PfEMP1-DBL1α amino acid motifs in severe disease states of *Plasmodium falciparum* malaria

Johan Normark[†‡], Daniel Nilsson[†‡§], Ulf Ribacke[†‡], Gerhard Winter[†‡], Kirsten Moll[†‡], Craig E. Wheelock[†‡], Justus Bayarugaba[¶‖], Fred Kironde[††], Thomas G. Egwang[‡‡], Qijun Chen[†‡], Björn Andersson[§], and Mats Wahlgren[†‡§§]

[†]Department of Microbiology and Tumor and Cell Biology (MTC), Karolinska Institutet and [‡]Swedish Institute for Infectious Diseases Control (SMI), Box 280, Nobels väg 16, SE-171 77 Stockholm, Sweden; [§]Programme for Genomics and Bioinformatics, Department of Cell and Molecular Biology, Karolinska Institutet, Berzeliusväg 35, SE-17177 Stockholm, Sweden; Departments of [¶]Paediatrics and [‖]Biochemistry, Mulago Hospital, Kampala, Uganda; [††]University of Makerere, Box 7072, Kampala, Uganda; and [‡‡]Medical Biotech Laboratories, Ssekindi Close, Tank Hill Bypass, Box 9364, Kampala, Uganda

Edited by Louis H. Miller, National Institutes of Health, Rockville, MD, and approved August 13, 2007 (received for review November 27, 2006)

An infection with *Plasmodium falciparum* may lead to severe malaria as a result of excessive binding of infected erythrocytes in the microvasculature. Vascular adhesion is mediated by *P. falciparum* erythrocyte membrane protein-1 (PfEMP1), which is encoded for by highly polymorphic members of the *var*-gene family. Here, we profile *var* gene transcription in fresh *P. falciparum* trophozoites from Ugandan children with malaria through *var*-specific DBL1α-PCR amplification and sequencing. A method for subsectioning region alignments into homology areas (MOTIFF) was developed to examine collected sequences. Specific PfEMP1-DBL1α amino acid motifs correlated with rosetting and severe malaria, with motif location corresponding to distinct regions of receptor interaction. The method is potentially applicable to other families of variant proteins and may be useful in identifying sequence-phenotype relationships. The results suggest that certain PfEMP1 sequences are predisposed to inducing severe malaria.

homology areas | antigenic variation | rosetting

Plasmodium falciparum malaria infection is one of the primary contributors to childhood mortality in many developing countries. Despite exhaustive research efforts, no vaccine capable of conferring an adequate level of immunity has been developed to date. Vaccine development is encouraged by the fact that children attain conditional immunity to severe malaria after relatively few infections. However, a lack of insight into the molecular interplay between the parasite and the host continues to thwart efforts at vaccine development. One of the key factors in this process is the ability of the parasite to continually express a plethora of antigenic variants of gene families, whose selection appears to correspond with interactions with the host immune system.

Severe malaria, a highly lethal form of the disease, is, in part, attributable to the sequestration of *P. falciparum*-infected erythrocytes (IE) and uninfected erythrocytes in postcapillary venules of the brain, the lungs, or other organs (1). *P. falciparum* expresses a number of proteins at the erythrocyte surface that are closely involved in the excessive sequestration characteristic of severe malaria. The polypeptide most closely scrutinized to date, PfEMP1, is a large 200- to 400-kDa clonally variant antigen encoded by a repertoire of ≈60 *var* genes (2). The *var* genes present a two-exon structure encoding a conserved C terminus that contains a predicted transmembrane region and a polymorphic submodular N terminus. This N terminus region possesses a number of cysteine-rich domains that are intimately involved in the sequestration of the parasite in the microvasculature (3–5). The Duffy binding-like domain-α (DBL1α) located in the N-terminal head structure of PfEMP1 mediates rosetting and endothelial binding of IE. These binding processes occur by means of a range of different receptors, including heparan sulfate, complement receptor 1, and/or the blood group A antigen (6–9). Rosetting and endothelial binding are intimately associated with severe falciparum malaria (10–14). Subsequently, the mechanisms of antigenic variation and the potential to cause severe disease reside in the same molecule. This phenomenon illustrates why much of the malaria research to date has focused on this protein family to explain parasite pathogenicity.

The *var* genes can be subdivided into five distinct classes (A, B, C, D, and E) depending on the 5′ upstream sequences (15, 16) and groups A–C may be further subclassified depending on domain structure (16). Group A *var* gene transcription is more frequent in rosetting parasites, whereas both groups A and B are commonly transcribed in parasites obtained from children with severe malaria (17–19). Recent evidence indicates that transcription patterns vary with the severe malaria patient disease state (20). The determination of the tertiary structure of two DBL domains orthologous to PfEMP1-DBL1α has revealed similarities in protein folding and binding mechanisms (21, 22). To explore the relationships between sequence and disease states or parasite phenotype, we acquired a set of DBL1α sequences from children with severe and mild malaria and developed specific algorithms to identify sequence-motifs (MOTIFF) and examined them for correlations with observed phenotypes. The range of motifs identified in this study were placed into a structural context and found to correlate with specific parasite and/or disease phenotypes, suggesting that disease state and progression may correlate with the expression of certain PfEMP1 species and motifs.

Results

Rosetting and Severe Malaria. The rosetting rates (RR) of the IE of 93 fresh clinical isolates of Ugandan children with severe or mild malaria were determined [supporting information (SI) Table 1]. The mean RR of the IE from the severe patients was higher than that of the IE of the 42 patients with mild malaria (x = 14% vs. 4.4%, respectively; Mann–Whitney rank sum test $P < 0.001$; SI Fig. 4*A*). Also giant rosettes were observed in a significantly higher frequency in the severe isolates (Fisher exact test $P < 0.001$). These data suggested that the factors responsible for rosetting were also correlated with disease severity in Ugan-

Author contributions: J.N., G.W., J.B., F.K., T.G.E., B.A., and M.W. designed research; J.N., D.N., U.R., K.M., and Q.C. performed research; J.N., D.N., G.W., K.M., J.B., F.K., T.G.E., Q.C., and B.A. contributed new reagents/analytic tools; J.N., D.N., G.W., C.E.W., and M.W. analyzed data; and J.N., D.N., C.E.W., and M.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Abbreviations: DBL1α, Duffy binding-like domain 1α; PfEMP1, *Plasmodium falciparum* erythrocyte membrane protein 1; RR, rosetting rate; IE infected erythrocytes; GAG, glycosaminoglycan; NUD, non ultra descriptus.

Data deposition: The sequences reported in this paper were deposited in dbEST (GenBank accession nos. EE572314–EE589229).

[§§]To whom correspondence should be addressed. E-mail: mats.wahlgren@ki.se.

This article contains supporting information online at www.pnas.org/cgi/content/full/0610485104/DC1.

© 2007 by The National Academy of Sciences of the USA

MEDICAL SCIENCES

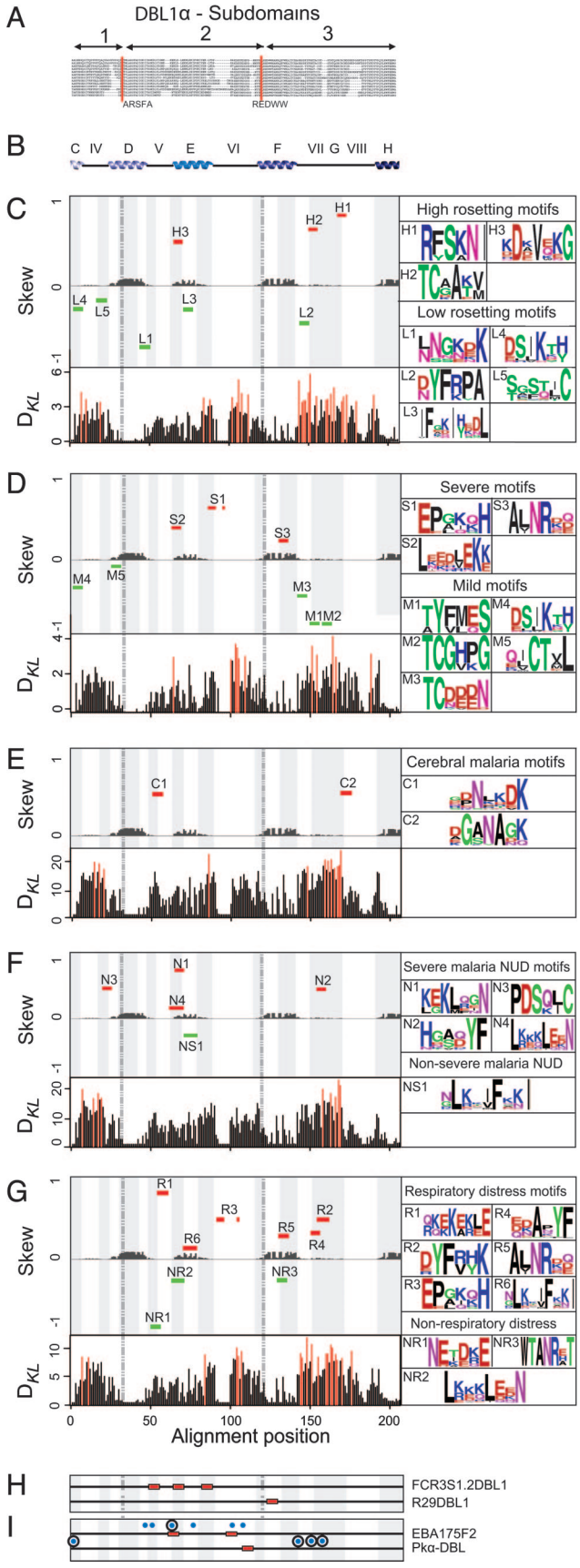

Fig. 1. Schematic of PfEMP1-DBL1α and the degenerate amino acid motifs identified in *P. falciparum*-infected erythrocytes of Ugandan children associated with different disease states. (*A*) An example alignment is displayed with the PCR primer-defined area subdividers (areas 1–3) in red. According to earlier classification, area 1 represents the end of homology block C, hypervariable block 4; area 2 represents homology blocks D and E, hypervariable blocks 5 and 6; and area 3 represents homology blocks F and H as well as hypervariable blocks 7 and 8 (27). (*B*) The part of the FCR3S1.2DBL1α model represented in the analysis has been linearized. The α-helices are shown in blue, and the loop regions are shown in black. The regions are annotated according to earlier classification. (*C*) Distributions of degenerate motifs statistically overrepresented in the high- and low-rosetting isolates. (*D*) Distributions of degenerate motifs statistically overrepresented in severe and mild cases. (*E*) Distributions of degenerate motifs statistically overrepresented in cerebral malaria. (*F*) Distributions of degenerate motifs statistically over- and underrepresented in severe malaria NUD patients. (*G*) Distributions of degenerate motifs statistically over- and underrepresented in patients with respiratory distress. (*H*) Schematic of the DBL1α regions found in *var* genes from *in vitro*-cultivated parasites, with putative GAG-binding domains marked with red boxes. (*I*) Schematic linear alignment of orthologous DBL domains for which the tertiary structures have been solved. Amino acids involved in receptor binding are indicated by blue dots, and amino acids essential for binding are represented by circled blue dots. (*C–G*) The divider between the high- and low-rosetting groups shows the high- and low-homology areas in the DBL1α fragment. Kullback–Liebler distances are stated for all positions in each alignment, statistically significant ($P < 0.05$) distance bars are shown in red; nonsignificant distance bars are shown in black. H refers to motifs found overrepresented in highly rosetting parasites (RR >5%); L, motifs in the low-rosetting category (RR ≥5%); S, severe motifs; M, mild motifs; C, cerebral malaria motifs; N, severe malaria NUD motifs; NS, nonsevere malaria NUD motifs; R, respiratory distress motifs; NR, nonrespiratory distress motifs.

dan *P. falciparum* infections, prompting focused studies on the composition of rosetting parasites.

***var* Gene Transcription.** To further our understanding of the adhesive events responsible for malaria severity, we investigated the *var* gene transcription of the collected IE. Because trophozoite-IE primarily express only one dominant transcript that translates into the relevant PfEMP1 species (23, 24) and because frozen and thawed parasites may switch the *var* gene expressed (25), we chose to examine *var* gene expression in purified fresh trophozoites. The primers α-AF and α-BR were used to amplify *var* transcripts as described (26). To ensure coverage of the majority of *var* genes, we developed a second set of primers, nDBLf and nDBLr, which targeted conserved areas in the DBL1α domain. In addition, we used a combination of the nDBLf and α-BR primers (see *Materials and Methods*). The primers provide high *var* gene coverage (86%; SI Fig. 5).

***var* Transcription in Wild Isolates.** A total of 26,784 sequencing reactions were performed from the amplified trophozoite cDNA, and assembled into 1,435 contigs (SI Fig. 6 and SI Table 2). Every isolate demonstrated a range of transcribed *var* genes, from a minimum of 9 (UAM5) to a maximum of 72 (UAS06). The three most dominant sequences from each isolate were then used for further analysis as described in *Materials and Methods*.

The assembled contigs showed an extreme diversity in sequence reflecting the high recombination and mutation rates occurring in the DBL1α domain. After assembly and translation, the amino acid sequences were assessed for distinguishing features according to their general amino acid composition. The PCR-primer sequences were used as dividing borders generating three subregions of DBL1α (areas 1–3; Fig. 1*A*), which are comparable to blocks C to H of Smith *et al.* (27) (Fig. 1*B*). A number of distinct patterns between sequence composition and phenotype were observed. Most importantly, low cysteine content of area 3 correlated with high RR ($r_s = -0.216$, $P < 0.001$; SI Table 3).

Motifs in DBL1α. Because the low-homology areas of the DBL1α sequence exhibited a number of distinguishing features, our

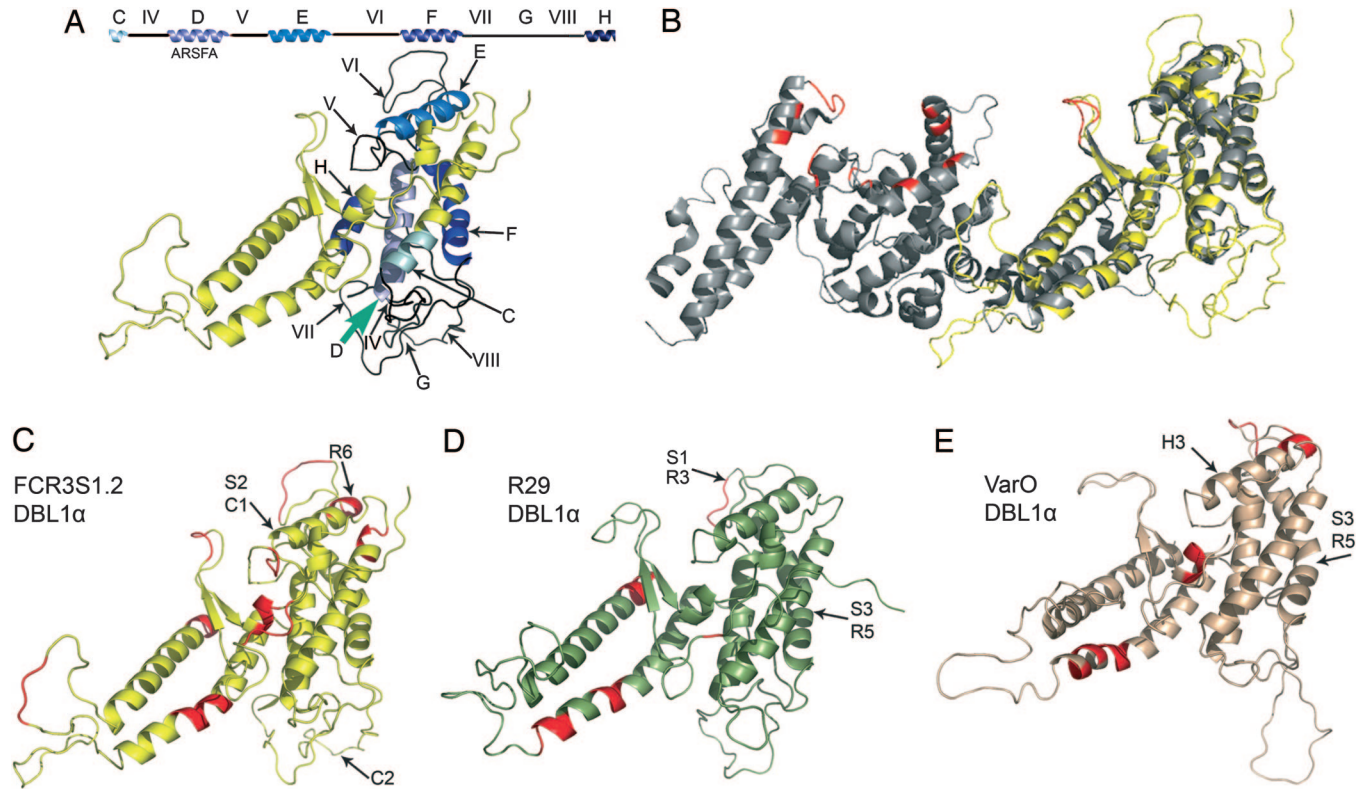

Fig. 2. Models of PfEMP1-DBL1α and the mapping of degenerate DBL1α amino acid motifs identified in *P. falciparum*-infected erythrocytes of Ugandan children. (*A*) The DBL1α region of FCR3S1.2, where α-helices are shown in blue and loop regions in black. The homology and hypervariable block regions are indicated by black arrows, except for the highly conserved ARSFA sequence region (homology block D), which is shown in green. (*B*) The FRC3S1.2DBL1α model is shown in yellow overlaid with the EBA-175 structure in gray. The positions of vital amino acids involved in binding to αSialyl-lactose are displayed in red. (*C*) FCR3S1.2DBL1α model. (*D*) R29DBL1α model. (*E*) VarODBL1α model. In *C–E*, putative GAG-binding motifs are indicated in red, and MOTIFF motifs identified in each model are indicated by arrows. H, high RR motif; S, severe malaria; R, respiratory distress; C, cerebral malaria; N, malaria NUD motifs.

strategy was to extract biochemically similar sequence motifs from these regions. Existing software for motif determination (e.g., MEME, MAST) had insufficient resolution for our purposes. It was therefore necessary to create a custom motif finder that was able to identify smaller motifs in very degenerate amino acid sequence sets. This algorithm was named "MOTIFF," because it is designed as a motif finder. Degenerate PfEMP1 motifs were classified by using disease severity and RR, with differential motif representations observed on a phenotype-dependent basis (hypergeometric $P < 0.05$; Fig. 1; SI Tables 4 and 5).

Motifs overrepresented in the *var* genes of low-rosetting isolates ($\leq$5%) were distributed over the entire length of the amplified DBL1α fragments (Fig. 1*C*). The sequences were often present in DBL1α domains containing four cysteine residues (cys4). In contrast, the motifs found significantly overrepresented and skewed for rosetting parasites (H1–H3; Fig. 1*C*) grouped into two distinct sections of DBL1α, with one cluster in area 2 and the other in area 3 (Fig. 1*C*). Area 2 is particularly rich in basic amino acids, with a 23% median molar percentage compared with 15% in area 3 (Mann–Whitney rank sum $P < 0.001$). The individual sequences constituting the motifs originate mainly from genes containing two cysteine residues (cys2), which are most likely group A *var* genes.

Motifs associated with the severity of the infection were similarly identified (S1–S3; Fig. 1*D*). The motifs were found to be clustered in areas 2 and 3 of DBL1α as with the high-rosetting motifs, albeit in different homology blocks. Only two motifs overlapped, mild M4 and low-rosetting L4 (Fig. 1 *B–D*).

The motif-finder software MOTIFF was additionally used to investigate motif overrepresentation in parasites of patients with different disease states. The severe patients were separated into three major groups: cerebral malaria, respiratory distress, and severe malaria non ultra descriptus (NUD) (SI Tables 6–8). It should be emphasized that separating the severe patients into subsets resulted in small sample sizes, which affected the confidence of the identified motifs. The cerebral malaria group, which was the smallest, generated two motifs (C1 and C2; Fig. 1*E*). Six motifs were identified in the DBL1α of the respiratory distress group (Fig. 1*G*), two of which (R3 and R5; Fig. 1*G*) were also present in the DBL1α general severe group (S1 and S3; Fig. 1*D*). Four motifs were found in the DBL1α of the NUD group, one of which (N4) was also present in sequences from the nonrespiratory distress group (NR2; Fig. 1*F*), and NS1 was also found as R6 (Fig. 1 *F* and *G*). The number of severe anemia cases as well as the circulatory collapse and multiphenotype groups was too small to generate statistically significant results.

3D Model Construction. Recent studies have shown that the receptor-binding residues of DBL1α lie in the central parts of the domain (9). Parallel experiments with other DBL orthologs have also identified receptor affinity in this region (28, 29), indicating the importance of this domain in binding to host receptors. The recently elucidated structure of DBL of erythrocyte-binding antigen 175 (EBA-175) (21) was here used to place the current study into a structural context by using characterized DBL1α sequences of FCR3S1.2, R29, and varO (6, 7, 30). The ultra-conserved ARSFA sequence (homology block D) was, as expected, found to be in the center of the DBL1α (Fig. 2*A*; green arrow). ARSFA is predicted to be surrounded by semiconserved

α-helices (homology blocks C, and E–G, Fig. 2*A*) that are in turn connected to flexible loops (hypervariable blocks IV, V, VI, VII, VIII; Fig. 2*A*), the latter being composed of highly variable amino acid sequences of DBL1α. The FCR3S1.2DBL1α model lacks the disulfide bond located in the predicted β-pleated sheet region. In R29 and varO, 8 of the 10 cysteine residues involved in disulfide bonding in the EBA-175 could be aligned and modeled. Two cysteines were lost because of truncation. A number of putative glucosamino-glycan (GAG) binding motifs (31) were identified in the models (Fig. 2 *C–E*).

The 3D models generated proved to be of high quality in the regions, matching well to the EBA-175F1 template, and evaluation with VERIFY3D evidenced high model quality in the areas with high template homology (score >0.1). In contrast, all of the loop regions of DBL1α scored poorly. Optimal models could only be created if the last 70 aa of the DBL1α were truncated from the molecules. The sequence identities of the models and the EBA-175 molecule were between 25% and 27% after truncation.

DBL1α Motifs Placed in the 3D Models. Motifs from the different MOTIFF analysis as well as potential GAG-binding motifs were identified in DBL1α FCR3S1.2, R29, and varO and placed in the models (SI Tables 9–13). As can be seen in Fig. 2, DBL1α of FCR3S1.2 was found to possess three MOTIFF motifs (S2, C1, and R6) that were localized to the same predicted α-helical structure (domain 2/homology region E). The 3D location of the H3 motif present in varODBL1α was also in the same predicted α-helical structure (Fig. 2*E*). On the opposite face of the molecule in the DBL1α of R29 and varO a motif (S3 and R5; Fig. 2*D*) was found in a predicted α-helix (homology block F) as well as a second motif (S1 and R3; Fig. 2*D*), located in an adjacent loop (hypervariable block VI) in R29DBL1α. One MOTIFF motif was found in the long-predicted flexible loop structure in area 3 (blocks VII, G, VIII) of FCR3S1.2DBL1α (C2, Fig. 2*C*).

Kullback–Liebler Distances. Kullback–Liebler distances (D_{KL}) were calculated for each amino acid position to complement the analysis (32). Regions with significant D_{KL} could in this way be colocalized with identified motifs. The Kullback–Liebler plots made for each disease state dichotomy pair evidenced significant alignment positions overlapping a majority of the areas where the motifs appear (Fig. 1). The blocks VII, G, and VIII showed the highest overall frequency of significantly distant alignment positions (Fig. 1 *A–G*).

Immunofluorescence. The key DBL1α motifs identified in this study were also observed in the DBL1α sequences of the *in vitro* strains (SI Tables 9–13). This finding prompted us to study the immune recognition of the Ugandan isolates by using two different sera raised to DBL1α of FCRS1.2 that contained the severe malaria motif S2, several other disease state-specific motifs (C1, C2, R6; SI Tables 9–13), and low-rosetting motifs. The mean serum reactivity to the IE surface (AntiR1, AntiR2, Fig. 3) was higher with the isolates in the severe relative to mild category. Averaging the two biological replicates for each isolate demonstrated significantly higher recognition [*t* test, $t = 2.195$; degrees of freedom (df) = 13; $P_t = 0.047$]. For the separate tests, AntiR2 demonstrated statistical significance (AntiR2: *t* test, $t = 2.413$; df = 13; $P_t = 0.031$; power = 0.524; Kolmogorov–Smirnov (K–S) test shows normal distribution of data: K–S distance = 0.232 and 0.292 in the two groups; AntiR1 not significant; SI Fig. 7). However, no significant correlation between RRs and antibody reactivity was observed in the material.

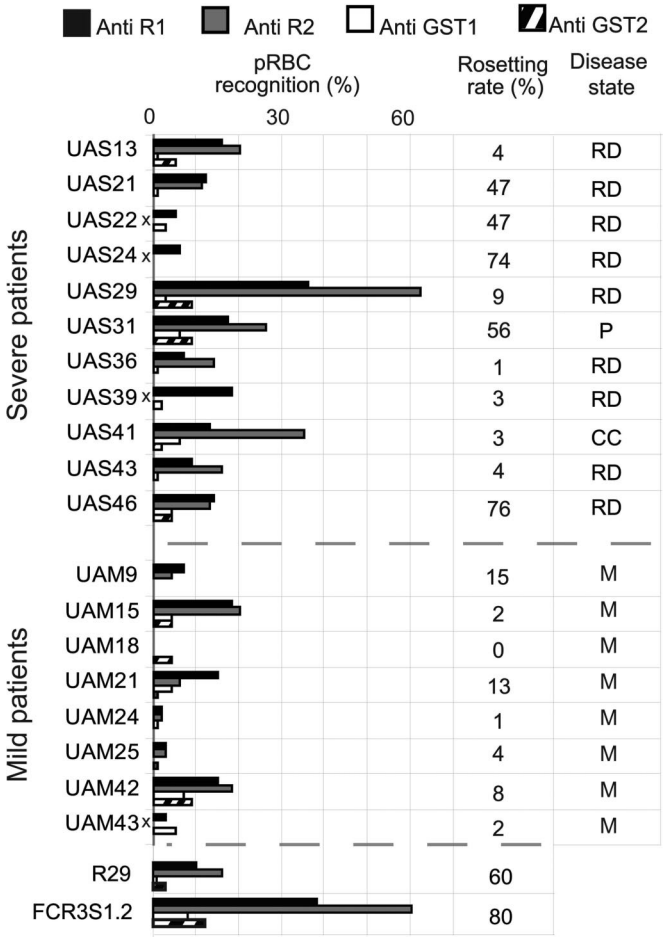

Fig. 3. Indirect immunofluorescence. The surface fluorescence rate in percentage of fluorescing pRBCs of the IE of severe (S) and mild (M) isolates. AntiR1 and AntiR2 denote the two antisera to DBL1α of PfEMP1var1 of FCR3S1.2DBL1. AntiGST1 and AntiGST2 denote the two control sera. AntiR1 is shown in black, AntiR2 is shown in gray, AntiGST1 is indicated in white, and AntiGST2 is indicated in hatched black and white. Isolates not tested with AntiR2 are marked with an X. The RR is stated in percent. The disease states of the patients are shown, with P signifying prostration, RD for respiratory distress, and CC for circulatory collapse.

Discussion

The ability of the parasite *P. falciparum* to undergo antigenic variation as a way to evade immune recognition by the host is well documented. However, an important remaining challenge in the study of antigenic variation is the ability to extract and analyze information from vast bulk sequence data sets. We here demonstrate a scheme to identify DBL1α sequence motifs pertinent to disease states and parasite phenotypes based on an algorithm named MOTIFF. By applying this method to *var* gene sequence data from 93 patients with well characterized disease states, we have identified motifs and differential amino acid contents in distinct areas of DBL1α that correlate to the observed RR and disease characteristics. These findings will be of relevance for vaccine development and understanding disease pathogenesis.

Amplifying the *var* transcripts encoding the relevant DBL1α domains responsible for the observed disease states relies heavily on the PCR primers demonstrating as unbiased of a selection of *var* species as possible. Three complimentary primer sets were therefore used to try to maximize selection and minimize primer bias. Still, even with a large number of reads sequenced from 3D7 genomic DNA, full coverage could not be attained (86%; SI Fig. 4). Accordingly, further work remains to be done to hone the methodology of assessing *var* gene transcript-dominance in an isolate.

Previous studies have demonstrated a relationship between the expression of *var* genes/ DBL1α domains with a lesser number of cysteines (Cys-2), an amino acid tag (REY*), and a high RR (17, 33). The correlation between the RR and the presence of cys2 transcripts was also found in our analysis, whereas we did not observe a correlation between the presence of the Cys-2/Rey* tag and a high RR (RR >5%, χ^2 test = 0.251, df = 1, P = 0.616). On the other hand, we identified a total of 15 degenerate motifs using MOTIFF that were associated with severe malaria and three motifs that were associated with high rosetting. Motif distribution along DBL1α was not random and found mainly in areas 2 and 3 (blocks D–H). There was still little sequence overlap between rosetting motifs and the severe motifs, even though they were frequently part of the same DBL1α subdomain. These data merely suggest the presence of a correlation between rosetting and severe disease, which is by no means certain. In addition, the parameter for high rosetting was arbitrarily set at >5%, which may also have influenced the distribution of the isolates in the analysis. Finally, the observation may also be a reflection of diversity within the severe phenotype.

To strengthen the data generated with the MOTIFF algorithm, patient isolates were analyzed for immune recognition by using sera raised to a DBL1α possessing motifs S2, C1, C2, and R6 (FCR3S1.2DBL1α). Sera raised according to the same methodology as herein have previously been shown to significantly reduce rosetting in FCR3S1.2 parasites (34). The subset of isolates studied originated in Apac and consisted of mainly respiratory distress patients, because this was the major patient category at the site. The use of an *in vitro* assay with live IE provided an authentic setting for assessing recognition of PfEMP1 epitopes on the IE. The results suggest that there is significant serological cross-reactivity in-between the isolates. Taken together, these data advocate that (*i*) similar epitopes in DBL1α domains expressed on the IE surface are recognized by the sera, (*ii*) cross-reactive immune responses in between distinct IE can be raised, and (*iii*) IE of children with severe malaria are, in part, antigenically distinct from those with mild disease.

From this work, it can be surmised that a well conserved DBL1α scaffold can be modeled, giving regions and constraints to the spatial location of motifs. The loop-region structures cannot be correctly predicted apart from their obvious location in the scaffold. In our detailed analysis of the four-cysteine FCR3S1.2DBL1α, a domain very well characterized for heparin binding, the location of putative GAG-binding motifs in region 2 (homology domain D) mimic the carbohydrate-binding pocket in EBA175F2. It also retains a GAG-binding loop aligning with the carbohydrate-binding motif in F1 of EBA175F2 (high-variable region V of DBL1α). When the severe S2 motif that is present in FCR3S1.2DBL1α is plotted onto the model, it is located in the semivariable helix adjacent to the GAG-binding motifs. Additional MOTIFF motifs (C1, R6) of FCR3S1.2DBL1α, were also identified in the same α-helical structure. However, the situation is slightly different in R29DBL1α, a two-cysteine domain, because the putative GAG-binding motifs are fewer and more dispersed, and the severe S1 and S3 motifs are located on the opposite side of the molecule F (Fig. 2 *D* and *E*). Interestingly, the amino acid positions in a related DBL domain (DBL3X) found to be skewed in the primi- and multigravidae groups by Dahlbäck *et al.* (32) overlap spatially with the S1/R3 and S3/R5 motif locations in the R29DBL1α model. This fold region in Pkα-DBL is also involved in DARC recognition (22).

As expected, the motifs extracted by MOTIFF from the sequence analyses generally appear in the predicted semiconserved α-helices exposed in the exterior of the molecule. These α-helical regions probably have morphological constraints but are likely to be exposed to the immune system and possibly available to function as adhesins. Furthermore, the location of the degenerate motifs in DBL1α correspond to the regions of fold-conservation and receptor-interaction in the DBL domains of EBA-175 and Pkα-DBL, suggesting that PfEMP1-DBL1α has retained a similar conserved structure and area for receptor binding (9). However, it has to be stressed that, although DBL1α is a hypervariable protein domain, it contains regions with high conservation. Because DBL1α binds to multiple receptors, it is possible that binding sites exist on either one or both faces of the molecule. The three domains (PfEMP1-DBL1α, DBL-EBA-175, and Pkα-DBL) are related, albeit the true structural relationships remain to be elucidated. These data therefore corroborate purported ligand–receptor interactions in reported crystal structures and offer a method to identify regions in the DBL1α molecule that are especially important in disease and parasite phenotypes. Our method of relating phenotype information to sequence may also be useful for identifying other specific motifs involved in biological interactions. Candidates for investigation may include the envelope proteins in HIV and variable protein families, such as *vmp* genes in *Borrelia* spp., *vsg* genes in *Trypanosoma* spp., and *als* genes in *Candida* spp.

The data presented here support the current hypothesis that interactions between residues on the DBL1α domain and ligands in the host contribute to disease manifestation. In addition, this mechanism is consistent with the currently available structural evidence.

Materials and Methods

For further details on *P. falciparum* culture, RNA extraction, cloning and sequencing, primer validation, sequence annotation, transcript dominance, tertiary structure modeling of DBL1α, Kullback–Liebler distances, animal immunizations, and indirect surface fluorescence assay see *SI Methods*.

Ugandan Study Population. A total of 93 children under the age of five, with active *P. falciparum* infection, were included in this study. Patients were recruited in two locations in Uganda: at the district hospital in Apac, which is situated in a malaria holoendemic area (35) 250 km north of Kampala, and at the Mulago hospital, located in the capital. The patients were grouped according to WHO guidelines (36) and the modified Blantyre score (37) by the medical officer in charge. Fifty-one patients were included in the severe malaria group (38 from Apac, 13 from Kampala) and 42 patients were included in the mild malaria group. The severe cases were further sub grouped in disease states: respiratory distress (n = 28), cerebral malaria (n = 7), malaria NUD (n = 10), (which included patients with convulsions, prostration, hyperparasitemia, and hyperpyrexia), circulatory collapse (n = 3), sole severe anemia (n = 2), and one patient with a multiphenotype including cerebral malaria, severe anemia, and respiratory distress (n = 1). Informed consent was obtained from the parents of the patients. The study was approved by Karolinska Institutet's Regional Ethical Review Board (permission 03/095) and the Uganda National Council for Science and Technology (permission MV 717).

RT-PCR and PCR of *var* Sequences. The GeneAMP RNA PCR kit (Applied Biosystems, Foster City, CA) was used to reverse transcribe total RNA. The DBL1α domain of *var*-genes/PfEMP1 was amplified from the cDNA template by PCR by using three primer sets. The α-AF and α-BR primers (26) (SI Fig. 5). The nDBLf (5′-TKGCAG CMAAWTA YGARGS-3′) and nDBLr (5′-KTCCAC CAATCT TCYCT-3′) generated 250–350 bp products, using the same PCR conditions as the α-AF/α-BR primers. Cycling conditions were a 3-min denaturation step, followed by 35 cycles of 30 sec at 45°C, 45 sec at 60°C, and 15 sec at 94°C and terminated with an elongation step of 7 min at 72°C. The nDBLf as the forward primer and the α-BR as the reverse primer were used in the last PCR. The conditions were exactly

the same as with the nDBLf/nDBLr primer set and generated products of 450–550 bp. Data processing pipeline perl and shell scripts are available upon request; see also www.varDB.org.

Bioinformatic Tools. Sequence reads were base-called by using phred (version 020425.c). The reads were clustered by using phrap (version 0.990319), with retain_duplicates, minmatch 20 and repeat_stringency 0.9, and, otherwise, default settings. The translations of cluster consensus sequences were scanned for the presence of the well conserved RSFADIGDI motif by using fuzzpro, EMBOSS 3.0.0 (38). Nucleotide and peptide alignments of the retained clusters were made by using CLUSTALW, (version 1.83). The fraction $f_{ij} = r_{ij}/n_i$, where r_{ij} denotes the number of sequences for patient i in cluster j and n_i the total number of reads from each patient i, was used to score *var* gene dominance. The f_{ij} were ranked to establish a transcriptional *var* gene dominance order in each patient.

Motif Finding (MOTIFF). The data consist of a protein alignment and dominance count data. The alignment was regionalized stepwise. A representative seed sequence was chosen for each region. PSIBLAST [National Center for Biotechnology Information (NCBI) blastpgp v2.2.10] with the BLOSUM62 substitution matrix, multipass inclusion threshold 10, expectation value 100, and hit-extension threshold 4 was used to iteratively identify other PfEMP1 members containing that motif. The motifs can either be constructed from any previously unhit part of the aligned sequences (nonstrict), or only against the columns being investigated and a small window surrounding them (strict). A position-specific score matrix motif database was established. The dominance labeled motif member clusters were counted. The hypergeometrical probability P_k of finding that level of overrepresentation or more by chance in a motif k was used to rank the motifs. Motifs were located in new sequences by means of position-specific scoring matrices (PSSMs) by using the prophecy and prophet programs of the EMBOSS package.

We thank Lena Lundin and Hamid Darban for technical assistance, Dr. Judy Orikiriza for assistance in the clinic, the staff at Apac and Mulago hospitals, and the parents and children who participated in the study. This work was supported in part by grants from the Swedish International Development Authority (Sida/SAREC), the Swedish Research Council (VR), and the European Commission (BioMalPar).

1. Miller LH, Good MF, Milon G (1994) *Science* 264:1878–1883.
2. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, *et al.* (2002) *Nature* 419:498–511.
3. Baruch DI, Pasloske BL, Singh HB, Bi X, Ma XC, Feldman M, Taraschi TF, Howard RJ (1995) *Cell* 82:77–87.
4. Smith JD, Chitnis CE, Craig AG, Roberts DJ, Hudson-Taylor DE, Peterson DS, Pinches R, Newbold CI, Miller LH (1995) *Cell* 82:101–110.
5. Su XZ, Heatwole VM, Wertheimer SP, Guinet F, Herrfeldt JA, Peterson DS, Ravetch JA, Wellems TE (1995) *Cell* 82:89–100.
6. Rowe JA, Moulds JM, Newbold CI, Miller LH (1997) *Nature* 388:292–295.
7. Chen Q, Barragan A, Fernandez V, Sundstrom A, Schlichtherle M, Sahlen A, Carlson J, Datta S, Wahlgren M (1998) *J Exp Med* 187:15–23.
8. Vogt AM, Barragan A, Chen Q, Kironde F, Spillmann D, Wahlgren M (2003) *Blood* 101:2405–2411.
9. Mayor A, Bir N, Sawhney R, Singh S, Pattnaik P, Singh SK, Sharma A, Chitnis CE (2005) *Blood* 105:2557–2563.
10. Carlson J, Helmby H, Hill AV, Brewster D, Greenwood BM, Wahlgren M (1990) *Lancet* 336:1457–1460.
11. Newbold C, Warn P, Black G, Berendt A, Craig A, Snow B, Msobo M, Peshu N, Marsh K (1997) *Am J Trop Med Hyg* 57:389–398.
12. Roberts DJ, Pain A, Kai O, Kortok M, Marsh K (2000) *Lancet* 355:1427–1428.
13. Rowe JA, Rogerson SJ, Raza A, Moulds JM, Kazatchkine MD, Marsh K, Newbold CI, Atkinson JP, Miller LH (2000) *J Immunol* 165:6341–6346.
14. Fairhurst RM, Baruch DI, Brittain NJ, Ostera GR, Wallach JS, Hoang HL, Hayton K, Guindo A, Makobongo MO, Schwartz OM, *et al.* (2005) *Nature* 435:1117–1121.
15. Voss TS, Thompson JK, Waterkeyn J, Felger I, Weiss N, Cowman AF, Beck HP (2000) *Mol Biochem Parasitol* 107:103–115.
16. Lavstsen T, Salanti A, Jensen AT, Arnot DE, Theander TG (2003) *Malaria J* 2:27.
17. Bull PC, Berriman M, Kyes S, Quail MA, Hall N, Kortok MM, Marsh K, Newbold CI (2005) *PLoS Pathog* 1:e26.
18. Kaestli M, Cockburn IA, Cortes A, Baea K, Rowe JA, Beck HP (2006) *J Infect Dis* 193:1567–1574.
19. Rottmann M, Lavstsen T, Mugasa JP, Kaestli M, Jensen AT, Muller D, Theander T, Beck HP (2006) *Infect Immun* 74:3904–3911.
20. Kyriacou HM, Stone GN, Challis RJ, Raza A, Lyke KE, Thera MA, Kone AK, Doumbo OK, Plowe CV, Rowe JA (2006) *Mol Biochem Parasitol* 150:211–218.
21. Tolia NH, Enemark EJ, Sim BK, Joshua-Tor L (2005) *Cell* 122:183–193.
22. Singh SK, Hora R, Belrhali H, Chitnis CE, Sharma A (2006) *Nature* 439:741–744.
23. Chen Q, Fernandez V, Sundstrom A, Schlichtherle M, Datta S, Hagblom P, Wahlgren M (1998) *Nature* 394:392–395.
24. Kyes S, Pinches R, Newbold C (2000) *Mol Biochem Parasitol* 105:311–315.
25. Southwell BR, Brown GV, Forsyth KP, Smith T, Philip G, Anders R (1989) *Trans R Soc Trop Med Hyg* 83:464–469.
26. Taylor HM, Kyes SA, Harris D, Kriek N, Newbold CI (2000) *Mol Biochem Parasitol* 105:13–23.
27. Smith JD, Subramanian G, Gamain B, Baruch DI, Miller LH (2000) *Mol Biochem Parasitol* 110:293–310.
28. Ranjan A, Chitnis CE (1999) *Proc Natl Acad Sci USA* 96:14067–14072.
29. VanBuskirk KM, Sevova E, Adams JH (2004) *Proc Natl Acad Sci USA* 101:15754–15759.
30. Russell C, Mercereau-Puijalon O, Le Scanf C, Steward M, Arnot DE (2005) *Mol Biochem Parasitol* 144:109–113.
31. Cardin AD, Weintraub HJ (1989) *Arteriosclerosis* 9:21–32.
32. Dahlback M, Rask TS, Andersen PH, Nielsen MA, Ndam NT, Resende M, Turner L, Deloron P, Hviid L, Lund O, *et al.* (2006) *PLoS Pathog* 2:e124.
33. Kirchgatter K, Portillo Hdel A (2002) *Mol Med* 8:16–23.
34. Chen Q, Pettersson F, Vogt AM, Schmidt B, Ahuja S, Liljestrom P, Wahlgren M (2004) *Vaccine* 22:2701–2712.
35. Yeka A, Banek K, Bakyaita N, Staedke SG, Kamya MR, Talisuna A, Kironde F, Nsobya SL, Kilian A, Slater M, *et al.* (2005) *PLoS Med* 2:e190.
36. World Health Organization (2000) *Trans R Soc Trop Med* 94:1–90.
37. Molyneux ME, Taylor TE, Wirima JJ, Borgstein A (1989) *Q J Med* 71:441–459.
38. Rice P, Longden I, Bleasby A (2000) *Trends Genet* 16:276–277.