan, X. Sun, J. Wang, T. F. McCutchan and W. Pan, *PLoS One*, 2011, **6**, e20591.

65. G. M. Warimwe, T. M. Keane, G. Fegan, J. N. Musyoki, C. R. Newton, A. Pain, M. Berriman, K. Marsh and P. C. Bull, *Proc Natl Acad Sci U S A*, 2009, **106**, 21801-21806.

66. D. Angeletti, L. Albrecht, K. Blomqvist, P. Quintana Mdel, T. Akhter, S. M. Bachle, A. Sawyer, T. Sandalova, A. Achour, M. Wahlgren and K. Moll, *PLoS One*, 2013, **7**, e50758.

67. P. C. Bull, S. Kyes, C. O. Buckee, J. Montgomery, M. M. Kortok, C. I. Newbold and K. Marsh, *Mol Biochem Parasitol*, 2007, **154**, 98-102.

68. J. Normark, D. Nilsson, U. Ribacke, G. Winter, K. Moll, C. E. Wheelock, J. Bayarugaba, F. Kironde, T. G. Egwang, Q. Chen, B. Andersson and M. Wahlgren, *Proc Natl Acad Sci U S A*, 2007, **104**, 15835-15840.

Figure Legends

Fig. 1: Subdomain architecture with secondary structure analysis of PfEMP1 DBLα domain.

Fig. 2: Multiple sequence alignment of DBLα domain sequences from three Plasmodium isolates with template.

Fig. 3: Phylogenetic analysis of all three isolates of DBLα domain sequences.

Fig. 4: Phylogenetic analysis based on subdomain wise classification of DBLα sequence.

Fig. 5: Structures, their epitopes and sequence logos.

Fig. 6: Representation of B-cell Epitope Prediction analysis by BepiPred and Discotope servers for all three isolates (3D7, IGH*var*, RAJ116*var*).

Fig. 7: Representation of RMSD (nm) with time (ns) simulation graph for nine variants from three isolates.

Fig. 8: Docking Interaction Studies with (A) Heparin and (B) Heparan Sulfate with all simulated DBLα domain variants proteins structures.

Table 1 Distribution of Cysteine residues in subdomains

Subdomains	Domains	Cysteine Bridges	Canonical residues
SD1	PD1	Cys22-**Cys62**	Cys(1)- Cys(4)
		Cys38-**Cys53**	Cys(2)- Cys(3)
	PD2	-	-
SD2	MD1	**Cys112**-Cys222	Cys(5)- Cys(6)
	PD3	-	-
SD3	MD2	**Cys251-Cys376**	Cys(7)- Cys(14)
		Cys265-Cys300	Cys(8)- Cys(12)
		Cys274-Cys297	Cys(9)- Cys(11)
		Cys281-Cys406	Cys(10)- Cys(16)
		Cys304-Cys403	Cys(13)- Cys(15)

The bold residues were the residues present in the fully or partially conserved domains whereas rest of the cysteine residues were involved in the network formation.

Table 2 Comparison of isolates Identity, Z-dope score and UPS groups

Protein Name	Identity (%)	Template	Z-dope score	UPS group
gi_86171174	44	2YK0	-0.81	C
gi_124512768	26	2YK0	-1.65	B
gi_124505159	54	2XU0	-1.09	A
IGHvar05	44	2YK0	-0.54	B
IGHvar26	58	2XU0	-1.11	A
IGHvar34	46	2XU0	-0.74	C
RAJ116var06	42	2XU0	-0.9	B
RAJ116var19	64	2XU0	-1.22	A
RAJ116var28	41	2YK0	-0.44	C

Table 3 Structure quality estimation analysis

Protein	Structure Quality Estimation analysis Predicted Structure						
	ProCheck				G factor	Verify 3D	Errate
	Core region	Allowed region	gener allowed region	disallowed region			
gi_86171174	88.3	10.80	0.5	0.3	-0.04	86.67	73.48
gi_124512768	91.7	7.1	0.9	0.3	-0.02	82.7	73.368
gi_124505159	90.7	8.2	0.8	0.3	0.04	86.22	87.918
IGHvar05	88.5	8.8	0.8	1.9	-0.01	72.68	69.797
IGHvar26	91	8.8	0	0.3	0.04	80.8	81.88
IGHvar34	85.8	11.5	1.4	1.4	-0.14	67.25	70.361
RAJ116var06	87.8	10.6	1.1	0.5	-0.05	83.21	74.129
RAJ116var19	91.2	7.6	0.8	0.3	0.03	90.59	86.053
RAJ116var28	86.3	11.2	1.5	1	-0.09	85.35	66.042

Table 4 Comparison of RMSD values in PfEMP1 and its conserved residues

Protein Name	Full Protein		CTVLARSFADIGDIVRG KDLY (MD1)		LDYVPQYLRWFEEWA EDFCR (MD2)	
	C-alpha	Main Chain	C-alpha	Main Chain	C-alpha	Main Chain
gi_86171174	2.759	2.775	2.738	2.716	2.662	2.639
gi_124512768	1.882	1.91	1.091	1.087	1.485	1.481
gi_124505159	2.808	2.819	2.834	2.807	2.743	2.719
IGHvar05	2.475	2.502	1.988	1.983	1.955	1.955
IGHvar26	2.656	2.667	2.702	2.696	2.078	2.081
IGHvar34	2.388	2.409	1.726	1.717	1.813	1.778
RAJ116var06	2.055	2.074	1.903	1.907	1.598	1.597
RAJ116var19	2.314	2.356	2.22	2.237	1.781	1.786
RAJ116var28	2.815	2.823	2.589	2.583	2.562	2.612

Table 5 RMSD values in Å after protein structure simulation

Protein Name	UPS group	After Simulation RMSDÅ		
		C-Alpha	Main-chain	Side-chain
gi_86171174	C	3.868	3.828	4.726
gi_124512768	B	4.199	4.181	4.672
gi_124505159	A	1.094	1.137	1.499
IGHvar05	B	3.560	3.541	4.475
IGHvar26	A	1.751	1.787	2.687
IGHvar34	C	3.376	3.335	4.512
RAJ116var06	B	3.084	3.222	3.469
RAJ116var19	A	2.858	2.934	3.740
RAJ116var28	C	2.883	2.590	4.744

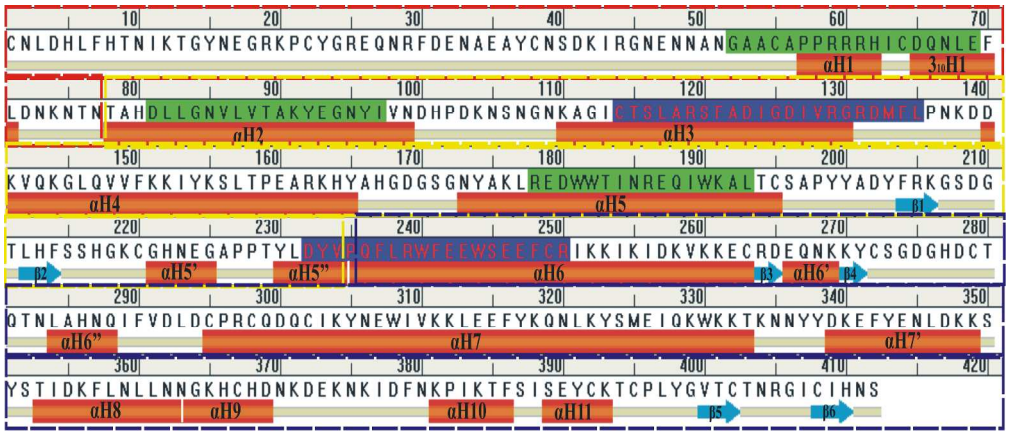

Fig. 1: Subdomain architecture with secondary structure analysis of PfEMP1 DBLα domain. The red color framed region was subdomain 1 while yellow and purple framed regions depict subdomain 2 and 3, respectively. The partially conserved regions have been highlighted with green box whereas fully conserved regions with purple box.

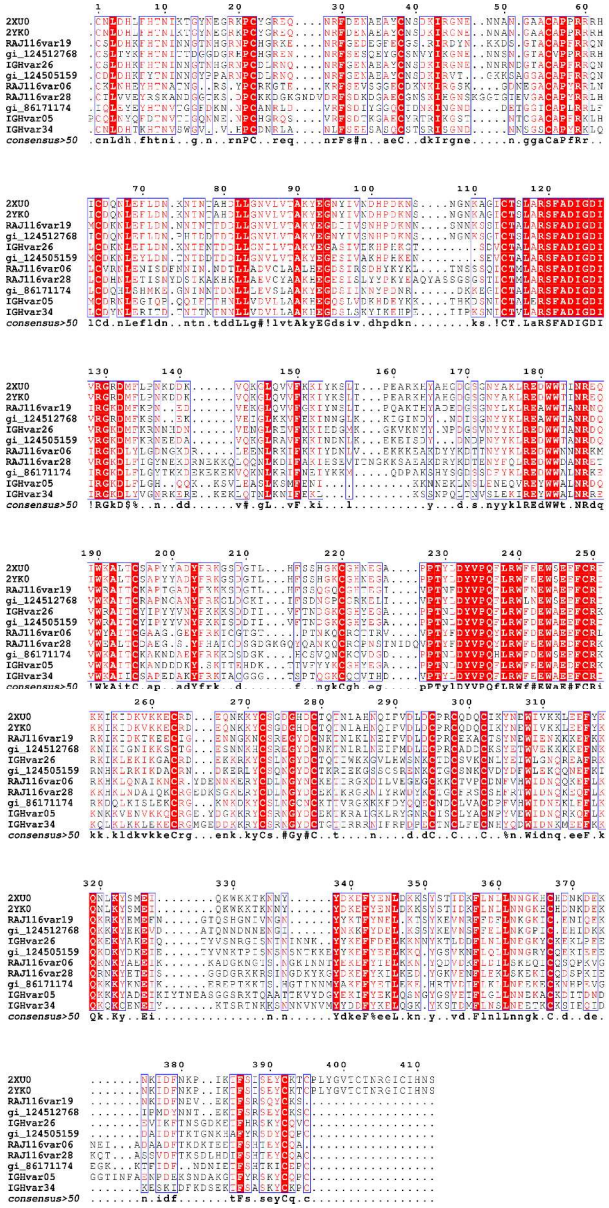

Fig. 2: Multiple sequence alignment of DBLα domain sequences from three Plasmodium isolates with template. The most conserved residues have been highlighted with red color, while partially conserved residues pink color.

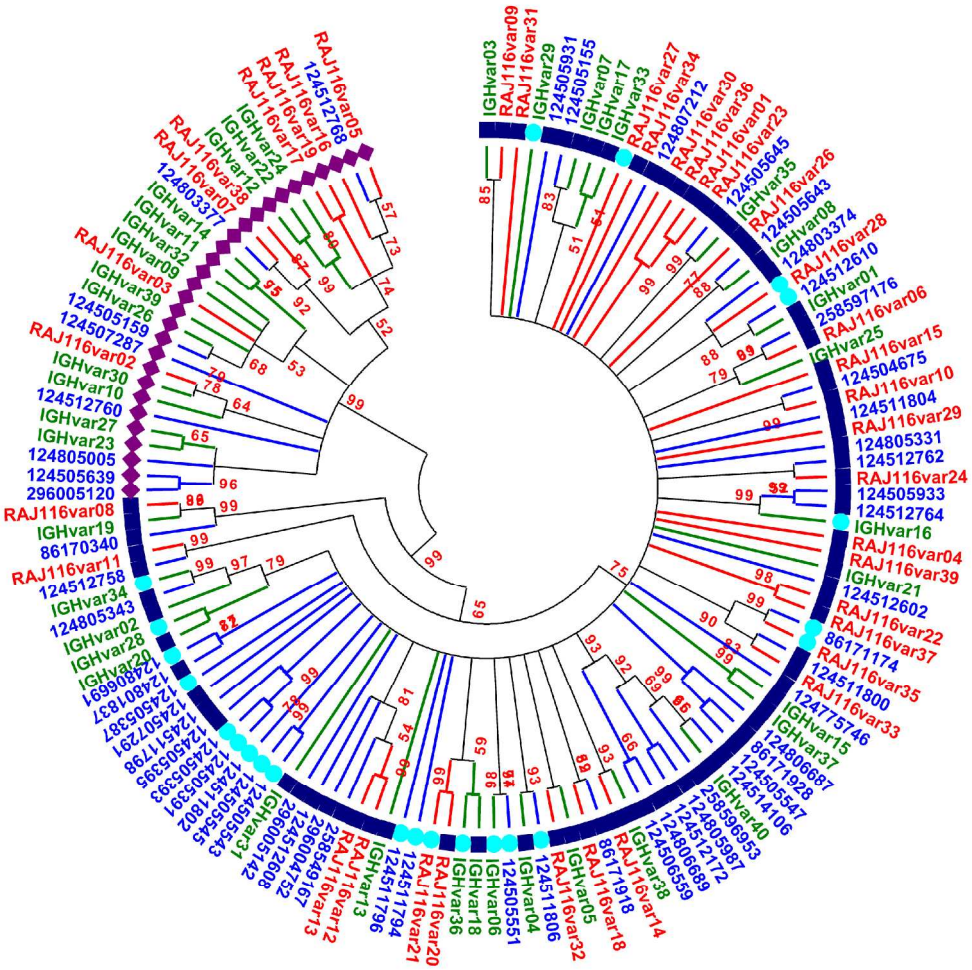

Fig. 3: Phylogenetic analysis of all three isolates of DBLα domain sequences. The 3D7 DBLα domain sequences are highlighted with blue color, RAJ116var with red and IGHvar with green color. The purple diamonds shows UPSA group whereas blue and cyan indicated UPSB and C group sequences.

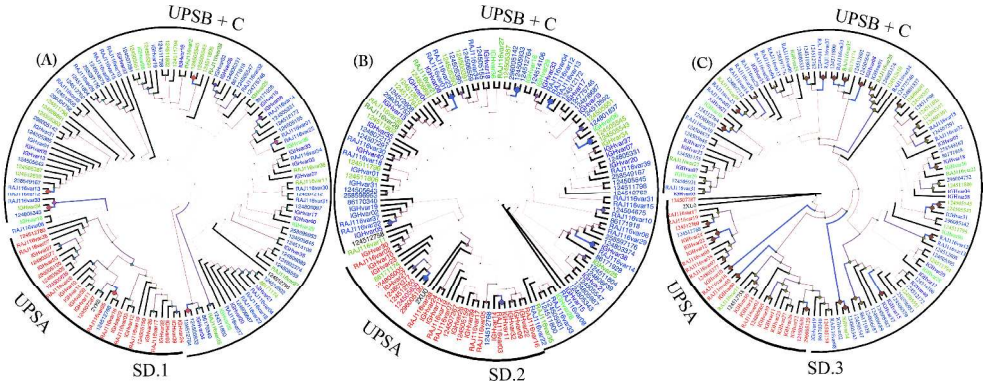

Fig. 4: Phylogenetic analysis based on subdomain wise classification of DBLα sequence. Phylogenetic analysis of subdomains SD1, SD2 and SD3 of DBLα sequences is depicted. UPSA group is shown in red color sequences from UPSB group in blue color and UPSC in green color respectively.

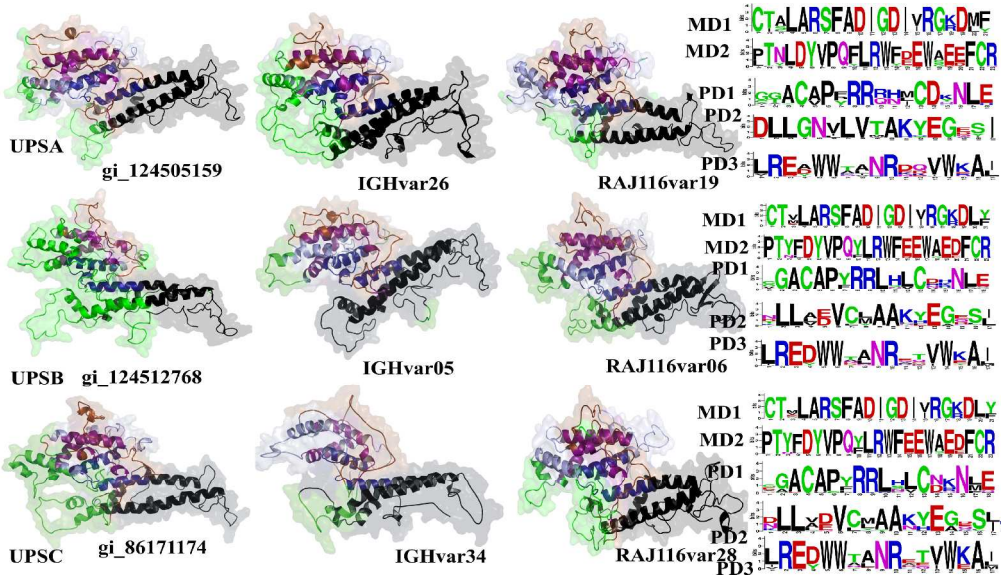

Fig. 5: Structures, their epitopes and sequence logos. The DBLα sequence consisted of 3 sub domains i.e. SD1, SD2 and SD3. SD1 has been represented in brown color, SD2 in light blue and SD3 in black color. The epitopes are seen in green color. The fully conserved domains (MD1, MD2) and partially conserved domains (PD1, PD2, and PD3) have been represented by density and deep purple colors, respectively. The sequence logos of fully conserved (MD1, MD2) and partially conserved domains (PD1, PD2, PD3) are also shown.

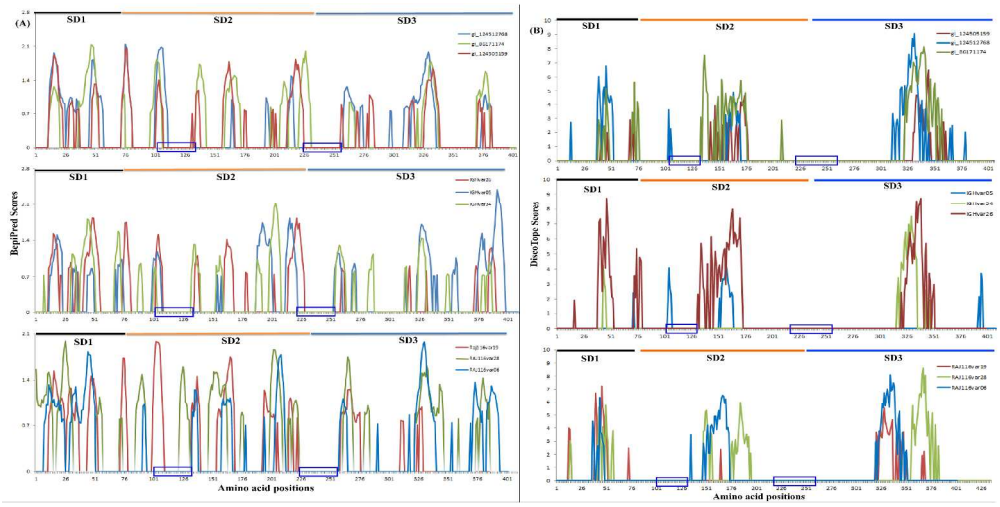

Fig. 6: Representation of B-cell Epitope Prediction analysis by BepiPred and Discotope servers for all three isolates (3D7, IGHvar, RAJ116var). (A) Linear B-cell epitope prediction by BepiPred, Values above 0.7 were considered as epitopes. (B) Conformational B-cell epitope prediction by DiscoTope and values above 1.9 has been considered as epitopes. The two highly conserved regions have been framed by blue color boxes and different domains were presented by SD1, SD2 and SD3 whereas UPSA group proteins are represented in red color, UPSB in blue and UPSC were in green color.

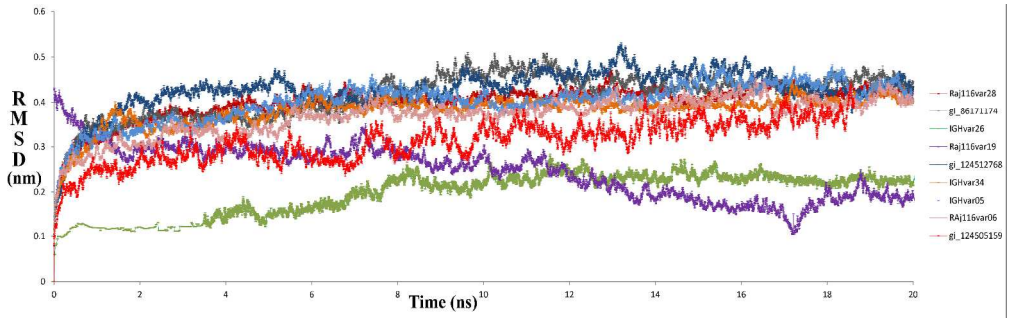

Fig. 7: Representation of RMSD (nm) with time (ns) simulation graph for nine variants from three isolates.

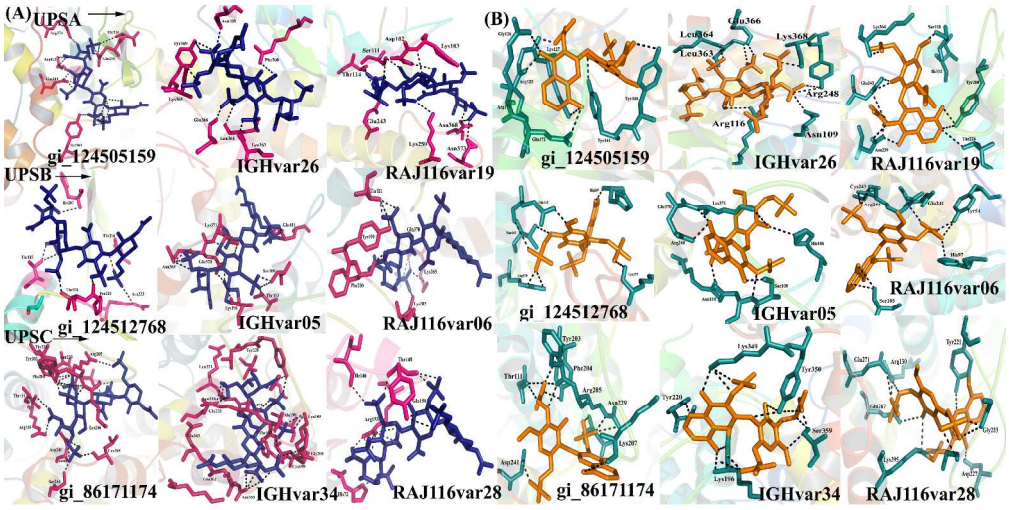

Fig. 8: Docking Interaction Studies with (A) Heparin and (B) Heparan Sulfate with all simulated DBLα domain variants proteins structures.