Epitope mapping and topographic analysis of VAR2CSA DBL3X involved in P-falciparum placental sequestration

Dahlback, Madeleine; Rask, Thomas Salhøj; Andersen, Pernille; Nielsen, Morten A; Ndam, Nicaise T.; Resende, Mafalda; Turner, Louise; Deloron, Philippe; Hviid, Lars; Lund, Ole; Pedersen, Anders Gorm; Theander, Thor G.; Salanti, Ali

Published in:
PLoS Pathogens

Link to article, DOI:
10.1371/journal.ppat.0020124

Publication date:
2006

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

Technical University of Denmark

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Epitope Mapping and Topographic Analysis of VAR2CSA DBL3X Involved in *P. falciparum* Placental Sequestration

Madeleine Dahlbäck¹wik, Thomas S. Rask²wik, Pernille H. Andersen², Morten A. Nielsen¹, Nicole T. Ndam¹,³, Mafalda Resende¹, Louise Turner¹, Philippe Deloron³, Lars Hviid¹, Ole Lund², Anders Gorm Pedersen², Thor G. Theander¹, Ali Salanti¹*

¹Centre for Medical Parasitology, University of Copenhagen and Copenhagen University Hospital (Rigshospitalet), Copenhagen, Denmark, ²Center for Biological Sequence Analysis, BioCentrum, Technical University of Denmark, Kongens Lyngby, Denmark, ³Institut de Recherche pour le Développement, Faculté de Pharmacie, Paris, France

Pregnancy-associated malaria is a major health problem, which mainly affects primigravidae living in malaria endemic areas. The syndrome is precipitated by accumulation of infected erythrocytes in placental tissue through an interaction between chondroitin sulphate A on syncytiotrophoblasts and a parasite-encoded protein on the surface of infected erythrocytes, believed to be VAR2CSA. VAR2CSA is a polymorphic protein of approximately 3,000 amino acids forming six Duffy-binding-like (DBL) domains. For vaccine development it is important to define the antigenic targets for protective antibodies and to characterize the consequences of sequence variation. In this study, we used a combination of in silico tools, peptide arrays, and structural modeling to show that sequence variation mainly occurs in regions under strong diversifying selection, predicted to form flexible loops. These regions are the main targets of naturally acquired immunoglobulin gamma and accessible for antibodies reacting with native VAR2CSA on infected erythrocytes. Interestingly, surface reactive anti-VAR2CSA antibodies also target a conserved DBL3X region predicted to form an α-helix. Finally, we could identify DBL3X sequence motifs that were more likely to occur in parasites isolated from primi- and multigravidae, respectively. These findings strengthen the vaccine candidacy of VAR2CSA and will be important for choosing epitopes and variants of DBL3X to be included in a vaccine protecting women against pregnancy-associated malaria.


Introduction

Individuals living in areas with high *Plasmodium falciparum* transmission acquire immunity to malaria over time and adults have markedly reduced risk of getting severe disease [1]. Pregnant women constitute an important exception to this rule, and this has severe consequences for both mother and child [2]. Pregnancy-associated malaria (PAM) is characterized by selective accumulation of *P. falciparum* in the intervillous blood spaces of the placenta [3,4]. The main pathophysiological consequences of PAM are delivery of low birth weight babies and maternal anaemia [5]. In areas of high parasite transmission PAM affects mainly primigravidae as immunity is acquired as a function of parity [2]. Parasite sequestration in the placenta is mediated by an interaction between chondroitin sulphate A (CSA) on the syncytiotrophoblasts and proteins expressed on the surface of infected erythrocytes [6]. VAR2CSA, a single and uniquely structured molecule belonging to the *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) family, is currently believed to be the main parasite ligand for placental binding [7]. Var2csa is markedly up-regulated in *P. falciparum* selected in vitro to bind to CSA [7] and in parasites isolated from the placenta [8]. Antibodies to the surface-expressed VAR2CSA are acquired by women exposed to malaria during pregnancy [9,10], and high levels of anti-VAR2CSA antibodies at delivery are associated with protection from low birth weight [9]. Furthermore, it has been demonstrated that targeted disruption of var2csa results in the loss of [11], or a marked reduction [12] in the ability of parasites to adhere to CSA. Based on these findings, VAR2CSA is recognized as the leading PAM vaccine candidate; however, var2csa is a polymorphic gene and the sequence variation between genes from different parasites ranges from 10%–30% at the nucleic acid level [7,13]. It is thus a major challenge for vaccine development to characterize the importance of the sequence variation and to define smaller epitopes that can be used in a vaccine to protect women against PAM. This study had two objectives. Firstly, to characterize the epitopes of the CSA-binding Duffy-binding-like (DBL) 3X domain of VAR2CSA, which are recognized by naturally acquired antibodies.
Synopsis

Pregnancy-associated malaria caused by *Plasmodium falciparum* is characterized by the accumulation of parasite-infected red blood cells in the placenta and is a major health problem in Africa. VAR2CSA is a parasite protein expressed on the surface of malaria-infected red blood cells and mediates the binding to the placental receptor, chondroitin sulphate A. It is believed that a vaccine based on VAR2CSA will protect pregnant women against the adverse effects of pregnancy-associated malaria. However, due to the size and polymorphism of VAR2CSA it is required to define smaller regions that can be included in a vaccine and to analyze the degree and consequences of sequence variation to ensure a broadly protective immune response. The authors have characterized the chondroitin sulphate A-binding DBL3X domain of VAR2CSA with regard to epitopes targeted by naturally acquired antibodies and the influence of sequence variation by bioinformatics and experimental data based on a VAR2CSA peptide array. They identify both variable and conserved surface-exposed epitopes that are targets of naturally acquired immunoglobulin gamma in pregnant women with placental malaria. These findings will be imperative for choosing epitopes and variants of DBL3X to be included in a vaccine protecting pregnant women against malaria.

Results/Discussion

Recombinant VAR2CSA DBL3X Binds CSA and the Affinity Depends on the Primary Amino Acid Sequence

It has previously been shown that DBL3X expressed on the surface of Chinese hamster ovary cells binds CSA in vitro [14]. However, it is important to test the CSA-binding properties of secreted VAR2CSA proteins produced in expression systems that could allow for larger-scale production of a vaccine. For this study, recombinant HIS-tagged proteins were produced in *Baculovirus*-transfected insect cells and binding to CSA was determined in an enzyme-linked immunosorbent assay (ELISA) system. It has previously been suggested that the FCR3 DBL3X domain binds CSA, whereas the 3D7 DBL3X domain does not [14]. We found that both variants had affinity to CSA and that the 3D7 variant exhibited the strongest binding. Binding of both FCR3 and 3D7 DBL3X was concentration-dependent (Figure 1A) and could be inhibited by soluble CSA in a dose-dependent manner (Figure 1B).

The Structure of the VAR2CSA DBL3X Domain Can Be Modeled on the Basis of the Structure of the DBL Domains of Erythrocyte-Binding Protein-175

Recently, the structures of the two DBL domains (F1 and F2) of erythrocyte-binding protein (EBA)-175 and a DBL domain in *Plasmodium knowlesi* (Pk3-DBL) were solved [15,16]. Using the crystal structure of EBA-175 F1 (Protein Data Bank code 1ZRO) as template, a three-dimensional model of DBL3X VAR2CSA was constructed by comparative modeling on the basis of the 3D7 sequence (Figure 1C). The sequence identity between DBL3X and EBA-175 F1 was 28%. From sequence alignments of DBL3X, Pk3-DBL, and EBA-175 (Figure S1) it was apparent that a number of cysteines were conserved, and ten of these from DBL3X aligned with cysteines which form disulfide bridges in the determined structures of EBA-175 [15] and Pk3-DBL [16]. In addition, we identified 34 buried hydrophobic residues in EBA-175 F1, EBA-175 F2, and Pk3-DBL, which corresponded to hydrophobic residues buried in the DBL3X model (Figure S1). Compared to EBA-175 F1, a number of insertions were found in DBL3X, and the majority of these were predicted to have coil secondary structure (Figure 1C). One of the insertions in DBL3X (L1: R56-I63) was found to align structurally next to a region in Pk3-DBL, which is described as a flexible linker with an experimentally determined proteolytic cleavage site [16]. A second insertion (L2: N1417-E1430) was aligned in a loop region where two residues were missing structural information in the structure of EBA-175 F1 [15], which also indicates high flexibility. Thus, the alignment of DBL3X to the solved DBL domain structures seems to match characteristics such as disulfide-bridges, stabilizing hydrophobic interactions, and flexible loop regions. These findings suggest that VAR2CSA DBL3X has the same basic structure as the solved DBL structures [15–17] in spite of the extensive sequence variation. The ability of proteins to form similar structures despite marked sequence variation has also been described for the VSG molecules covering the surface of *Trypanosoma* [18,19].

Sequence Variation within the var2csa Family Is Present as Small Hypervariable Blocks and Parasite Diversity Is Similar on a Local and Global Scale

*Var2csa* is a relatively conserved *var* gene carried by all *P. falciparum* genomes; however, sequence polymorphisms are present in the gene. In a previous study, it was established that *var2csa* is transcribed at high levels by placental parasites isolated at delivery from Senegalese women [8,20]. Using cDNA from 24 placentas from that study, the region encoding DBL3X of *var2csa* was cloned and sequenced. A multiple alignment of 43 sequences showed that the average nucleotide diversity was low ($\pi = 0.0048 \pm 0.0037$, see Materials and Methods for details) reflecting a limited inter-parasite diversity in these isolates collected from a geographically small and well-defined area of West Africa. To test whether the Senegalese placental DBL3X sequences represented a monophyletic group compared to non-Senegalese DBL3X sequences, a phylogenetic tree, which included VAR2CSA DBL3X sequences available in GenBank, was constructed (Figure 2). These non-Senegalese sequences included four lab-strain parasites with different geographic origins (DD2 from Laos, MC from Thailand, IT4 from Brazil, and 3D7 with unknown origin) and 21 sequences from a sequencing study from Malawi [21]. The four database sequences and the Malawi sequences were scattered evenly in the tree indicating that the Senegalese sequences were representative for the *var2csa* repertoire in general and that parasite nucleotide diversity is similar on a local and global scale. It is also apparent that there is no clear subgrouping within the DBL3X sequences. The protein alignment of the DBL3X sequences (Figure 3A) suggested that DBL3X could be divided into four relatively conserved regions (C1–4) separated by three shorter variable regions (V1–3). When the variable regions were mapped to the 3-D model (Figure 3B), V1 and
V3 were part of flexible loops, whereas V2 included another flexible loop but also extended into a helical region. The length of V1 and V3 varied between sequences and the 3D7 sequence was relatively short in both regions.

**VAR2CSA Sequence Motifs Can Be Linked to the Parity of the Infected Women**

We have previously shown that the expression of certain var genes is associated with severe malaria in young children [22] and suggested that var gene expression is hierarchically structured. This could occur because the progeny of parasites expressing var gene products that mediate the most efficient sequestration outgrow the progeny of parasites expressing a molecule mediating less efficient binding [22–24]. A similar process could shape the expression of molecules mediating binding in the placenta. If that was the case, it would be predicted that a systematic difference between molecules mediating binding in primigravidae and multigravidae women could be detected in areas of high *P. falciparum* transmission where women are exposed to several parasite clones during pregnancy [25]. This was addressed by calculating the Kullback-Leibler distance between 21 VAR2-CSA sequences originating from primigravidae and 21 sequences from multigravidae. The overall cumulated Kullback-Leibler distance (*DKL*) between sequences from the primigravidae and the multigravidae was higher than for
randomly chosen groupings of the sequences \( p = 0.0075 \). The DKL was also calculated for each position in the alignment to determine which polymorphisms contributed to the difference between the sequence of parasites from primi- and multigravidae women and visualized in a Kullback-Leibler sequence logo (Figure 4A). This implicated two stretches of amino acids in positions 135–175 and 233–245 located in the V2 and V3 regions, respectively. The region from position 158 to 162 was of special interest since the motif "EIEKD" was mainly found in primigravidae and the motif "GIEGE" mainly in the multigravidae (Table 1). Intermediate motifs like "(G/E)IERE" where the fourth position had changed from lysine to arginine were also found, implying that the following evolutionary pathway may have been operating: lysine (AAA or AAG) $\rightarrow$ arginine (AGG) $\rightarrow$ glycine (GGG). The change from glutamate and lysine/arginine, which have large charged side chains, into glycine without a side chain, could result in marked functional and antigenic changes. Thus, it was interesting that positions 1, 4, and 5 in the motif (position 158, 161, and 162) were predicted to be surface-exposed in the structural model, which was also the case for all of the other amino acid positions in V2 that differed significantly in the Kullback-Leibler sequence logo (Figure 4B and 4C). The finding that parasites expressing the EIEKD motif were more prevalent in primigravidae than in multigravidae women (Table 1, Chi-square test: \( p < 0.001 \)) could indicate that these parasites have a biological advantage in women experiencing their first pregnancy and that parasites expressing the other motifs (G/EIERE or GIEGE) have an advantage in women who have been pregnant before. This could arise because parasites carrying the EIEKD motif are the most efficient mediators of binding and therefore dominate in women with limited
immunity against PAM. As immunity develops against these parasite forms, parasites expressing other motifs that are less efficient in binding but not serologically cross-reactive take over. Interestingly, a monoclonal Ghanaian field isolate undergoing full genome sequencing at the Sanger Institute has two copies of var2csa, one with the EIEKD motif and one with the GIEGE. Although the functional background for the observed phenomenon is presently not clear, the systematic sequence variation at positions predicted to be surface-exposed between parasites from primi- and multigravidae women strengthen the concept that VAR2CSA is the main parasite ligand for sequestration of malaria parasites in the placenta.

VAR2CSA Is Under both Positive and Purifying Selection

A recent study has suggested that sequence polymorphisms in a region of VAR2CSA upstream to DBL3X largely are due to positive natural selection pressure [13]. To further investigate the nature of sequence diversity in var2csa, the dN/dS ratios (dN, rate of non-synonymous mutations per non-synonymous site and dS, the rate of synonymous mutations per synonymous site) for DBL3X were calculated on the basis of the sequenced DBL3X domains. A dN/dS < 1 indicates that the position is under negative or purifying selection pressure leading to conservation of the residue, while dN/dS > 1 suggests positive or diversifying selection pressure, and suggests that amino acid changes are evolutionarily advantageous at the position. Purifying selection was mainly found in the conserved regions C1–4 and it was especially pronounced in the C2 and C4 regions, although single sites were observed to be under diversifying selection (Figure 5A). Several blocks appeared to be under strong diversifying selection and these appeared most prominently in regions V1, V2, and V3. It is interesting to note that residues under diversifying selection mainly were situated in regions predicted to be surface-exposed and concentrated on one side of the molecule (Figure 5D and 5E). The two DBL domains of EBA-175 are predicted to form a reverse handshake dimer with the F1 domain of each molecule interacting with F2 of the other [15]. In this four-domain DBL structure, we replaced one of the EBA-175 F1 domains with our VAR2CSA DBL3X model (Figure 5F). It was noticeable that the largely conserved C2 and C4 regions of DBL3X were predicted to take part in the lining of the central cavity of the
four-domain structure and form the region next to the cavity facing the membrane in native configuration. The model presented here is unlikely to faithfully reflect the structure of the native molecule, but the positioning of conserved DBL3X regions in the model makes biological sense, and the finding underscores the need to obtain knowledge about possible interactions between VAR2CSA DBL domains. The lack of amino acid positions under diversifying selection in the regions adjacent to the cavity might be due to the possibility that these sites are involved in ligand binding and thus functional constraint. Another explanation could be that the regions forming the predicted cavity and the area predicted

Table 1. Sequence Signature in DBL3X Position 158–162 of VAR2CSA Expressed by 42 Parasites Infecting either Primigravidae or Multigravidae

<table>
<thead>
<tr>
<th>Sequence Motif in DBL3X</th>
<th>Primigravidae (n = 21)</th>
<th>Multigravidae (n = 21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EIEKD</td>
<td>16*</td>
<td>4</td>
</tr>
<tr>
<td>EIERE/EIEGE/GIERE</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>GIEGE</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>Gap</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

The motif “EIEKD” is overrepresented in primigravidae and underrepresented in multigravidae. The intermediate motifs and “GIEGE” are mainly present in multigravidae. One of the sequences from multigravidae contains a gap in the region.

*Significantly higher represented in primigravidae, p < 0.001, Chi-square.

doi:10.1371/journal.ppat.0020124.t001
to face the membrane are not accessible for antibodies in natively folded molecules.

Recombination Is a Factor in the Generation of var2csa Sequence Variation

Previous studies have reported that frequent recombination events generate sequence diversity in the PfEMP1 family [26–28]. To determine the role of recombination for DBL3X sequence variation, we estimated the population recombination rate ρ defined for partially inbreeding haploid species by the compound parameter $2N_e r (1 - f)$, where $N_e$ represents the effective population size, $r$ is the per-generation cross-over recombination rate per base pair (bp), and $f$ is the inbreeding coefficient. Variations in $\rho$ across DBL3X correspond to variations in the recombination rate $r$, as $N_e$ and $f$ are constant for the dataset. Two recombinational hotspots were observed at bp positions 138–178 and 704–730 (Figure 5B). The two best defined recombination breakpoints were present in the C1/V1 borderline at bp 177–179 and within V3 at bp 728–730. Both the V1 and V3 region harbored major deletions/insertions in several of the sequences, which may

Figure 5. Sequence Variation and B Cell Epitopes in DBL3
(A) dN/dS ratio per amino acid based on an alignment of 47 DBL3X sequences. dN/dS ratio higher than the dashed line indicates that the position is under positive selection. V1–V3 and C1–C4 are indicated with solid bars. It should be noted that the inference of selection pressure is less reliable in positions with gaps, which are mainly found in the variable regions (see Figure 3).
(B) The population recombination rate $\rho = 2N_e r (1 - f)$ estimated per bp across DBL3X. Circles indicate the SNPs for which $\rho$ is estimated, linear interpolation is shown between SNPs. Two hotspots corresponding to a local raise in $r$ are observed in association with the variable regions V1 and V3.
(C) B cell epitope predictions of the 3D7 sequence performed by the BepiPred server. DD2 and IT-4 were used to fill 3D7 sequence gaps.
(D) Cartoon representation of the DBL3X model showing amino acid positions under diversifying selection in red (dN/dS ratios > 1).
(E) DBL3X model with surface topography prediction showing amino acid positions under diversifying selection in red (dN/dS ratios > 1).
(F) The homo-dimer of the two domains of RII of EBA-175 has been shown to form a handshake structure with F1 of one molecule dimerizing with F2 of another molecule and vice-versa. Here VAR2CSA DBL3X has been superimposed on the F1 domain of one of the EBA-175 molecules in the dimer. The F1 domain is shown in green, F2 domains are blue, and the linker regions are gold. Amino acids under diversifying selection in DBL3X are shown in red (dN/dS ratios > 1). It is clear that most residues inward of the dimer are not under positive selection.
(G) DBL3X model showing residues with BepiPred scores higher than 0.9 in green. Locations of V1–V3 are indicated on the model. GB indicates the glycan-binding site of EBA-175.
doi:10.1371/journal.ppat.0020124.g005
have arisen as a consequence of unequal cross-over during recombination at the hotspots. Unequal cross-over results in either deletions or insertions of variable number of tandem repeats (VNTR), and the DBL3X V1 region did indeed contain a high amount of VNTR. The loop region of V2 contained a small VNTR insert in some sequences, while VNTR were less obvious in the V3 region. In a recent study of the P. falciparum Chromosome 3, high recombination rates were found in sub-telomeric regions, and African P. falciparum strains showed a much higher population recombination rate than strains from other regions of the world [28]. The overall population recombination rate for the DBL3X region was estimated to \( \rho = 0.54 \) per bp (95% CI: 0.41–0.90). This is in accordance with a recent report for a VAR2CSA region upstream to DBL3X around interdomain 1, where the rate was estimated to 0.71 per bp [13], and in agreement with recombination rates in Chromosome 3 of African parasite lines summarized as \( \rho > 0.1 \) per bp [28]. The high population recombination rate in DBL3X combined with the observed sequence variation adjacent to the detected recombination hotspots, suggests that recombination is an important factor in generating sequence variation. The V1 region that seems to be most strongly affected by recombination is predicted to form a structurally unrestricted flexible loop allowing for sequence variation, and it is possible that the whole V1 region is under diversifying selection pressure exerted by the immune system, even though this could not be predicted by the dN/dS method due to gaps. This notion is supported by the findings discussed below, showing that this region is part of a major B cell epitope.

**VAR2CSA DBL3X B Cell Epitope Prediction**

VAR2CSA epitopes exposed on the surface of the protein and accessible for immunoglobulin gamma (IgG) binding could be under diversifying selection resulting in escape mutations and high dN/dS ratios, whereas residues involved in protein folding, stability, and anchoring could be under purifying selection with low dN/dS ratios. To predict linear B cell epitopes, the 3D7 sequence was submitted to the BepiPred server [29] and seven epitopes were predicted within the DBL3X sequence (Figure 5C). Some of these epitopes were located in areas with high dN/dS ratios and there was a weak but statistically significant association between the BepiPred score and the dN/dS ratio (Pearson’s \( r = 0.18 \) and \( p = 2.5 \times 10^{-3} \)). The reason for this weak association could be that antibody epitopes are situated in regions that are functionally constrained or that positive selection pressure is also driven by other forces like MHC-2 binding and T helper cell activation as found for HIV-1 [30]. Furthermore, the BepiPred algorithm predicts linear epitopes, and some of these could be located in parts of the molecule that are not accessible to antibodies in the native folded molecule. Nevertheless, most of the predicted epitopes (Figure 5G) were situated in surface-exposed loop regions of the model, and one of the highest scoring epitopes was in the V1 region. Residues that align directly to the glycan-binding residues of EBA-175 F1 and to the Duffy antigen receptor for chemokines (DARC) binding site in Pκx-DBL were not predicted to be part of epitopes. However, a part of the V2 region in proximity to the putative glycan-binding loop was predicted to be targeted by antibodies and had high dN/dS values.

**Fine Epitope Mapping of VAR2CSA Antibodies Acquired during Pregnancy**

To verify the above bioinformatical predictions, we evaluated the fine specificity of naturally acquired human antibodies to VAR2CSA DBL3X in a peptide array consisting of 442 overlapping 31mer peptides covering exon 1 of VAR2CSA of the 3D7 sequence. Antibody reactivity to individual amino acids was assigned on the basis of an algorithm based on the observation that a major part of antibody binding motifs in a set of conformational epitopes are from two to seven amino acids long, containing either two or three defined residues spaced by undefined amino acids [31]. The VAR2CSA of 3D7 contains 31,149 such motifs and each was assigned an average PepScan value by adding the measured reactivity from the 31mers in which the motif was present, and dividing with the number of times the motif occurred. The method, described in Materials and Methods, was validated by testing serum from rabbits immunized with a VAR2CSA construct, which showed that both the measurements based on individual peptide readings and the algorithm described above, mapped antibody responses to regions present in the antigen used for immunization (Figure S2). Furthermore, there was a high concurrence between antibody peaks defined by the reactivity of the individual peptides and the peaks defined by the algorithm defining single amino acid scores (Figure S2). Plasma from individuals not exposed to malaria did not react with any of the peptides in the array (unpublished data). The IgG reactivity in the plasma of eight Ghanaian women with a known history of a placental malaria infection was analyzed and the peptide array data was visualized on the DBL3X model (Figure 6). The regions with the highest reactivity were on the side of the domain where glycan-binding is found in EBA-175 F1 (Figure 6A and 6B) and therefore IgG reactivity was visualized on a model positioned with this side in the front (Figure 6A and 6C–I). It was clear that the majority of the individuals had specific IgG against the variable regions, V1 or V2. V3 is partly deleted in 3D7, and IgG reactivity could thus not be measured in the peptide array based on the 3D7 sequence. Remarkably, a short \( z \)-helix (Figure 6A, arrow 1) in proximity to the loop for glycan-binding in EBA-175 F1 showed the highest antibody reactivity in all serum samples, despite the fact that this region had very low dN/dS ratios (Figure 5A, positions 120–132). Another \( z \)-helix was also often recognized by antibodies (Figure 6A, arrow 2). This helix was predicted to contain a B cell epitope by the BepiPred algorithm and had polymorphic residues (Figure 5A and 5C, positions 251–281). The region corresponding to the loops containing the EBA-175 F1 glycan-binding sites was not recognized by any of the serum samples. Conserved regions (Figure 6A, arrow 1) targeted by naturally acquired IgG are of considerable interest in the search for vaccine constructs which could elicit a broad protective immune response.

**PepScan Analysis of Affinity-Purified Antibodies**

During infection, VAR2CSA will be degraded and antibodies will be acquired against epitopes that are not accessible for antibodies when the protein is in its natural conformation. It is therefore possible that some of the antibody reactivities measured in the peptide array were directed against such epitopes. To address this question, analysis was performed on plasma, which had been affinity
purified on recombinant DBL3X or antibody-depleted by incubation with erythrocytes infected with VAR2CSA-expressing parasites.

Plasma from a rabbit immunized with recombinant DBL3X and plasma from women who had suffered a placental infection were affinity purified on the recombinant CSA-binding DBL3X protein and analyzed by the peptide array. Before affinity purification, the plasma pool from Ghanaian women showed reactivity corresponding to eight peaks distributed throughout the domain (Figure 7A). In contrast, the reactivity of the affinity-purified IgG was concentrated to three peaks (SE1–SE3) in the C1, V1, and C2/V2 regions. The immunized rabbit had strong IgG reactivity against the epitope in the C2/V2, which was also affinity purified from the female plasma pool (Figure 7B), but the rabbit had not raised an IgG response against the epitopes in C1 and V1. A plasma pool from Tanzanian women was also analyzed and in this case surface reactive antibodies were depleted from the pool by incubation with VAR2CSA-expressing infected erythrocytes. In this pool the main reactivity was against the surface-exposed epitopes (SE1–SE3) defined in the Ghanaian plasma, and the depletion experiment indicated that absorption with infected erythrocytes caused a marked reduction in the reactivity against these epitopes (Figure 7C). These findings indicate that the three identified regions were accessible to antibodies on the native protein and that the folding of the Baculovirus-produced recombinant protein was close to the natural configuration. SE2 and SE3 corresponded to loop regions V1 and V2, which are both protruding from the structure of the DBL3X model (Figure 7D). Interestingly, SE1 and SE3, which are located in separate parts of the primary structure of the domain, form a continuous region.
Figure 7. Reactivity of Plasma Affinity Purified on Recombinant DBL3X and Human IgG Depleted for Surface Reactivity by Incubation with Infected Erythrocytes Expressing VAR2CSA

(A) PepScan IgG reactivity in a Ghanaian pool of pregnancy plasma before (red) and after (blue) affinity purification on recombinant DBL3X protein. Variable regions V1 and V2 are indicated with a black line. Arrows indicate three surface-exposed regions (SE). The x-axis is amino acid position in 3D7 VAR2CSA.

(B) PepScan IgG reactivity in plasma from a rabbit immunized with DBL3X before (red) and after (blue) affinity purification on recombinant DBL3X protein. Arrow indicates the surface-exposed region. N- and C-terminal parts of the recombinant proteins did also appear to be surface-exposed; this could be due to improper folding of the N- and C-terminal parts of the protein.

(C) PepScan IgG reactivity in a Tanzanian pool of pregnancy plasma before (red) and after (blue) depletion of antibodies directed against surface-exposed VAR2CSA regions by incubation with infected erythrocytes expressing VAR2CSA on the surface. Arrows indicate three regions where the reactivity was markedly reduced after depletion largely corresponding to SE1, SE2, and SE3 on (A).

(D) DBL3X model showing the location of surface-exposed epitopes, SE1 (blue), SE2 (red), and SE3 (green).

(E) DBL3X surface topography showing SE1–SE3. From the model it is apparent that the SE1 and SE3 form a continuous epitope.

(F) VAR2CSA DBL3X superimposed on the F1 domain of one of the molecules in the EBA-175 dimer. The F1 domain is shown in green, F2 domains are blue, and the linker regions are gold. On the DBL3X domain the three experimentally verified surface-exposed regions are shown in similar coloring as (D and E) The three regions are mainly predicted to be on the opposite side of the central cavity of the dimer, and parts of SE3 correspond to amino acid positions directly involved in the dimerization of EBA-175.

doi:10.1371/journal.ppat.0020124.g007
on the surface of the predicted DBL3X protein structure (Figure 7E). The model also predicts that all surface-exposed sites are located on one part of the domain indicating that the other parts are buried or engaged in the intact PEEMP1 molecule expressed on the surface of the infected erythrocyte (Figure 7E). Unexpectedly, the highly conserved part of C2, which was well recognized in the peptide array by all women, corresponded to a part of the surface-exposed epitope SE3.

When exchanging the F1 domain of one of the EBA-175 molecules in the EBA-175 dimer with VAR2CSA DBL3X and mapping SE1–SE3 to the model, it is apparent that the surface-exposed regions are on the opposite side of the central cavity of the dimer (Figure 7F). However, the part of DBL3X that may be directly involved in the dimerization extends into the SE3 region (shown in green), indicating that in this model the potential dimerization motifs are accessible for antibodies. EBA-175 is suggested to dimerize upon ligand interaction and it could be that SE3 was “unengaged” due to the lack of CSA during antibody depletion. If several domains need to interact to form a buried CSA-binding site, antibodies targeting conserved residues, like the region in C2 that forms parts of SE3, could function by inhibiting the dimerization of domains.

Our results demonstrate that both conserved and polymorphic surface-exposed regions are targets for VAR2CSA DBL3X antibodies acquired during pregnancy by malaria-infected women. This opens for vaccine, strategies similar to those predicted to be generated from the sequence polymorphisms present in regions SE1 and SE2. The human antibody pools used in this study to identify surface-exposed antigenic targets inhibit parasite binding to CSA in vitro (unpublished data). However, the molecular targets of the inhibitory antibodies have not been identified and knowledge is required about the antigenic targets for antibodies on the other VAR2CSA DBL domains. The high similarity between the DBL structures of EBA-175 [15], Pkz-DBL [16], and the VAR2CSA DBL3X model suggests that the DBL structures in Plasmodium are relatively conserved and that the antigenic characteristics of the DBL3X might be comparable to those of the remaining VAR2CSA DBL domains. However, a more comprehensive analysis of sequence variation, antibody epitopes, and structure of the VAR2CSA DBL domains not belonging to DBL3X is needed to establish the extent of the structural conservation between the domains.

Development of PAM vaccines requires a much better understanding about the molecular interaction between placental parasites and the ligand on the syncytiotrophoblasts, as well as knowledge about the fine specificity of the targets for antibodies inhibiting binding. Native PEEMP1 molecules are difficult to isolate and with current technologies it is difficult to produce correctly folded recombinant material in quantities allowing structure elucidation by crystallography. In this paper, we have combined in silico methods such as model building and sequence analysis and the analysis of antibody reactivity to obtain new information and generate hypotheses about the structure and functional relationship of VAR2CSA.

Materials and Methods

Cloning and expression of VAR2CSA domains. DBL3X and DBL4e of VAR2CSA were amplified from FCR3 and 3D7 genomic DNA with the following primers: FCR3 DBL3X – 5' CG GAA GTA TCG ACC AAT ATT AAT AAA AGT GAA and 3' ATT TGC GGC GCC CAG CAT TAT TAT ATT TGT A; 3D7 DBL4e - 5' CG GAA GTA TCG ACC AAT ATT AAT AAA AGT GAA and 3' ATT TGC GGC GCC CAG CAT TAT TAT ATT TGT A. The amplification reactions were digested with EcoRI and NotI and ligated into either the Baculovirus vector, pAcGP67-A (BD Biosciences, http://www.bdbiosciences.com), which was modified to contain the V5 epitope upstream of a histidine tag in the C-terminal end of the constructs. Linearised Bacpak6 Baculovirus DNA (BD Biosciences) was co-transfected with pAcGP67-A into Sf9 insect cells for generation of recombinant virus particles.

Recombinant protein was purified on Co2+ metal-chelate agarose columns as secreted histidine-tagged proteins from the supernatant of infected High-Five insect cells. Recombinant protein was purified on Co2+ metal-chelate agarose columns as secreted histidine-tagged proteins from the supernatant of infected High-Five insect cells.

CSA binding assay. Binding to CSA (C9819, Sigma-Aldrich, http://www.sigmaaldrich.com) was determined in an ELISA system. ELISA plates (Falcon 351172) were coated overnight with CSA (50 µg/ml) in PBS at 4°C. Coating with 1% BSA in PBS (blocking buffer) was used as negative control. Plates were incubated with blocking buffer for 1 h at room temperature (RT) to inhibit non-specific adsorption to the plate. VAR2CSA proteins were diluted in blocking buffer (1–10 µg/ml protein), added to the wells, and incubated for 1 h at RT. For the inhibition assays, proteins (7 µg/ml) were pre-incubated with different concentrations of soluble CSA for 30 min. Plates were washed between times in PBS, and specific binding was visualized by adding an HRP-conjugated antibody (R960-25, Invitrogen, http://www.invitrogen.com) targeting the V5 epitope of the constructs. Plates were incubated for 1 h with the anti-V5 antibody diluted 1:3000 in blocking buffer. The color reactions were developed for 15 min by the addition of o-phenylenediamine substrate and stopped by adding 2.5 M H2SO4. The optical density was measured at 490 nm.

Cloning and sequencing of placental var2csa genes. All DBL3X sequences were obtained from cDNA, whereas DBL2X and the overlap region of DBL4e and DBL5e of isolated mRNA from genomic DNA of placental parasites. In brief, parasites were dissolved in Trizol LS (Invitrogen) and RNA was prepared according to the manufacturer’s instruction. RNA pellets were dissolved in 10 µl of RNase-free water and treated with DNaseI (Sigma-Aldrich) for 25 min at RT, followed by 10 min heat inactivation at 65 °C. All RNA samples were subsequently tested in real-time PCR for contamination with genomic DNA using a primer set for the housekeeping gene, seryl-tRNA synthetase (Superscript II, Invitrogen) and random hexamer primers as described by the manufacturer. Following primer sets were used for cloning DBL3X from cDNA into either the Baculovirus vector, pAcGP67-A (BD Biosciences) or the pCR2.1-TOPO vector (Invitrogen): p505 5' CG GAA GTA TCG ACC AAT ATT AAT AAA AGT GAA and p510 3' ATT TGC GGC GCC CAG CAT TAT TAT ATT TGT A. The amplification reactions were digested with EcoRI and NotI and ligated into either the Baculovirus vector, pAcGP67-A (BD Biosciences) or the pCR2.1-TOPO vector (Invitrogen): p505 5' CG GAA GTA TCG ACC AAT ATT AAT AAA AGT GAA and p510 3' ATT TGC GGC GCC CAG CAT TAT TAT ATT TGT A. The amplification reactions were digested with EcoRI and NotI and ligated into either the Baculovirus vector, pAcGP67-A (BD Biosciences) or the pCR2.1-TOPO vector (Invitrogen): p505 5' CG GAA GTA TCG ACC AAT ATT AAT AAA AGT GAA and p510 3' ATT TGC GGC GCC CAG CAT TAT TAT ATT TGT A. The amplification reactions were digested with EcoRI and NotI and ligated into either the Baculovirus vector, pAcGP67-A (BD Biosciences) or the pCR2.1-TOPO vector (Invitrogen): p505 5' CG GAA GTA TCG ACC AAT ATT AAT AAA AGT GAA and p510 3' ATT TGC GGC GCC CAG CAT TAT TAT ATT TGT A.

DNA-free RNA samples were used for synthesis of cDNA by reverse transcriptase (Superscript II, Invitrogen) and random hexamer primers as described by the manufacturer. Following primer sets were used for cloning DBL3X from cDNA into either the Baculovirus vector, pAcGP67-A (BD Biosciences) or the pCR2.1-TOPO vector (Invitrogen): p505 5' CG GAA GTA TCG ACC AAT ATT AAT AAA AGT GAA and p510 3' ATT TGC GGC GCC CAG CAT TAT TAT ATT TGT A. The amplification reactions were digested with EcoRI and NotI and ligated into either the Baculovirus vector, pAcGP67-A (BD Biosciences) or the pCR2.1-TOPO vector (Invitrogen): p505 5' CG GAA GTA TCG ACC AAT ATT AAT AAA AGT GAA and p510 3' ATT TGC GGC GCC CAG CAT TAT TAT ATT TGT A. The amplification reactions were digested with EcoRI and NotI and ligated into either the Baculovirus vector, pAcGP67-A (BD Biosciences) or the pCR2.1-TOPO vector (Invitrogen): p505 5' CG GAA GTA TCG ACC AAT ATT AAT AAA AGT GAA and p510 3' ATT TGC GGC GCC CAG CAT TAT TAT ATT TGT A. The amplification reactions were digested with EcoRI and NotI and ligated into either the Baculovirus vector, pAcGP67-A (BD Biosciences) or the pCR2.1-TOPO vector (Invitrogen): p505 5' CG GAA GTA TCG ACC AAT ATT AAT AAA AGT GAA and p510 3' ATT TGC GGC GCC CAG CAT TAT TAT ATT TGT A.
regions. All PCR products were cloned into the pcCR2.1-TOPO vector and the inserts sequenced on a 3100-Avant Genetic Analyzer (Applied Biosystems, http://www.appliedbiosystems.com). The origin of parasites is described in [20].

Phylogenetic reconstruction. The alignment of 45 placental and four database VAR2CSA DBL3X sequences, covering the 3D7 amino acid positions 1217–1255, was constructed using RevTrans [36], and subsequently manually corrected for errors. To cover more of the DBL3X domain, an alignment of 17 database sequences was constructed in the same way, covering the 3D7 positions 1327–1355. For the phylogenetic tree, 21 Malawian sequences [21] were aligned with the sequences mentioned above, again using RevTrans and manual correction, resulting in a 690-bp alignment. The program MrModeltest version 2.2 [57] was used to find the most appropriate nucleotide substitution model based on the Akaike information criterion [38]. Phylogenetic trees based on the above alignments were constructed using Bayesian inference by the program MrBayes version 3.1.1 [39]. In all cases Markov chain Monte Carlo (MCMC) sampling was performed for 10,000,000 generations with eight chains. Convergence was confirmed by comparing the results of two independent runs. Burn-in was determined using Tracer [40] and 50% majority rule consensus trees were constructed.

Model fitting and Akaike weighted dN/dS average. The program codeml from the PAML package version 3.14 [41] was used to fit a range of codon-based evolutionary models to the VAR2CSA DBL3X region using the alignments and Bayesially mentioned above. All 11 codon-based models were tested using codeml: M0, M1a, M2a, M3 (with either 3, 4, 5, 6, or 7 site categories), M5, M7, and M8 [42–44]. All 11 models were fitted using the F3x4 (different nucleotide frequencies for each codon position) approach for estimating codon frequencies. Convergence was confirmed by comparing the results of several runs started from different parameter values.

The Akaike information criterion (AIC) was used to assess model fits [38,45,46]. Briefly, AIC estimates the expected relative Kullback-Leibler distance (i.e., AIC is an estimate of the amount of information that is lost when a given model is used to approximate the full truth). AIC is a function of the maximized log-likelihood (lnL) and the number of estimated parameters (K) for a model. Specifically, $AIC = -2\ln L + 2K$ where lower AIC values indicate better models. From AIC it was possible to compute Akaike weights, which can be interpreted as the conditional probability of the model given the data and the set of initial models [38,46]. On this basis, dN/dS ratios for codon positions were calculated as an average of the dN/dS ratios estimated from each of the 11 models, weighted by the Akaike weights for the corresponding model.

Recombination and mutation rates, diversity, and sequence logo creation. The population recombination rate $\rho$ was estimated for the VAR2CSA DBL3X domain in LDhat version 2.0 [47], which is based on population genetics uses coalescent methods specially adapted to account for the possibility of recurrent or back mutation and for an A1-rich genome such as that of P. falciparum [48]. As argued by Mu et al. [28], the coalescent approach can, for particular haploid species such as P. falciparum, be interpreted as the compound parameter $\rho = 2Ner(1-r)$, where $N$ represents the effective population size of the DBL3X population, $r$ denotes the rate of recombination, and $er$ is the inbreeding coefficient. The effective population size should be thought of as the size of an ideal population [48,49] with the same magnitude of random genetic drift as the actual population with size $N$ [50].

To test if the placental DBL3X sequence data showed evidence for deviation from the neutral model of evolution assumed by the coalescent approach we calculated Tajima’s D statistic [31] and Fu and Li’s D* and F* statistic [32]. All three statistics are insignificantly different from zero ($p > 0.1$ in all cases), indicating that the coalescent approach could be applied. The hypothesis of no recombination was rejected using the likelihood permutation test [48] with 10,000 permutations of segregating sites, of which none produced a higher maximum composite likelihood than for the DBL3X data. The hypothesis of a constant recombination rate across the analyzed region was also rejected ($p = 0.048$ with 10,000 permutations of segregating sites, of which none produced a higher maximum composite likelihood than for the DBL3X data). The cumulated Kullback-Leibler distance was calculated as

$$D_{KL} = \sum (p - p') \cdot \frac{\log(p)}{p}$$

(1)

where $p$ and $p'$ are the frequencies of an amino acid type in each of the compared samples, and $p$ is the frequency of that amino acid type. The cumulated Kullback-Leibler distance was calculated as the sum of $D_{KL}$ for all positions in the alignment. Gaps in the alignment were in this analysis assigned the letter “O” and treated as an amino acid class. To test if the mentioned grouping according to peptide was significantly different for the three groups we created 10,000 random groupings and for each of these summed the $D_{KL}$ over all amino acid positions. Similarly for the individual positions in the logo, the distribution of $D_{KL}$ for 10,000 random shuffles of sequence grouping was noted specifically for each position, and the $p$-values were based on these distributions.

Peptide motif analysis. 442 overlapping 31mer peptides covering the exon 1 of 3D7 VAR2CSA were synthesized as solid phase peptide synthesis (SPPS) with a stepwise addition of the different amino acids attached to a solid resin. The long peptides were synthesized with a cysteine at amino acid (aa) position 15 allowing some secondary structure. This approach thus includes short peptides that cannot be mapped using short, linear peptides (Peptide Systems, the Netherlands). The raw data from a PepScan experiment consists of figures measuring the amount of IgG bound to each of the overlapping 31mer peptides. We used the overlap in primary sequence to determine more specifically what the antibodies have affinity for.

The motif analysis is based on the concept that polyclonal IgG response consists of subpopulations of monoclonal antibodies each binding a certain set of amino acid sequence motifs. We then made a list of all possible binding motifs and transferred the information from the peptide array to these, giving each of these a score indicating the IgG affinity for the motif. The PepScan assay is designed primarily to determine linear epitopes, and thus we are mainly interested in short binding motifs and gaps. On this basis we performed the PepScan motif analysis, using motifs containing either two or three defined residues spaced by undefined amino acids up to a certain maximal length of 5, 7, 10, 15, 20, or 25 residues. We found that the different maximal motif lengths gave similar results, but with different detail resolutions (long motifs had a smoothing effect), and a binding motif with a maximal length of seven residues was selected as being most informative. The results are based on the assumption that motifs are two to seven amino acids long, and contain either two or three defined residues spaced by undefined amino acids, e.g., a possible motif could be “WXXDXEXE” or simply “KX”. The method was validated by comparison to the raw data, where rabbit serum from a rabbit immunized with a DBL5 construct was used in the PepScan assay (Figure S2). The figure shows that the motif analysis produces peaks approximately at the same positions as in the raw data, and that the analysis does not introduce bias in the other regions of the protein. As control for the human IgG used in Figure 6, we used non-immune Dutch serum as well as a malaria-exposed nulliparous female individual.

Affinity purification of antibodies and depletion of serum on parasites. Affinity purification of antibodies was done according to manufacturer’s instructions. In brief, 1 mg of recombinant protein was coupled against 0.2 M NaHCO3, 0.5 M NaCl (pH 8.3), 1 ml applied to a NHS-activated HiTrap 1 ml column (GE Healthcare, http://www.gehealthcare.com) that had been equilibrated with 3 × 2 ml 4°C 1 mM HCl. After coupling, the columns were washed with 0.5 M ethanolamine, 0.5 M NaCl (pH 8.3), 0.1 M acetate, 0.5 M NaCl (pH 4), and then washed with PBS (pH 7.4) to 1 ml of a plasma pool (28 women from Ghana) was then applied to the column. After washing in 10 ml PBS, affinity-bound antibodies were eluted with CH3COONH4 (pH 3) and neutralized in 1 M Tris (pH 7.5). The specificity of the purified antibodies was tested in ELISA against (1) the domain used for affinity purification, (2) other VAR2CSA domains, and (3) glutamine rich (GLURP) domain. After antibodies used for PepScan analysis were all negative in ELISA against control proteins and positive against the homologous domain (unpublished data).

Surface reactive antibodies in a pool of pregnancy plasma (from 15
pregnant women from Korogwe, Tanzania) were depleted using a parasiteline selected for VAR2CSA expressionusing VAR2CSA specific antibodies [9]. In brief, 40 μl of the plasma pool were incubated with 2.0 × 10^6 MACs purified intact late stage trophozoite- and schizont-infected erythrocytes for 20 min at 4 °C. Hereafter, the cells were centrifuged at 800 g for 8 min, and the supernatant used to suspend a new pellet of infected erythrocytes. This procedure was repeated four times. The depletion of surface reactive antibodies from the plasma pool was confirmed using a flow cytometry assay [9] after the final depletion.

3-D modeling of the 3D7 DBL3X domain. The 3-D structure of the 3D7 sequence (PFL0030c aa 1217–1559) was modeled using the HHpred server with default settings [56]. Briefly, the HHpred method is specialized in remote homology detection using hidden Markov models (HMMs) built from PSI-BLAST profiles and secondary structure predictions [57]. The crystal structure of EBA-175 F1 (PDB code 1ZRO chain A, [15]) was used as template and had the highest sequence and secondary structure alignment scores. The HHpred HMMs for DBL3X and the template continued to match. Finally the corrected alignment was used to generate a 3-D model using MODELLER [57] with the protocol setup in the MAMMOTH-mult alignment server [59]. The resulting alignment was inspected to identify conserved, buried, and hydrophobic positions shown in blue, conserved disulfide bonds highlighted in bright yellow, buried hydrophobic residues (RSA < 30%). Structural visualizations were made using PyMol [60].

**Supporting Information**

**Figure S1.** Multiple Structural Alignment of VAR2CSA DBL3X, EBA-175 F1, F2, and Pkz-DBL domains. Cysteines in conserved disulfide bonds are highlighted in blue, conserved tryptophan buried in the structures are shown in red, and other conserved, buried, and hydrophobic positions are shown in green.

**References**